Cholinergic Striatal Interneurons and the Midbrain Dopamine System Control the Learning and Unlearning of Procedural Skills

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

in

Psychology

by

Matthew J. Crossley

Committee in charge:

Professor F. Gregory Ashby, Chair
Professor Scott T. Grafton
Professor Aaron Ettenberg
Professor Skirmantas Janusonis

2011
The dissertation of Matthew J. Crossley is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

Chair

University of California, Santa Barbara

2011
DEDICATION

To two, the loneliest number since the number one.
# TABLE OF CONTENTS

Signature Page ................................................................. iii
Dedication ................................................................. iv
Table of Contents ............................................................ v
List of Figures ............................................................... vii
List of Tables ............................................................... ix
Acknowledgements .......................................................... x
Vita and Publications ....................................................... xi
Abstract ........................................................................ xii

Chapter 1  Introduction ...................................................... 1
  1.1 Learning and Unlearning .............................................. 1
  1.2 The Basal Ganglia and the Striatum ............................... 3
  1.3 DA and ACh - Synaptic Plasticity in the Striatum ............. 6
  1.4 Instrumental Conditioning .......................................... 8
  1.5 II Category Learning .................................................. 12
  1.6 Aims and Scope of the Model .................................... 14

Chapter 2  Instrumental Conditioning Applications ................. 15
  2.1 Reacquisition ......................................................... 15
      2.1.1 Introduction ................................................... 15
      2.1.2 Methods ....................................................... 16
      2.1.3 results ....................................................... 24
      2.1.4 Discussion ................................................... 28
  2.2 Renewal ............................................................... 29
      2.2.1 Introduction ................................................... 29
      2.2.2 Methods ....................................................... 30
      2.2.3 results ....................................................... 33
      2.2.4 Discussion ................................................... 39

Chapter 3  Category Learning Applications ............................ 42
  3.1 Experiment 1 ......................................................... 43
      3.1.1 Introduction ................................................... 43
      3.1.2 Method ....................................................... 45
      3.1.3 Results ....................................................... 47
      3.1.4 Discussion ................................................... 53
3.2 Experiment 2 ........................................... 60
  3.2.1 Introduction ........................................ 60
  3.2.2 Method ............................................. 60
  3.2.3 Results ............................................. 61
  3.2.4 Discussion .......................................... 66
3.3 A Theoretical Account of Experiments 1 and 2 .......... 71
  3.3.1 Methods ............................................ 72
  3.3.2 Results ............................................. 79
  3.3.3 Discussion .......................................... 82

Chapter 4 Discussion ....................................... 85
  4.1 Introduction to Discussion ............................ 85
  4.2 Scope and Limitations of the Model ................. 86
  4.3 Relationship to other contemporary theories ....... 87
  4.4 Relationship to historically relevant theories .... 89
  4.5 Clinical Relevance .................................... 89

Appendix A ................................................. 92
  A.1 Fit Parameters ......................................... 92
  A.2 Decision Bound Models ............................... 93
    A.2.1 Rule-Based Models ............................... 93
    A.2.2 Information-Integration Models ............... 94
    A.2.3 Random Responder Models .................... 94

Bibliography .............................................. 96
LIST OF FIGURES

Figure 1.1: A few examples of stimuli that might be used in an information-integration category learning task. ........................................ 13

Figure 2.1: Model Architecture used to fit Woods and Bouton (2007). .... 17

Figure 2.2: Behavioral results from Woods and Bouton (2007) Experiment 1. Right: Behavioral results obtained from the model. See text for more details. .................................................. 25

Figure 2.3: Top left: Average DA release. Bottom left: Average proportion responses emitted. Top right: Average synaptic strength of the CTX-MSN synapse. Bottom right: Average synaptic strength of the Pf-TAN synapse. ...................................... 26

Figure 2.4: Model Architecture used to fit Bouton et al. (2011) ........... 31

Figure 2.5: Top and Middle panels: Behavioral results from Bouton et al. (2011). Bottom: Behavioral results obtained from the model. See text for more details. ........................................ 34

Figure 2.6: A) Average DA release. B) Average synaptic strength of the CTX-MSN synapse. C) Average proportion responses emitted. D) Average synaptic strength of the Pf-TAN synapse. ....... 36

Figure 2.7: A) Average DA release. B) Average synaptic strength of the CTX-MSN synapse. C) Average proportion responses emitted. D) Average synaptic strength of the Pf-TAN synapse. ....... 38

Figure 2.8: A) Average DA release. B) Average synaptic strength of the CTX-MSN synapse. C) Average proportion responses emitted. D) Average synaptic strength of the Pf-TAN synapse. ....... 40

Figure 3.1: The stimuli used in Experiments 1 and 2 ......................... 44

Figure 3.2: The left column shows results obtained from all participants, and the right column shows results obtained by exclusively considering participants best fit by a decision-bound model that assumed II responding. A and B) The mean accuracy across all blocks of the experiment in both the Reacquisition condition and the Meta-Learning condition. C and D) The mean accuracy in the Reacquisition condition during each phase of the experiment overlaid on top of each other. E and F) The mean accuracy in the Meta-Learning condition during each phase of the experiment overlaid on top of each other. ............. 48
Figure 3.3: The left column shows results obtained from all participants, and the right column shows results obtained by exclusively considering participants best fit by a decision-bound model that assumed II responding. A and B) The mean accuracy across all blocks of the experiment in both the Reacquisition condition and the Meta-Learning condition. C and D) The mean accuracy in the Reacquisition condition during each phase of the experiment overlaid on top of each other. E and F) The mean accuracy in the Meta-Learning condition during each phase of the experiment overlaid on top of each other.

Figure 3.4: Model Architecture used to fit the data from Experiments 1 and 2

Figure 3.5: A) Human behavioral results from Experiment 1. B) Model account of Experiment 1. C) Human behavioral results from Experiment 2. D) Model account of Experiment 2.

Figure 3.6: A) Average DA release. B) Average synaptic strength of the CTX-MSN synapse. C) Average proportion responses emitted. D) Average synaptic strength of the Pf-TAN synapse.
LIST OF TABLES

Table 3.1: Number of participants in the two conditions of Experiment 1 whose responses were best accounted for by a model that assumed an information-integration (II) decision strategy, a rule-based (RB) strategy, or random guessing, and the percentage of responses accounted for by those models ........................................ 52

Table 3.2: Number of participants in the two conditions of Experiment 2 whose responses were best accounted for by a model that assumed an information-integration (II) decision strategy, a rule-based (RB) strategy, or random guessing, and the percentage of responses accounted for by those models ................................. 65

Table A.1: Fit Parameters ......................................................... 92
ACKNOWLEDGEMENTS

Thanks to whoever deserves credit for coffee.
VITA

2003 B. S. in Physics, University of California, Santa Barbara

2011 Ph. D. in Psychology, University of California, Santa Barbara

PUBLICATIONS


Procedural skills (i.e., skills that are learned incrementally through trial-and-error) represent a huge and important subset of behavior that underpin many of the most fundamental aspects of how organisms survive and prosper. However, procedural skills are also thought to underly maladaptive states such as addiction. There is overwhelming evidence that once a procedural skill is learned, it is extremely difficult to unlearn. In this dissertation, we propose and test a biologically-detailed theory of how such skills are learned, and why they are so difficult to unlearn. The theory we propose in this dissertation is characterized by a number of key features. First, procedural skills are learned at cortico-striatal synapses. Second, cholinergic interneurons in the striatum known as TANs (i.e. Tonically active
neurons) tonically inhibit cortical input to striatal output neurons, and thereby
gate the learning and expression of striatal dependent behaviors. The TANs are
driven by cells in the parafascicular (Pf) nucleus of the thalamus, which in turn
are broadly tuned to features of the environment. Finally, the model assumes that
learning at all cortical-striatal and Pf-TAN synapses is driven by a dopamine signal
that is sensitive to action-outcome contingencies. We examine the ability a formal
computational model based on these features to account for key results from two
well known classes of procedural skills (i.e., appetitive instrumental conditioning
and information integration category learning).
Chapter 1

Introduction

1.1 Learning and Unlearning

The world spends huge amounts of money every year on unlearning programs in prisons, psychotherapy clinics, drug treatment facilities, and schools. Yet virtually all of these programs must be classified as failures because recidivism rates are high in each of these fields. Why are behaviors so difficult to unlearn? This manuscript proposes a neurobiological theory of why one important class of behaviors namely, those acquired via procedural learning is so resistant to unlearning. We apply this theory in two important ways: 1) We successfully apply this theory to a number of important results from the instrumental conditioning literature, and 2) we report the results of new, theoretically-motivated behavioral experiments that demonstrate promising initial results at inducing true unlearning. Together, these applications offer exciting insight into why unlearning is so difficult to achieve, and provides initial guidelines to develop more effective unlearning training protocols.

To begin, it is important to distinguish between learning and performance. Researchers have long recognized that although the expression of a skilled behavior generally indicates that the behavior has been learned, the absence or failure to produce the behavior after initial learning does not necessarily indicate that the behavior was unlearned. By unlearning we mean the erasing or obliterating of the memory traces that encode the behavior. An animal might fail to produce a
behavior for many reasons, only one of which is unlearning.

This distinction between learning and performance is well known in the conditioning literature. The best known example occurs during extinction, in which the removal of reinforcement causes a previously reinforced behavior to disappear. There is overwhelming evidence that extinction does not imply unlearning (e.g., Bouton, 2002; Konorski, 1948; Pearce & Hall, 1980; Pavlov, 1927; Rescorla, 2001) because a variety of experimental manipulations can cause the behavior to quickly reappear. For example, if the reinforcements are reintroduced after extinction the behavior is reacquired much faster than during original acquisition (Woods & Bouton, 2007; Bullock & Smith, 1953). This observation is commonly referred to simply as fast reacquisition and is widely recognized to suggest that the original learning was preserved during the extinction period, even though the behavior disappeared. There are many other behavioral phenomena that imply a failure of unlearning during extinction (Brooks & Bouton, 1993; Estes, 1955; Pavlov, 1927; Bouton & Bolles, 1979; Bouton & King, 1983; Nakajima, Tanaka, Urshihara, & Imada, 2000).

One influential account of extinction is that removing reinforcements causes new learning rather than unlearning (e.g., Bouton, 2002, 2004). The idea is that during original acquisition the animal learns the behavior, and also the context under which the behavior is appropriate (i.e., reinforced). During extinction the animal then learns about a context in which the behavior is inappropriate (i.e., not reinforced). So according to this account, extinction is not primarily an unlearning phenomenon, but rather an example of context learning. The theory proposed in this article is conceptually similar to this context learning account of extinction. In fact, our theory could be interpreted as a neurobiological instantiation of this context learning account.

The theory that we propose is restricted exclusively to behaviors acquired via procedural learning and that depend on the striatum (a major input structure within the basal ganglia). Often such behaviors are referred to as skills or habits. A convenient operational definition is that such behaviors improve incrementally and require extensive practice with feedback (Willingham, Wells, Farrell,
Procedural skills cannot be mastered via observation or by listening to a lecture. Prototypical examples include athletic skills and playing a musical instrument, but many cognitive skills also meet these criteria.

1.2 The Basal Ganglia and the Striatum

The basal ganglia are a collection of subcortical structures that have been implicated in nearly all forms of learning and behavior, and are especially important for the type of procedural learning we are chiefly concerned about in this manuscript. The basal ganglia are comprised of two major input structures (the striatum and the subthalamic nucleus), two major output structures (the internal segment of the globus pallidus and the substantia nigra pars reticulata), and the external segment of the globus pallidus which is connected primarily to other basal ganglia nuclei. The basal ganglia send and receive massive anatomical projections to and from nearly the entire cortex, thalamus, and brainstem. In the classical view, these connections are organized into three main pathways (e.g., the direct pathway, the indirect pathway, and the hyper direct pathway), all of which are organized into multi-synaptic loops that start in the cortex, are routed through the basal ganglia to thalamic nuclei, and are subsequently returned back to cortex (Alexander et al., 1986; Parent & Hazrati, 1995; Middleton & Strick, 2000). The direct pathway - which is essentially a loop containing cortex, striatum, GPi/SNr, and thalamus - forms the foundation of the theory developed in this dissertation.

The striatum is known to contribute to many aspects of motor, cognitive, and limbic processing, and a huge literature suggests that the striatum is critically important in procedural skill learning (for reviews, see e.g., Ashby & Ennis, 2006; Doyon & Ungerleider, 2002; Packard & Knowlton, 2002; Yin & Knowlton, 2006). In humans, approximately 96% of all striatal neurons are medium spiny neurons (MSNs; Yelnik, Francois, Percheron, & Tande, 1991), which receive cortical input and send axons out of the striatum to basal ganglia output structures. The rest of the striatum is composed of a variety of interneurons. The TANs, which are large aspiny cholinergic interneurons that have extensive axon fields allowing them
to project to large striatal regions, are of primary interest to us (e.g., Calabresi, Centonze, Gubellini, Pisani, & Bernardi, 2000; Kawaguchi, Wilson, Augood, & Emson, 1995). The TANs are the striatum's primary source of ACh and yet the striatum has one of the densest supply of acetylcholine (ACh) in the entire brain (Hebb & Silver, 1956; Yamamura, 1974).

The TANs are tonically active in their resting state and they have a prominent modulatory effect on MSNs (Akaike, Sasa, & Takaori, 1988; Akins, Surmeier, & Kitai 1990; Dodt & Misgeld, 1986; Gabel & Nisenbaum, 1999; Pakhotin & Bracci, 2007). These effects are both pre- and post-synaptic and cannot be precisely summed up as simply excitatory or inhibitory. With respect to cortical input however, the predominant effect of TAN activity on MSN activation is inhibitory. For example, Pakhotin and Bracci (2007) reported that a single TAN spike caused a significant reduction in the excitatory postsynaptic current induced by cortical (glutamatergic) input. Based on these and other results, they concluded that following a TAN pause, MSNs "will transiently become much more responsive to cortical inputs" (p. 399), and that the resumption of TAN firing "will cause an abrupt reduction of MSN excitation" (p. 399).

Thus, MSNs are likely to be especially responsive to cortical input during TAN pauses. To understand the behavioral significance of this phenomenon, it is critical to study the environmental conditions that cause TANs to pause. In fact, it is well established that TANs pause to the delivery of reward and to stimuli that predict the delivery of reward (Apicella, Legallet, & Trouche, 1997; Apicella, Scarnati, & Schultz, 1991; Aosaki et al. 1994b; Kimura, 1992). They also pause to novel stimuli (Blazquez, Fujii, Kojima, & Graybiel, 2002). Another important result is that whereas most MSNs fire to a restricted set of stimuli from a single sensory modality (e.g., Caan, Perrett, & Rolls, 1984), many TANs respond to stimuli from a number of different modalities (Matsumoto, Minamimoto, Graybiel, & Kimura, 2001). Thus, a TAN might respond to the discriminative cue associated with reward, but it is also likely to respond to other visual, auditory, and olfactory cues that occur coincidentally with reward delivery.

The TANs receive their strongest excitatory glutamatergic input from the
caudal intralaminar nuclei of the thalamus (Cornwall & Phillipson, 1988; Sadikot, Parent, & Francois, 1992; Smith, Raju, Pare & Sidibe, 2004), which includes the centremedian (CM) and the parafascicular (Pf) nuclei. The CM-Pf complex receives input from a number of places, including orbitofrontal cortex, pedunculo-pontine tegmental nucleus (PPTN), and the ascending reticular activating system (Van der Werf, Witter, & Groenewegen, 2002) structures that are well known to participate in reward processing and arousal.

The TANs are also prominent targets of substantia nigra dopamine (DA) cells. While the effect of DA on TAN behavior is multifaceted, there are two features of this input that are relevant to the models proposed in this manuscript. First, dopamine cell responses and TAN pauses are temporally coincident under a variety of behavioral conditions (Cragg, 2006; Morris, Arkadir, Nevet, Vaadia, & Bergman, 2004). Second, long-term potentiation (LTP) in TANs requires elevated levels of dopamine (Aosaki, Graybiel, & Kimura, 1994a; Suzuki, Miura, Nishimura, & Aosaki, 2001). These results suggest that TANs may learn to pause to cues that signal reward via reinforcement learning at CM-Pf TAN synapses. In support of this idea, simultaneous single-unit recordings from CM-Pf neurons and TANs show that although an intact CM-Pf response is required for the TANs to pause, the CM-Pf response to environmental cues is relatively constant, regardless of the reward contingencies of the task, whereas the TANs pause primarily to reward-predicting cues (Matsumoto et al., 2001). Since TAN pauses are primarily driven by the CM-Pf complex, it seems reasonable that plasticity at CM-Pf TAN synapses allows the TANs to learn to pause in the presence of cues that predict reward.

The evidence is also good that the learning of procedural skills depends critically on the basal ganglia, and in particular on the striatum. This is true for instrumental conditioning tasks (Barnes, Kubota, Hu, Jin, & Graybiel, 2005; Divac, Rosvold, & Szwarcbart, 1967; Konorski, 1967; ODoherty et al., 2004; Yin, Ostlund, Knowlton & Balleine, 2005) and for II category learning tasks (Ashby & Ennis, 2006; Filoteo, Maddox, Salmon, & Song, 2005; Knowlton, Mangels, & Squire, 1996; Nomura et al., 2007; Waldschmidt & Ashby, 2011).

In the following chapters we use these ideas to formally construct a com-
putational model with the basic architecture shown in Figures 2.1, 2.4 and 3.4. The idea is that, in the absence of CM-Pf input, the TANs high spontaneous firing tonically inhibits the MSN response to cortical input. When cells in the CM-Pf complex fire, DA-dependent reinforcement learning at the CM-Pf – TAN synapse quickly cause the TAN to pause when in a rewarding environment. This releases the MSNs from tonic inhibition, thereby allowing them to respond to cortical inputs and thus to gate learning and behavioral expression at cortical-striatal synapses. As we will see in Chapters 2 and 3, the choice of DA model greatly influences the conditions under which the TANs pause or do not pause. In this way, our model assumes that the TANs and the DA system interact systematically to learn procedural skills, and protect them from unlearning.

1.3 DA and ACh - Synaptic Plasticity in the Striatum

The types of procedural behaviors that we are chiefly concerned with in this manuscript are thought to depend heavily on long term synaptic plasticity at synapses between the cortex and the striatum. The striatum is richly populated with dopamine, cannabinoid, serotonin and acetylcholine signals. All of these neurotransmitters have been implicated to some degree as important factors in cortico-striatal synaptic plasticity (for a review, see Lerner & Kreitzer, in press), yet no existing formally developed learning theory has included them all. This is, at least in part, due to fact that while we are beginning to understand the physiological significance of a broad array of neurotransmitters in cortico-striatal plasticity, we know very little about the behavioral significance of each neurotransmitter. DA and ACh are exceptions to this fact, since much is known about how the sources of DA and ACh in the striatum respond to behaviorally relevant events (Apicella, 2007; Glimcher, 2011). As such the theory developed here assigns key roles in learning and behavior only to dopaminergic inputs from the midbrain, and acetylcholine (as determined by the TANs pattern of activity).

DA has long been implicated in striatal-based learning and behavior and is
known to influence the striatum in two principle ways: 1) DA is known to enhance the signal to noise ratio in cortico-MSN signaling (Ashby & Casale, 2003), and 2) DA is known to directly influence the degree and direction of long term changes in synaptic efficacy (Reynolds & Wickens, 2002; TJArbuthnott, Ingham, & Wickens, 2000; Calabresi, Pisani, Mercuri, & Bernardi, 1996). We only model the second of these effects. The main findings are that DA levels must be elevated above baseline for LTP to occur, they must be depressed below baseline for LTD to occur, and neither LTP nor LTD occur if DA does not deviate from baseline (Reynolds & Wickens 2002; Arbuthnott, Ingham, & Wickens, 2000; Calabresi, JPisani, Mercuri, & Bernardi, 1996).

More is known about the firing properties of DA during the learning and expression of procedural skills than perhaps any other neurotransmitter. The major characteristics of midbrain DA cell firing that we incorporate into our model are as follows: 1) they are known to fire to unexpected rewarding events, 2) they reduce their firing to unexpected aversive events, 3) the magnitudes of the fluctuations to unexpected rewarding or aversive events are well described by models of reward prediction error \(^1\) (RPE; Schultz, Dayan, & Montague, 1997; Glimcher, 2010). The basic idea here is that brain tries to predict the time, valance, and magnitude of rewarding and aversive events and DA cell firing represents the error in this prediction.

The evidence is good that ACh also modulates cortico-striatal LTP and LTD (Bonsi et al., 2008; Centonze, Gubellini, Bernardi, & Calabresi, 1999; Wang et al., 2006). In vitro results seem to suggest that 1) steady-state ACh levels are required for normal cortico-striatal LTP and 2) reduced ACh levels are required

\(^1\)The RPE signal that DA neurons appear to encode is qualitatively similar to the RPE signal that computer science researchers studying reinforcement learning have been using for decades (Sutton & Barto, 1998). Consequently, it is common for reinforcement learning researchers and theoretical neurobiologists alike to treat the basal ganglia as a reinforcement learning module and treat DA as the learning signal in that module. However, the degree of actual biological constraint in these modeling efforts has varied widely. The theory we develop is conceptually resonant with the notion that the basal ganglia is a reinforcement learning module with DA as the learning signal, however, we have placed our DA model (which captures the reinforcement learning appearance of our model) into a biologically realistic neural network. We then allow our DA model to explicitly influence learning at cortico-striatal synapses within this network. Given the multiple layers of our model (network dynamics, DA signal, and learning equations), it is unclear to what degree our model can be generalized as a reinforcement learning module.
for LTD. Any model of basal ganglia function must resolve the apparent paradox implied by this first result. An obvious assumption is that that a TAN pause is associated with reduced ACh levels, so these results seem to imply that corticostriatal LTP can not occur during a TAN pause, only LTD. The paradox is that the environmental conditions that are known to cause TANs to pause (e.g., the appearance of cues that predict reward) are exactly the same as the conditions thought to promote cortico-striatal LTP. For example, in conditioning tasks an animal is rewarded for associating a motor response with a sensory cue. Many such studies have shown that MSNs learn to fire a burst to the presence of the cue (e.g., Carelli, Wolske, & West, 1997; Barnes, Kubota, Hu, Jin, & Graybiel, 2005), and presumably this increase in MSN activation is mediated by LTP at cortico-striatal synapses.

One possibility is that a TAN pause may not cause a simple reduction in striatal ACh levels. The TAN response to sensory cues associated with reward is multi-phasic. Frequently the TAN pause is preceded by an initial burst and also followed by a rebound burst (Apicella, 2007). Thus, ACh levels may fluctuate rapidly during the course of a TAN pause. As a result, an informed model of the role that ACh plays in cortico-striatal synaptic plasticity may require a better understanding of the temporal dynamics of the ACh signal and its physiological effects. Lacking such data, the theory we develop in this manuscript ignores the influence of ACh in the learning equations, and focuses only on the presynaptic inhibitory component of the TANs on MSN excitability. As we will see, this alone is enough to capture the simple data sets that we set our attention to.

1.4 Instrumental Conditioning

A huge part of the conditioning literature has focused on phenomena that indicate that initial learning in conditioning tasks is very resistant to erasure (for review, see Bouton, 2004). There is also good evidence that many conditioning tasks recruit the basal ganglia (Divac, Rosvold, & Szwarcbart, 1967; Konorski, 1967) and the striatum in particular (Barnes, Kubota, Hu, Jin, & Graybiel, 2005; O'Brien, 1974; 1989, 1990, 1992).
et al., 2004; Yin, Ostlund, Knowlton & Balleine, 2005). As such they seem like ideal examples of the procedural skills we are focused on in this manuscript. However the overall conditioning literature is very heterogeneous, and significantly varies along several key dimensions. We must carefully choose among these dimensions in order to find data sets most appropriate to our modeling efforts.

The first important dimension along which conditioning tasks vary has to do with action-outcome contingencies. There are two main classes of task along this dimension, classical conditioning and instrumental conditioning. The participants of a typical instrumental conditioning task must learn to perform a particular behavior (e.g., press a lever) in order to obtain a reinforcer (e.g., to obtain a pellet of food or to avoid an electric shock). This is in contrast to classical conditioning tasks in which the animal passively experiences all stimuli in the experiment (i.e., the experimental events occur independent of the animals behavior). This means that in instrumental conditioning, the outcome of behavioral events is contingent on the participants behavior. In classical conditioning, on the other hand, the outcome of behavioral events is completely non-contingent on the participants behavior. Although classical and instrumental conditioning share many of the same basic results and appear to rely on some of the same basic neural mechanisms, there are important differences from both a psychological and neurobiological perspective (Dayan & Balleine, 2002). Since we are chiefly concerned with procedural skill learning (i.e., learning to perform a certain action to obtain a reinforcer) we focus exclusively on instrumental learning.

The second important dimension along which conditioning tasks vary has to do with the valence and nature of the reinforcer. If the reinforcer is rewarding (e.g., food pellets or drops of juice) then the task is said to be an appetitive conditioning task, and if the reinforcer is aversive (e.g., electric shock) then the task is said to be an aversive conditioning task. There is good reason to suspect that appetitive conditioning and aversive conditioning rely on at least partially distinct neural systems, with the basic idea being that appetitive learning is strongly influenced by dopamine while aversive learning is strongly influenced by serotonin (Fletcher & Korth, 1999; Fletcher, Ming, & Higgins, 1993; Fletcher, Tampakeras, & Yeomans,
Since our theory is chiefly concerned with DA and ACh contributions to learning and behavior we focus exclusively on appetitive learning.

Conditioning tasks also significantly vary with regard to the schedules of reinforcement used throughout the various phases of an experiment. A schedule of reinforcement is simply the process by which the availability of reinforcements are determined. There are far too many types of possible schedules to discuss here, so we limit our discussion to the three main types that are especially relevant to the data we test our theory against in later chapters: 1) Continuous schedules of reinforcement deliver a reinforcer after every response, 2) Variable interval (VI) schedules of reinforcement deliver a reinforcer after the first response after a variable time interval, regardless of the number of intervening responses, and 3) Variable ratio (VR) schedules of reinforcement deliver a reinforcer after a variable number of responses have been made, regardless of the time elapsed between reinforcers. In Chapter 2 we test our models against behavioral data that was collected under VI schedules of reinforcement, yet the models developed here are only capable of imitating VR schedules. There is evidence that the schedule of reinforcement used can influence the behavioral and neurobiological properties of learning and behavior (Yin & Knowlton, 2006), however none of this evidence suggests that the qualitative results of the data we fit would be altered under VR schedules of reinforcement.

There are many classic findings in the conditioning literature that illustrate the robustness of initial learning to erasure (for review, see Bouton, 2004). Of these, we have chosen two (fast reacquisition and renewal) for which there exist clear appetitive instrumental data sets for us to model. The typical fast reacquisition experiment has three phases: an acquisition phase, an extinction phase and a reacquisition phase. Participants are trained to perform an instrumental behavior in order to obtain a reinforcer according to some schedule of reinforcement during the acquisition phase. The performance of the instrumental behavior is then abol-

---

2This is a consequence of the discrete-trial structure we have used to program our models (i.e., we do not include inter-trial-intervals in the model)

3although not nearly as many that are specifically appetitive and instrumental in design
ished (or at least greatly reduced) by either removing reinforcements altogether or by greatly reducing the frequency of reinforcements during the extinction phase. Finally, the reinforcements are reinstated to a much greater degree (although not necessarily as great as the acquisition phase). The result is that the performance of the instrumental behavior reappears during the reacquisition phase more quickly than it originally appeared during the acquisition phase.

The typical renewal experiment also has three phases that are highly similar to the phases in a fast reacquisition experiment. There are two key differences: 1) The three phases occur in potentially different environments (or contexts), and 2) reinforcements are not reinstated during the third phase (i.e., the third phase is another extinction phase, but is often referred to as the renewal phase). There are three types of renewal that illustrate the robustness of initial learning against erasure: ABA, AAB, and ABC (The letters refer to training environments so that in ABA renewal, for example, the animal is trained in environment A, extinguished in environment B, and finally returned to environment A for the renewal phase). The finding in all of these types of renewal is that the transfer to the second extinction environment is characterized by a brief renewal of instrumental responding.

The conditioning literature has naturally focused on how the strength of the association between the instrumental behavior and the reward varies during these different conditions. However, fast reacquisition and renewal are difficult results to reconcile under this theoretical framework. This is because any model that assumes that the instrumental behavior is purely a result of the strength of stimulus-response associations can only account for extinction performance by erasing the initial learning of these associations. One more recent and influential account of extinction is that removing reinforcements causes new learning rather than unlearning (e.g., Bouton, 2002, 2004). The idea is that during original acquisition the animal learns the behavior, and also the context under which the behavior is appropriate (i.e., reinforced). During extinction the animal then learns about a context in which the behavior is inappropriate (i.e., not reinforced). So according to this account, extinction is not primarily an unlearning phenomenon, but rather an example of context learning. The theory proposed in this article is
conceptually similar to this context learning account of extinction. In fact, our theory could be interpreted as a neurobiological instantiation of this context learning account.

### 1.5 II Category Learning

II categorization tasks are those where accuracy is maximized only if information from two or more non-commensurable stimulus dimensions is integrated at some pre-decisional stage (Ashby & Gott, 1988). Perceptual integration could take many forms—from computing a weighted linear combination of the dimensional values to treating the stimulus as a Gestalt. Typically, the optimal strategy in II tasks is difficult or impossible to describe verbally (which makes it difficult to discover via logical reasoning). An example of an II task is shown in Figure 1.1. In this case the four categories are each composed of single black lines that vary in length and orientation. The diagonal lines denote the category boundaries. Note that no simple verbal rule correctly separates the lines into the four categories. Nevertheless, many studies have shown that with enough practice, people reliably learn such categories (e.g., Ashby & Maddox, 2005).

II categorization tasks are often contrasted with rule-based tasks, in which the categories can be learned via some explicit reasoning process. Frequently, the rule that maximizes accuracy (i.e., the optimal strategy) is easy to describe verbally. In the most common applications, there are two contrasting categories, only one stimulus dimension is relevant, and the subjects task is to discover this relevant dimension and then to map the different dimensional values to the relevant categories. However, there is no requirement that rule-based tasks be one-dimensional. For example, a conjunction rule (e.g., respond A if the stimulus is small on dimension x and small on dimension y) is a rule-based task because it can be discovered through logical reasoning (e.g., a conjunction is easy to describe verbally). Many real-world skills require II categorization to master. For example, radiologists can make initial progress in categorizing x-rays as according to whether or not they display a tumor by receiving explicit instruction (e.g., via book or lecture), but
Figure 1.1: A few examples of stimuli that might be used in an information-integration category learning task.
expertise in this skill requires years of hands-on training (e.g., during a residency).

Many studies have documented a wide variety of qualitative differences in how rule-based and II tasks are initially learned. For example, delaying feedback by a few seconds (Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005), switching the location of the response keys (Ashby, Ell, & Waldron, 2003; Maddox, Bohil, & Ing, 2004), or informing participants of the category label before the stimulus rather than after the response (Ashby, Maddox, & Bohil, 2002) all interfere with performance in II tasks much more than in rule-based tasks. In contrast, adding a secondary (dual) task (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006) or reducing the time available to process the feedback (Maddox, Ashby, Ing, & Pickering, 2004) interferes with performance in rule-based tasks much more than in II tasks. These differences are all consistent with the hypothesis that learning in rule-based tasks is mediated by declarative memory systems, whereas learning in II tasks is mediated by procedural memory (Ashby & O’Brien, 2005). The models we develop here focus exclusively on II category learning tasks.

1.6 Aims and Scope of the Model

Our primary goal is to develop a model of basal ganglia function that is accurately constrained by known neuroanatomy and neurophysiology, and simultaneously accounts for key behavioral results from instrumental conditioning and II category learning. Specifically we are interested in modeling how the striatum learns procedural skills, and how these skills are protected when action-outcome contingencies are changed. We do not intend to capture every known biological detail about striatal and basal ganglia function. Similarly, we do not hope to generate a model that can account for every known behavioral result in instrumental conditioning or category learning. We are interested specifically in developing a model that exists in the middle of these two extremes. We will limit our attention to a few key behavioral results, and use only the biological constraints needed to account for these results.
Chapter 2

Instrumental Conditioning
Applications

2.1 Reacquisition

2.1.1 Introduction

Woods and Bouton (2007) performed an instrumental conditioning experiment on rats that illustrated a clear effect of the schedule of reinforcement used during extinction and reacquisition on the rate of reacquisition performance recovery. Their design was composed of four conditions called Ext2, Ext8, Prf2, and Prf8. The participants in every condition received a VI-30 second schedule of reinforcement during the acquisition phase. Reinforcements were completely removed during the extinction phase for the Ext2 and Ext8 groups, and they were greatly reduced (to about a VI-16 minute schedule) for the Prf2 and Prf8 groups \(^1\). The groups were farther divided during the reacquisition phase. Groups Ext2 and Prf2 received a VI-2 minute schedule of reinforcement and groups Ext8 and Prf8 received a VI-8 minute schedule of reinforcement.

It is difficult to assess whether any condition exhibited true fast or slow reacquisition (relative to acquisition) because Woods and Bouton (2007) only reported

\(^1\)Actually, the schedule of reinforcement during the extinction phase began at VI-4 and was gradually reduced until VI-32 minutes was reached
the level of performance during the last block of acquisition and only collected data from two reacquisition sessions. There are, however, several key qualitative results apparent in the data of Woods and Bouton (2007) that we have attempted to capture: 1) All groups showed approximately equal performance during the Acquisition phase, 2) The Ext2 and Ext8 groups extinguished to the same level and the Prf2 and Prf8 groups extinguished to the same level, 3) the Ext2 and Ext8 groups extinguished to a lower asymptotic level of responding than the Prf2 and Prf8 groups, 4) the Ext2 and Prf2 groups reacquired more quickly than the Ext8 and Prf8 groups, and within that division the Ext groups reacquired faster than the Prf groups.

2.1.2 Methods

Model Architecture

The architecture and activation equations of the model we use to fit the Woods and Bouton (2007) data is identical to the model defined by Ashby and Crossley (2011). For convenience, we describe them here in complete detail. The architecture of this model is shown in Figure 2. Every trial was initiated by activity in the sensory association cortex, which is then propagated through the direct pathway of the basal ganglia to the ventral anterior (VA) and ventral lateral (VL) nuclei of the thalamus, which in turn projects to presupplementary motor cortex (pre-SMA) and supplementary motor cortex (SMA). The model also includes the Pf thalamic nucleus which projects to the TAN unit in the striatum. We modeled each of these areas with a single unit in the model (each unit is modeled as a single spiking neuron), but is intended to reflect the average activity of a population of neurons within each area. Finally, the model includes DA inputs to the Pf-TAN synapse and the cortico-striatal synapse.

Activation Equations

The activation of all sensory cortical units and CM-Pf units was either off (with activation 0) or on (see table A.1 for amplitude values) during the duration
Figure 2.1: Model Architecture used to fit Woods and Bouton (2007).

of stimulus presentation (note that we did not specifically model sensory inputs to the CM-Pf). Our model of changes in the membrane potential of striatal MSNs was adapted from a model proposed by Izhikevich (2007). The model includes two coupled differential equations for the medium spiny unit. The first equation models fast changes in membrane potential (measured in mV), and the second equation models slow changes in the activation and inactivation of various intracellular ion channels (e.g., Na+ and K+). We supplement the Izhikevich (2007) model by assuming that the key inputs to the medium spiny cells include 1) excitatory inputs from sensory cortex, and 2) presynaptic inhibitory input from the TAN. Specifically, our complete medium spiny unit model assumes that the membrane potential in the striatal unit at time $t$, denoted $S(t)$, is determined by:

$$
50 \frac{dS(t)}{dt} = w(n)[I(t) - \beta_sT(t)] + [S(t) + 80][S(t) + 25] + E_s - u_s(t) + \sigma_s \epsilon(t) \quad (2.1)
$$
100 \frac{d u_s(t)}{dt} = -20[S_j(t) + 80] - u_s(t) \tag{2.2}

where $\beta_s$, $E_s$, and $\sigma_s$ are constants, $w(n)$ is the strength of the synapse between the sensory cortical unit and the striatal unit on trial $n$, $T(t)$ in the membrane potential of the TAN at time $t$, and $\epsilon(t)$ is white noise. The third term is the quadratic integrate-and-fire model (Ermentrout, 1996). To produce spikes, when $S(t) = 40$ mV then $S(t)$ is reset to $S(t) = -55$ mV. The last term models noise. Equation 2 models the slow changes in various intracellular ion channels. When Equation 1 produces a spike (i.e., when $S(t) = 40$ mV), $u_S(t)$ is reset to $u_S(t) + 150$. All specific numerical values in Equations 1 and 2 and the numerical values used in the resetting procedures are taken from Izhikevich (2007). The model described by Equations 1 and 2 accurately accounts for patch-clamp data collected from MSNs in the rat (i.e., see Fig. 8.37, Izhikevich, 2007), in the sense that it displays both the up and down states that characterize MSN firing patterns, and it displays realistic spiking behavior.

The TANs are more challenging to model, because of their unusual dynamics. For example, when excitatory input is delivered to the TANs, they fire an initial burst and then pause (Kimura, Rajkowski, & Evarts, 1984; Reynolds, Hyland, & Wickens, 2004). We developed a model of TAN firing that displays these same qualitative properties by modifying the Izhikevich (2003) model of intrinsically bursting cortical neurons. Specifically, we assumed that changes in the TAN membrane potential at time $t$, denoted $T(t)$, are described by the following two coupled equations.

\begin{align*}
100 \frac{dT(t)}{dt} &= v(n)Pf(t) + 1.2[T(t) + 75][T(t) + 45] + 950 - u_T(t) \tag{2.3} \\
100 \frac{du_T(t)}{dt} &= 5[T(t) + 75] - u_T(t) + 2.7v(n)R(t) \tag{2.4}
\end{align*}

where $v(n)$ is the strength of the synapse between the Pf and the TAN on trial $n$, and $Pf(t)$ is the input from the CM-Pf at time $t$. The constant 950 models spontaneous firing, and the function $R(t) = Pf(t)$ up to the time when CM-Pf
activation turns off, then $R(t)$ decays exponentially back to zero (with rate .0018). To produce spikes, when $T(t) = 40$ mV then $T(t)$ is reset to $T(t) = -55$ mV and $u_T(t)$ is reset to $u_T(t) + 150$. The dynamical behavior of this model that allows it to mimic the unusual firing properties of TANs is described in Ashby & Crossley (2011).

Note that we modeled the effects of CM-Pf activation as purely excitatory. In fact, the evidence is good that glutamate inputs from the CM-Pf also synapse on GABAergic interneurons in the striatum, which then synapse on TANs. As a result, CM-Pf activation can also induce an inhibitory input to the TANs (Suzuki et al., 2001; Zackheim & Abercrombie, 2005). We chose not to model this indirect inhibitory effect because TANs pause when positive current is injected directly into the cell (Reynolds et al., 2004). Thus, whereas the inhibitory input may potentiate the TAN pause, it is apparently not necessary to induce the pause.

For all other units in the model, we excluded the slow regulatory term $u(t)$, and instead modeled membrane potential with the standard quadratic integrate-and-fire model. For example, activation in the globus pallidus at time $t$, denoted by $G(t)$, is described by:

$$15 \frac{dG(t)}{dt} = -\alpha_G f[S(t)] + 71 + 0.7[G(t) + 60][G(t) + 40]$$ (2.5)

where $\alpha_G$ is a constant. The first term models the inhibitory input from the striatum, the second term ensures a high tonic firing rate, and the last term is the quadratic integrate-and-fire component that is the same as in Equations 1 and 4. Spikes are produced after $G(t) = 35$ by resetting to $G(t) = -50$. The function $f[S(t)]$ in Equation 5 is called the alpha function and is a standard method for modeling the postsynaptic effects of a spike in a presynaptic cell (e.g., Rall, 1967). When a presynaptic cell generates an action potential, synaptic vesicles open, neurotransmitter is released, it diffuses across the synapse, and binds to postsynaptic receptors, which initiates events that eventually effect the membrane potential of the postsynaptic cell. The alpha function models the time course of these effects. The idea is that every time the presynaptic cell spikes, the following input is delivered to the postsynaptic cell:
\[ \alpha(t) = \frac{t}{\lambda e^{\lambda t}} \]  

(2.6)

This function has a maximum value of 1.0 and it decays to .01 at \( t = 7.64\lambda \).

Activation in the thalamus at time \( t \) is given by

\[
\frac{dV(t)}{dt} = -\beta f[G(t)] + 71 + 0.7[V(t) + 60][V(t) + 40]
\]  

(2.7)

where \( \beta_T \) is a constant, and \( f[G(t)] \) is the alpha function from Equations 1 and 6. Spikes are produced after \( V(t) = 35 \) by resetting to \( V(t) = -50 \). The first term models the inhibitory input from the globus pallidus. The ventral anterior and ventral lateral thalamic nuclei of the thalamus also receive a variety of excitatory inputs (e.g., cerebellum, PFC). We modeled these via the constant 71. For our purposes the most important of these excitatory inputs may be from PFC (e.g., Anderson & DeVito, 1987). The PFC input is critical because it is thought that striatal firing, by itself, does not trigger a motor response. When the striatum fires, it disinhibits the thalamus (i.e., by reducing pallidal inhibition), but it does not excite the thalamus. For this reason, random sensory stimuli that are encountered as one moves through the world could cause the striatum to fire, but this firing will typically not elicit an unintended motor response. In a skill-learning task, instructions from an experimenter about how to respond could cause the cortical input to thalamus to increase, thereby priming the relevant response goals. Because of the tonic inhibition from the globus pallidus however, this cortical input is not enough to trigger a response. Instead, the striatum must first inhibit the globus pallidus, an event that would allow the thalamus to trigger one of the primed motor response goals.

Activation in the premotor unit at time \( t \), denoted by \( C(t) \), is given by

\[
\frac{dC(t)}{dt} = \beta_C f[V(t)] + 69 + 0.7[C(t) + 60][C(t) + 40] + \sigma_C \epsilon(t)
\]

(2.8)

where \( \beta_C \), and \( \sigma_C \) are constants, and \( \epsilon(t) \) is white noise. As in other units, spikes are produced after \( C(t) = 35 \) by resetting to \( C(t) = -50 \).
Learning Equations

Following standard models, we assume that synaptic plasticity at the cortical-striatal synapse and at the CM-Pf TAN synapse is modified according to reinforcement learning that requires three factors: 1) strong pre-synaptic activation, 2) post-synaptic activation that is strong enough to activate NMDA receptors, and 3) dopamine levels above baseline (e.g., Arbuthnott, Ingham, & Wickens, 2000; Calabresi, Pisani, Mercuri, & Bernardi, 1996; Reynolds & Wickens, 2002). If post-synaptic activation is strong, but dopamine is below baseline then the synapse is weakened. The synapse is also weakened if there is post-synaptic activation that is below the NMDA threshold (regardless of dopamine levels). We model this by including a second lower threshold that roughly corresponds to activation of AMPA receptors (a low-threshold glutamate receptor).

Let $w(n)$ denote the strength of the synapse on trial $n$ between the cortical unit and the striatal unit. We model reinforcement learning as follows:

$$w(n + 1) = w(n) + \alpha_w \int I(t)dt \left[ \int [S(t)]^+ dt - \theta_{NMDA} \right]^+ [D(n) - D_{base}]^+[w_{max} - w(n)]$$

$$- \beta_w \int I(t)dt \left[ \int [S(t)]^+ dt - \theta_{NMDA} \right]^+ [D_{base} - D(n)]^+w(n)$$

$$- \gamma_w \int I(t)dt \left[ \theta_{NMDA} - \int [S(t)]^+ dt \right]^+- \theta_{AMPA} \right]^+ w(n)$$

(2.9)

The function $[g(t)]^+ = g(t)$ if $g(t) > 0$, and otherwise $g(t) = 0$. The integrals in Equation 10 are all over the time of stimulus presentation. Thus, $\int [S(t)]^+ dt$ is the total positive medium spiny activation during stimulus presentation. $D_{base}$ is the baseline dopamine level, $D(n)$ is the amount of dopamine released following feedback on trial $n$, and $\alpha_w, \beta_w, \gamma_w, \theta_{NMDA}, \theta_{AMPA}$ are all constants. The first line describes the conditions under which LTP occurs (striatal activation above the threshold for NMDA receptor activation and dopamine above baseline) and lines two and three describe conditions that produce LTD. The first possibility (line 2) is that post-synaptic activation is above the NMDA threshold
but dopamine is below baseline (as on an error trial), and the second possibility is
that striatal activation is between the AMPA and NMDA thresholds. Note that
synaptic strength does not change if post-synaptic activation is below the AMPA
threshold.

We assume that learning at cortical-MSN synapses and at Pf-TAN synapses
are both mediated by this same model. We allow the learning rates to differ at
these two synapse types, but we assume the same numerical values for $\theta_{NMDA}$ and
$\theta_{AMPA}$. The numerical values for all parameters are given in the Appendix (i.e.,
see Table A.1).

**Dopamine Model**

The Equation 9 model of reinforcement learning requires that we specify the
amount of dopamine released on every trial in response to the feedback signal [the
$D(n)$ term]. The more that the dopamine level increases above baseline ($D_{\text{base}}$),
the greater the increase in synaptic strength, and the more it falls below baseline,
the greater the decrease.

Although there are a number of powerful models of dopamine release, Equation 8 requires only that we specify the amount of dopamine released to the feed-
back signal on each trial. The key empirical results are (e.g., Schultz, Dayan, &
Montague, 1997; Tobler, Dickinson, & Schultz, 2003): 1) midbrain dopamine cells
fire tonically, 2) dopamine release increases above baseline following unexpected re-
ward, and the more unexpected the reward the greater the release, and 3) dopamine
release decreases below baseline following unexpected absence of reward, and the
more unexpected the absence, the greater the decrease. One common interpreta-
tion of these results is that over a wide range, dopamine firing is proportional to
the reward prediction error ($RPE$):

$$RPE = ObtainedReward - PredictedReward$$

A simple model of dopamine release can be built by specifying how to
compute $ObtainedReward$, $PredictedReward$, and exactly how the amount of
dopamine release is related to the RPE. Our solution to these three problems is as follows.

**Computing Obtained Reward** We define the obtained reward $R_n$ on trial $n$ as $+1$ for reinforced responses, and 0 otherwise.

**Computing Predicted Reward** We compute $PredictedReward$ on trial $n$, which we denote as $P_n$ by,

$$P_{n+1} = P_n + \alpha (R_n - P_n) \quad (2.11)$$

It is well known that when computed in this fashion, $P(n)$ converges exponentially to the expected reward value and then fluctuates around this value until reward contingencies change.

**Computing Dopamine Release from the RPE** We assume that the amount of dopamine release is related to the RPE in the manner reported by Bayer and Glimcher (2005). Specifically, we assumed that

$$DA(n) = \begin{cases} 1 & \text{if } RPE > 1 \\ 0.8RPE + 0.2 & \text{if } -0.25 < RPE \leq 1 \\ 0 & \text{if } RPE < 0.25 \end{cases} \quad (2.12)$$

Note that the baseline dopamine level is .2 (i.e., when the $RPE = 0$) and that dopamine levels increase linearly with the $RPE$. However, note also the asymmetry between dopamine increases and decreases. As is evident in the Bayer and Glimcher (2005) data, a negative RPE quickly causes dopamine levels to fall to zero, whereas there is a considerable range for dopamine levels to increase in response to positive RPEs.
Simulation Details

The specific parameter values used for the simulations we report here are summarized in table A.1. Each simulation of the experiment was composed of 300 trials of acquisition, 300 trials of extinction, and 300 trials of reacquisition, in that order. During acquisition, the models responses were rewarded with probability 0.25 for all groups. Rewards were completely removed during extinction for groups Ext2 and Ext8. The models responses in the Prf2 and Prf8 groups were rewarded with probability 0.14 for the first 10 trials of extinction, probability 0.12 for the following 10 trials, probability 0.094 for the following 10, and then 0.047 for the remaining extinction trials. The models responses during reacquisition were rewarded with probability 0.19 in the Ext-2 and Prf-2 groups, and they were rewarded with probability 0.047 in the Ext-8 and Prf-8 groups. The main parameters that were manipulated to achieve our results were the learning rates (i.e., $\alpha_w$ and $\beta_w$ from equation 2.9, for both the Pf unit and the MSN unit).

2.1.3 results

The data used to generate the results shown in Figure 2.2 and Figure 2.3 were generated by simulating the model 100 times and averaging the results across simulations. Figure 2 was processed in such a way as to mimic the style of the results presented in Woods and Bouton (2007) as closely as possible. Specifically, we split the 900 trials of the averaged simulation into 18 blocks of 50 trials each. Next, we transformed each proportion of responses emitted value contained in our 18 blocks into a percentage of responses emitted during the last block of acquisition. Finally, we only display the extinction blocks and the first two reacquisition blocks.

Note that the model correctly captured almost all of the major qualitative features in the Woods and Bouton (2007) data. The main difference is that the difference between the Ext8 and Prf8 groups in reacquisition didnt emerge until the second reacquisition block in the model, but was present from the first reacquisition block in the data of Woods of Bouton (2007).

Figure 3 shows the underlying mechanics of the models behavior. The top left panel shows the average DA release, the bottom left panel shows the average
Figure 2.2: Behavioral results from Woods and Bouton (2007) Experiment 1. Right: Behavioral results obtained from the model. See text for more details.

proportion of responses emitted, the top right panel shows the average synaptic strength of the CTX-MSN synapse, and the bottom right panel shows the average synaptic strength of the Pf-TAN synapse. We make extensive use of this figure in the following discussion.

Acquisition

The mechanics during the acquisition phase (trials 1 through 300) were identical for all conditions. Panel A of Figure 2.3 shows that the average DA signal fluctuated above baseline for the duration of the acquisition phase. The DA signal reached it’s peak by about trial 100, and then began to decrease as reinforced responses became more expected. Panel C shows that trial 100 also marks a decrease in the increase of the proportion of responses emitted (i.e., the proportion of responses emitted begins to asymptote around trial 100). Panels B and D show the synaptic strengths of the CTX-MSN and Pf-TAN synapses, respectively. Note that the Pf-TAN synaptic strength increases quickly, while the CTX-MSN synapse increases more gradually. The basic idea is that the TANs must learn to pause (i.e., Pf-TAN synapse must be high) before cortical activity can get through to the MSN unit.
Figure 2.3: Top left: Average DA release. Bottom left: Average proportion responses emitted. Top right: Average synaptic strength of the CTX-MSN synapse. Bottom right: Average synaptic strength of the Pf-TAN synapse.
Extinction

The mechanics during the extinction phase (trials 301 through 600) differed between the Ext groups and the Prf groups. Specifically, the Pf-TAN synaptic weight was not reduced during the Extinction phase to the same degree in the Prf groups as it was in the Ext groups. This is because the Ext groups never received a reinforced response during extinction. As panel A shows, this caused the DA signal to depress below baseline maximally, thereby inducing maximal unlearning at the Pf-TAN synapse. The Prf groups, however, received occasional reinforced responses during the extinction phase. This reduced the magnitude of the DA depression (panel A), thereby reducing the amount of unlearning experienced at Pf-TAN synapses (panel B). The net effect of this differential unlearning at the Pf-TAN synapse between conditions was that the TANs pause response was not abolished as completely in the Prf groups as it was in the Ext groups. Consequently, the CTX-MSN synapse was less protected by the TANs in the Prf groups, and was therefore subject to more unlearning (panel D). The fact that the TANs pause response was not completely abolished in the Prf groups is apparent in the observation that the Prf groups respond more frequently during extinction than the Ext groups (panel C). This is because the TANs don’t completely inhibit cortico-striatal transmission in the Prf groups.

Reacquisition

The mechanics during the reacquisition phase (trials 601 through 900) differed between groups Ext2 and Prf2 and groups Ext2 and Prf8. The key to understanding these mechanics is to remember that the Prf groups have experienced more LTD at CTX-MSN synapses than the Ext groups (panel B). This means that they will, in general, have to grow more during reacquisition in order for responding to reappear. This makes reacquisition in the Prf2 group slower than in the Ext2 group and reacquisition in the Prf8 group slower than in the Ext8 group. Now, recall that the Ext2 and Prf2 groups received reinforced instrumental responses with probability 0.19, while the Ext8 and Prf8 received reinforced instrumental responses with probability 0.047 groups. This caused the Pf-TAN synapse to grow,
and consequently the TAN unit to start pausing again, more rapidly in the Ext2 and Prf2 groups than it did in the Ext8 and Prf8 groups. Consequently the Ext2 and Prf2 groups experienced LTP more often than the Ext8 and Prf8 groups, and therefore reacquired more quickly.

2.1.4 Discussion

Woods and Bouton (2007) focus their discussion on the idea that the schedule of reinforcement partly define the conditioning context. One way to incorporate this idea is to treat the schedule of reinforcement as a contextual cue (e.g., a Pf unit in our model). In this view, the acquisition, extinction, and reacquisition phases would be more distinct from one another in the Ext experimental groups than they would be in the Prf experimental groups. The idea would then be that the increased distinctiveness of the training phases (or training contexts) would increase the renewal effect at the beginning of reacquisition. However, it is somewhat awkward to include the schedule of reinforcement as a contextual cue because the schedule of reinforcement must be estimated from the outcome of many instrumental responses. Other contextual cues more traditionally used in renewal paradigms are static features of the environment and require only perception to recognize. In any case, since Woods and Bouton (2007) offer no precise description of how a context is defined, or how the similarity of differing contexts influences renewal, so the viability of this explanation is unclear.

Woods and Bouton (2007) also discuss the trial-signaling view (Bouton et al., 2004), which is conceptually resonant with Capaldi’s sequential view (Capaldi, 1967). The basic idea behind both of these views (in the case of instrumental conditioning) is that the outcome of a particular action comes to predict more of those outcomes. In this way, partial reinforcement delivered during extinction in the Prf conditions allows for more generalization from extinction to reacquisition. This would basically make extinction less distinct from reacquisition, thereby slowing reacquisition. As Woods and Bouton (2007) point out, this is very similar to the explanation of the partial reinforcement extinction effect (PREE) provided by Capaldi’s sequential theory, where the idea is that a partial reinforcement schedule
used during acquisition creates more generalization from acquisition to extinction, thereby slowing the loss of responding during extinction.

Our account of Woods and Bouton (2007) does not fit neatly with any of these explanations. We do not explicitly include the schedule of reinforcement as a contextual cue (that is to say it is not explicitly represented as a Pf unit). However, the Pf-TAN synaptic weight is effected by the schedule of reinforcement (because all learning in the model is DA dependent and DA depends on the schedule used). There is also no clear analog in the model for the concept of generalization. The model accounts for the Woods and Bouton (2007) data because the partial reinforcement used during extinction in the Prf groups prevented the TANs from unlearning their pause response as thoroughly as they did in the Ext groups. This left the instrumental response (which was learned at cortico-MSN synapses) subject to more unlearning, and therefore slowed reacquisition in the Prf groups relative to the Ext groups.

2.2 Renewal

2.2.1 Introduction

Bouton et al. (2011) performed a set of experiments that demonstrated ABA, AAB, and ABC renewal in appetitive instrumental conditioning (see section 1.4 for a description of basic renewal experimental designs). The participants in every condition (ABA, AAB, and ABC) received a VI-30 second schedule of reinforcement during the acquisition phase and reinforcements were completely removed during both extinction phases. The only difference between conditions was the context (i.e., conditioning chamber) in which each phase occurred. In the ABA group acquisition occurred in context A, extinction occurred in context B, and renewal occurred in context A. In the AAB group acquisition and extinction occurred in context A, and renewal occurred in context B. In the ABC group acquisition occurred in context A, extinction occurred in context B, and renewal occurred in context C.

Bouton et al. (2011) was characterized by several key qualitative attributes:
1) Performance during acquisition was equal across groups, 2) ABA and ABC groups extinguished more quickly than the AAB group, 3) Group ABA exhibited the strongest renewal effect and groups AAB and ABC exhibited considerably smaller renewal effects, and 4) Although the magnitude of the renewal effect (i.e., the difference in mean responding between the extinction context and the renewal context) was similar in groups AAB and ABC, group AAB showed slightly higher responding in both the extinction context and the renewal context.

2.2.2 Methods

Model Architecture

Figure 2.1 shows the architecture of the model that we use to fit the data of Bouton et al. (2011). Note that this model is nearly identical to the model used to fit the data of Woods and Bouton (2007) in the previous section, with the only difference being that the present model included 36 Pf units (as opposed to the single Pf unit used in the Woods and Bouton (2007) application). 24 of these units were associated with a particular context (A, B, or C), and the remaining 12 were associated with all three contexts (we refer to these Pf units as the “overlap” Pf units). The idea was that Pf units respond uniquely to certain features in the environment, and that different training environments are made up of mostly unique features, but that there is some overlap in features as well. Every Pf unit projected to the TAN, and every Pf-TAN synapse was plastic and governed by equation 2.9.

Activation Equations

The activation of all Pf units is context-dependent. Pf units in their associated context are either off (with activation 0) or on (with activation 0.05) during the duration of stimulus presentation. Pf units that are not associated with the present context are always off (with activation 0). Formally, the activation of the \( j \)th Pf unit in the \( i \)th context denoted by \( Pf_{i,j} \) was given by,
Figure 2.4: Model Architecture used to fit Bouton et al. (2011)

\[ Pf_{i,j} = \begin{cases} 
0.05 & \text{if } i \text{ is the active context} \\
0.05 & \text{if } j \text{ is an overlap cell} \\
0 & \text{otherwise} 
\end{cases} \quad (2.13) \]

The addition of many Pf units to the model means that we must redefine how we incorporate the influence of the Pf on the TAN unit. The activation of the TAN unit in this application was computed by the following coupled differential equations,

\[
100 \frac{dT(t)}{dt} = \sum_{i=1}^{3} \sum_{j=1}^{12} v_{i,j}(n) Pf_{i,j}(t) + 1.2[T(t)+75][T(t)+45] + 950 - u_T(t) \quad (2.14)
\]

\[
100 \frac{du_T(t)}{dt} = 5[T(t)+75] - u_T(t) + 2.7R(t) \sum_{i=1}^{3} \sum_{j=1}^{12} v_{i,j} Pf_{i,j}(t) \quad (2.15)
\]

where \( v_{i,j} \) is the synaptic strength of the synapse between the TAN and the \( j \text{th} \) Pf unit associated with context \( i \). \( R(t) \) was defined the same way as in the
previous section. All other activation equations were the same as those described in the previous section.

**Learning Equations**

The learning equations we use in this application are identical to those described by equation 9. All 36 Pf units and the CTX-MSN synapse are subject to learning. For clarity, we formally specify the learning equation used to update the Pf-TAN synapses. Let \( v_{i,j}(n) \) denote the strength of the synapse between the Pf unit and the TAN unit on trial \( n \). Then,

\[
v_{i,j}(n+1) = v_{i,j}(n) + \alpha_v \int P_{f_{i,j}}(t) dt \left[ \int [T(t)]^+ dt - \theta_{NMDA} \right]^+ [D(n) - D_{base}]^+ [v_{max} - v_{i,j}(n)]
- \beta_v \int P_{f_{i,j}}(t) dt \left[ \int [T(t)]^+ dt - \theta_{NMDA} \right]^+ [D_{base} - D(n)]^+ v_{i,j}(n)
- \gamma_v \int P_{f_{i,j}}(t) dt \left[ \theta_{NMDA} - \int [T(t)]^+ dt \right]^+ \theta_{AMPA}^+ v_{i,j}(n)
\]

(2.16)

**Dopamine Model**

The dopamine model was identical to that described in the previous section.

**Simulation Details**

The specific parameter values used for the simulations we report here are summarized in table A.1. Each simulation of the experiment was composed of 300 trials of acquisition, 300 trials of extinction, and 300 trials of reacquisition. During acquisition, the models responses were rewarded with probability 0.25 for all groups. During extinction and renewal, the model was never rewarded. The main parameters that were manipulated to achieve our results were the learning rates (i.e., \( \alpha_v \) and \( \beta_v \) from equation 2.16, for both the Pf unit and the MSN unit).
2.2.3 results

The results shown in Figures 2.5 through 2.8 were generated by simulating the model as just described 100 times and average the results across simulations. Figure 2.5 was processed in such a way as to mimic the style of the results presented in Bouton et al. (2011) as closely as possible. Specifically, we split the 900 trials of the averaged simulation into 18 blocks of 50 trials each. Finally, in the right most section of panel C, we computed "extinction" performance by taking the average of the last two extinction blocks and we computed "renewal" performance by taking the average of the first two renewal blocks.

Note that the model captured nearly all ordinal qualitative results of Bouton et al. (2011). The model results are: 1) the model generated equal acquisition performance across all groups, 2) AAB and ABC groups extinguished more quickly than the ABA group, and 3) Group ABA exhibited the strongest renewal effect and groups AAB and ABC exhibited considerably smaller renewal effects. The main apparent difference between the modeling results and the data of Bouton et al. (2011) is seen by comparing the AAB and ABC renewal effects. Where the data of Bouton et al. (2011) show slightly higher responding in the AAB group as compared to the ABC group in both the extinction and the renewal contexts, the model showed essentially identical performance between these two groups.

There are two main criteria that must be met in order for the model to respond: 1) there must be adequate learning at the CTX-MSN synapse to drive activity through the basal ganglia and thalamus and up to the motor response unit, and 2) the TANs must pause in order for cortical input to get to the MSNs at all. In the model, the magnitude of the TANs pause is a function of the net Pf-TAN synaptic strength. The specific mechanism behind AAB, ABA, and ABC renewal are all different, however, the key in all three is the dynamic of the net Pf-TAN synaptic strength.

ABA Renewal

Figure 2.6 shows the model mechanics underlying ABA renewal. Since the acquisition phase (trials 1 to 300) occurs in context A, only the Pf units sensitive to
Figure 2.5: Top and Middle panels: Behavioral results from Bouton et al. (2011). Bottom: Behavioral results obtained from the model. See text for more details.
features present in context A are active. This means that only the Pf-TAN synapses that correspond to these Pf units experience any synaptic plasticity, and only these units contribute to the learning and expression of the instrumental response. This is apparent in panel B, which shows the average synaptic strengths of the Pf-TAN synapses and the CTX-MSN synapse. Note that the context A Pf-TAN synapses grow considerably higher than the other Pf-TAN synapses. The only context A and context B Pf-TAN synapses that grow are the overlap Pf units. By the end of the Acquisition phase the context A Pf-TAN synapses have grown large enough for a robust TAN pause and the CTX-MSN synapse has grown enough to consistently drive activity from sensory cortex through the basal ganglia and thalamus and up to the response cortical unit.

Since the extinction phase (trials 301 to 600) occurs in context B, only the Pf units sensitive to features present in context B are active. Note that the only learning these synapses experienced during the acquisition phase occurred at Pf-TAN synapses that correspond to the overlap Pf units (as is represented in panel B by the context B and context C Pf-TAN weights). This small amount of learning was enough to facilitate a TAN pause for a short while during extinction, but it doesn’t take much unlearning at these Pf-TAN synapses to quickly extinguish the TAN pause and thereby extinguish instrumental responding. Note that this is why extinction proceeds more quickly in the ABA and ABC designs then it does in the AAB design (i.e., the switch to a novel context during the extinction phase instantly reduces the magnitude of the TAN pause). This is apparent in panel D, which shows the net effective Pf-TAN synaptic weight, which is an average of all Pf-TAN synapses weighted according to the active synapse. The net effective Pf-TAN synaptic strength determines whether or not the TANs pause.

Since the renewal phase (trials 601 to 900) occurs in the original training context (i.e, context A), the Pf-TAN synapses driving the TANs are immediately switched back to those sensitive to context A. Note that the only unlearning these synapses experienced during the extinction phase again occurred at the Pf-TAN synapses that correspond to the overlap Pf units. This means that the learning at the remaining context A Pf-TAN synapses was preserved throughout extinction.
Since these synapses have already grown to near maximal levels, the TANs pause robustly at the onset of the renewal phase, and continue to do so until the Pf-TAN synapses experience enough LTD to extinguish the TAN pause. This apparent in panel D, which shows that the net effective Pf-TAN synaptic strength abruptly jumps up at the beginning of the renewal phase, and then gradually declines until responding extinguishes.

![Model Mechanics: ABA](image)

**Figure 2.6:** A) Average DA release. B) Average synaptic strength of the CTX-MSN synapse. C) Average proportion responses emitted. D) Average synaptic strength of the Pf-TAN synapse.
AAB Renewal

Figure 2.7 shows the model mechanics underlying ABA renewal. Everything about the acquisition phase is identical to our analyses for ABA renewal.

Since the extinction phase occurs in the same context as acquisition (i.e., context A), the same Pf units that were active during the acquisition phase are active again during the extinction phase. This means that there is no immediate decrement of the TAN pause at the beginning of extinction, and the TANs take longer to stop pausing. This is apparent in panel D, which shows that the net effective Pf-TAN synaptic weight smoothly decreases from the onset of extinction to the end of extinction. This smooth decrease means that the TANs take slightly longer to abolish their pause response, which leaves the CTX-MSN synapse vulnerable to more unlearning than it was vulnerable to in ABA renewal.

Since the renewal phase occurs in context B, only the Pf units sensitive to features present in context B are active. Panel B shows that these Pf units are associated with slightly more net Pf-TAN synaptic strength than those sensitive to context A. This because the context A Pf-TAN synapses were driven below baseline during extinction, while those sensitive to context B (not including the overlap units) were preserved exactly at baseline. As a result, the net effective Pf-TAN synaptic strength (panel D) jumps up slightly at the onset of the renewal phase. Although this isn’t adequate for a large TAN pause, it is adequate for a small (easily extinguished) TAN pause. This small TAN pause, in combination with the rather large CTX-MSN synaptic strength developed during acquisition is enough for a small renewal effect.

ABC Renewal

Figure 2.8 shows the model mechanics underlying ABA renewal. Everything about the acquisition and extinction phases is identical to our analyses for ABA renewal.

Since the renewal phase occurs in context C, only the Pf units sensitive to features present in context C are active. Panel B shows that these Pf units are associated with slightly more net Pf-TAN synaptic strength than those sensitive
Figure 2.7: A) Average DA release. B) Average synaptic strength of the CTX-MSN synapse. C) Average proportion responses emitted. D) Average synaptic strength of the Pf-TAN synapse.
to context B (i.e., the extinction context). This is because the Pf-TAN synapses sensitive to context B were driven below baseline during extinction, while those sensitive to context C (not including the overlap units) were preserved exactly at baseline. As in AAB renewal, this isn’t adequate for a large TAN pause, but it is adequate for a small, easily extinguished, TAN pause. As a result, the net effective Pf-TAN synaptic strength (panel D) jumps up slightly at the onset of the renewal phase. As in AAB renewal, this small TAN pause, in combination with the rather large CTX-MSN synaptic strength developed during acquisition, is enough for a small renewal effect. However, note that the ABC renewal generated by the model cannot be smaller than AAB renewal. This is because the CTX-MSN synapse in AAB renewal unlearns more during the extinction phase than it does in ABC renewal, but the net effective Pf-TAN synaptic strength is the same.

2.2.4 Discussion

As we have discussed, Renewal indicates that extinction did not completely erase the learning that occurred during acquisition. Bouton et al. (2011) primarily discuss two main mechanisms by which initial learning could be protected. The first possibility is that the context is directly conditioned as part of the stimulus. This would mean that extinction training in a different context (as in ABA and ABC renewal) would be conditioning with a different stimulus and therefore would not interfere with the initial learning. The renewal effect observed in AAB and ABC paradigms would then presumably be explained by some type of generalization of the instrumental response to novel contexts. The second possibility is that the context “sets the occasion” for the active reinforcement contingencies. The idea here is that there are two learning processes occurring simultaneously. One of these processes learns about the instrumental response, and the other of these processes learns about the context that the instrumental response is valid in.

The model we proposed here is resonant with the latter of these ideas. The idea is that the instrumental response is learned at cortico-striatal synapses, and the context is learned at Pf-TAN synapses. It is noteworthy, however, that it would be fairly straightforward to incorporate contextual cues into the stimulus repre-
Figure 2.8: A) Average DA release. B) Average synaptic strength of the CTX-MSN synapse. C) Average proportion responses emitted. D) Average synaptic strength of the Pf-TAN synapse.
sentation (i.e., absorb the context information into the cortico-striatal pathway). This would, however, defeat the purpose of the TANs, since they require contextual input in order to learn to pause or not pause in order to protect cortico-striatal learning.

Several other neural network models have been proposed similar mechanisms to the occasion setting view proposed by Bouton and the context-learning account provided by our TANs model. These other models can account for renewal by assuming that extinction is a process of learning that the environmental context has changed (Gershman, Niv, & Blei, 2010; Redish et al., 2007). These models assume two separate processes: a situation recognition process that learns to recognize the current environmental context, and a standard temporal difference reinforcement learning component. The models are not neurobiologically detailed, although Redish et al. (2007) and Gershman et al. (2010) both speculate that the locus of their context-learning module is within prefrontal cortex and/or the hippocampus.
Chapter 3

Category Learning Applications

As we have discussed, II category learning seems like an ideal example procedural skill learning. However, we know of no existing data that have investigated if II category learning is as resistant to unlearning as other procedural skills (e.g., instrumental conditioning). As such, we performed two experiments that attempted to induce unlearning in an II category learning task\(^1\). Both experiments included acquisition, extinction, and reacquisition phases. During the acquisition and reacquisition phases, participants learned II categories using standard feedback-based training. In Experiment 1 the extinction phase used random feedback in an attempt to induce unlearning of the category structures. In Experiment 2 the extinction phase used a combination of random feedback and true feedback (which we will refer to as semi-random feedback) in attempt to induce unlearning of the category structures\(^2\). In Experiment 1, reacquisition performance recovered more quickly than it originally developed during the acquisition phase. In Experiment 2, reacquisition performance never recovered to the levels present during acquisition. These results imply that the extinction phase in experiment 1 failed to induce unlearning, but that the extinction phase in experiment 2 successfully induced at least some unlearning of the category structures. We will

\(^1\) Both category learning experiments discussed in this chapter were run by Todd W. Maddox in his laboratory at the University of Texas, Austin. The results presented here, both empirical and theoretical, are contained in Crossley, Ashby and Maddox (in preparation).

\(^2\) Note that this design was motivated by the theoretical interpretation of the empirical results from Experiment 1. See the “A theoretical Account of Experiments 1 and 2” later in this chapter.
show that these results are inconsistent with all current category-learning models. For example, all current models incorrectly predict that random feedback should induce unlearning. Finally, we propose a new neurobiologically detailed theory of II category learning (based on Ashby & Crossley, 2011) that successfully accounts for the reacquisition effects observed in Experiments 1 and 2.

3.1 Experiment 1

3.1.1 Introduction

Experiment 1 included acquisition, extinction, and reacquisition phases of 300 trials each. Every stimulus in all three phases was a line (as in Figure 1.1) that varied across trials in length and orientation. Identical II category structures were used in all three phases. These are represented abstractly in Figure 3.1. Note that these categories are similar to the categories shown in Figure 1.1, except with more exemplars in each category. Also note that the categories overlap, so perfect accuracy is impossible. In fact, the best possible accuracy with these categories is 95%.

The acquisition, extinction, and reacquisition phases were identical except in the nature of the feedback provided after each response. During acquisition, feedback indicated whether each response was correct or incorrect. During extinction, the feedback was random. On each trial, participants were informed that their response was correct with probability $\frac{1}{4}$ and incorrect with probability $\frac{3}{4}$, regardless of what response they actually made. The transition from acquisition to extinction occurred without the participants knowledge. On trial 301 the feedback rule simply changed without warning. Reacquisition began after the extinction phase was complete, again without warning. During the reacquisition phase, feedback was again veridical. There were two conditions. In the Reacquisition condition, the category structures remained the same as during acquisition and extinction. In the Meta-Learning Control condition, the stimuli were the same but the category labels were switched so that stimuli originally assigned to categories A, B, C, and D were now assigned to categories B, A, D, and C, respectively. Previous
Figure 3.1: The stimuli used in Experiments 1 and 2
research shows that a label switch of this type (i.e., without an intervening extinction phase) causes performance to drop almost to the same level as the first block of acquisition and that learning then proceeds at about the same rate as original acquisition (Maddox, Glass, OBrien, Filoteo, & Ashby, 2010). This control condition was included to ensure that fast reacquisition was not due to some sort of meta-learning. For example, because participants will have had 600 prior trials to familiarize themselves with the stimuli and task instructions, it is feasible that this experience would facilitate their learning even of novel categories. Thus, the critical test of fast reacquisition will be to compare the reacquisition performance of participants in the Reacquisition condition to the reacquisition performance of participants in the Meta-Learning Control condition.

3.1.2 Method

Participants

There were 19 participants in the Reacquisition condition and 20 participants in the Meta-Learning Control condition. All participants completed the study and received course credit for their participation. All participants had normal or corrected to normal vision. To ensure that only participants who performed well above chance were included in the post-acquisition phase, a learning criterion of 40% correct (25% is chance) during the final acquisition block of 100 trials was applied. Using this criterion, we excluded 3 participants in the Reacquisition condition and 2 participants in the Meta-Learning Control condition from further analyses.

Stimuli

Stimuli were black lines that varied only in length (pixels) and orientation (radians). The stimuli are displayed in Figure 3.1, and were generated by drawing 25 random samples from each of four bivariate normal distributions along the two stimulus dimensions with means along the x dimension of 72, 100, 100, and 128 and along the y dimension of 100, 128, 72, and 100 for categories A to D, respectively.
The variance along the x and y dimension was 100 and the covariance was 0 for all categories. The random samples were linearly transformed so that the sample means and variances equaled the population means and variances. Each random sample \((x, y)\) was converted to a stimulus by deriving the length (in pixels) as \(l = x\), and orientation (in degrees counterclockwise from horizontal) as \(o = y - 30\). These scaling factors were chosen to roughly equate the salience of each dimension. Optimal accuracy was 95%. The 100 stimuli were randomized separately for each participant in each block.

**Procedure**

Participants in the Reacquisition condition were told that they were to categorize lines on the basis of their length and orientation, that there were four equally-likely categories, and that high levels of accuracy could be achieved. At the start of each trial, a fixation point was displayed for 1 second and then the stimulus appeared. The stimulus remained on the screen until the participant generated a response by pressing the Z key for category A, the W key for category B, the / key for category C, or the P key for category D. None of these four keys were given special labels. Rather, the written instructions informed participants of the category label to button mappings, and if any button other than one of these four was pressed, an invalid key message was displayed.

During the 3 100-trial acquisition blocks, the word correct was presented if their response was correct or the word incorrect was presented if their response was incorrect. Once feedback was given, the next trial was initiated. During the 3 100-trial extinction blocks, the feedback was random. During the 3 100-trial reacquisition blocks feedback was again veridical.

The acquisition and extinction procedures in the Meta-Learning Control condition were identical to those from the Reacquisition condition. The reacquisition phase was replaced with 3 100-trial blocks of a category-label switch (Maddox et al., 2010). During the category-label switch, the association between stimulus clusters and category labels was changed so that stimuli originally assigned to categories A, B, C, and D were now assigned to categories B, A, D, and C,
respectively.

### 3.1.3 Results

**Accuracy-based results**

The left panel of Figure 3.2 shows the mean accuracy for every 25-trial block of each condition (ignore the right panel for now). During extinction, a response was coded as correct if it agreed with the category membership shown in Figure 3.1. Recall that the categories and feedback were identical in the two conditions until the beginning on the reacquisition phase. Note that participants from both conditions were able to learn the categories, reaching their peak accuracy near the end of acquisition, before falling to near chance during extinction. As expected, there are only minor differences between participants in the Reacquisition and Meta-Learning Control conditions during the acquisition and extinction phases of the experiment. During reacquisition however, the two learning curves diverge. Participants in the Reacquisition condition show fast reacquisition while participants in the Meta-Learning Control condition show interference (i.e., slow reacquisition).

To test these conclusions formally we conducted a 2 conditions (Reacquisition versus Meta-Learning Control) X 36 blocks repeated measures ANOVA. We found no significant effect of condition \([F(1, 32) = 1.19, p = 0.28]\), but the interaction [Greenhouse-Geisser corrected \(F(7, 224) = 3.71, p < 0.001\)] and the effect of block [Greenhouse-Geisser corrected \(F(7, 224) = 3.62, p < 0.05\)] were both significant. We then conducted several 2 conditions (Reacquisition versus Meta-Learning Control) X 12 blocks repeated measures ANOVAs, where the 12 blocks corresponded to the acquisition, extinction, or reacquisition phase. In the ANOVA corresponding to the acquisition phase, we found no effect of condition \([F(1, 32) = 2.07, p = 0.16]\), and the interaction was not significant [Greenhouse-Geisser corrected \(F(10, 205) = 1.03, p = 0.41\)], but the effect of block was significant [Greenhouse-Geisser corrected \(F(10, 205) = 21.24, p < 0.001\)]. In the ANOVA corresponding to the extinction phase we found no effect of condition \([F(1, 32) = 0.09, p = 0.76]\), and no interaction [Greenhouse-Geisser corrected \(F(8, 256) = \)].
Figure 3.2: The left column shows results obtained from all participants, and the right column shows results obtained by exclusively considering participants best fit by a decision-bound model that assumed II responding. A and B) The mean accuracy across all blocks of the experiment in both the Reacquisition condition and the Meta-Learning condition. C and D) The mean accuracy in the Reacquisition condition during each phase of the experiment overlaid on top of each other. E and F) The mean accuracy in the Meta-Learning condition during each phase of the experiment overlaid on top of each other.
0.96, \( p = 0.46 \), but the effect of block was significant [Greenhouse-Geisser corrected \( F(8, 256) = 4.52, p < 0.001 \)]. In the ANOVA corresponding to the reacquisition phase, we found a significant effect of condition [\( F(1, 32) = 6.40, p < 0.05 \)], and block [Greenhouse-Geisser corrected \( F(5, 188) = 9.17, p < 0.001 \)], but the interaction was not significant [Greenhouse-Geisser corrected \( F(5, 188) = 1.60, p = 0.15 \)]. The key result from these analyses is that there was only a significant difference between conditions during the reacquisition phase.

Next, we computed several repeated measures t-tests to compare performance between phases (i.e., acquisition, extinction, and reacquisition). For the Reacquisition condition, mean acquisition performance (across blocks) and mean reacquisition performance were both significantly better than mean performance during extinction [acquisition: \( t(191) = 12.32, p < 0.001 \); reacquisition: \( t(191) = 17.95, p < 0.001 \)]. A more important result was that reacquisition was significantly better than acquisition [\( t(191) = -5.11, p < 0.001 \)]\(^3\). This difference is clearly seen in the middle and bottom panels of Figure 3.2, which superimpose the acquisition and reacquisition curves from the Reacquisition condition.

Inspection of Figure 3.2 suggests that the difference between the acquisition and reacquisition phases is primarily due to the first five blocks of each phase. To test this hypothesis we first computed a repeated measures t-test on the first five blocks of acquisition pooled together and the first five blocks of reacquisition pooled together. This test revealed a significant difference [\( t(79) = -6.08, p < 0.001 \)]. We then computed a repeated measures t-test on the remaining blocks pooled together, and found no significant difference [\( t(111) = -1.08, p = 0.28 \)]. Within the Meta-Learning Control condition, acquisition performance was significantly better than extinction [\( t(215) = 19.22, p < 0.001 \)], reacquisition was significantly better than extinction [\( t(215) = -9.15, p < 0.001 \)], reacquisition was significantly worse than acquisition [\( t(215) = 8.85, p < 0.01 \)].

In summary, for both the Reacquisition and Meta-Learning Control conditions, performance during extinction was significantly worse than performance during acquisition and reacquisition. For the Reacquisition condition, performance

\(^3\)We also computed these t-tests by pooling the data from the acquisition phase across conditions. Since the results were qualitatively identical, we did not include them.
during reacquisition was significantly better than performance during acquisition, but this difference was only driven by the first five blocks of each phase. For the Meta-Learning Control condition, performance during reacquisition was significantly worse than performance during acquisition.

Next, we performed two separate tests to examine extinction. First, we compared the first block of acquisition to the last block of extinction (i.e., acquisition - extinction). These tests revealed that these two blocks were not significantly different from each other in either the Reacquisition condition \( t(15) = 0.61, p = 0.55 \), or the Meta-Learning Control condition \( t(17) = 1.16, p = 0.26 \). Second, we pooled the data from the last two extinction blocks from both conditions and computed a t-test on the null hypothesis that these data were generated from a distribution with mean 0.25 (i.e., the accuracy we would expect if performance was truly at chance). This t-test revealed that the last two blocks of extinction were significantly different from chance \( t(67) = 2.94, p < 0.001 \). In summary, extinction eventually reduced accuracy to the level present in the first 25 trials of acquisition training (i.e., before much learning could have occurred), but did not reduce performance to chance levels.

**Model-based results**

The accuracy-based results show fast reacquisition in the Reacquisition condition. However, this could either be because participants truly reacquired the strategy they learned during acquisition, or it could be because they switched away from the II strategy they learned during acquisition in favor of a simpler rule-based strategy. To test this hypothesis, we partitioned the data from each participant into blocks of 100 trials and fit different types of decision bound models (e.g., Ashby, Waldron, Lee, & Berkman, 2001; Maddox & Ashby, 1993) to each block of data from each participant. One type assumed a rule-based decision strategy, one type assumed an II strategy, and one type assumed random guessing. See Appendix 1 for details.

\(^4\)We choose to fit the models to blocks of 100 trials instead of the blocks of 25 trials used for the accuracy analyses because the reliability of the fits are greatly improved by a larger sample size.
Table 3.1 shows the number of participants in the two conditions best fit by a model of these three types. In the final block of acquisition, about 2/3 of all participants from the Reacquisition condition and slightly less than 2/3 of all participants from the Meta-Learning Control condition were best fit by models that assumed information integration. The remaining participants were best fit by a model that assumed a rule-based strategy. In the first block of reacquisition, about 3/4 of all participants in the Reacquisition condition and about 1/2 of all participants in the Meta-Learning Control condition were best fit by a model that assumed an II strategy. The remaining participants from the Reacquisition condition were all best fit by a model that assumed a rule-based strategy, whereas most of the remaining participants from the Meta-Learning Control condition were best fit by a model that assumed a guessing strategy. Thus, the proportion of participants fit best by a model that assumed information integration slightly increased from the last block of acquisition to the first block of reacquisition in the Reacquisition condition, and slightly decreased in the Meta-Learning control condition. Even so, neither the slight increase in the Reacquisition condition or the slight decrease in the Meta-Learning condition was significant [Reacquisition condition: $t(30) = -0.82, p = 0.21$; Meta-Learning condition: $t(34) = 0.67, p = 0.25$].

The reacquisition responses of more participants in the Reacquisition condition were best fit by a model that assumed an II strategy than in the Meta-Learning Control condition. This supports fast reacquisition because if the extinction period had caused unlearning then the reacquisition modeling results should have been the same for the two conditions. Also note that the number of participants in the Reacquisition condition whose responses were best fit by a model that assumed an II strategy remained roughly constant from acquisition to reacquisition. These results suggest that strategy switches from II to rule-based strategies were not driving fast reacquisition.

The fact that slightly less than 2/3 of all participants in both conditions were best fit by a rule-based model during the last block of acquisition is not unusual for an II experiment, but it raises the question of whether the fast reacquisition we observed could have been driven by rule use. In other words, it is
important to verify that participants who used an II strategy during reacquisition also showed fast reacquisition. To examine this question we repeated all the accuracy-based analyses reported in the previous section using only the data from the participants from each condition that were best fit by an II strategy during the last block of acquisition and during the first block of reacquisition.

The right panel of Figure 3.2 shows the mean accuracy of the II users for every 25-trial block of both conditions. Note that these data look much the same as the data from all participants shown in the left panel. For example, the II users from both conditions were able to learn the categories, reaching their peak accuracy near the end of the acquisition phase, and fell to near chance during extinction blocks. During reacquisition, the two learning curves do not diverge as obviously as they do in the left panel. Even so, II users in the Reacquisition condition appear to show fast reacquisition, whereas II users in the Meta-Learning Control condition appear to show only a small interference.

The results of the statistical tests were also similar to those reported earlier for all participants. First, the major difference in the ANOVA results was that during reacquisition the main effect of condition was no longer significant \([F(1,13) = 0.70, p = 0.42]\). Even so, by a sign test, reacquisition performance in the Reacquisition condition was still better than in the Meta-Learning Control condition (12 successes in 12 tests, \(p < .01\)). Second, a repeated measures t-test on the difference scores between all acquisition blocks pooled together and all reacquisition blocks pooled together revealed that performance during reacquisition was significantly better than during acquisition for the Reacquisition condition \([t(119) =

Table 3.1: Number of participants in the two conditions of Experiment 1 whose responses were best accounted for by a model that assumed an information-integration (II) decision strategy, a rule-based (RB) strategy, or random guessing, and the percentage of responses accounted for by those models

<table>
<thead>
<tr>
<th></th>
<th>Reacquisition Condition</th>
<th>Meta-Learning Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A3</td>
<td>R1</td>
</tr>
<tr>
<td>II</td>
<td>11</td>
<td>68.3</td>
</tr>
<tr>
<td>RB</td>
<td>5</td>
<td>64.4</td>
</tr>
<tr>
<td>Guessing</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>
-3.16, p < 0.005]. This difference is easily seen in the right panel of Figure 3.2, which superimposes the acquisition and reacquisition curves for the II users from the Reacquisition condition. Subsequent analyses confirmed that this difference was primarily driven by the first four blocks of each phase. Finally, as in the earlier analyses, performance in the first block of acquisition did not differ significantly from performance in the last block of extinction in either the Reacquisition condition [t(9) = 0.84, p = 0.42], or the Meta-Learning Control condition [t(4) = 1.10, p = 0.33]. Also, performance during the last two blocks of extinction was still significantly above chance [t(29) = 0.65, p = 0.52]. In summary, accuracy analyses using only the data from participants that were best fit by a model that assumed an II strategy during the last acquisition and the first reacquisition block replicated the results of our earlier analyses that used all the data.

3.1.4 Discussion

The results from Experiment 1 indicate fast reacquisition in the Reacquisition condition and interference in the Meta-Learning Control condition. It is critical to note that both of these results suggest a lack of complete unlearning during extinction, despite the fact that accuracy fell nearly to chance in both conditions5. If the extinction phase caused complete unlearning of the category structures that were learned during acquisition, then reacquisition should proceed at the same pace as original acquisition and changing the category structures as in the Meta-Learning Control condition should not produce interference. By definition, interference means that there must be some prior learning that is interfering with new learning.

At first glance, our results echo a classic result within the instrumental conditioning literature. Insofar as II category learning and instrumental conditioning are mediated by similar reinforcement learning mechanisms, our finding

5The fact that accuracy wasn’t reduced all the way to chance during extinction also reflects the lack of complete unlearning during extinction. It is important to note that this does not complicate the finding of fast reacquisition in the Reacquisition condition, nor does it complicate the interference found in the Meta-Learning Control condition. This is because performance dropped to the same level in each condition, and performance by the end of extinction was about the same as it was during the first block of acquisition.
that learning is mostly preserved during extinction with random feedback may not be that surprising\textsuperscript{6}. However, in addition to theoretical difficulties with this view that will be discussed later, there are a number of important differences between category learning and instrumental conditioning that need to be appreciated before settling on this conclusion. Aside from being a more cognitively complex task, category learning and instrumental conditioning differ fundamentally in their response characteristics. Instrumental conditioning is free response, of which there is rarely more than one response to make, and so learning is characterized by an abundance of responses. During extinction, removing reward is sufficient to make the behavior disappear. Category learning, on the other hand, is forced choice and participants must choose between at least two responses on every trial. Learning is characterized by an abundance of correct responses. However, since category learning is forced choice, extinction cannot be characterized by the simple absence of responding. Furthermore, unpublished data from our lab indicates that simply removing feedback during extinction in category learning is not sufficient to make accurate responding disappear. In light of these fundamental differences, the finding of fast reacquisition after extinction with random feedback during category learning is quite unexpected. As we will show in the next few sections, they also present a severe challenge to existing theoretical accounts of category learning.

\textbf{Strategy Recall}

An obvious possibility is that during reacquisition participants simply recalled the categorization rule that they were using at the end of the acquisition.

\textsuperscript{6}The random feedback used in Experiment 1 bears some similarity to random sequence blocks used to assess learning in the serial reaction time (SRT) task (Nissen & Bullemer, 1997). In this procedure a fixed sequence of stimuli is suddenly replaced with a random sequence. When the fixed sequence is later reintroduced, recovery of fast responding occurs more quickly than during original learning. Despite these surface similarities, the use of random sequence blocks in SRT tasks has important differences from our use of random feedback in Experiment 1. Most importantly, in the SRT task no feedback is typically given because accuracy is essentially perfect. Furthermore, during random feedback blocks in the SRT task the correct response to each stimulus is the same as during fixed sequence blocks (i.e., press the key associated with the presented stimulus). In contrast, in the extinction phase of Experiment 1 the feedback becomes random, but the stimuli do not change in any way. Because of these significant differences, it is not clear that the apparent similarities between the effects of random sequence blocks in SRT and random feedback blocks in II categorization are driven by a common or by different mechanisms.
period. Thus, no new learning was required, only memory retrieval. This would be a viable hypothesis if the categorization task was rule based, but for several reasons it cannot account for the results of Experiment 1. First, in II categorization tasks, participants do not have conscious awareness of their categorization strategy (e.g., Ashby & Maddox, 2005). In particular, participants cannot recall the strategy that they used immediately after II categorization training, so there is no reason to expect them to be able to do this during the reacquisition phase. Second, participants do not learn rules during II categorization, even if rule is interpreted in the most abstract, mathematical sense. Ashby and Waldron (1999) presented strong evidence that participants do not learn decision bounds in II categorization tasks or any other decision rule that can be specified by some specific mathematical function. Thus, there is no rule to recall in II categorization.

**Single Category-Learning Systems Accounts**

We know of no existing single-system theories of category learning that can account for fast reacquisition. For example, consider exemplar theory (Brooks, 1978; Estes, 1986, 1994; Hintzman, 1986; Lamberts, 2000; Medin & Schaffer, 1978; Nosofsky, 1986). Exemplar theory assumes that feedback is used to associate a category label with every exemplar encountered. When a new stimulus is seen, its similarity is computed to the memory representation of every previously seen exemplar from each potentially relevant category. The stimulus is then assigned to the category for which the sum of these similarities is greatest. The extinction phase in Experiment 1 randomly reassigns stimuli to each of the four contrasting categories. Exemplar theory predicts that performance will drop to chance during extinction when every stimulus is equally similar to the stored exemplars that have been associated with each of the four categories. During reacquisition the random category assignments that were learned during extinction will impair re-learning. In fact, exemplar theory naturally predicts that reacquisition will be slower than original acquisition (i.e., in Experiment 1) because during initial acquisition there are no random category assignments that must be overcome.

Other single-system accounts of category learning have an equally difficult
time accounting for our results (e.g., the striatal pattern classifier of Ashby & Waldron, 1999). This is because in each of these, the extinction period either causes unlearning or random learning. Reacquisition is then either a process of starting over from scratch (in the case of unlearning) or overcoming all the new random associations. In the former case reacquisition should occur at the same pace as original acquisition and in the later case reacquisition should be slower than acquisition.

Multiple Category-Learning Systems Accounts

Existing multiple-systems models of category learning are equally challenged by fast reacquisition. Included in this list are ATRIUM (Erickson & Kruschke, 1998) and COVIS (Ashby et al., 1998). Both models assume two systems one that is rule based and one that is similarity based. ATRIUM assumes that the similarity-based system is a standard exemplar model, whereas COVIS assumes that it is the striatal pattern classifier. During acquisition and reacquisition, ATRIUM and COVIS both predict that the similarity-based system will dominate performance (because the categories cannot be learned with a simple rule). For both models there are two possibilities for extinction. One is that the similarity-based system will continue to dominate. In this case the models essentially reduce to single-system accounts because they would predict that the similarity-based system would dominate all phases of the experiment. As single-system models they both fail to predict fast reacquisition for the reasons described in the preceding section.

A second possibility is that ATRIUM and COVIS might be able predict that participants switch to rule-based strategies during extinction. The problem here is that both models predict independent learning in the two systems. Thus, even if participants switch to rule-based strategies during extinction the models predict that the random feedback will cause unlearning in the similarity-based system. Thus, during reacquisition the similarity-based system will have to overcome this unlearning, which will prevent fast reacquisition.
Other Cognitive Accounts

A variety of mechanisms have been proposed in the cognitive literature that could theoretically account for fast reacquisition by postulating that different cognitive processes are in operation during the extinction phase as compared to the acquisition and reacquisition phases. For example, several categorization models postulate highly flexible attention mechanisms that can be modulated up and down depending on feedback (for a review, see Kruschke, 2011). If the gain on attention was low during extinction, then the stimuli presented during the extinction phase would have little impact on the category representation and if the gain was turned back up during reacquisition, then fast reacquisition would result. A similar account is provided by knowledge partitioning, which is the phenomenon in which people break down a task into subtasks, and apply a unique strategy in each subtask that is not influenced by the strategies used in the other subtasks (Lewandowsky & Kirsner, 2000; Yang & Lewandowsky, 2004). If participants are able to create a partition that includes acquisition and reacquisition on one side and extinction on the other then fast reacquisition should occur in the Reacquisition condition of Experiments 1 and 2.

The attention learning and knowledge partitioning accounts are similar in that both require a cognitive switch to be flipped at the appropriate times during training. In the attention-learning account, flipping the switch at the beginning of extinction would reduce attentional gain, whereas in the knowledge partitioning account it would create a new knowledge structure. Flipping it again at the beginning of reacquisition would either turn the attentional gain back up or cause a switch back to the original knowledge structure. The challenge for all switch-based accounts is to specify a mechanism that would allow the cognitive switch to be flipped during the beginning of extinction and to be flipped again at the start of the reacquisition phase. To our knowledge, none of the attention learning or knowledge partitioning models predicts that a change to random feedback will cause the switch to be flipped. For example, each of the attention learning models assumes that changes in the allocation of attention are mediated by attempts to reduce error (e.g., as in back ‘propagation). Since the feedback given during the
extinction phase in Experiment 1 was random, there is no reallocation of attention that will reduce error and therefore, presumably there would be no attentional relearning (i.e., since the gradient is flat). Similarly, most experiments that study knowledge partitioning include one or more cues that signal the participant where to make the partition. In the present studies, the only such cue was a change in the validity of the feedback. Ashby and Crossley (2010) reported the results of several categorization experiments in which knowledge partitioning would have allowed perfect accuracy. Participants in these studies could have used the validity of feedback to construct the partition, or the value of one of the two stimulus features (i.e., large versus small). Despite a variety of attempts to induce knowledge partitioning, only 2 of 53 participants in the Ashby and Crossley (2010) experiments showed any evidence that knowledge partitioning was successful. Thus, the available evidence suggests that the validity of feedback alone is not enough to induce knowledge partitioning.

In summary, although current attention learning and knowledge partitioning models are incompatible with our results, it appears that their primary shortcoming is that they lack a mechanism that would flip the switch at the appropriate times during extinction and reacquisition. In a later section of this article, we propose such a mechanism.

**General Learning Theories**

Given the preceding discussion, it is not surprising that fast reacquisition has also posed a difficult challenge for learning theories in general. For example, as mentioned above, fast reacquisition disconfirms any theory that assumes learning is purely a process of strengthening associations between stimuli and responses (e.g., Reddish, Jensen, Johnson, & Kurth-Nelson, 2007). Partly for this reason, some conditioning researchers have proposed that extinction is not a process of unlearning, but rather a process of new learning (Bouton, 2004; Rescorla, 2001). In particular, Bouton (2004) suggested that conditioned responding is highly context-specific. According to this view, a new context is learned during extinction.

---

7Bouton (2004) proposed that even the presence or absence of reinforcement partially defines a context. In this way, even extinction in the same conditioning chamber causes a new context
leaves the original learning in the acquisition context intact. This explanation accounts for fast reacquisition as well as a number of other important conditioning phenomena. Although this explanation offers no real insight into the mechanism of where or how such context learning is mediated, Bouton and colleagues have recently reviewed evidence suggesting that the hippocampus and the amygdala may play critical roles (Bouton et al., 2006).

**Computational and Neurobiological Accounts**

There are several computational and neurobiological accounts of fast reacquisition. OReilly and Munakata (2000) proposed a computational model in which learning strengthens a connection (i.e., weight) between critical units in a neural network model and extinction weakens this weight. Once the strength falls low enough for the behavior to disappear, however, this weight is no longer weakened by further extinction trials. This allows the model to predict fast reacquisition, because the first rewarded trial following extinction brings the connection strength above threshold, and therefore reinstates the behavior.

Several neural network models have been proposed that can account for fast reacquisition by assuming that extinction is a process of learning that the environmental context has changed (Gershman, Niv, & Blei, 2010; Redish et al., 2007). These models assume two separate processes—a situation recognition process that learns to recognize the current environmental context, and a standard temporal difference reinforcement learning component. The models are not neurobiologically detailed, although Redish et al. (2007) and Gershman et al. (2010) both speculate that the locus of their context-learning module is within prefrontal cortex and/or the hippocampus. In addition, both models have only been applied to standard extinction paradigms, and thus, it is not clear whether they would be able to account for the effects of random feedback in Experiment 1.
3.2 Experiment 2

3.2.1 Introduction

The results of Experiment 1 suggest that random feedback does not induce true unlearning. As we discuss in the following section, A Theoretical Account of Experiments 1 and 2, our model accounts for this failure by assuming that the TANs do not pause during extinction. The idea is that even though random feedback provides relatively frequent positive feedback, it does not provide feedback that is contingent on the subjects behavior (i.e., there is no stable action-outcome contingency during extinction). We hypothesize that the dopamine system is highly sensitive to action-outcome contingency. This suggests that one possible method to induce true unlearning might be to make the feedback partially contingent on the behavior of the subject. One way to do this is to include some accurate feedback trials among the random feedback trials that define the unlearning period. The idea is to make the contingency between feedback and behavior high enough to keep the TANs paused but low enough to induce true unlearning. To test this hypothesis Experiment 2 includes a Partially-Contingent condition. The acquisition and reacquisition phases are the same as in Experiment 1, but during extinction random feedback will be given on 75% of the trials and true (accurate) feedback will be given on 25% of the trials. The hope is that the true feedback trials will be frequent enough to keep the TANs paused throughout the extinction phase and the random feedback will overwrite the relevant memory traces with random associations. Experiment 2 also includes a meta-learning control condition that is similar to the one used in Experiment 1.

3.2.2 Method

Participants

There were 32 participants in the Partially-Contingent condition, 27 participants in the Meta-Learning Control condition, and 29 participants in the Partially-Contingent-Blocked condition. All participants completed the study and received
course credit for their participation. All participants had normal or corrected to normal vision. To ensure that only participants who performed well above chance were included in the post-acquisition phase, a learning criterion of 40% correct (25% is chance) during the final acquisition block of 100 trials was applied. Using this criterion, we excluded 4 participants in the Partially-Contingent condition and 3 participants in the Meta-Learning Control condition from further analyses.

Stimuli

The stimuli were identical to those from Experiment 1.

Procedure

The procedure was identical to that from Experiment 1 except for the nature of the feedback during the extinction phase. In Experiment 1, during the three 100-trial extinction blocks, the feedback was random. In Experiment 2, during each of the three 100-trial extinction blocks, veridical feedback was provided on a random 25% of trials and random feedback was given on the remaining trials.

3.2.3 Results

Accuracy-based results

The left panel of Figure 3.3 shows the mean accuracy for every 25-trial block of each condition (ignore the right panel for now). During extinction, a response was coded as correct if it agreed with the category membership shown in Figure 3.1. Recall that the categories and feedback were identical in the two conditions until the beginning of the reacquisition phase. Note that participants from both conditions were able to learn the categories, reaching their peak accuracy near the end of acquisition, before falling to about 40% correct during extinction. As expected, there are only minor differences between participants in the Reacquisition and Meta-Learning Control conditions during the acquisition and extinction phases of the experiment. Interestingly, the two learning curves do not diverge
during reacquisition. Participants in both conditions show an interference (i.e., slow reacquisition).

To test these conclusions formally we conducted a 2 conditions (Reacquisition versus Meta-Learning Control) X 36 blocks repeated measures ANOVA. We found no significant effect of condition \( F(1, 50) = 0.26, p = .61 \), nor interaction [Greenhouse-Geisser corrected \( F(7, 382) = 1.49, p = .162 \)], but the effect of block was significant [Greenhouse-Geisser corrected \( F(7, 382) = 15.94, p < .001 \)]. We then conducted several 2 conditions (Reacquisition versus Meta-Learning Control) X 12 blocks repeated measures ANOVAs, where the 12 blocks corresponded to the acquisition, extinction, or reacquisition phase. In the ANOVA corresponding to the acquisition phase, we found no effect of condition \( F(1, 50) = 0.15, p = .70 \), or interaction [Greenhouse-Giesser corrected \( F(6, 302) = 1.34, p = .24 \)], but the effect of block was significant [Greenhouse-Geisser corrected \( F(6, 302) = 17.75, p < .001 \)].

In the ANOVA corresponding to the extinction phase we found no effect of condition \( F(1, 50) = 2.74, p = .10 \), and no interaction [Greenhouse-Giesser corrected \( F(5, 251) = 0.79, p = .56 \)], but the effect of block was significant [Greenhouse-Geisser corrected \( F(5, 251) = 2.30, p < 0.05 \)]. In the ANOVA corresponding to the reacquisition phase, we found no significant effect of condition \( F(1, 50) = .09, p = .77 \), but the interaction and block effects were both significant [Interaction: Greenhouse-Geisser corrected \( F(5, 297) = 3.40, p < .005 \), Block: Greenhouse-Geisser corrected \( F(5, 297) = 7.21, p < .001 \)]. The key result from these analyses is that there was only a significant difference between conditions during the reacquisition phase.

Next, we computed several repeated measures t-tests to compare performance between phases (i.e., acquisition, extinction, and reacquisition). For the Reacquisition condition, mean acquisition performance (across blocks) and mean reacquisition performance were both significantly better than mean performance during extinction [acquisition: \( t(335) = 13.96, p < 0.001 \); reacquisition: \( t(335) = 10.05, p < 0.001 \)]. Interestingly, reacquisition was significantly worse than acquisition [\( t(335) = 3.04, p < 0.005 \)]\(^8\). This difference is clearly seen in the middle

\(^8\)We also computed these t-tests by pooling the data from the acquisition phase across conditions. Since the results were qualitatively identical, we did not include them.
Figure 3.3: The left column shows results obtained from all participants, and the right column shows results obtained by exclusively considering participants best fit by a decision-bound model that assumed II responding. A and B) The mean accuracy across all blocks of the experiment in both the Reacquisition condition and the Meta-Learning condition. C and D) The mean accuracy in the Reacquisition condition during each phase of the experiment overlaid on top of each other. E and F) The mean accuracy in the Meta-Learning condition during each phase of the experiment overlaid on top of each other.
and bottom panels of Figure 3.3, which superimpose the acquisition and reacquisition curves from the Reacquisition condition. The qualitative pattern of results for the Meta-Learning condition was the same. Specifically, acquisition and reacquisition were better than extinction [acquisition: \( t(287) = 11.68, p < 0.001 \); reacquisition: \( t(287) = 5.24, p < 0.001 \)], and reacquisition was worse than acquisition [\( t(287) = 5.01, p < 0.001 \)]. In summary, for both the Reacquisition and Meta-Learning Control conditions, performance during extinction was significantly worse than performance during acquisition and reacquisition, and performance during reacquisition was significantly worse than performance during acquisition.

Next, we performed two separate tests to examine extinction. First, we compared the first block of acquisition to the last block of extinction (i.e., acquisition - extinction). These tests revealed that these two blocks were not significantly different from each other in either the Reacquisition condition [\( t(27) = 0.46, p = 0.64 \)], or the Meta-Learning Control condition [\( t(23) = 1.25, p = 0.22 \)]. Second, we pooled the data from the last two extinction blocks from both conditions and computed a \( t \)-test on the null hypothesis that these data were generated from a distribution with mean 0.25 (i.e., the accuracy we would expect if performance was truly at chance). This \( t \)-test revealed that the last two blocks of extinction were significantly different from chance [\( t(103) = 7.44, p < 0.001 \)]. In summary, extinction eventually reduced accuracy to the level present in the first 25 trials of acquisition training (i.e., before much learning could have occurred), but did not reduce performance to chance levels.

**Model-based results**

The accuracy-based results show slow reacquisition in both conditions. This could either be because participants never reacquired the strategy they learned during acquisition, or it could be because they applied the II strategy they learned during acquisition less successfully. To explore these possibilities, we partitioned the data from each participant into blocks of 100 trials and fit different types of decision bound models (e.g., Ashby, Waldron, Lee, & Berkman, 2001; Maddox & Ashby, 1993) to each block of data from each participant. One type assumed a rule-
based decision strategy, one type assumed an II strategy, and one type assumed random guessing. See Appendix 1 for details.

Table 3.2 shows the number of participants in the two conditions best fit by a model of these three types. In the final block of acquisition, about 2/3 of all participants from both the Reacquisition condition and about 1/2 the Meta-Learning Control condition were best fit by models that assumed information integration. The remaining participants were best fit by a model that assumed a rule-based strategy, with one subject best fit by a model that assumed a guessing strategy. In the first block of reacquisition, about 1/2 of all participants in the Reacquisition condition and about 1/2 of all participants in the Meta-Learning Control condition were best fit by a model that assumed an II strategy. The remaining participants from the Reacquisition condition were about evenly split between a model that assumed a rule-based strategy and a model that assumed a guessing strategy. A similar pattern held true for the Meta-Learning Control, although there were slightly fewer subjects best fit by a model that assumed a guessing strategy. Thus, the proportion of participants fit best by a model that assumed an information-integration strategy was decreased from the last block of acquisition to the first block of reacquisition in both the Reacquisition condition and the Meta-Learning control condition, although this decrease was only significant for the reacquisition condition [Reacquisition condition: $t(54) = 1.88, p < 0.05$; Meta-Learning condition: $t(46) = 0.58, p = 0.28$].

Table 3.2: Number of participants in the two conditions of Experiment 2 whose responses were best accounted for by a model that assumed an information-integration (II) decision strategy, a rule-based (RB) strategy, or random guessing, and the percentage of responses accounted for by those models

<table>
<thead>
<tr>
<th></th>
<th>Reacquisition Condition</th>
<th>Meta-Learning Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A3</td>
<td>%</td>
</tr>
<tr>
<td>II</td>
<td>19</td>
<td>73.0</td>
</tr>
<tr>
<td>RB</td>
<td>9</td>
<td>65.2</td>
</tr>
<tr>
<td>Guessing</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

Since the model developed here deals exclusively with an II-based memory system, we repeated all the accuracy-based analyses reported in the previous
section using only the data from the participants from each condition that were best fit by an II strategy during the last block of acquisition. The right panels of Figure 3.3 shows the mean accuracy of the II users for every 25-trial block of both conditions. Note that these data look much the same as the data from all participants shown in the left panel. II users in both the Reacquisition condition and the Meta-Learning condition appear to show slow reacquisition (i.e., an interference).

The results of the statistical tests were qualitatively identical to those reported earlier for all participants. First, a 2 condition X 36 blocks repeated measures ANOVA again gave no significant effect of condition [Greenhouse-Geisser corrected F(1,31) = 0.39, p = 0.537], or interaction [Greenhouse-Geisser corrected F(8, 263) = 1.19, P = 0.31]. The effect of block was again significant [Greenhouse-Geisser corrected F(1, 263) = 14.7, P < 0.001]. Second, a repeated measures t-test on the difference scores between all acquisition blocks pooled together and all reacquisition blocks pooled together revealed that performance during reacquisition was significantly worse than during acquisition in the Reacquisition condition [t(215) = 2.14, p < 0.05] but not in the Meta-Learning condition [t(179) = 0.46, p = 0.63]. This results are clearly seen in the right panel of Figure 3.3, which superimposes the acquisition and reacquisition curves for the II users from the Reacquisition condition. Finally, as in the earlier analyses, performance in the first block of acquisition did not differ significantly from performance in the last block of extinction in either the Reacquisition condition [t(17) = 0.50, p = 0.62], or the Meta-Learning Control condition [t(14) = 1.36, p = 0.19]. Also, performance during the last two blocks of extinction was still significantly above chance [t(65) = 5.29, p < 0.001]. In summary, accuracy analyses using only the data from participants that were best fit by a model that assumed an II strategy during the last acquisition and the first reacquisition block replicated the results of our earlier analyses that used all the data.

3.2.4 Discussion

The results from Experiment 2 indicate interference in both the Reacquisition condition and the Meta-Learning Control condition. It is critical to note that
while these results imply some degree of unlearning (or perhaps random learning), they do not imply complete unlearning. If the extinction phase caused complete unlearning of the category structures that were learned during acquisition, then not only should reacquisition proceed at the same pace as original acquisition, but changing the category structures as in the Meta-Learning Control condition should not produce interference. These data do not meet either of these criteria for unlearning. Additionally, performance did not drop to near chance levels during the extinction phase.

Our results bear some resemblance to the data of Woods and Bouton (2007) that we discussed in chapter 2. Woods and Bouton (2007) found that using a lean schedule of reinforcement during the extinction phase could slow the rate of reacquisition. Ostensibly, a lean schedule of reinforcement in instrumental conditioning is like the mixture of random and true feedback used in our experiment, so there seems to be some resonance between these results. However, as we discussed in Experiment 1, there are important fundamental differences between category learning and instrumental conditioning that seem significant enough to obfuscate any predictions one might draw from one data set to the other.

**Single Category-Learning Systems Accounts**

Single-system theories of category learning can account for the interference that we observed in Experiment 2. For example, consider exemplar theory (Brooks, 1978; Estes, 1986, 1994; Hintzman, 1986; Lamberts, 2000; Medin & Schaffer, 1978; Nosofsky, 1986). Exemplar theory assumes that feedback is used to associate a category label with every exemplar encountered. When a new stimulus is seen, its similarity is computed to the memory representation of every previously seen exemplar from each potentially relevant category. The stimulus is then assigned to the category for which the sum of these similarities is greatest. The extinction phase in Experiment 2 was a mixture of two components. One component (75% of the mixture) randomly reassigned stimuli to each of the four contrasting categories, and the other component (25% of the mixture) provided valid feedback. Exemplar theory predicts that performance will drop, but not to chance, since
the presence of some valid feedback during extinction will prevent every stimulus from becoming equally similar to the stored exemplars that have been associated with each of the four categories. However, the random category assignments that were learned during extinction will impair re-learning during reacquisition. As we stated in our previous analysis (see Experiment 1 discussion), exemplar theory naturally predicts that reacquisition will be slower than original acquisition (i.e., in Experiment 1) because during initial acquisition there are no random category assignments that must be overcome. Exemplar theory also naturally predicts the higher accuracy levels observed during extinction. However, exemplar theory also predicts that reacquisition in Experiment 2 should be faster than in Experiment 1 (i.e., it predicts there should have been an even more substantial interference in Experiment 1). The striatal pattern classifier of Ashby & Waldron (1999) has similar difficulties. That is, it naturally predicts the interference observed in Experiment 2, but predicts that the interference should be smaller than in Experiment 1. This is because in both models, the random learning during extinction in Experiment 1 is more severe than the random learning in Experiment 2.

**Multiple Category-Learning Systems Accounts**

Existing multiple-systems models of category learning also struggle with the results of Experiment 2. As in our discussion of Experiment 1, one possibility with these models is that they never switch out of their similarity-based system. In this case our analysis reduces to our analysis of single category-learning systems given above. In the case that these models switch to their rule-based systems during the extinction phase, then they can certainly account for the accuracy levels observed during extinction in Experiment 2. This because there are a huge number of possible rules with a practically continuous goodness of fit to the category structures used here. It is not clear, however, that either of these models is capable of simultaneously fitting the extinction accuracy levels in both Experiment 1 and Experiment 2. Since both Atrium and COVIS assume independent learning in each system, both models would experience semi-random learning in the similarity-based system during extinction. This means that while each of these models naturally accounts
for the interference observed in Experiment 2, they predict a larger interference for Experiment 1 than Experiment 2.

Other Cognitive Accounts

In the discussion of Experiment 1 we determined that both knowledge partitioning and attentional modulation could possibly account for fast reacquisition after an extinction period of random feedback, but that such an account lacked a well specified mechanism for creating the partitions or modulating the attentional gain in such a way to protect learning during extinction. A similar analysis holds for the results of Experiment 2. The relatively high accuracy levels observed during extinction indicate that attention must not have been dialed back as completely as it was in Experiment 1 during this period. This would imply that the synapses driving the behavior would be more subject to semi-random learning during extinction, and reacquisition would be slow. A similar explanation holds for knowledge partitioning. It is reasonable to suppose that semi-random feedback (Experiment 2) would be harder to partition then completely random feedback (Experiment 1), and so creating a unique partition for the extinction phase would be more difficult in Experiment 2 than in Experiment 1. This difficulty could lead to failure, leaving the relevant synapses subject to semi-random learning. However, it is critical to note that these explanations still lack a clearly specified mechanism to control the partitioning or the attentional modulation in the context of the task.

General Learning Theories

Boutons (2004) idea that a new context is learned during extinction resonates quite well with both the results of Experiment 1, Experiment 2, and the theory that we propose in the following section. The basic idea is that a random-feedback extinction period (Experiment 1) is different enough from the true feedback present during acquisition that a new context is learned (very much like our discussion of knowledge partitioning). The key to Boutons idea is that learning is context specific, so that extinction learning does not effect acquisition learning. In this view, the semi-random feedback used during the extinction period of Experi-
ment 2 would not be different enough from the feedback given during acquisition for a new context to be learned. This would allow acquisition learning to be overwritten with semi-random random learning during extinction. This explanation, however, has not been formally developed.

Computational and Neurobiological Accounts

There are several computational and neurobiological accounts of fast reacquisition that might be adapted to account for the slow reacquisition we observed in Experiment 2. OReilly and Munakata (2000) proposed a computational model in which learning strengthens a connection (i.e., weight) between critical units in a neural network model and extinction weakens this weight. Once the strength falls low enough for the behavior to disappear, however, this weight is no longer weakened by further extinction trials. This allows the model to predict fast reacquisition, because the first rewarded trial following extinction brings the connection strength above threshold, and therefore reinstates the behavior. Slow reacquisition could be accounted for simply by setting the parameters such that semi-random feedback does not cause enough unlearning to cross the threshold.

Several neural network models have been proposed that can account for fast reacquisition by assuming that extinction is a process of learning that the environmental context has changed (Gershman, Niv, & Blei, 2010; Redish et al., 2007). These models assume two separate processes—a situation recognition process that learns to recognize the current environmental context, and a standard temporal difference reinforcement learning component. The models are not neurobiologically detailed, although Redish et al. (2007) and Gershman et al. (2010) both speculate that the locus of their context-learning module is within prefrontal cortex and/or the hippocampus. In addition, both models have only been applied to standard extinction paradigms, and thus, it is not clear whether they would be able to account for the effects of random feedback (i.e., fast reacquisition) in Experiment 1 or the effects of semi-random feedback (i.e., slow reacquisition) in Experiment 2.
3.3 A Theoretical Account of Experiments 1 and 2

COVIS (Ashby et al., 1998; Ashby & Waldron, 1999) provides a neurobiologically detailed model of II category learning which assumes that II tasks are learned by a procedural-learning system that gradually associates categorization responses with regions of perceptual space via reinforcement learning. The key structures in the COVIS procedural learning system are the putamen and the premotor cortex (i.e., the supplementary motor area SMA). Early versions of COVIS assumed that the striatal regions most critical to the procedural system were the body and tail of the caudate nucleus (Ashby et al., 1998). More recent evidence however, suggests that the procedural system has a strong motor association (Ashby et al., 2003; Maddox et al., 2004), which caused the focus to switch to the putamen. Recent neuroimaging data supports this hypothesis (Waldschmidt & Ashby, 2011). The key site of learning in this model is at cortical-striatal synapses, and this synaptic plasticity is presumed to be facilitated by a dopamine-mediated reinforcement training signal from the substantia nigra pars compacta. Ashby, Ennis, and Spiering (2007) developed a computational version of this model. As discussed in the previous two sections, COVIS does not account for the results of Experiment 1 Experiment 2. The model predicts that random feedback will weaken appropriate cortical-striatal synapses and strengthen inappropriate synapses, and as a result, learning would have to begin anew during reacquisition.

Ashby and Crossley (2011) proposed a similar model that accounts for fast reacquisition in traditional instrumental learning paradigms in which extinction is defined by the absence of all feedback, rather than by random feedback as in Experiment 1. When applied to category learning, the Ashby and Crossley (2011) model is characterized by a number of key features. First, as in COVIS, category-response associations in II tasks are learned at cortical-striatal synapses. Second, the theory assumes that TANs tonically inhibit cortical input to striatal output neurons (MSNs). The TANs are driven by cells in the centremedian and parafascicular (CM-Pf) nuclei of the thalamus, which in turn are broadly tuned to features of the
environment. In rewarding environments the TANs learn to pause to stimuli that predict reward, which releases the striatal output neurons from inhibition. This allows striatal output neurons to respond to excitatory cortical input, thereby facilitating cortical-striatal plasticity. In this way, TAN pauses facilitate the learning and expression of striatal-dependent behaviors. When rewards are no longer available, the TANs cease to pause, which protects striatal learning from decay. Third, dopamine-dependent reinforcement learning occurs at cortical-striatal and CM-Pf TAN synapses. Fourth, dopamine release is modeled discretely on a trial-by-trial basis and is proportional to the reward prediction error. Ashby and Crossley (2011) showed that a computational version of this theory predicts fast reacquisition in instrumental conditioning paradigms at the same time that it correctly accounts for a wide variety of single-unit recording data. Included in this list are single-unit recordings from medium spiny neurons during acquisition, extinction, and reacquisition phases of instrumental conditioning and during category learning. In the following subsections, we use the general framework provided by Ashby and Crossley (2011) to account for the results of Experiment 1 and Experiment 2.

3.3.1 Methods

Model Architecture

The overall architecture of the model proposed by Ashby and Crossley (2011) is illustrated in Figure 3.4. Note that this is essentially the architecture used to fit the Woods and Bouton (2007) data, except it is generalized for a task that requires 4 response options. Each response option gets its own pathway, and each pathway receives a copy of the projection from the sensory association cortex. The overall mechanics of the model are the same as in previous applications. Specifically, in the absence of strong CM-Pf input, the TANs high spontaneous firing tonically inhibits the striatal medium spiny neuron (MSN) response to cortical input. When cells in the CM-Pf complex fire, learning at the CM-Pf TAN synapse quickly causes the TAN to pause when in a rewarding environment. This releases the MSNs from tonic inhibition, thereby allowing them to respond to cortical inputs and thus to gate learning at cortical-striatal synapses.
Figure 3.4: Model Architecture used to fit the data from Experiments 1 and 2

Activation Equations

The activation equations we use to determine the network behavior of the Figure 3.4 model are simply the equations presented in section 2.1.2 generalized to account for the extra units required by a task with 4 response options. In most cases, this generalization is trivial. Nevertheless, we fully describe them in this section for convenience.

The activation of all CM-Pf units was either off (with activation 0) or on (see table A.1 for amplitude values) during the duration of stimulus presentation. We model sensory cortex in the same way as in Ashby et al. (2007). Briefly, this means we assumed an ordered array of 40,000 units in sensory cortex, each tuned to a different stimulus. We assumed that each unit responds maximally when its preferred stimulus is presented, and that its response decreases as a Gaussian function of the distance in stimulus space between the stimulus preferred by that unit and the presented stimulus. For the present applications, it sufficed to assume that activation in each unit was either 0 or equal to some positive constant value during the duration of stimulus presentation. Specifically, we assumed that when
a stimulus is presented, the activation in sensory cortical unit K at time t is given by

\[ I_k(t) = \alpha e^{\frac{d(K,\text{stimulus})^2}{2\beta^2}} \]  

where \( \alpha \) and \( \beta \) are constants and \( d(K,\text{stimulus}) \) is the distance (in stimulus space) between the stimulus preferred by unit K and the presented stimulus. Equation 1 is a popular method for modeling the receptive fields of sensory units, both in models of categorization (e.g., Kruschke, 1992) as well as other tasks (e.g., Joo-Er, Wu, Lu, & Toh, 2002; Oglesby & Mason, 1991; Riesenhuber & Poggio, 1999; Rosenblum, Yacoob, & Davis, 1996).

The activation in striatal unit \( j \) at time \( t \), denoted \( S_j(t) \), was determined by the following coupled differential equations:

\[
50 \frac{dS_j(t)}{dt} = \sum_k w_{k,j}(n) [I_k(t) - \beta T(t)] - \gamma_s \sum_{m \neq j} f[S_m(t)] + [S_j(t) + 80][S_j(t) + 25] + E - u_s(t) + \sigma_s \epsilon(t)
\]

\[ 3.2 \]

\[ 100 \frac{du_s(t)}{dt} = -20[S_j(t) + 80] - u_s(t) \]  

\[ 3.3 \]

The function \( f[S_m(t)] \) is defined in the paragraph preceding equation 2.6. The activation in the TAN unit is perfectly described by equations 2.3 and 2.4. Activation in globus pallidus unit \( j \) at time \( t \), denoted by \( G_j(t) \), is described by:

\[
15 \frac{dG_j(t)}{dt} = -\alpha_G f[S_j(t)] + 71 + 0.7[G_j(t) + 60][G_j(t) + 40]
\]

\[ 3.4 \]

where \( \alpha_G \) is a constant. The first term models the inhibitory input from the striatum, the second term ensures a high tonic firing rate, and the last term is the quadratic integrate-and-fire component that is the same as in Equations 2.5. Spikes are produced after \( G_j(t) = 35 \) by resetting to \( G_j(t) = -50 \).

Similarly, activation in thalamus unit \( j \) at time \( t \) is given by

\[
\frac{dV_j(t)}{dt} = -\beta f[G_j(t)] + 71 + 0.7[V_j(t) + 60][V_j(t) + 40]
\]

\[ 3.5 \]
where $\beta_T$ is a constant, and $f[G_j(t)]$ is the alpha function from the paragraph preceding equation 2.6. Spikes are produced after $V_j(t) = 35$ by resetting to $V_j(t) = -50$. The first term models the inhibitory input from the globus pallidus. The constant 71 models excitatory input not explicitly included in the model. See section 2.1.2 for a complete discussion of this term.

Activation in the $j$th unit in premotor cortex at time $t$, denoted by $C_j(t)$, is given by

$$\frac{dC_j(t)}{dt} = \beta_C f[V_j(t)] - \gamma_C \sum_{k \neq j} f[C_k(t)] + 69 + 0.7 \sum_{k \neq j} f[C_k(t)] + 60 \sum_{k \neq j} [C_j(t) + 40] + \sigma_C \epsilon(t) \quad (3.6)$$

where $\beta_C$, $\gamma_C$, and $\sigma_C$ are constants, and $\epsilon(t)$ is white noise. As in other units, spikes are produced after $C_j(t) = 35$ by resetting to $C_j(t) = -50$. The second term on the right models lateral inhibition in the same way as in Equation 1. In tasks with two possible responses, evidence suggests that cortical units in premotor areas are sensitive to the cumulated difference in evidence favoring the two alternatives (e.g., Shadlen & Newsome, 2001). We used a more biologically plausible method that is known to simulate this difference process that is, we placed a separate threshold on the activation of each unit, but included lateral inhibition between the units (Usher & McClelland, 2001).

**Learning Equations**

The learning equations used in this application were the same as those described by equation 2.9. The only difference being that their were 4 sets of cortico-striatal synapses (one for each MSN in the model) each represented by a 40000 element vector. Each synapse in this vector was subject to the equation 2.9 learning rule.

**Dopamine Model**

Learning at the CM-Pf TAN synapse, and at all cortical-striatal synapses in the Ashby and Crossley (2011) model was a function of presynaptic activity, postsynaptic activity, and the dopamine released on each trial. The model of dopamine
release proposed by Ashby and Crossley (2011) was developed to be consistent with a broad array of dopamine cell firing data. In particular, the model was developed to account for three well replicated results (e.g., Schultz, Dayan, & Montague, 1997; Tobler, Dickinson, & Schultz, 2003): 1) midbrain dopamine cells fire tonically, 2) dopamine release increases above baseline following unexpected reward, and the more unexpected the reward the greater the release, and 3) dopamine release decreases below baseline following unexpected absence of reward, and the more unexpected the absence, the greater the decrease. One common interpretation of these latter two results is that over a wide range, dopamine firing is proportional to the reward prediction error (RPE):

\[ RPE = ObtainedReward - PredictedReward \] (3.7)

Ashby and Crossley (2011) built a simple model of dopamine release by specifying how to compute Obtained Reward, Predicted Reward, and exactly how the amount of dopamine release is related to the RPE. Although the Ashby and Crossley (2011) model easily accounts for fast reacquisition following a traditional no feedback extinction period, it fails to account for fast reacquisition after a period of random feedback in category learning (as was observed in experiment 1). This is because random feedback provides sufficient reward (on average, once every four trials) to keep the dopamine system fluctuating significantly above baseline. This causes the TANs to maintain their pause response. When the TANs are paused, the model unlearns the category structures, and as a result reacquisition is slow.

The problem is that the model was developed exclusively from data collected in traditional instrumental conditioning paradigms in which no feedback of any kind is given during extinction (see Chapter 1). The model predicts that fast reacquisition will occur in any paradigm in which dopamine cell firing eventually remains at baseline following responses during the extinction period. With random feedback, it seems likely that humans eventually will come to realize that the feedback is no longer reliable. In this case they may still monitor the feedback, but they are likely to give it less attention than in the past. With less attention, dopamine fluctuations could decrease. We know of no current models of DA release
that display this subtlety.

In Experiment 1, two cues signal participants that the feedback has become random. One is that the probability of receiving positive feedback is at chance. A second is that reward valence is no longer contingent on behavior. For several reasons, we believe the latter of these two cues is more important than the former. First, the probability of receiving positive feedback is also at chance when the experiment begins and at the beginning of reacquisition. Yet in both of these cases, accuracy steadily increases, which by all current dopamine-mediated learning models can only occur if dopamine fluctuations are significant. Second, any inference that performance is at chance requires a mental model of the testing conditions. For example, a positive feedback rate of 25% signals chance performance only under the realization that there are four equally likely response alternatives. In contrast, even in the absence of any such model, determining that the feedback is noncontingent on behavior is always a matter of determining whether reward valence is correlated with expectation. Furthermore, there is evidence that animals are highly sensitive to feedback contingency. In fact, in instrumental conditioning tasks, extinction can be induced simply by suddenly making the rewards noncontingent on behavior (Woods & Bouton, 2007; Nakajima, Urushihara, & Masaki, 2002). Additionally, fMRI studies in humans have shown that activity in the dorsal striatum is correlated with RPE when feedback is contingent on behavior but not when feedback is independent of behavior (ODoherty et al., 2004; Harunu & Kawato, 2005).

Based on this reasoning, we modified the Ashby and Crossley (2011) model of dopamine release in the following way. First, Ashby and Crossley used a simple model of predicted reward that depended only on past reinforcement history. This is sufficient for instrumental conditioning tasks with a single cue and a single response, but not for categorization tasks. One of the most ubiquitous results in the categorization literature is that accuracy, response confidence, and response time are all strongly correlated with the distance from the stimulus to the category decision bound (e.g., Ashby, Boynton, & Lee, 1994). Stimuli far from the bound are easier to categorize and response confidence is higher than for stimuli near the
bound. Thus, even if past reinforcement history is identical, reward expectation will be higher for stimuli far from the bound than for stimuli near the bound. In the Figure 3.4 model, response confidence is determined by the difference between the output magnitudes of the competing motor units. When one unit is much more active than the others, then the model is confident that the response associated with the activated unit is correct. In this case, the expectation of reward is high. If two units are almost equally active however, then the model is undecided which of two responses to emit and consequently, reward expectation will be low. Thus, we propose defining predicted reward on trial \( n \) in Equation 1, which we denote by \( P4(n) \), as the normalized difference between the maximum outputs of the two most active motor units. Specifically,

\[
P(n) = \frac{M_1(n) - M_2(n)}{M_1(n)}
\]

where \( M_1(n) \) is the maximum output from the most active motor unit on trial \( n \) and \( M_2(n) \) is the maximum output from the second most active motor unit on that trial. Equation 3.8 is sensitive both to past reinforcement history and to the identity of the stimulus.

Second, we assume that the amount of dopamine release is modulated by the correlation between reward expectation (i.e., \( P(n) \)) and the valence of obtained feedback. If these are uncorrelated then for example, the probability of receiving positive feedback does not depend on the response of the model because positive feedback is equally likely on trials when model uncertainty is high as when it is low. We assume that in such cases, dopamine fluctuations decrease in magnitude.

Ashby and Crossley (2011) assumed that the amount of dopamine released on trial \( n, D(n) \), was a linear function of the reward prediction error (RPE) between minimum and maximum values of 0 and 1, respectively, with a baseline level of 0.2. This model was chosen to closely match dopamine single-unit firing data reported by Bayer and Glimcher (2005). We propose to generalize this model as follows:
\[
DA(n) = \begin{cases} 
1 & \text{if } RPE > 1 \\
4r(n)RPE + 0.2 & \text{if } -0.25 < RPE \leq 1 \\
0 & \text{if } RPE < 0.25 
\end{cases}
\]  

(3.9)

where \( r(n) \) is a measure of the correlation between the predicted (i.e., \( P(n) \)) and obtained (i.e., 1 if feedback was positive and 0 if feedback was negative) rewards as of trial \( n \). Note that \( P(n) \) is a continuous variable and obtained reward is binary (0 or 1). Thus, to compute the correlation we used a variation of the point-biserial correlation coefficient. Specifically, we defined \( r(n) \) as

\[
r(n) = |\overline{P}_+ - \overline{P}_-| 
\]  

(3.10)

where \( \overline{P}_+ \) is the mean response confidence on trials that received positive feedback, and \( \overline{P}_- \) is the mean response confidence on trials that received negative feedback. Note that if feedback is noncontingent on behavior then \( \overline{P}_- = \overline{P}_+ \) and consequently, \( r(n) = 0 \). On the other hand, if confident responses are always rewarded with positive feedback then and the correlation will be large. To implement this model we assumed an intermediate value of \( r(n) = 0.4 \) during the first 25 trials and thereafter computed \( r(n) \) using Equation 3.10 with a window size of 25 trials.

### 3.3.2 Results

The data shown in Figure 3.5 and Figure 3.6 were generated by simulating the model 50 times per condition and averaging the results across simulations. Figure 3.5 was additionally processed in the same way the as the human behavioral data. Specifically, we split the 900 trials of the mean simulation results into blocks of 25 trials, and computed the standard error for each of these blocks. Figure 3.5 shows the predictions of this model (panels B and D) as compared to human behavioral (panels A and C). The top panel shows categorization accuracy in each 25-trial block. For comparison, the empirical learning curves from Experiments 1
and 2 are also shown (panels A and C respectively). Note that the model effectively captures the major qualitative properties of the data from both experiments that is, the model learns the categories, performance drops well below acquisition performance, and reacquisition is faster than acquisition (i.e., the average slope of dashed line is steeper than the average slope of the solid line) in the Experiment 1 account and slower than acquisition (i.e., the average slope of dashed line is shallower than the average slope of the solid line) in the Experiment 2 count. However, the model does not account for the asymptotic levels of accuracy observed in the human data. Specifically, human participants did not reach peak accuracy levels during reacquisition that were higher than peak accuracy levels during acquisition. The model, however, consistently reaches peak accuracy levels during reacquisition, regardless of the condition. In fact, reacquisition performance for participants in Experiment 2 was notably worse than their acquisition performance. One possibility is that human performance asymptotes because learning at cortico-striatal synapses is progressively attenuated through the experiment. In fact, there is evidence that the NMDA threshold for LTP is increased (i.e., LTP is more difficult to induce) after periods of high activity (Kirkwood, Rioult, & Bear, 1996; Bienenstock, Cooper, and Munro 1982). The model developed here was not equipped with a dynamic NMDA threshold (i.e., $\theta^{NMDA}_{\Delta t} = \text{constant across all trials}$), and it seems likely that adding such a feature would improve the quality of the fits.

Figure 3.6 shows the underlying mechanics of the model when applied to Experiments 1 (the black lines) and 2 (the grey lines). Panel A shows the mean DA release, panel B shows the mean strength of the synapses between each visual cortical neuron and the MSN associated with the correct response to the visual stimulus that most strongly excites this cortical cell\textsuperscript{9}, panel C shows the mean proportion correct, and panel D shows the mean CM-Pf–TAN synaptic strength. Note that the DA signal (panel A) quickly falls during the extinction period under both random and semi-random feedback, but that the DA levels are consistently lower under fully random feedback than they are under semi-random feedback. This difference emerges because the correlation between feedback and response

\textsuperscript{9}Actually, this mean only included visual cortical cells that significantly responded to stimuli in the category distribution.
**Figure 3.5**: A) Human behavioral results from Experiment 1. B) Model account of Experiment 1. C) Human behavioral results from Experiment 2. D) Model account of Experiment 2.
confidence (see equation 3.10) is necessarily higher under semi-random feedback than it is under fully random feedback.

Every other factor in the model follows the basic pattern set by the DA signal. Specifically, the CM-Pf TAN synapse (panel D) rapidly decreases in both conditions, but averages a lower value in the fully random condition than in the semi-random condition. This means that the TANs pause less (i.e., don’t protect the CTX-MSN synapses) in the fully random condition than in the semi-random condition. Note that this is also why the model responds at near chance levels during extinction (i.e., since the cortical input to the striatum is heavily attenuated by the TANs, the only activity in the premotor unit is due to noise) under fully random feedback but less so under semi-random feedback. The difference at CTX-MSN synapses (panel B) is obscured by the fact that the CTX-MSN synapse in the semi-random condition reached higher levels by the end of the acquisition phase than it did in the full random condition. Nevertheless, under careful inspection it is clear that the CTX-MSN synapse decreased more under semi-random feedback than it did under fully feedback. During reacquisition, the feedback again becomes contingent on behavior (in both conditions), dopamine fluctuations increase, and the CM-Pf TAN synaptic strength quickly grows strong enough to induce a TAN pause. At this point the behavior reappears because of the incomplete decay in cortical-MSN synaptic strengths during the random feedback period.

### 3.3.3 Discussion

The model developed in the last section has a number of attractive features. First, it correctly accounts for the behavioral results from II category-learning experiments. The great majority of these experiments include only an acquisition phase where veridical feedback is given after every response. In such experiments, the addition of the TANs do not change the predictions of the model in any significant way, since the model predicts that the TANs would quickly learn to pause and that they would remain paused throughout the course of the experiment. Without the TANs, the model reduces to a biologically detailed version of the striatal pattern classifier (Ashby & Waldron, 1999), which has successfully accounted for
Figure 3.6: A) Average DA release. B) Average synaptic strength of the CTX-MSN synapse. C) Average proportion responses emitted. D) Average synaptic strength of the Pf-TAN synapse.
II category-learning data for many years (Ashby & Maddox, 2005). Second, the model is consistent with a wide variety of neuroscience data. For example, 1) it is based on known neuroanatomy, 2) the units in each brain region accurately model single-unit recordings from these same regions (e.g., see Ashby & Crossley, 2011), and 3) the model is roughly consistent with the available neuroimaging data from II tasks for example, by postulating a key role for the putamen and SMA (Waldschmidt & Ashby, 2011). Third, of course, the model successfully accounts for the reacquisition effects observed in Experiments 1 and 2. As we have discussed, this is a significant accomplishment that makes the model unique among all other category learning models. As an added benefit, the model specifies a biological mechanism (i.e., the TANs) that protects learning during periods when action-outcome contingencies are changed (e.g., when the potential for rewarded behavior is suddenly reduced as they are in typical extinction protocols).
Chapter 4

Discussion

4.1 Introduction to Discussion

In this article, we showed how cholinergic interneurons in the striatum might protect cortical-striatal synapses during periods action-outcome contingencies are changed. The idea is that the TANs exert a tonic inhibitory influence over cortical input to striatal MSNs that prevents the execution of striatal-dependent actions. However, the TANs learn to pause in rewarding environments, and this pause releases the striatal output neurons from inhibition, thereby facilitating the learning and expression of striatal-dependent behaviors. When rewards are no longer available, the TANs cease to pause, which protects striatal learning from decay. Ashby and Crossley (2011) showed that the resulting model was consistent with a variety of single-cell recording data, and that it also predicted some classic behavioral phenomena, including fast reacquisition following extinction. The applications in this dissertation show that the model is capable of accounting for a broad array of behavioral results from instrumental conditioning (e.g., several reacquisition effects and several types of renewal effect) and II category learning (e.g., fast and slow reacquisition following extinction with random and semi-random feedback). The version of the model developed for the II category learning applications was driven by a novel DA model sensitive to the correlation between actions and outcomes. We know of no other model of DA that possesses this feature.
4.2 Scope and Limitations of the Model

Our primary goal was to develop a model of basal ganglia function that is accurately constrained by known neuroanatomy and neurophysiology, while simultaneously accounting for key behavioral results from instrumental conditioning and II category learning. We did not intend to capture every known detail about striatal and basal ganglia function. Similarly, we did not hope to generate a model that could account for every known behavioral result in instrumental conditioning or category learning. We were interested specifically in developing a model that exists in the middle of these two extremes. Nevertheless, we think it's important to acknowledge the rather severe limitations this imposes on our model.

First, with respect to its interactions with cortex, it is important to note that the striatum is organized into a set of functionally separate, parallel loops (Alexander, Delong, & Strick, 1986). Which loop a particular subregion of striatum is in, is determined primarily by the cortical regions that project to it. The theory developed here concerns the learning of stimulus-response associations, and therefore applies to striatal regions receiving input from sensory areas of cortex (e.g., see Figure 1). This excludes anterior regions of striatum, which are innervated primarily by areas of frontal cortex. For example, tasks that activate prefrontal cortex (PFC) also commonly activate the head of the caudate nucleus, since the PFC projects strongly to this anterior region of the striatum. The PFC and its striatal targets (e.g., dorsal striatum, head of the caudate nucleus) have their own role in context processing, which is beyond the scope of this article. For example, there is considerable evidence that a PFC head of the caudate circuit plays a critical role in attentional switching between different contexts (e.g., Robbins, 2007). There is also evidence that the TANs play a critical role in this process (Ragozzino, 2003), so a theory that attempts to account for such switching may postulate a role for the TANs similar to the one proposed here.

It is also important to note that the theory proposed here is meant to apply only to initial skill (or habit) learning. With overtraining, skills eventually come to be executed automatically. Following similar suggestions in the literature (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Miller, 1981, 1988), Ashby et al. (2007)
proposed a model in which the development of automaticity is mediated by a transfer of control from the cortical-striatal pathways emphasized here to cortical-cortical pathways from the relevant areas of sensory cortex to the areas of premotor and motor cortex that mediate the selection and execution of the appropriate motor program. For example, several studies have reported evidence that with overtraining, skills of the type modeled here become independent of both dopamine and the striatum (e.g., Bespalov et al., 2007; Carelli et al., 1997; Choi, Balsam, & Horvitz, 2005; Turner, McCaurn, Simmons, & Bar-Gad, 2005). It would be straightforward to augment the present model with the cortical-cortical pathways proposed by Ashby et al. (2007) (with cortical-cortical plasticity mediated by Hebbian learning). This augmented model should be used to make predictions about the effects of overtraining on the tasks considered in this article.

Third, we have greatly oversimplified the neuroanatomy of the basal ganglia, omitting for example, striosomes (i.e., patch compartments), the ventral striatum, and the indirect and hyperdirect pathways. However, rather than build the most complete model of the basal ganglia that was possible, our goal instead was to focus on the effects of the TANs on MSNs. For all regions downstream of the striatum, we simply tried to construct the simplest reasonable model that could account for the limited behavioral phenomena considered in this article. It seems likely that if the model was extended to more complex behaviors, then more biological detail would be needed in downstream areas. This generalization will be a goal of future research.

4.3 Relationship to other contemporary theories

There have been a number of other proposals that the TANs learn to pause in environments associated with reward (e.g., Apicella, 2007; Sardo, Ravel, Legallet, & Apicella, 2000; Shimo & Hikosaka, 2001). However, to our knowledge none of these have been developed into predictive theories.

There have also been several computational models that have provided accounts of acquisition and extinction based on a computational model of dopamine
release that is similar to the dopamine model used here (Kakade & Dayan, 2002; OReilly & Munakata, 2000; Redish et al., 2007). The former two of these models assume extinction is exclusively an unlearning phenomenon, and do not account for fast reacquisition. In the OReilly and Munakata (2000) model however, once the strength of what in our model is the cortical-striatal synapse falls low enough for the behavior to disappear, this synaptic strength is no longer weakened by further extinction trials. This allows the model to predict fast reacquisition, because the first rewarded trial following extinction brings the synaptic strength above threshold, and therefore reinstates the behavior. The TANs endow our model with a similar property.

Tan and Bullock (2008) recently developed a Hodgkin-Huxley-type computational model of TAN firing that is more detailed than the model proposed here. For example, Tan and Bullock specifically modeled changes in several specific ion concentrations, along with the effects on TAN activation of dopamine and of GABAergic interneurons. On the other hand, they did not specifically model activity in striatal MSNs, nor did they model activity in any cells outside the striatum. Thus, their empirical applications are limited to data collected from single TAN units (e.g., no behavioral data were modeled). The major difference between their model and ours is that they account for intrinsic and learned TAN responses via modulation of TAN activity by GABAergic and dopaminergic input, rather than by synaptic plasticity (as we assume). Since the evidence of LTP at TAN synapses is good (Aosaki et al., 1994a; Suzuki et al., 2001), it seems likely that TAN pauses are modulated by all of these factors. In any case, Tan and Bullocks (2008) model may best be seen as a detailed theory of TAN responses within classical conditioning paradigms. Our model however, is primarily concerned with how TAN responses act to gate learning at cortico-striatal synapses, and how this function influences behavior in instrumental conditioning paradigms and more general striatal-dependent behaviors.
4.4 Relationship to historically relevant theories

The models proposed in this dissertation are conceptually resonant with a set of early learning theories starting with the concept of drive and habit formation first described by Hull in 1943 in his book, "Principles of Behavior." Hull’s work lead the way to several influential models of simple reinforcement-based learning including Bush and Mosteller’s (1951) simple model of instrumental conditioning, Estes (1950) and Estes and Burke’s (1953) famous stimulus sampling theory, and later the famous Rescorla-Wagner (1972) generalization of Bush and Mostellar’s earlier model. Each of these models differs from the next in terms of the clarity and rigor their mathematical development, their underlying psychological assumptions, and the particulars of their chosen lexicon. They also differ in terms of whether they were primarily motivated and tested against behavioral data from classical conditioning, or by behavioral data from instrumental conditioning. Nevertheless, they share the critical assumption that behavior is derived from stimulus-response associations that are learned on a trial-by-trial basis derived from feedback or reinforcement of some kind. It is interesting to note that this assumption, when placed in a dopamine mediated cortico-basal-ganglia network, perfectly defines the models described in this dissertation. In this light, it is perhaps not surprising that the mathematical form of the learning signal in our model (i.e., DA as it depends on predicated reward) is quite similar to the mathematical forms used to guide learning (or the development of habit strength) in each of the aforementioned models.

4.5 Clinical Relevance

Given that we have a biologically detailed model of procedural skill learning that specifies a mechanism that protects learning, an obvious question that has both theoretical and practical significance is: how can we use this model to design an effective unlearning protocol? The model hypothesizes that the key to unlearning in procedural skill tasks are the TANs. Specifically, the model predicts that unlearning can be effective only if the TANs pause during the unlearning training.
If the TANs do not pause, then the medium spiny neurons in the striatum will never see the stimuli (because the cortical input will be inhibited presynaptically), and the unlearning training will have no effect.

We propose two different methods for maintaining a TAN pause during unlearning training—one pharmacological and one behavioral. The pharmacological method would be to direct a muscarinic M2 receptor antagonist at the relevant striatal territory. Such an antagonist would bind to the relevant M2 receptors but have no postsynaptic inhibitory effect. The unlearning protocol could include either a traditional extinction phase in the case of instrumental conditioning, or a random feedback phase in the case of cognitive tasks like the one used in Experiment 1. In either case, the TANs would presumably cease pausing during the early unlearning trials. The increase in TAN firing associated with pause cessation would cause acetylcholine to be released at cortical-TAN synapses, but the critical M2 receptors would all be occupied by the antagonist. As a result, no binding sites for the acetylcholine would be available, so the TAN firing would have no inhibitory effect on the cortical input to the medium spiny neurons. Thus, an M2 antagonist would mimic a TAN pause, and the model therefore predicts that in the presence of such an antagonist, unlearning training should be successful. A behavioral consequence is that an M2 antagonist directed at the striatum should abolish fast reacquisition. To our knowledge this prediction has not been tested.

A number of studies have looked at the effect of M2 antagonists on various aspects of learning and memory (e.g., Buresova, Bures, Bohdanecky & Weiss, 1964), but we know of no studies that have asked whether an M2 antagonist directed at the striatum prevents fast reacquisition of a previously extinguished instrumental behavior.

Even if future pharmacological studies are consistent with our hypothesis, it will certainly be impractical to administer drugs directly into the striatum of human patients. Fortunately, the TANs are sensitive to the reward structure of the environment, so it might be possible to construct treatment protocols that are designed in such a way as to keep the TANs pausing throughout the extinction training. As we have seen in this dissertation, this is a promising approach both
from the instrumental conditioning and the II category learning perspectives. The key results from each of these fields seems to suggest that maintaining some degree of contingent rewarded responding during the extinction phase is paramount to true unlearning. Along these lines, Bouton, Woods, Moody, Sunsay, and Garca-Gutirrez (2006) reviewed a broad array of experimental manipulations designed to increase the effectiveness of extinction training (i.e., reduce relapse effects such as fast reacquisition and renewal). Although the majority of the evidence they reviewed was derived from classical fear-conditioning procedures, the insights they derived bear striking resemblance to the predictions derived from our models. Overall, it seems that the neurobiology detailed by our theory could be a very reasonable target for future unlearning therapy.
## Appendix A

### A.1 Fit Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation 2.1</td>
<td></td>
</tr>
<tr>
<td>$\omega(0)$</td>
<td>0.2</td>
</tr>
<tr>
<td>$\beta_s$</td>
<td>125</td>
</tr>
<tr>
<td>$E_s$</td>
<td>125</td>
</tr>
<tr>
<td>$\sigma_s$</td>
<td>5</td>
</tr>
<tr>
<td>Equation 2.5</td>
<td></td>
</tr>
<tr>
<td>$\alpha_G$</td>
<td>0.2</td>
</tr>
<tr>
<td>Equation 2.6</td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>100</td>
</tr>
<tr>
<td>Equation 2.7</td>
<td></td>
</tr>
<tr>
<td>$\beta_V$</td>
<td>0.275</td>
</tr>
<tr>
<td>Equation 2.8</td>
<td></td>
</tr>
<tr>
<td>$\beta_C$</td>
<td>0.275</td>
</tr>
<tr>
<td>$\omega_C$</td>
<td>0.275</td>
</tr>
<tr>
<td>$\sigma_C$</td>
<td>0.275</td>
</tr>
<tr>
<td>Equation 2.9</td>
<td></td>
</tr>
<tr>
<td>$\alpha_w$</td>
<td>0.275</td>
</tr>
<tr>
<td>$\beta_w$</td>
<td>0.275</td>
</tr>
<tr>
<td>$\gamma_w$</td>
<td>0.275</td>
</tr>
<tr>
<td>Equation 2.11</td>
<td></td>
</tr>
<tr>
<td>$\alpha_P$</td>
<td>0.275</td>
</tr>
</tbody>
</table>
A.2 Decision Bound Models

A.2.1 Rule-Based Models

The General Conjunctive Classifier (GCC). Three versions of the GCC (Ashby, 1992) were fit to the data. One version assumed that the rule used by participants is a conjunction of the type: Respond A if the length is short and the orientation is shallow (e.g., less than 45 degrees), respond B if the length is short and the orientation is steep (e.g., greater than 45 degrees), respond C if the length is long and the orientation is shallow, or respond D if the length is long and the orientation is steep. This version has 3 parameters: one for the single decision criterion placed along each stimulus dimension (one for orientation and one for bar width), and a perceptual noise variance. A second version assumed that the participant sets two criteria along the length dimension partitioning the lengths into short, medium, and long, and one criterion along the orientation dimension partitioning the orientations into shallow and steep. The following rule is then applied: Respond A if the length is short, respond B if the length is short and the orientation is steep, respond C if the length is short and the orientation is shallow, or respond D if the length is long. A third version assumed that the participant sets two criteria along the orientation dimension partitioning the orientations into shallow (e.g., less than 30 degrees), intermediate (e.g., between 30 and 60 degrees), and steep (e.g., greater than 60 degrees), and one criterion along the length dimension partitioning the lengths into long and short. The following rule is then applied: Respond A if the length is short and the orientation is intermediate, respond B if the orientation is steep, respond C if the length is long and the orientation is intermediate, or respond D if the orientation is shallow. The latter two versions of the model have four parameters: three decision criteria, and a perceptual noise variance. The assignments of category labels to response regions were modified in the appropriate manner when being applied to the label switch condition.
A.2.2 Information-Integration Models

The Striatal Pattern Classifier (SPC; Ashby & Waldron, 1999) has provided good fits to information-integration categorization data in a variety of previous studies (e.g., Ashby et al., 2001; Maddox, Molis, & Diehl, 2002). The model is outlined in detail in Ashby and Waldron (1999; Ashby et al., 2001), but is briefly reviewed here. Following COVIS (Ashby et al., 1998), the SPC assumes that information-integration category learning is a process of gradually associating a response with groups of cells in visual cortex. The response units are assumed to be located in the striatum. It is well established that there is a many-to-one mapping of cortical cells onto cells in the striatum (Wilson, 1995), and thus in the SPC the resolution of response space is considerably coarser than the resolution of perceptual space. The SPC is a simple model of information-integration category learning that is based on this architecture. It is important to be clear that the SPC is a computational model that is inspired by what is known about the neurobiology of the striatum. Even so, the striatal units should be considered hypothetical and could be interpreted within the language of some other computational model (e.g., as prototypes in a multiple prototype model). In the present applications, the SPC assumed one response unit per category for a total of 4 units and that each of these is associated with a single point in the two dimensional length-orientation space. Because the location of one unit can be set arbitrarily, the model has 6 free response-unit parameters. One additional noise parameter is also included for a total of 7 parameters. The optimal model is a special case of the SPC in which the striatal units are placed in such a way that the optimal decision bounds are used. The optimal model contains only one parameter (i.e., noise variance). Random Guessing Models

A.2.3 Random Responder Models

Fixed Random Responder Model This model assumed that the participant guessed randomly and that all responses were equally likely. Thus, the predicted probability of responding A, B, C, or D was .25. This model had no free parameters.
**General Random Responder Model** This model assumed random guessing, but that some responses were more likely than others. Thus, the predicted probabilities of responding A, B, C, and D were parameters that were constrained to sum to 1 (i.e., so this model had three free parameters).
Bibliography


