A NEUROCOMPUTATIONAL MODEL OF THE FUNCTIONAL ROLE OF DOPAMINE IN STIMULUS-RESPONSE TASK LEARNING AND PERFORMANCE

George L. Chadderdon III

Submitted to the faculty of the University Graduate School
In partial fulfillment of the requirements
For the degree
Doctor of Philosophy
In the Department of Psychological and Brain Sciences
And the Cognitive Science Program of
Indiana University
March, 2009

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____

Olaf Sporns, Ph.D.

_____

Joshua Brown, Ph.D.

_____

James Townsend, Ph.D.

_____

Peter Todd, Ph.D.

February 27, 2009

Acknowledgements

As the long, arduous process of preparing this doctoral thesis draws to a close, I am grateful for the support, ideas, patience, and encouragement of my committee, the university, and my family. Above all, I am grateful to my advisor Olaf Sporns for being patient over the years, and for providing, for me, an ideal balance of freedom to pursue my research interests with generous guidance, advice, and critical feedback. I would also like to thank Josh Brown: for letting me audit his relevant and useful class on cognitive control in the brain, for joining my committee at the last minute, and for considerable time and contribution of ideas during the stages of development of the model in this dissertation. I would also like to thank Jim Townsend for his continued support and encouragement, and Peter Todd who joined my committee when one of its previous members left IU. Much of my broad understanding of the gestalt organization of the brain I owe to Gary Lucas and his useful course co-taught with Bill Timberlake. My experience at Indiana University has been a positive one, and I am grateful for the friendly and intellectually stimulating environment provided by the Bloomington campus.

To my parents, I am grateful always for their moral and financial support, and for their enthusiastic encouragement of my pursuit of my many diverse interests.

Finally, I would like to thank the National Institute of Mental Health and IU's Psychological and Brain Sciences department and Cognitive Science program for their financial support during my stay at IU. For two years, I was funded as a fellow on the NIMH Math Modeling Training Grant T32MH019879-11. Over two summers, I was supported under an IU Cognitive Science Program Summer Research Fellowship.

George L. Chadderdon III

A NEUROCOMPUTATIONAL MODEL OF THE FUNCTIONAL ROLE OF DOPAMINE IN STIMULUS-RESPONSE TASK LEARNING AND PERFORMANCE

The neuromodulatory neurotransmitter dopamine (DA) plays a complex, but central role in the learning and performance of stimulus-response (S-R) behaviors. Studies have implicated DA's role in reward-driven learning and also its role in setting the overall level of vigor or frequency of response. Here, a neurocomputational model is developed which models DA's influence on a set of brain regions believed to be involved in the learning and execution of S-R tasks, including frontal cortex, basal ganglia, and cingulate cortex. An 'actor' component of the model is trained, using 'babble' (random behavior selection) and 'critic' (rewarding and punishing) components of the model, to perform acceptance/rejection responses upon presentation of color stimuli in the context of recently presented auditory tones. The model behaves like an autonomous organism learning (and relearning) through 'trial-and-error'. The focus of the study, the impact of hypo- and hyper-normal DA activity on this model, is investigated by three different dopaminergic pathways—two striatal and one prefrontal cortical—being manipulated independently during the learning and performance of the color response task. Hypo-DA conditions, analogous to Parkinsonism, cause slowing and reduction of frequency of learned responses, and, at extremes, degrade the learning (either initial or reversal) of the task. Hyper-DA conditions, analogous to psychostimulant effects, cause more rapid response times, but also can lead to perseveration of incorrect learning of response on the task. The presence of these effects often depends on which DA-ergic pathway is manipulated, however, which has implications for interpretation of the pharmacological

experimental data. The proposed model embodies an integrative theory of dopamine function which suggests that the base rate of DA cell activity encodes the overall 'activity-oriented motivation' of the organism, with hunger and/or expectation of reward driving both response vigor and tendency to generate an explorative 'babble' response. This more 'tonic' feature of DA functionality coexists naturally with the more extensively-studied 'phasic' reward-learning features. The model may provide better insights on the role of DA system dysfunction in the cognitive and motivational symptoms of disorders such as Parkinsonism, psychostimulant abuse, ADHD, OCD, and schizophrenia, accounting for deficits in both learning and performance of tasks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Motivation and Background

1.1 Summary of Investigation

1.1.1 Task-Oriented Behavior Selection and Volition

Production of behavior that meets survival and reproductive needs of the organism is the fundamental goal of all animal cognition. Organisms in their environments seek to engage in behaviors which lead to rewards (e.g. food, drink, mating) and avoid punishments (e.g. pain, injury). Positive *outcomes* (O; all abbreviations listed in Table 1.1) are often contingent on making a correct *response* (R) to particular *stimulus* (S) cues in the environment. The essential problem of *volition / executive control* is how to most *adaptively*—meaning most advantageously for the organism—select an R from the animal's entire repertoire of candidate behaviors.

In addition to immediate S cues, however, the organism's best R also tends to depend on the current internal state of the organism: its bodily state or the goal it is currently trying to accomplish (e.g. nest-building, hunting, fleeing from predators). The selection of an appropriate behavior in the context of the organism's current environment and internal state is the essence of *adaptive decision-making*. In previous work (2008), the author has defined 'volition' as "the capacity for adaptive decision-making" and suggested that this capacity is possessed in varying degrees by organisms. Increasing volition means an increasing flexibility of (and internalization of) control of behaviors.

In (Chadderdon, 2008), the author presents an ordinal scale intended to measure volition in organisms (either natural or artificial). Levels of this scale are characterized by the degree of flexibility of control the organism/system has. At a low level of volition are systems that have only "hard-wired" reflex behaviors. Systems capable of learning

from experience, however, possess a higher degree of volition. Associative learning allows learning of new adaptive behavior patterns and also the informational structure of the environment. Possession of working memory (as will be discussed in more detail) confers still more flexibility, as does the faculty of long-term memory. The ability to deliberate on candidate Rs and contemplate their likely Os before committing to an action is a feature that adds significant volitional capability to an organism. Finally, environmental bootstrapping—including social behavior, object manipulation, symbolic communication, and external storage of symbols—allows still higher levels of volition.

Even within the same organism, however, behaviors may be subject to varying levels of control. Simple "hard-wired" fixed stimulus-response (S-R) reflexes, such as spinal pain limb-withdrawal reflexes, or swallow responses, represent a low level of volitional control, behavior control through collections of instinctual tropisms. Associative learning, however, allows a higher level of volitional control because it allows learning of novel S-R mappings attuned to the conditions in the organism's environment. The inherent plasticity of neurons endows even evolutionarily primitive animals such as sea-slugs with some degree of S-R learning capacity.

The ability to learn new S-R mappings is an important component of an animal's ability to adapt to its environment. However, there are complicating factors in how the world may distribute rewards that may make simple S-R learning insufficient. First, an S cue that prompts an appropriate R may be transient. For example, a predator may spot a prey animal only to have it disappear into hiding nearby. It would be highly maladaptive for the predator to forget about the existence of the prey animal, and would be more

adaptive to engage some kind of searching or tracking behavior, at very least searching where the quarry was last seen.

Another possible complication is that there may be a many-to-many mapping between particular S cues and most-adaptive Rs. When seeing a prowling hyena, a monkey's best escape R may depend on whether there are trees around or whether they are out in the open. The most adaptive R, given the immediate S, will depend on other cues, sometimes called *discriminative stimuli* (S^Ds): the location of the monkey in the previous example. In a more difficult case, these cues may themselves be transient, as when the vocalization of a predator is heard. These types of responding require some kind of internal representation to be formed and held active during the time after critical environmental cues have disappeared. These are the kinds of reinforcement conditions in the environment that likely create the selective pressure for the development of *task-oriented behavior selection* (TOBS).

We may imagine an experiment in which a rat learns a simple S-R mapping in which food-pellets are always delivered after it presses a lever (R) in response to the onset of a light (S). However, we could imagine making the reward contingency more complex by adding two auditory tones (high vs. low) which change whether the rat is administered a food-pellet (reward), or a footshock (punisher) in response to its lever press at the onset of the light. The tones are played between blocks of light/lever trials to signal changes in reward contingency state in the environment. These tones may be thought of as signaling for the rat two distinct *tasks*: PRESS-LEVER-ON-LIGHT, or AVOID-LEVER-ON-LIGHT. Some form of what has been called *working memory* (Baddeley, 2003) is required for the rat to maintain a representation of the current reward

contingency state of the environment. The selection of a behavior contingent on a current task is what is meant by TOBS (Chadderdon & Sporns, 2006). Working memory may also be used to temporarily store *task parameters*: including information specifying subgoals, stimulus targets, or specific means or manner of performing the task.

If we consider the kinds of long-range behaviors humans engage in, we see that TOBS is a critical component in adaptation to civilization. For example, when we are in our cars, we need to have some kind of representation of our intended destination, lest we end up somewhere else by default. When engaged in a sequential task such as adding two numbers together in our heads, we need to be able to remember what steps we've completed as well as the final goal. Generally, humans are goal-oriented creatures whose environment requires the kind of behavior involving the engaging and disengaging of task states that override default behavior.

A laboratory example of TOBS in humans which illustrates the problem of many-to-many S-to-best-R mappings is the Stroop task in which the subject is instructed to either read the text of words presented in colored ink or the color of the ink instead (Stroop, 1935). If the word 'blue' is written in red ink, then the correct response will depend on which task (READ-WORD vs. READ-COLOR) is active. Clearly, the READ-WORD task is more of a default task and must be overridden in order for correct responses to be made, a fact which causes errors and slower performance of the READ-COLOR task when its answers conflict with those of READ-WORD. The ability to use internally maintained task context to direct behavior represents an increase in behavior flexibility over fixed S-R mapping. Failure of the working memory system that would

allow task/goal representation would result in more 'stimulus-driven' behavior such as has been observed in patients with frontal-lobe damage (Miller, 2000) .

1.1.2 Reinforcement Learning

Seeing that TOBS is an important feature of animal behavior, we may wonder how an organism learns the particular S-R mappings for each task and the discriminative stimulus cues (S^Ds) that signal switching to/between the tasks. Animals can be trained to perform behaviors through conditioning procedures. *Classical (Pavlovian) conditioning* procedures (Rescorla, 1988) traditionally involved learning to associate a previously neutral stimulus, a *conditioned stimulus* (CS) with another stimulus, the *unconditioned stimulus* (US) which tends to produce a corresponding *unconditioned response* (UR). Through this association, the CS is able to trigger a *conditioned response* (CR). More recent conceptualizations of Pavlovian learning (Rescorla, 1988, 1991) characterize it as stimulus-outcome (S-O) learning whereby an association is learned between one event (the S) and a subsequent event (the O) that it predicts/precedes. Presumably, it is the O that drives the conditioned R, though some stimulus context may determine the exact nature of the R. For example, a dog in the classic Pavlov experiment is learning to map a bell (S) to an expectation of food (O) which leads to salivation (R).

When the O does not automatically follow the S, however, but is contingent on a particular R, then *instrumental (operant) conditioning* (Rescorla, 1991) procedures are used to train the subject. For example, a rat may be required to press a lever (R) in order to receive a food-pellet reward (O). Essentially, instrumental learning involves delivery (or omission) of rewarding and/or punishing stimuli (Os) to "stamp in" or "stamp out" an R (Thorndike, 1911), potentially in some environmental context (S). The basic pattern is

as follows. First, the subject engages spontaneously in a behavior (R) during a particular environmental context (S). Then, the rewarding or punishing consequences (O) of that action are relayed to the subject. According the Law of Effect (Thorndike, 1911), if a reward (punisher) is delivered, then the likelihood of the subject choosing that particular behavior in a similar environmental context is increased (decreased). So the subject learns to choose Rs that maximize reward and minimize punishment. Sutton and Barto (1998) have characterized *reinforcement learning* (a theoretical abstraction of animal instrumental learning) as "…learning what to do—how to map situations to actions—so as to maximize a numerical reward signal."

Instrumental learning has been theorized to involve learning of a hierarchical S-(R-O) mapping (Rescorla, 1991), but this dissertation investigates the neural mechanisms that might underlie the (conceptually more traditional) theory of S-R mappings being trained by the Os. There is evidence, in fact, that some behaviors (though not as many as was originally believed) are cast as direct S-R mappings, rather, that are independent of any immediate outcome expectancy (Kirsch, Lynn, Vigorito, & Miller, 2004) . These behaviors are *habitual*, and resistant to *extinction* (unlearning caused by omission of rewards). Thus, the model to be offered in this dissertation may be considered a preliminary model of habit-learning, albeit habit learning that relies on working memory traces to keep track of a task state.

It is proposed here that both fixed S-R mappings and more context-dependent mappings, such as would be required for TOBS, may be learned via the same learning mechanisms. In essence, the main stimulus (S) and the task context (S^D) may be considered a conjunctive stimulus (S′) such that each stimulus/context conjunction may

be mapped to an R. A neurocomputational model has been constructed that models the learning (and unlearning) of TOBS S-R mappings. Rewards and punishers, respectively, allow learning and unlearning of stimulus/context conjunction-to-response mappings, but effects of extinction (omission of reward) are not modeled.

1.1.3 Dopamine

Neuromodulatory neurotransmitters such as dopamine (DA) provide a global mechanism by which network- and system-level neural dynamics can be altered. (Marder & Thirumalai, 2002) contains a broad discussion of the effects of neuromodulators on the intrinsic firing properties and the synaptic plasticity of neurons. Neuromodulators adjust the maximum conductances for particular ion channels in cells which can, for example, convert them from tonically inactive cells that require afferent input to tonically active cells or bursting cells. Synaptic strength can also be affected by either pre- or post-synaptic effects of neuromodulator release. Multiple neuromodulators can interact with one another leading to hard-to-predict results. Overall, the computational consequences of neuromodulator effects are bewilderingly variable and the effects of the individual transmitters are in principle not easily disentangled.

However, there do seem to be certain large-scale functions associated with DA in mammalian brains. DA is a critical component in the circuitry allowing voluntary behavior, as is evidenced by the difficulty that Parkinson's patients—who suffer DA depletion due to death of dopaminergic cells in midbrain nuclei—have initiating and rapidly executing movements (Gauntlett-Gilbert & Brown, 1998; Muller et al., 1999; Schultz et al., 1989). There is evidence that heightened levels of DA, such as may be caused by psychostimulant (e.g. cocaine or amphetamine) use, may lead to both shorter

reaction times (Halliday et al., 1994; Hienz, Spear, & Bowers, 1994) and a greater degree of explorative activity (Carr & White, 1987; Ikemoto & Panksepp, 1999). DA activity is involved with working memory maintenance (Durstewitz, Seamans, & Sejnowski, 2000; Sawaguchi & Goldman-Rakic, 1994; Zahrt, Taylor, Mathew, & Arnsten, 1997) which is a critical component of executive control. Finally, DA seems to play a significant role in reinforcement learning as is evidenced both by cellular studies of how DA modulates synaptic plasticity (Reynolds & Wickens, 2002; Shen, Flajolet, Greengard, & Surmeier, 2008) and by the conditions which prompt DA cell firing. Studies suggest that DA cells tend to fire in response to novelty and the unexpected delivery of rewards, but their firing tends to be suppressed by omission of expected rewards and aversive stimuli (Schultz, 1998, 2007; Ungless, Magill, & Bolam, 2004). More details about the likely neural mechanisms underlying many of these effects will be discussed in Section 1.2.3.

Caution is warranted when attempting to generalize about the function of any particular neurotransmitter such as DA, but it may be illuminating to try to integrate what is understood about the disparate functions DA is involved in and what is known about the circuitry the DA cells are a part of. Certain generalizations can and have been made about neurotransmitter functions which add greatly to our intuitions about their likely roles in the brain. For example, catecholamines such as norepinephrine (NE) and DA seem to be involved in arousal, whereas indoleamines such as serotonin (5-HT) seem to, generally speaking, inhibit or modulate the effects of arousal (Panksepp, 1986, 1998).

One of the larger aims of this dissertation is to attempt to make such a generalization about the function of DA: namely, it proposes that *DA cell base-rate activity correlates with an organism's current 'activity-oriented motivation' and upon*

this tonic *signal is superimposed a* phasic *reinforcement signal wherein large bursts of activity, relative to the baseline firing, signal associative learning through reward (or novelty); and significant dips of activity signal associative unlearning through punishment*. The superposition of these signals has certain implications for predictions we might make about the effects of abnormal levels of DA and their effects on cognitive and executive function.

Chapter 8 of Jaak Panksepp's *Affective Neuroscience* (Panksepp, 1998) suggests a precedent for such a unified understanding of the role of DA. Panksepp proposes that DA is the key neurotransmitter involved in the activation of a SEEKING affective system, i.e., an appetitive motivational emotional circuit involved in a generalized energizing of explorative and foraging behavior. The lateral hypothalamus (LH) signals physiological needs such as hunger or thirst, and these trigger activation of DA cells in the ventral tegmental area (VTA), one of the important midbrain DA nuclei. VTA activity targets areas such the ventral striatum (nucleus accumbens) and this leads to an increase in explorative behaviors, such as sniffing and forward locomotion in rats. Activation of this circuit is believed to be associated with a subjective experience of anticipatory excitement, rather than hedonic pleasure that would be associated with consumption/consummation. Physiological needs, such as hunger, may activate this system, but it is also possible for the system to learn to activate in response to initially neutral stimulus cues, e.g. a tone preceding feeding. Drug-cravings in humans may work via such a mechanism with particular environmental cues such as the sight of drug-use paraphernalia potentially triggering the activation of this expectancy/anticipatory emotional state, even after physical withdrawal has been overcome (Hyman & Malenka,

2001). The SEEKING system activation is non-drive-specific: hunger, thirst, and sexual desire all trigger the same system, and it has been observed that activation through one drive (e.g. hunger) can lead to increased consummatory behaviors related to other drives (e.g. thirst).

While the other major dopaminergic pathway to the basal ganglia, the substantia nigra-to-dorsal striatal pathway (to be discussed later) is not active in exactly the same circumstances as the ventral striatal pathway, the evidence regarding effects of Parkinson's suggests that DA facilitates motor activity through this pathway also, and as depletion of the ventral striatal pathway leads to decreased exploratory behavior (Ikemoto & Panksepp, 1999), depletion of dorsal striatal pathway leads to decreased performance of more stimulus-specific learned behaviors (Packard & Knowlton, 2002; Yin, Knowlton, & Balleine, 2004). In both cases, base-rate of DA cell activity seems to correlate with a drive to be active, rather than inactive. It is as if the base-rate of DA cell activity signals the overall need of the organism to engage in purposive, voluntary behavior rather than remaining quiescent. Under this assumption, low dopamine should correlate with states of both affective and psychomotor sluggishness, whereas high DA states correlate with heightened subjective sense of excited expectancy and the motivation to "do something" and with physical hyperactivity. The phasic learning signals, then, allow behaviors (Rs) to be associated with cues (Ss) that lead to positive outcomes (Os), and also allow S-R mappings that lead to negative Os to be unlearned. Thus, the same DA cells that energize the organism for voluntary action provide a reinforcement learning signal for training the organism to perform those actions. An interesting implication is that in the energized state there is a bias towards reward learning, whereas in the 'down' state there is a bias

towards punishment learning, a proposition that has been made by Frank and colleagues in their modeling (Frank & O'Reilly, 2006; Frank, Seeberger, & O'Reilly, 2004).

The model developed in this research attempts to formulate a preliminary understanding of dopamine's role in reinforcement learning and execution of TOBS. Of particular interest are the effects of too little or too much dopamine activity in the neural substrates of TOBS. More details of the mechanisms of dopamine are discussed in Section 1.2.3, and wider implications of the theory are discussed in Section 4.1.4.

1.1.4 Objectives and Research Questions

From the outset, the research documented in this dissertation has sought an explanatory computational simulation model of TOBS. Two main objectives are:

1. Creation of a large-scale neurocomputational model of the neural mechanisms and pathways involved in the learning of S-R tasks;
2. Modeling DA's role in modulating learning and performance of S-R tasks with emphasis on examining the effects of DA agonism (hyper-DA) and antagonism (hypo-DA).

Three research questions, essentially, are investigated:

1. What is the neural substrate of TOBS?
2. How are TOBS behaviors learned by this substrate?
3. What role does the neurotransmitter DA play in the learning and performance of these behaviors?

In Section 4.1, a preliminary theory, suggested by the operation of the model, will be offered addressing these questions.

1.1.5 Overview of Research Approach

Understanding the neural and functional mechanisms of mammalian behavior and cognition is one of the larger goals of neuroscience and psychology, since much of this understanding may reveal the physiological and informational basis of human cognition and behavior. A plethora of techniques exists for collecting relevant data including neuroimaging (e.g., EEG, MEG, PET, fMRI), lesion case studies, and, more recently, transcranial magnetic stimulation in humans; and cell recordings, lesions, and pharmacological manipulations in non-human animals (Gazzaniga, Ivry, & Mangun, 2002). Unfortunately, each of these techniques suffers from a limitation of scope or viewpoint. Cell recordings provide excellent temporal resolution, but give only a sense of the behavior of a few sampled cells, rather than a network as a whole. fMRI provides an overall collective view of brain activity, but with a relatively poor temporal resolution and the spatial resolution is also much coarser than the level of individual cells. Lesions suggest localized correlations of damage with particular psychological and behavioral dysfunctions, but they may also cause disruptions by interfering with functionality of surrounding neural tissue. Currently, no one empirical technique presents a complete enough pattern to build detailed theories of mechanism on.

Because of this situation, the study of brain mechanisms underlying mental process is still at an essentially pioneering, exploratory stage. A vast amount of data exists that needs to be integrated into at least provisional theories that can provide a coherent explanatory model of functionality. Theoretical neuroscience needs to construct schemas, and assign functions, at least tentatively, to specific anatomical areas. These integrative theories, in turn, may suggest hypotheses that can be tested by the empirical

neuroscience methods, and may allow researchers to cross-validate findings in the disparate methodologies.

Computational modeling is a theoretical method that has been effectively used in recent decades and continues to be a promising approach to neuroscientific inquiry (Arbib, Erdi, & Szentagothai, 1998; Churchland & Sejnowski, 1992; Dayan & Abbott, 2001; O'Reilly & Munakata, 2000). Computational simulation provides theorists with an excellent medium for formulating their theoretical constructs. Unlike actual neural systems, neurocomputational models allow perfect information of their internal state. Activation and synaptic strengths during the execution of the model can be recorded for all neurons in the model (assuming sufficient data storage capacity) for analysis. In single-cell recording or neuroimaging, by contrast, one is limited by the particular neurons sampled from on the one hand, and by the spatial coarseness of the measured aggregate activity on the other.

Forced attention to mechanistic details is another advantage. In order to create a working simulation, details of mechanism that might have been overlooked or ignored may need to be fleshed out, and this may provide a highly concrete structure whose validity can be tested quantitatively. An additional benefit of a neurocomputational modeling approach is that it suggests specific mechanisms of cognition and behavior in artificial systems, potentially yielding significant advances in the field of artificial intelligence. This specificity of implementation is expensive, perhaps, in design time and perhaps immediately in terms of how well the model may accurately reflect brain functionality, for each proposed specific mechanism adds to the likelihood of the model disagreeing on some points with later empirical findings. However, these disadvantages

are more than offset by the fact that the specific mechanisms hypothesized may suggest later candidate mechanisms that are closer to the true ones.

A good computational model may serve as a theoretical guidepost, even if it is wrong on some of the details. As the field of cognitive science is in its youth, a proliferation of models and theories seems appropriate and useful, with the proviso that the theories should be considered as works in progress. As the field matures and more data is assimilated, the population of theories should show some convergence. The final theories in this process, the ones which hopefully explain the detailed functionality of brain process, seem likely to disagree in some fashion with most current models, but are likely to at least retain some of the elements of those theoretical constructs.

In light of the advantages of using a neurocomputational modeling approach, this research has aimed to create a neurocomputational simulation model of learning and performance of TOBS. The design of the model draws upon the knowledge of what is known about the neural pathways of executive control and reinforcement learning, the cellular mechanisms of learning, and DA's role in modulating behavior and learning. Using this model, the dissertation attempts to formulate a preliminary theory of TOBS, and to explain and make predictions about effects of hypo- and hyper-DA pharmacological manipulations on TOBS learning and performance.

1.2 Review of Neuroscience Literature

Before creating a model, it was first necessary to review relevant literature regarding neural mechanisms of executive control and reinforcement learning, and the role DA plays in these mechanisms. Broadly speaking, the modeling efforts have focused on four regions of mammalian brain that are involved with TOBS: the frontal

lobe of the cerebral neocortex, the basal ganglia, the midbrain DA nuclei, and the anterior cingulate cortex (ACC). The theory proposed in Section 4.1, based on the model, is primarily a theory of the "division of labor" between these areas in mediating TOBS.

Before discussing each of these regions in more detail, however, a summary may be given of the role of each of these areas proposed by the model of the dissertation. Frontal neocortex is the main component of the 'actor' of the model, maintaining working memory states about present and remembered stimuli and tasks, and activating learned task-appropriate behaviors: i.e., performing the S-R mapping. The midbrain dopamine nuclei act as a 'critic', signaling rewarding and punishing events that train the actor when it performs a correct or incorrect response. They also set of the level the *activity-oriented motivation* of the organism. A portion of ACC monitors the overall satisfaction state of the organism, firing increasingly when the organism is becoming more frustrated with not being rewarded. Another portion of ACC is also involved with initiation of random, exploratory 'babble' behaviors when the organism is in a state of high activity-oriented motivation. Finally, the basal ganglia provides a gating mechanism for the actor and babble pathways, with the permissiveness of the gate set by the DA-signaled activity-oriented motivation.

1.2.1 Frontal Neocortex

1.2.1.1 Overall Structure and Function of Frontal Neocortex

A decorticate animal—i.e., one with its neocortex removed—is capable of classical and operant conditioning and able to perform complex instinctual behaviors such as grooming or copulation, but has difficulty learning complex discriminations, planning, or learning to navigate a complex environment (Kolb & Whishaw, 2003). The

neocortex may be thought of as an evolutionarily later layer of the brain which allows increased flexibility in adapting more instinctive behaviors to novel situations, permitting what is generally considered "higher" cognitive functioning (MacLean, 1990).

While the posterior portions of neocortex and the temporal lobe seem to chiefly be involved with sensory processing and memory, the frontal cortex seems to be devoted to matters related to control of behavior: motor processing and executive control (Kolb & Whishaw, 2003; Luria, 1973). The frontal cortex can be thought of as divided into a primary motor, a premotor, and a prefrontal component.

1.2.1.2 Primary and Premotor Cortices

Primary motor cortex (M1), located in Brodmann area (BA) 4 (see Figure 1.1 for BA map), synapses directly with spinal neurons and is responsible for the execution of simple motions (Kolb & Whishaw, 2003). It is organized in a coarsely somatotopic mapping in broad face, upper limb, and lower limb regions, but there is not a one-to-one mapping within these regions, but evidence suggests, rather, a distributed convergence-divergence pattern between M1 neurons and the controlled muscles within an extremity (Schieber, 2001).

The premotor region of the frontal cortex consists of BA 6, with the lateral portion being considered the premotor cortex (PMC) and the medial portion being the supplementary motor area (SMA); and BA 8, whose lateral and medial portions, respectively, are the frontal eye fields (FEF) and supplementary eye fields (SEF) and are involved with execution and planning of eye-movements. Whereas M1 is believed to be more associated with simple, immediate movements, PMC and SMA are more associated with control and coordination of movements. PMC is believed to be involved in

preparation for and sensory guidance of movement (Wise, 1985). Evidence suggests that the dorsal portion of PMC may be involved with translating (visually cued) working memory instructions into motor sequences (Ohbayashi, Ohki, & Miyashita, 2003). For PMC, the emphasis seems to be on learning mappings between external cues and behaviors (Deiber et al., 2004; Mitz, Godschalk, & Wise, 1991). By comparison, SMA is believed to be more involved with internally cued, self-paced, voluntary behaviors (Cunnington, Bradshaw, & Iansek, 1996; Deiber et al., 2004; Passingham, 1993).

1.2.1.3 Prefrontal Cortex

There seems to be some variance in the literature on how the remaining portion of frontal cortex, the prefrontal cortex (PFC), is anatomically defined and subdivided (see (Kolb & Whishaw, 2003; Miller & Cohen, 2001; Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004) for some example schemes; see Figure 1.2 for Miller and Cohen's (2001)). One major difference is that some schemes include the anterior cingulate cortex (ACC), BA 24, 25, and 32, in PFC (Kolb & Whishaw, 2003; Ridderinkhof, van den Wildenberg et al., 2004), and others do not (Miller & Cohen, 2001). This dissertation will largely follow Ridderinkof and colleagues' (2004) conventions. Thus, the lateral portion of the PFC is divided up into a dorsolateral portion (dlPFC: BA 9 and 46), a ventrolateral portion (vlPFC: BA 44 and 45), and an inferior frontal junction portion (IFJ: junction of BA 8, 6, and 44). The ventromedial portion of PFC, the orbitofrontal cortex (OFC) consists of BA 10 (also called the frontopolar cortex), BA 11, 13, 14, and 47/12. The remaining medial portions of PFC consist of the dorsomedial PFC (dmPFC: medial part of BA 9), and the ACC (BA 24, 25, and 32).

PFC is generally said to be involved with executive control and working memory (Funahashi, 2001; Funahashi, Bruce, & Goldman-Rakic, 1989; Fuster, 1973; Miller, 2000; Miller & Cohen, 2001; Miller, Erickson, & Desimone, 1996). For this role, it is advantageously centrally connected to read in a wide array of inputs, and exert an influence on behavior through premotor and motor outputs (see Figure 1.2). The diverse inputs to PFC (Kaufer & Lewis, 1998; Kolb & Whishaw, 2003; Miller & Cohen, 2001) include

- posterior parietal cortex (visual spatial location)
- inferior temporal cortex (visual object/feature identity)
- superior temporal gyrus (auditory information)
- caudal parietal lobe (somatosensation)
- gustatory cortex in insula
- olfactory regions of the pyriform cortex
- rostral superior temporal sulcus (multimodal representations, maybe semantic memory)
- hippocampus (episodic and semantic memory)
- amygdala (emotional and internal drive information).

In addition, PFC outputs to secondary motor areas, which means that, because of reciprocal connections, the secondary motor areas can inform the PFC of ongoing motor plans. So PFC is well-placed to represent abstractions of conjunctions of stimuli, internal state, recalled declarative memory traces, and ongoing motor behavior.

OFC is involved with context-dependent mapping of stimuli to reinforcements (S-O learning), and so allows the organism to better control reward- and punishment-related

behavior according to the current environmental context (Rolls, 1999). Of greater interest in this dissertation, however, are the more lateral components of PFC (dlPFC and vlPFC) for these receive visual and auditory information from the parietal and temporal lobes (see Figure 1.2). (ACC's role will be treated separately in Section 1.2.4.)

PFC cellular activity is believed to be involved with the online (i.e. through working memory activation) representation of abstract rules (Miller & Cohen, 2001; Wallis, Anderson, & Miller, 2001) and Rougier and colleagues (Rougier, Noelle, Braver, Cohen, & O'Reilly, 2005) have developed a reinforcement learning model where PFC cells learn rules for which stimulus feature to attend to, and these, through top-down activation, bias the activation in the posterior cortex-to-motor output pathway of the model. It is an example of what Miller and Cohen (2001) have suggested is the fundamental mechanism of PFC influence: delay (i.e., working memory) activity in PFC cells, through top-down feedback connections to other areas of brain, including the posterior cortex and secondary motor areas, adds a biasing activation that temporarily changes the input-output mapping. Working memory keeps track of changing cognitive context state and biases the ordinarily more automatic input-output pathways, so that, for example, in the Stroop task, the color-reading instruction may bias the input-output pathways temporarily to facilitate the correct responses instead of the default word-reading responses.

1.2.2 Basal Ganglia

1.2.2.1 Basal Ganglia as Crucial Area for Behavioral Organization

Beneath the neocortex lie evolutionarily older regions of brain that are critical to mammalian behavior, both instinctual and learned. The basal ganglia (BG; see Figure

1.3) are a set of structures that are critically involved with motor control in vertebrates (Grillner, Hellgren, Ménard, Saitoh, & Wikström, 2005; Redgrave, Prescott, & Gurney, 1999a), even ones as primitive as lampreys (Grillner, 2003). Diencephalic animals—those deprived of both neocortex and basal ganglia, but retaining the diencephalon, including thalamus and hypothalamus—exhibit affective displays and responses to stimuli, and hyperactivity, including hyperactivity in locomotion, and can be fed with effort, but generally their behavior is aimless and uncoordinated (Kolb & Whishaw, 2003). Decorticate animals with basal ganglia intact, however, are able to learn to crudely forage and link adaptive behaviors together into sequences. Paul MacLean (1990) in his triune brain theory regarded the BG as the chief component of the 'reptilian brain', more fundamental than the limbic system areas that are involved in mammalian emotion. Panksepp (1998, p. 70) quotes an early neurophilosopher (unspecified) as saying, "the royal road to the soul goes through the corpus striatum." The BG seems to be important both for the execution of instinctual, stereotypic fixed action patterns (Berridge, Aldridge, Houchard, & Zhuang, 2005; Greenberg, 2003), and for learned habitual behaviors (Graybiel, 1998; Packard & Knowlton, 2002).

1.2.2.2 Action Selection and the Braking Release Mechanism

The BG have been implicated in many functions, but one influential theory that seems to explain the generality of basal ganglia function is that it is involved in action selection: "[the selection of] some actions/motor programmes at the expense of others" (Redgrave et al., 1999a) so that conflicts are resolved between systems competing for the same output resource, as for example, when an organism has cues simultaneously to perform arm movements in different directions. The basal ganglia doesn't operate

through selective excitation of behavior, but rather seems to work through a peculiar "selective braking release" mechanism (Gurney, Prescott, & Redgrave, 2001; Mink, 1996; Wichmann & DeLong, 1996). Mink (1996) expresses the idea as follows: "The hypothesis states that the basal ganglia do not generate movements. Instead, when voluntary movement is generated by cerebral cortical and cerebellar mechanisms, the basal ganglia act broadly to inhibit competing motor mechanisms that would otherwise interfere with the desired movement. Simultaneously, inhibition is removed focally from the desired motor mechanisms to allow that movement to proceed."

Figure 1.4, borrowed from (Gurney et al., 2001), summarizes a possible anatomical circuit implementing the action selection described above. Inputs to the basal ganglia such as the cortex or limbic areas excite the input area of the BG, the striatum (caudate nucleus and putamen and ventral striatum). The main pathway, sometimes referred to as the "direct pathway" (Wichmann & DeLong, 1996) or "Go pathway" (Frank et al., 2004; O'Reilly & Frank, 2006), runs through a set of striatal cells inhibiting the output areas of the BG, the internal segment of the globus pallidus (GPi) and the substanti nigra pars reticulata (SNr). The GPi/SNr cells fire in a tonically inhibitory fashion, and all of the feedforward striatal cells are GABAergic (inhibitory), so the direct pathway disinhibits the areas innervated by the BG output cells, such as the thalamus or areas in the midbrain or brainstem. Thus, striatal activity through the Go pathway selectively disinhibits particular motor programs or actions. The "hyper-direct pathway" running from cortical, etc., into the subthalamic nucleus (STN) is excitatory and diffuse and leads to widespread excitation of the GPi/SNr cells which is probably used for an umbrella of 'default' inhibition around most of the competing behaviors in an active area

(Mink, 1996). Go pathway activation essentially "pokes a hole" in this inhibition for the desired behavior. The "indirect" or "NoGo" pathway, however, is likely to be used to "veto" undesired behaviors. This runs from afferents of the BG to another set of striatal cells which, when active, inhibit the external segment of the globus pallidus (GPe). These cells, by default, tonically inhibit both GPi/SNr and STN, so activation of the NoGo pathway disinhibits inhibition of efferents of the BG. Two different families of DA receptors, D1 and D2, are believed to modulate the activity of the Go and NoGo pathways, respectively (Gerfen, 1992). Approximately speaking, DA has an excitatory effect on the D1-dominated cells and an inhibitory effect on the D2-dominated cells (Gurney et al., 2001; Hernández-López, Bargas, Surmeier, Reyes, & Galarraga, 1997; Hernández-López et al., 2000). Because of this, dopamine release would tend to excite the Go pathway and suppress the NoGo pathway. According to this model, the akinesia and bradykinesia of Parkinson's can be explained, then, as depleted DA causing there to be sluggish Go pathway and disinhibited NoGo pathway activity (Wichmann & DeLong, 1996). A number of recent neurocomputational models of basal ganglia action selection make use of this Go/NoGo pathway principle (Brown, Bullock, & Grossberg, 2004; Frank et al., 2004; O'Reilly & Frank, 2006), and the model in this dissertation follows suit.

1.2.2.3 Diverse Parallel Corticostriatal Selection/Gating Pathways

An input-output circuit such as described above is instantiated myriad times in the BG, with different inputs and outputs. Some efferents of the BG outputs are subcortical: midbrain and even brainstem nuclei, such as are involved with basic motor programs (Grillner et al., 2005). The neocortex, however, is also thoroughly connected, through

the thalamus, to the basal ganglia, so that it, too, falls under the inhibitory-release control of the BG. As a result, the BG participates not only in motor control, but also cognitive and affective control (Middleton & Strick, 2000). The precise input-output mapping of these cortico-basal ganglial pathways it is still a topic of investigation, but one currently prevalent theory is that the BG consists of a number of segregated, parallel thalamocortical loops, each performing their action selection on a different information pathway (G. E. Alexander, DeLong, & Strick, 1986; Middleton & Strick, 2001). In (Middleton & Strick, 2001), there are proposed to be several sets of loops, based on anatomical tracing evidence. There are skeletomotor loops regulating the motor and premotor areas of the frontal cortex: M1, SMA, and PMC (the ventral component). There are oculomotor loops controlling FEF and possibly SEF. There are separate dorsolateral PFC loops that control BA 9 and BA 46, and these are likely to be important in planning and spatial working memory (Middleton & Strick, 2000). There are lateral orbitofrontal loops (mainly BA 12) which are probably associated with object working memory function. There are likely to be medial OFC loops (mainly BA 13), and ACC loops (both motor- (BA 24c) and limbic-related (BA 25,32)). Finally, there are likely to be BG loops associated with inferotemporal (IT) and posterior parietal cortical (PPC) areas which are involved in object recognition and spatial perception, respectively (Milner & Goodale, 1998).

Another way of characterizing the division of the basal ganglia (Joel & Weiner, 1994, 2000) is to divide it into motor, associative, and limbic components. The motor circuits mainly run through the (dorsal) putamen into motor and premotor areas. The associative circuits mainly run through the (dorsal) caudate nucleus into dorsal PFC. The

limbic circuits run through the ventral striatum—which includes the ventral putamen and caudate nucleus, the nucleus accumbens (NAc), and the olfactory tubercle—into medial PFC areas (e.g. OFC and portions of ACC: prelimbic cortex (BA 32) and infralimbic cortex (BA 25)) related to limbic processing. Joel and Weiner (1994, 2000) propose that there may not be a strict segregation of the thalamocortical loops, but that the loops may be interconnected via a branching between striatal areas and the distinct nigral/pallidal outputs.

The parallel, possibly segregated, striatal circuits, such as discussed earlier (G. E. Alexander et al., 1986; Middleton & Strick, 2001) are likely to be differentiated along a dorsolateral-to-ventromedial gradient (Voorn, Vanderschuren, Groenewegen, Robbins, & Pennartz, 2004). Typically, a distinction is drawn between dorsal and ventral striatum; Figure 1.5 shows the boundary usually drawn. The dorsolateral-to-ventromedial axis is depicted in Figure 1.6. The motoric components of the striatum lie in the dorsolateral portion (putamen and caudate nucleus) whereas the associative and limbic components lie in the more ventromedial regions (nucleus accumbens core and shell). The entirety of this striatal axis is innervated by the dopaminergic nuclei, as will be discussed soon.

1.2.2.4 Reinforcement Learning in the Gating Pathways: Actor/Critic Learning

So far, only the basic feedforward activity of the BG circuits has been discussed. However, an important question is how these action selection pathways develop in the first place. The likely answer is that DA-dependent corticostriatal plasticity allows reinforcement learning in the BG (Mahon, Deniau, & Charpier, 2004; Reynolds, Hyland, & Wickens, 2001; Reynolds & Wickens, 2002). The BG has been extensively implicated in reinforcement learning as well as action selection (Graybiel, 1998; Packard &

Knowlton, 2002). Lesions of striatal cells, particularly in the dorsolateral striatum (Faure, Haberland, Condé, & El Massioui, 2005; Packard & Knowlton, 2002; Yin et al., 2004), tend to disrupt the acquisition of habitual S-R responses. Dopamine depletion in the dorsal striatum, either through antagonist delivery or through lesioning of DA cells, also leads to impairment of S-R learning (Faure et al., 2005; Robbins, Giardini, Jones, Reading, & Sahakian, 1990).

While the dorsal striatum seems to be more involved with S-R learning, the ventral striatum appears to be more important in S-O learning (O'Doherty et al., 2004), but disruptions to plasticity or dopamine transmission there also interfere with instrumental learning (Hernandez, Sadeghian, & Kelley, 2002; Smith-Roe & Kelley, 2000). The ventral striatum's role in learning S-R mappings is a topic of interest in this dissertation. O'Doherty and colleagues (2004) suggest that the dorsal striatum is an 'actor' pathway, whereas the ventral striatum is a 'critic' pathway, responsible for recognizing the reward-potential of the situation and relaying this to the 'actor' to train the latter to respond correctly. For example, if a rat is learning to press a lever (R) in response to a light (S), for a food-pellet reward (O), the dorsal striatal actor, will be rewarded by the ventral striatal critic when the actor stumbles upon the correct behavior, and this will "stamp in" the response. But the critic may need to learn the mapping between the light (S) and the food-delivery (O) for the instrumental learning to take place. One possibility as to why could be that the unexpected delivery of the reward in the absence of an already-learned (S-O) association might not be enough to signal learning in the S-R actor. In other words, the ventral striatum's association is needed to train the dorsal striatum. Another possibility, however, might be that the organism is not

aroused enough to try a response until it learns a mapping between the light and potential reward; in other words, the ventral striatum may need to engage in *incentive learning* (Ikemoto & Panksepp, 1999) before the organism will be motivated to try a response for which it might be rewarded.

1.2.3 Midbrain Dopamine Nuclei

1.2.3.1 DA Receptor Types, Their Effects and Locations

The neurophysiology of DA is complex, but well-studied, with thousands of papers extant in the literature on the subject, and discoveries continuing to be made. Five (agreed upon) subtypes of DA receptors exist, designated D1-D5 (Bergson et al., 1995; Gardner & Ashby, 2000). These are grouped into two families, the D1-like family (D1 and D5), and the D2-like family (D2, D3, and D4). In this paper, henceforth, when D1 or D2 are referred to, it will refer to the family, not the receptor subtype. Both D1 and D2 (and the rest of their families) are slow-acting metabotropic G-protein receptors, but they have opposing effects because D1 activation stimulates the second messenger cAMP production, whereas D2 activation inhibits it (Greengard, 2001). In the striatum, the consequence is that D1 activation enhances striatal medium spiny neuron excitability (in relatively depolarized conditions) by enhancing L-type calcium currents, but D2 activation (in the NoGo striatal cells, at least) reduces neuron excitability by reducing these same currents (Greengard, 2001; Hernández-López et al., 1997; Hernández-López et al., 2000). Actually, D1 activation, in effect, heightens the signal-to-noise ratio of cell responsiveness, suppressing firing in relatively hyperpolarized cells, and expediting firing in relatively depolarized ones (Hernández-López et al., 1997; Schultz, 1998; Servan-Schreiber, Printz, & Cohen, 1990).

D2 activity should oppose this action, making the cell responsiveness more permissive, but less active during mid-range stimulation. However, in the striatum, D1 receptors seem to predominate in the medium spiny neurons of the direct (Go) pathway, whereas D2 receptors predominate in the neurons of the indirect (NoGo) pathway (Gerfen, 1992). In primate neocortex, both D1 and D2 receptors are present, with highest concentrations in the frontal lobe, but D1 is about 10-20 times more frequent than D2 (Lidow, Goldman-Rakic, Gallager, & Rakic, 1991). The neocortical D1 receptors are mainly located in extrasynaptic portions of the dendritic spines of pyramidal cells (Smiley, Levey, Ciliax, & Goldman-Rakic, 1994), which means that their DAergic activity is more likely driven by extrasynaptic concentrations of DA than by synaptically released DA. Evidence shows that D1 receptors are also extrasynaptic in the striatum and substantia nigra (Caillé, Dumartin, & Bloch, 1996).

1.2.3.2 Anatomical Efferent Connectivity of the Midbrain DA Nuclei

The major nuclei containing dopaminergic cells are all located in the midbrain, in neighboring areas (moving lateral-to-medial) designated A8 (the retrorubral area), A9 (the substantia nigra pars compacta; SNc), and A10 (the ventral tegmental area; VTA) (Oades & Halliday, 1987; Voorn et al., 2004). Connections from these nuclei to the striatum fall along the dorsolateral-to-ventromedial axis described by Voorn and colleagues (2004) and shown in Figure 1.6, with A8 connecting most dorsolaterally, VTA connecting most ventromedially, and SNc connecting in an intermediate fashion to both dorsal and ventral striatum, though primarily the former. The SNc pathway, which projects mostly to the (dorsal) caudate nucleus and putamen (collectively called the neostriatum), is generally referred to as the mesostriatal, or nigrostriatal, pathway

(Gardner & Ashby, 2000; Le Moal & Simon, 1991), though it includes, to a much lesser degree, connections from VTA. As somewhat of an artifact of terminology, the VTA connections to the nucleus acccumbens (a large portion of the ventral striatum) are considered part of the mesolimbic pathway which also innervates other limbic regions such as the amygdala, bed nucleus of stria terminalis, cingulate cortex (including ACC), parts of the hippocampal complex, etc. (Gardner & Ashby, 2000; Oades & Halliday, 1987). VTA also connects, through a mesocortical pathway, primarily to frontal portions of neocortex such as PFC and premotor regions, but also to sensory associational areas, including areas in the temporal cortex (Oades & Halliday, 1987; Schultz, 1998). Figure 1.7, taken from a recent review (Fields, Hjelmstad, Margolis, & Nicola, 2007), shows the major pathway connectivity of the VTA.

The model developed in this dissertation includes pathways that are likely to have correspondents in mammalian brain that fall within the SNc nigrostriatal pathway and within the VTA mesolimbic (mesoaccumbens in this case) and mesocortical pathways.

1.2.3.3 DA Cell Activity and Regulation of DA Release: Tonic vs. Phasic Mechanisms

DA cells that are active—maybe two-thirds of them on average at a given time—tend to fire single spikes at an irregular slow rate (interspike intervals: 200-250 ms), driven by endogenous "pacemaker" conductances (Bunney, Chiodo, & Grace, 1991; Grace, 2002). When depolarized further, they may switch to a bursting mode (average inter-burst interval around 350 ms; burst frequency > 12 Hz), and at even higher depolarization, become inactive again (Bunney et al., 1991; Ungless et al., 2004). Background ("default", unstimulated) activity of DA cells may be bursting or non-

bursting (Schultz, 2007). Bursting seems to depend upon excitatory activity of the pedunculopontine nucleus (PPN) and laterodorsal tegmental nucleus (LDT) (Schultz, 2007).

In the striatum, the dopamine transporter (DAT) very quickly reuptakes synaptic releases of DA (Grace, 1991, 2002; Schultz, 1998). The baseline concentration of DA, probably maintained by spontaneous "default" DA cell firing, in the striatum is about 5-10 nM (Schultz, 2007). This is enough to stimulate high-affinity D1 and D2 receptors, but not their low-affinity counterparts. In the striatum 80% of the D1 receptors are low-affinity, and over 80% of the D2 receptors are high-affinity (Schultz, 1998). As D1 receptors are primarily extrasynaptic, this means that they are (mostly) quiescent by default, whereas the D2 receptors, both synaptic and extrasynaptic, are likely to have some degree of default activity.

Under non-bursting DA firing, the DA concentrations may increase, even by a factor of 2 or 3, but this is probably not enough to engage the low-affinity receptors which require activation concentrations on the order of hundreds of nM (Schultz, 2007). However, since 20% of the extrasynaptic receptors are high-affinity, the rate of non-bursting firing might still affect a significant number of the D1 receptors, in addition to a majority of the D2 receptors. Some of the D2 receptors are release-inhibiting autoreceptors on the DA cell synaptic terminals, and these may be activated by this extracellular DA, leading to negative-feedback inhibition of DA cell dopamine release (Grace, 1991, 2000).

When burst-firing is encouraged, DA release is much augmented and overwhelms DAT reuptake, exciting the postsynaptic DA receptors, and also spilling out into the

extracellular area around the target synapses so that short-term peak concentrations of extrasynaptic DA briefly run over the low-affinity receptor threshold (Schultz, 1998). Thus, most of the D1 receptors would require burst firing to activate. The D2 autoreceptors would more strongly inhibit the DA release from the DA terminals.

Grace in his research has emphasized that tonic (i.e. extracellular) DA inhibits phasic (i.e. DA cell spiking-triggered) DA release. Both in striatum (Grace, 1991) and in PFC (Takahata & Moghaddam, 1998), there are also glutamatergic receptors on the DA cell axon terminals which afferent neurons (e.g., from the neocortex) may activate to promote (primarily extrasynaptic) release of DA, but such release seems to depend on some baseline DA cell firing perhaps providing available DA reserves to be released (Grace, 2002; Keefe, Zigmond, & Abercrombie, 1992). However, the glutamatergic afferents could thereby provide a mechanism by which PFC and other cortical areas could inhibit striatal (phasic) DA action, something which may have clinical consequences. Grace (1991) hypothesizes that in schizophrenia, low ambient PFC activity leads to reduced baseline extracellular tonic DA which disinhibits phasic DA action because of decrease of D2 autoreceptor action, and possibly because of up-regulation of postsynaptic DA receptors. By comparison, regarding alcohol and psychostimulant addiction, he hypothesizes (Grace, 2000) that the drugs tend to increase the general tonic level of DA in striatum, which causes down-regulation of response to phasic DA signals. Then, when the drugs are withdrawn, it takes time for the phasic DA response to recalibrate, and the subject feels a dysphoric state while the phasic DA response is hypoactive.

A number of situations can lead to long-term DA concentration increases for a duration on the order of minutes (as measured by microdialysis) where the increases are from 20%-100%, and sometimes up to 200% above the 5-10 nM baseline (Schultz, 2007). Again, these concentrations are probably not sufficient to activate the low-affinity receptors, so D2 receptors are probably primarily affected. Schultz (2007) suggests that the presynaptic glutamatergic release mechanism previously mentioned is more likely to be responsible for these long-term DA concentration changes than the phasic signal. However, as will be seen in Section 1.2.3.6, there is evidence of an example where the DA cell firing rate can dynamically modulate DA concentrations over a time period on the order of a few seconds.

The tonic/phasic interaction described above holds mainly for the striatum. The dynamics of DA release in PFC are likely to be different because there is little DAT, but DA is instead reuptaken by noradrenergic (NE) reuptake mechanisms (Grace, 2002). This likely leads to a much slower reuptake dynamics. In addition, the predominance of extrasynaptic D1 receptors which are believed to be critical for working memory function (Durstewitz & Seamans, 2002; Zahrt et al., 1997) suggests that the base rate of DA cell firing may be responsible for maintaining a tonic level of PFC dopamine that is necessary for proper PFC function, including cortical plasticity, as will be discussed soon.

A note of clarification would be useful at this point regarding the usage of 'tonic' and 'phasic', both in this dissertation, and in the literature in general. 'Tonic' DA effects might refer to the effects of extracellular concentrations of DA, as in Grace's papers, or it might refer to the effects of a baseline rate of firing of the DA cells. In this dissertation, the latter is mostly intended. The former and latter could be decoupled by, for example,

the inhibitory effects of the D2 autoreceptors, or by corticostriatal glutamatergic stimulation of the DA terminal boutons. By 'phasic' DA, Grace is referring to the DA cell spike-driven release of DA. In this dissertation, though, the term will be mainly used to designate short-term events (e.g. bursts or dips) that modify the base-rate of firing, or, alternatively, the rate of DA cell bursting since typical firing seems to be slow-rate firing punctuated with short bursts (Ungless et al., 2004). A central theme in this dissertation is that the tonic DA firing is a little like a carrier wave in radio communications, and the phasic changes are like frequency-modulation of this. The carrier wave itself broadcasts to DA targets a level of activity-oriented motivation in the system, whereas the superimposed frequency modulations signal events that should lead to (reward or punisher) reinforcement.

1.2.3.4 Conditions for DA Cell Phasic Bursts and Dips

Extensive study has been made by Schultz and colleagues (Fiorillo, Tobler, & Schultz, 2003; Hollerman & Schultz, 1998; Ljungberg, Apicella, & Schultz, 1992; Schultz, 1998, 2007; Tobler, Fiorillo, & Schultz, 2005; Waelti, Dickinson, & Schultz, 2001) of the conditions under which DA neurons fire in response to environmental events. Figure 1.8 shows their main results: DA cells fire with a phasic burst when there is an unexpected reward and when there is a stimulus cue that predicts a reward (Hollerman & Schultz, 1998; Ljungberg et al., 1992; Waelti et al., 2001), but firing is phasically suppressed when a reward that was predicted based on a stimulus cue is omitted (Hollerman & Schultz, 1998; Waelti et al., 2001). DA cells also respond to salient or novel events, though the response to these habituates rapidly (Ljungberg et al., 1992). There is also evidence that the amplitude of phasic bursts may encode the

amount/intensity of reward received or (during the conditioned stimulus) expected (Tobler et al., 2005). Evidence in this research has suggested that the responsive cells from the DA nuclei (including SNc and VTA) fire in synchrony, suggesting a coupling between the activation of the nuclei (Ljungberg et al., 1992). One possible mechanism suggested that might allow this is electrotonic coupling of the DA cells during burst activity, possibly due to the influence of the neuropeptide CCK (Bunney et al., 1991). Also, as previously mentioned, evidence suggests that excitatory activity of the PPN is important for triggering burst firing in the DA cells likely to be involved in phasic release (Schultz, 2007).

More recently, there has appeared evidence that DA cells are suppressed by aversive (punisher) stimuli. Ungless and colleagues (2004) discovered that earlier studies had mistakenly labeled aversive stimulus-responsive non-DAergic cells in VTA as being DA-releasing. Their evidence showed that the true DA cells were actually inhibited by aversive stimuli; both basal firing rate and rates of emitted bursts were slowed during the application of foot pinch to rats. Probably the VTA cells that were excited by aversive stimuli were inhibitory and responsible for the suppression of the nearby DA-ergic VTA cells. In addition to these cells, there are GABA-ergic cells in VTA which receive wide afferent input (from areas such as LH and PFC) and project to the DA-ergic VTA cells (Fields et al., 2007). These might also allow aversive cues to inhibit the DA cells.

The evidence, then, suggests that phasic DA signals may signal reward prediction error which can be used for reinforcement learning (Hollerman & Schultz, 1998; Schultz, Dayan, & Montague, 1997; Suri, 2002; Waelti et al., 2001). When an organism is not expecting a reward and it receives one, there is a positive prediction error, and a burst is

signaled. On the other hand, when the organism is expecting a reward and it's omitted, there is a negative prediction error, and there is a phasic dip (but see (Fields et al., 2007); evidence for reward omission dips have not been universally observed). When expectations and reward receipt are in alignment, there is zero prediction error and firing remains at baseline. There are additionally dips during aversive stimuli and bursts during novel or extremely salient stimuli (but see (Fields et al., 2007); rats have shown responses for fully predicted rewards). Redgrave and colleagues have challenged the conclusion that the DA signal represents reward prediction error (Redgrave, Prescott, & Gurney, 1999b), and favor the interpretation that phasic bursts signal a need to switch attention or select behaviors. But it seems possible that the phasic DA signal could play more than one role, signaling both prediction error and/or a need to attend to a stimulus. In the model in this dissertation, DA activity phasically increases during rewards, but also during novel stimuli. Under task learning conditions, novelty-induced phasic bursts actually cause the model to learn the task in cases where it would have otherwise failed by encouraging explorative behavior and simultaneously rewarding it, even when an actual reward is not delivered. Generally speaking, phasic bursts may, in addition to promoting learning, temporarily encourage initiation of new behaviors. Phasic dips may, in addition to promoting unlearning, temporarily discourage initiation of new behaviors. The model in this dissertation would suggest this to be plausible.

1.2.3.5 Afferent Regulation of DA Cell Firing Rate

Many researchers make a simplifying assumption about the homogeneity of DA cell function, but there are likely several potentially independent circuits, due to the heterogeneous structure of the DA nuclei (Fields et al., 2007; Gardner & Ashby, 2000).

The model developed in this dissertation recognizes a variety of efferent projections, but follows other researchers in treating SNc and VTA as a single compartment responding in a uniform way, as if their afferents were the same and exerted the same influence.

There is evidence for functional dependence of the SNc pathway on the VTA compartment (Joel & Weiner, 2000). The ventral pallidum which is the output part of the nucleus accumbens pathway projects to PPN (Chivileva & Gorbachevskaya, 2008). This, in turn, projects in an excitatory way (see below) to SNc. This suggests that SNc is under control of NAc activity through the standard Go pathway circuit shown in Figure 1.4. But NAc is dopaminergically innervated by VTA. The VTA compartment, then, probably enables ventral striatal-mediated release of SNc activity, though other afferents of the enabling NAc pathway, e.g. amygdala or orbitofrontal cortex, may also need to be active. Ikemoto and Panksepp (1999) have suggested that the ventral striatal pathway is more involved with flexible approach behavior and incentive learning whereas the dorsal striatal path is more involved with habitual reactive behavior. It seems likely that ventral striatal activity triggered by incentive cues (e.g. drug-use cues) represented in OFC and/or amygdala could disinhibit SNc unit activity which would promote habitual S-R responses mediated by the dorsal striatum, including drug use responses.

There are other important afferent neuroanatomical differences between the two DA nuclei: the SNc afferents seem to be primarily GABAergic (therefore inhibitory) whereas the VTA afferents are primarily glutamatergic (therefore excitatory) (Lee, Abercrombie, & Tepper, 2004). Focusing on the VTA compartment/s, there is also evidence that there are separate VTA circuits supplying DA to PFC and NAc because

intra-VTA administration of neurochemicals can lead to different DA concentration changes in PFC and NAc (Fields et al., 2007).

However, as Schultz's work has shown relative homogeneity of reward-related response by the separate compartments, they will be treated as the same in this dissertation, though future research may want to address more of the heterogeneities in the mesostriatal/mesocortical axis. The fact that the PPN innervates both SNc and VTA (Fields et al., 2007) and it is a likely excitatory source of the reward response (Schultz, 1998, 2007) suggests a mechanism by which both compartments could possibly synchronize in at least their reward-related responses. As mentioned, also, electrotonic coupling (i.e. gap junctions) could allow nearby DA neurons to synchronize under certain neurophysiological conditions (Bunney et al., 1991).

Figure 1.7 shows the major afferents (as well as efferents) of VTA (Fields et al., 2007). The connection to the lateral hypothalamus is one key excitatory afferent pathway. Cells in LH that release the neuropeptide orexin into areas such as VTA and NAc are believed to fire selectively in association to pursuit of consummatory rewards (e.g. food and drugs) (Harris, Wimmer, & Aston-Jones, 2005). Activation of LH is stronger when animals are searching for food and is suppressed when foraging is successful and the animal has switched to consumption (Panksepp, 1998).

Two neuromodulatory nuclei project to VTA: the NEergic locus coeruleus and the serotonergic dorsal raphe nucleus (Fields et al., 2007). NE release is associated with attentional arousal and, like the other major catecholamine, DA, increases the signal-to-noise responses in target neurons (Panksepp, 1986), so this connection could allow DA activity in VTA to increase during vigilance states. 5-HT release is largely emotionally

and behaviorally inhibitory (Panksepp, 1986), and dorsal raphe nucleus action inhibits DA cell activity (Schultz, 1998), so this connection may allow a pathway for DA cell suppression. An opponent relationship between 5-HT and DA has been suggested that may have an important bearing on the mechanisms of reinforcement learning (Daw, Kakade, & Dayan, 2002).

Another (likely excitatory) afferent of VTA is the amygdala (Schultz, 1998). The amygdala in general is involved with evaluative learning and its various nuclei are known to respond both to rewarding and aversive stimuli (Baxter & Murray, 2002). Other major afferents to the VTA include the laterodorsal tegmental nucleus (LDT), PPN (through an indirect pathway through SNc), and PFC (Fields et al., 2007). These exert various excitatory and inhibitory influences on VTA DA and GABA cells feeding into two distinct PFC and NAc projection circuits (Fields et al., 2007). A functional explanation of these different influences remains to be sorted out, but there are clearly several pathways by which other areas of the brain could potentially stimulate or restrict DA cell firing from VTA.

The other major DA compartment, the SNc, receives GABAergic connections from the striosomes (patch-cells) of the striatum (Gerfen, 1992), as well as from SNr (Lee et al., 2004). Recent evidence (Lee et al., 2004) suggests that pallidal chemical excitation leads to increased bursting of DA cells (in conjunction with only a mild increase of DA cell firing rate), and an elevation in neostriatal extracellular DA. By contrast, electrical stimulation of the same area leads to DA cell inhibition. This seeming contradiction is reported to probably be due to multiple pathways from GP to SNc through the SNr which have varying sensitivities. The PPN provides a major excitatory input (mixed cholinergic

and glutamatergic) to SNc (Blaha & Winn, 1993; Mena-Segovia, Bolam, & Magill, 2004). Brown and colleagues (Brown, Bullock, & Grossberg, 1999) have constructed a neurocomputational model of the learning dynamics described by Schultz and colleagues of the phasic DA signal involving a number of these pathways.

1.2.3.6 Behavioral and Cognitive Correlates of DA Cell Firing

As mentioned, evidence suggests that the switch of DA neurons from single-firing mode into a burst mode has a greater impact than a change in their overall firing rate on extracellular concentrations of DA in the striatum (Lee et al., 2004; Schultz, 1998). This is probably due to the fast DAT reuptake mechanisms in the striatum. Stimulus events associated with rewards may trigger bursting in the DA cells that release a large amount of DA. In addition to time (on the order of tens of ms) required to reuptake released DA, effects of DA receptor activation may persist over long-term period (hundreds of ms to minutes) because DA receptors are metabotropic, acting through much slower second-messenger pathways mechanisms than faster acting (c. 1 ms) ionotropic (AMPA or GABA) receptors (Greengard, 2001). Therefore, cues for phasic DA release could lead to behavioral and cognitive effects persisting over a second to minute duration. Such phasic signals are probably involved with the previously-described short-term reward signals that may allow reward error-prediction.

Rewards and other stimuli, however, may also trigger changes in dopamine action (measured by extracellular DA concentration) over a longer time course, such as might be measured with voltammetry (the order of seconds) or microdialysis (the order of minutes) (Schultz, 2007). Ikemoto and Panksepp (1999) review much of this evidence for the NAc (ventral striatum) DA. Long-term rises in extracellular DA have been observed during

both anticipatory and consummatory phases of foraging behavior. Novel and/or unusually tasty food also lead to increased DA. Interestingly, during operant tasks NAc DA rises when trials are begun and lowers back to a baseline after a session is completed. Within a session, there are often rises in DA during lever pressing and eating. Aversive stimuli, too, can lead to increased extracellular DA levels, which initially may seem puzzling, given DA cell suppression by aversive stimuli, but could be explained by engagement of glutamatergic stimulation mechanisms. As Ikemoto and Panksepp point out, the aversive stimulus DA may be involved in facilitating active avoidance behaviors.

Pharmacological microinjection and 6-OHDA lesions have been used to perform DA level manipulations (Ikemoto & Panksepp, 1999). The former involves agonist or antagonist injections into localized brain areas. The latter technique selectively lesions DA cells projecting to particular targets, though this selectivity in location has some significant limitations. Injections of DA and DA agonists into NAc tend to lead to heightened locomotor activity, specifically activity of an explorative and/or appetitive nature. DA depletion tends to reduce such locomotor activity, and has been shown to reduce hyperactivity associated with novelty. Some evidence suggests the shell portion of the NAc is more important than the core region for exploration. Raised DA levels also lead to increased responses to reward-predicting CSs. Disruption of NAc DA leads to decreased hoarding behaviors, but does not seem to disrupt the consumption of food. When a choice is available of eating a default food for free or having to lever-press for a better food, NAc (core subregion) DA disruption seems to reduce rats' likelihood of working for the better food (Sokolowski & Salamone, 1998). This suggests that NAc DA may stimulate instrumental responding, that S^Ds predicting reward facilitate responding

through a ventral striatal pathway. NAc DA also seems to be involved with learning and performance of active avoidance behaviors. VTA activation of the accumbal circuit appears to be sufficient to allow Pavolvian (S-O) incentive learning, though other pathways may allow Pavlovian learning as well (Fields et al., 2007).

In most of these cases, it is not clear whether the increases seen in long-term DA concentrations are caused by enhanced DA cell firing, or by glutamatergic stimulation at DA targets. (Fiorillo et al., 2003) presents a case, however, where DA cell effects are involved. As Figure 1.9 shows, under Pavlovian situations, the dynamics of the base-rate of DA cell firing seems to be affected by how probable the reward delivery is. A ramping up of DA cell activity is seen after the phasic onset spike occurs related to the CS. This ramping up is terminated when the reward is delivered and/or the CS is offset. Interestingly, the rate of ramp-up seems to correlate with not the probability of reward delivery per se, but rather with the uncertainty on whether the reward will be delivered or not. The ramp-up is maximal when $p = 0.5$, and minimal either when $p = 0$ or $p = 1$.

The cognitive/behavioral function of this signal and the potential mechanism allowing the uncertainty ramp-up response are unclear. Why should the tonic DA signal be maximal in more uncertain reward conditions? In the case of Pavlovian situations, it doesn't seem to make much sense, except maybe as a part of mechanism for updating the reward prediction error estimation (Schultz, 2007). Schultz (2007) points out at there have been difficulties detecting this uncertainty-related response for instrumental cues. However, such an uncertainty signal might be functionally useful during instrumental tasks to signal when it might be useful to try a novel response behavior. Presumably, under conditions where the animal is habitually performing the correct response, $p = 1$,

and it is less desirable to try new behaviors, so it would be advantageous for the DA signal to not increase over baseline. On the other hand, if $p = 0$, it means the animal probably hasn't been rewarded for any behaviors it's tried so far, so it is likely to mean that there is no useful relationship between the CS and any useful R that might lead to a positive O, so the animal should ignore the ineffectual S. But if the animal has been rewarded on some occasions for responding, but hasn't got the correct S-R mapping figured out yet, then $0 < p < 1$, and it may be useful for the animal to try some novel behavior to see if it improves its chances of success.

Assuming that reward uncertainty influences the DA signal according to the above rationale, what is the likely afferent pathway allowing the ramp-up excitation in the DA cells' activity? Two likely candidates (see Figure 1.7) might be the orbitofrontal cortex, which is considered a part of PFC, or the amygdala. Both of these areas have cells that selectively fire in response to cues that predict rewards or punishers. There is evidence, also, of reciprocal connections of the cingulate cortex to VTA (Oades & Halliday, 1987). Section 1.2.4 will focus on the possibility of cingulate cortex involvement which this dissertation favors as the likely pathway for representing uncertainty/frustration.

One (hard-to-test) hypothesis regarding the subjective experience correlating with the uncertainty ramp-up is that the animal becomes increasingly frustrated/anxious as time progresses in maximally uncertain reward conditions. If $p = 0$, the animal doesn't get excited because there is no expectation of impending reward. If $p = 1$, the animal doesn't get excited enough to change its behavioral course because it is confident it will be rewarded. The animal is maximally motivated to "do something" when chances of

reward are "up for grabs". There would be a ramp-up in desire to act in a novel way because, generally speaking, if an animal waits too long to act, the opportunity to engage in the action that leads to reward will pass.

In short, it would evolutionarily adaptive for a brain to have a signal that activates when a reward cue appears and ramps up as time goes by, finally shutting off when the promise of reward goes away or after the reward has been gained. Such a signal would be strongest, not when rewards are most likely to be gained by the usual responses, but rather when the novel behavior may be needed to determine what the best response is.

It is not immediately clear how the uncertainty measure would be calculated in the neural circuitry, but it could involve having a predictor attempt to guess whether a reward would or would not be delivered, each time a cue was presented. Another pathway mapping a cue to an estimate of uncertainty would be trained by this predictor pathway, with errors made by the predictive pathway increasing the synaptic weight (LTP) and correct responses decreasing it (LTD). Reward delivery might inhibit the output cell/s of this circuit, and disappearance of the cue would remove excitation from the circuit.

1.2.3.7 Striatal DA Modulation

The dynamics of striatal cell activation by its afferents (cortex, thalamus, etc.) is quite complicated because extracellular DA levels in the striatum modulate the dynamics through both D1 and D2 receptor activation (Grace, 2002). In the presence of DA, medium spiny neurons have three possible states: a 'down state' in which they are inactive and unresponsive to input, a inactive 'up state' in which they are highly depolarized, but not firing, and an active 'up state' in which they are firing (Grillner et

al., 2005). Either 'up' or 'down' states can last for periods on the order of minutes or hours (C. J. Wilson & Kawaguchi, 1996). D1 receptor activity seems to be required to facilitate transition from a 'down state' to an 'up state' (Grillner et al., 2005). Neurons in inactive 'up state' require only a little more afferent activation to induce to firing. As mentioned in Section 1.2.3.1, D1 agonism leads to an increase in signal-to-noise ratio of striatal cell activity in the Go units, meaning weak input is suppressed (perhaps because the cell is in 'down state'), but strong input strengthened (perhaps because the cell is in 'up state') (Hernández-López et al., 1997). By comparision, D2 agonism in the NoGo units is inhibitory (Hernández-López et al., 2000), and, moreover could inhibit DA cell firing through D2 autoreceptor activation.

Neural plasticity in the striatum, particularly in the corticostriatal synapses, is critical to the basal ganglia mechanisms of reinforcement learning. Both the Go and NoGo pathways need *long-term potentiation* (LTP)—increasing of synaptic weights—to allow the learning of response to stimuli, and *long-term depression* (LTD)—decreasing of synaptic weights—to unlearn responses. Hebb (1949) argued that, when two neurons fire at (roughly, at least) the same time (an event referred to in this dissertation as a *Hebbian event*), their synaptic weights increased between them (LTP). However, there also needs to be an "anti-Hebb" mechanism that allows unlearning of weights (LTD) when the pre-postsynaptic relationship is weak or otherwise functionally undesirable. It would be functionally useful for systems of neurons to have a *value-driven learning* (Almássy, Edelman, & Sporns, 1998; Sporns & Alexander, 2002) mechanism whereby a reinforcement learning signal could signal either a reward or punisher and implement

Thorndike's Law of Effect on the target synapses, whereby rewards trigger LTP and punishers trigger LTD on synapses where there have been Hebbian events.

While there may be many LTP/LTD mechanisms, dependent on the location of the neurons, one well-described hippocampal CA1 LTP/LTD mechanism involving NMDA and AMPA receptors (J. Lisman, 1989; Malenka, 2002) may be illustrative of at least a wide class of LTP/LTD mechanisms. Calcium ion (Ca^{2+}) concentrations in postsynaptic spines may determine whether the synapses there are weakened, strengthened, or kept at the same weight. At zero Ca^{2+} concentrations, there is no plasticity (LTP or LTD); at moderate concentrations, LTD; at larger concentrations, again no plasticity; at larger concentrations, LTP; and finally at huge concentrations, LTD (J. Lisman, 1989). Ca^{2+} ion concentration is mostly determined by activation of NMDA glutamate channels. These tend to be inactive, except when the cell is highly depolarized, and when they are active, permit both Na^{+} and Ca^{2+} entry. Postsynaptic or presynaptic activation alone may give rise to moderate levels of Ca^{2+} concentration leading to LTD, whereas simultaneous activation of the cells leads to high Ca^{2+} concentration and LTP. LTP potentiates glutamatergic AMPA receptors, which are reactive to glutamatergic excitation, even when the cell is relatively hyperpolarized, and LTD depotentiates them.

DA receptors may modulate this activity also. D1 receptors, for example, potentiate NMDA action, whereas D2 receptors depotentiate it; and there are also direct influences of the receptors on the activation of Ca^{2+} currents (Greengard, 2001). The complex striatal circuitry contains both D1 and D2 receptors whose activations may be

driven by the extracellular concentrations of DA, so unraveling the effects of DA on LTP and LTD is challenging, though much progress has been made.

Evidence from an *in vivo* study suggests that striatal synapses engage in LTD by default, when a Hebbian event is stimulated, but increased DA levels (dependent on D1 receptor activity) change this LTD to LTP instead (Reynolds et al., 2001). Reynolds and Wickens (2002) proposed that there is no LTP or LTD in the absence of a Hebbian event; and that, in the presence of a Hebbian event, low levels of DA lead to LTD, high levels to LTP, and intermediate levels to no change. They proposed that the high range corresponded to the range of DA concentration experienced during phasic reward bursts, that the low range corresponded to the concentration during phasic dips, and the intermediate range was the normal range of tonic concentration of DA in the absence of rewarding or punishing events. The model in this dissertation follows this pattern with the Go units, but reverses the ranges for the D2-receptor dominated NoGo units, i.e., making the low range the LTP zone, and the high range the LTD zone. Both the theorized function of the NoGo units and recent cellular mechanisms data support this.

Very recently, Shen and colleagues (Shen et al., 2008) have presented a review of a battery of *in vitro* experiment results revealing/reproducing a host of neurotransmitter influences on LTP and LTD. Figure 1.10 presents their summary figures showing the distinct sets of influences, DAergic and otherwise, on the Go and NoGo striatal units. In the Go units, LTD is not dependent on DA receptor activation, depending rather on glutamate and Cav 1.3 (a type of L-type Ca^{2+} channel) activity, as well, and endocannabinoid activity. LTP, however, is dependent on D1 and NMDA receptor activity. At extremely low DA levels, then, LTD activity would be expected to

predominate. At intermediate (tonic default) levels, the low affinity D1 receptors may begin to activate enough for the LTP and LTD effects to cancel each other. At high DA levels, the LTP effect would be strongly potentiated. In the NoGo units, however, LTP does not depend upon DA activity because an adenosine receptor, rather than D1, is responsible for the actions of the LTP mechanism. However, the LTD mechanism is like the LTD mechanism of the Go units, but additionally requires D2 receptor activation. So, at extremely low DA levels, we should have LTP by default. At 'default' DA levels, the LTD effect should perhaps cancel the LTP effect, and at highest levels, the LTD effect should dominate. None of the LTP or LTD mechanisms would be engaged, however, in the absence of a Hebbian event.

1.2.3.8 Prefrontal Cortex DA Modulation

The anatomical circuitry in PFC is vastly different from that of the striatum and this may result in different mechanisms and dynamics of DA modulation of PFC activation and plasticity. As already mentioned, there are far more D1 than D2 receptors in (primate) PFC (Lidow et al., 1991) and PFC lacks DAT with DA reuptake regulated instead by NE reuptakers (Grace, 2002). Most of the D2 receptors are located in layer V of the frontal, parietal, and occipital primate cortex, suggesting D2 may regulate cortical output (Lidow et al., 1991). In human PFC (BA 9), D1 receptors seem most concentrated in layer V (Lidow et al., 1991).

Like striatal cells, (layer V) pyramidal PFC cells seem to have 'up' and 'down' states *in vivo* (B. L. Lewis & O'Donnell, 2000). Default firing consists of alternations between the two states, with some firing during the 'up state'. Neither hippocampal or thalamic activation seems to trigger the 'up states', suggesting corticocortical stimulation

is involved. VTA pulse stimulation can prolong the 'up states' for periods on the order of seconds, and this seems to depend upon D1 receptors because D1 antagonists inhibit DAergic effects. D1 potentiates NMDA channel activity while mildly suppressing non-NMDA glutamatergic activity (Seamans, Durstewitz, Christie, Stevens, & Sejnowski, 2001). Since NMDA is only active at higher depolarization, switching the balance of glutamatergic response from AMPA channels to NMDA channels leads to D1-modulated increase of signal-to-noise ratio responses (Ashby & Casale, 2003). There is also, however, evidence that D2 receptor activation may lead to decreases in PFC cell firing in layer V PFC neurons (Gulledge & Jaffe, 1998).

D1 receptor activity is apparently important for proper working memory function, as both too little or too much D1 receptor activity can disrupt working memory performance (Zahrt et al., 1997). Working memory may be mediated by the D1 receptor stabilization of recurrent excitation in the deep layer (e.g. layer V) PFC pyramidal cells (Brunel & Wang, 2001; Durstewitz, Kelc, & Güntürkün, 1999; Durstewitz et al., 2000; Gao, Krimer, & Goldman-Rakic, 2001). Hypo-DA is likely to disrupt the NMDA channel activity that is used to sustain the recurrent excitation, whereas hyper-DA may potentiate GABAergic interneurons which also have D1 receptors (Brunel & Wang, 2001; Goldman-Rakic, Muly III, & Williams, 2000). The PFC extracellular level of DA may thus control the maintenance of working memory (Durstewitz et al., 1999), and the author of this dissertation in previous work has suggested that this extracellular DA level may be locally controlled by glutamatergic activation (by other PFC cells) of the terminal boutons of the VTA cells (Chadderdon & Sporns, 2006). Frank and colleagues (Frank, Loughry, & O'Reilly, 2001; O'Reilly & Frank, 2006) have suggested that the basal

ganglia may provide a dynamic gating mechanism for engaging and disengaging working memory. Perhaps this gating mechanism might regulate the extracellular DA levels in the target PFC zones, allowing multiple working memory traces to be maintained through dopamine stimulation, and released either through turning off of the bouton stimulation of DA release, or perhaps direct BG-mediated NoGo inhibiting of the working memory PFC regions.

DAergic modulation of synaptic plasticity in PFC is complex and still not well understood, though much data has been collected (Seamans & Yang, 2004). In *in vitro* PFC brain slices with depleted extracellular DA, DA seems to induce LTD in stimulated layer V cells (Law-Tho, Desce, & Crepel, 1995; Otani, Blond, Desce, & Crepel, 1998). This seems to depend upon DA receptors, and groups I and II metabototropic glutamate (mGluR) receptors, but not necessarily NMDA receptors (Otani, Auclair, Desce, Roisin, & Crépel, 1999). However, when the PFC slices are "primed" with a bath of DA for 12-40 minutes, NMDA-dependent LTP is induced instead of LTD when the PFC neurons are stimulated (Matsuda, Marzo, & Otani, 2006). Hippocampal-PFC synapse NMDA-dependent LTP appears to depend upon D1, but not D2 receptors (Gurden, Takita, & Jay, 2000). Maintenance of either LTP or LTD in layer V PFC neurons appears to depend on D1 receptor activation (Huang, Simpson, Kellendonk, & Kandel, 2004).

Together, this data seems supportive of a DA control of LTP and LTD which is similar to that of the striatal Go cells (see Figure 1.10 (top)), except LTD depends also on DA receptor activation. It would be expected, then, that zero concentration of DA would disable either LTD or LTP. Low extracellular levels of DA such as occur when DA-depleted PFC slices are bathed in DA would lead to LTD. Under high extracellular DA

conditions, however, LTP would occur. It may be the case that intermediate levels of DA would lead to cancellation of the LTP and LTD effects. More studies need to be done to test these hypotheses.

In the model in this dissertation, DA signal-to-noise modulation of basal PFC activation is not modeled. LTP and LTD in model frontal cortical units are regulated using the same mechanism that the D1-dominated model striatal Go cells use. As will be discussed in Sections 4.4 and 4.5, DA modulation of working memory is not modeled, though this would be a natural extension for future research.

1.2.3.9 DA and Response Vigor

The postsynaptic effects of DA on the excitability of its targets, such as the dorsal and ventral striatum and PFC, lead to a role in modulating *vigor* of response (Niv, Daw, & Dayan, 2005; Niv, Daw, Joel, & Dayan, 2006). Niv and colleagues suggest that the tonic DA signal encodes the *opportunity cost* of inaction which is similar and conceptually related to this dissertation's proposal of a signal of 'activity-oriented motivation'. ('Opportunity cost' in a term from economics meaning the reward that might have been gained by making an alternate choice.)

The level of DA often serves to set the level of effort the organism is willing to exert for rewards in its environment. As mentioned, the ventral striatal (nucleus accumbens) extracellular DA levels seem to correlate with exploratory activity motivated by drives, reward, and/or novelty cues; and sometimes activation of conditioned avoidance behaviors (Fields et al., 2007; Ikemoto & Panksepp, 1999). Antagonism of DA can lead to reduced stimulus-cued instrumental behavior (Dickinson, Smith, & Mirenowicz, 2000; Ikemoto & Panksepp, 1999). Within the NAc, lesions to the core (but

not the shell) lead to impairment of fixed-rate responding (for example when an animal is rewarded for every 3rd lever press), and also lead to decreased working for a more desirable food when a less desirable food is made immediately available (Sokolowski & Salamone, 1998). DA effects on the dorsal striatum can be ascertained by looking at the effects of early-stage Parkinsonism since the SNc cells, which innervate the dorsal striatum, are the first die off (Cools, 2006). DA depletions in PFC, as mentioned, as well as great excesses, leads to failures of working memory (Zahrt et al., 1997). Similar to the findings of Sokolowski and Salamone in the NAc, DA concentrations in ACC seem to be important for the choice of high-cost, high-reward actions vs. low-cost, lower-reward options with D1, but not D2, receptor antagonists decreasing the performance of the higher-effort behaviors (Schweimer & Hauber, 2006). Generally speaking, depletions of DA lead to hypo-functionality in their target pathways, and the specific consequences of that hypo-functionality depend on the targets (Le Moal & Simon, 1991). But it may be a valuable unifying hypothesis to suggest that the base-rate DA signal is a measure of activity-oriented motivation or opportunity cost, with the caveat that glutamatergic stimulation of extracellular DA release can lead to differences of DA target extracellular concentrations in the face of the same DA cell firing pattern, meaning, for example, that PFC-mediated working memory might be at a given time more potentiated than NAc-mediated exploratory motivation. Thus, the DA cell (SNc/VTA) signal becomes a "global" motivation signal which is modified locally at the DA targets to give several "local" motivation signals.

Results of the model in this dissertation will mainly be compared (in Chapter 3) with data related to Parkinson's disease (PD) and psychostimulant use. The former is a

canonical example of DA depletion (hypo-DA), whereas the latter is an example of excess DA activity (hyper-DA). The remainder of this section will consider the vigor modulating effects of hypo- and hyper-DA, whereas Section 1.2.3.10 will consider the effects on reinforcement learning.

Parkinson's is not merely a static hypo-DA lesion of the SNc, but often a progressive disorder in which the dopaminergic neurons in the midbrain nuclei gradually die off, adding new symptoms to the disorder as damage spreads from SNc to VTA, spreading DAergic denervation from the dorsal to the ventral striatum (Cools, 2006). Thus, the motor striatum is affected first, then the associative striatum, then finally the limbic striatum. In the striatal DA targets affected, the D1-dominated Go pathway will become hypo-active, and the D2 NoGo pathway will become hyper-active, although it's becoming apparent that the NoGo pathway, in particular, suffers a great deal of dendritic spine loss, probably due to moderately high NMDA-stimulated Ca^{2+} concentration which leads to chronic LTD (Day et al., 2006; Gerfen, 2006). The expectation, then, is that activity mediated by dorsal striatum should fail first, including motor program activation in the motor striatal loops, leading to the classic motor symptoms of delayed reaction time and slowed movement (Gauntlett-Gilbert & Brown, 1998; Muller et al., 1999; Schultz et al., 1989). Correspondingly, delivery of both D1 and D2 antagonists in a primate reaction task has been shown to slow reaction time (Weed & Gold, 1998). As the disease progresses and damage spreads towards VTA, the associative BG circuits should begin to be affected, leading to executive/cognitive control dysfunctions (S. J. G. Lewis, Dove, Robbins, Barker, & Owen, 2003; Woods & Tröster, 2003). Finally, when VTA and limbic striatum are affected, difficulties should be observed involving incentive

learning and exploratory drive. As damage to striatal DA innervation is uneven, the effects of L-dopa treatment of PD become uneven, potentially leading to hyper-DA in the striatal pathways that the primary disorder has yet left undamaged (Cools, 2006). It is therefore useful, in simulation modeling of corticostriatal function, to model distinct corticostriatal pathways, and this dissertation takes this approach. Selective hypo- or hyper-DA effects may be manifested independently in different pathways which complicates consequences of global drug delivery.

The effects of two psychostimulants are considered as models of hyper-DA in this dissertation: cocaine and amphetamine, although studies have also been done using D1 or D2 agonists. Both cocaine and amphetamine suppress the dopamine transporter DAT, leading to reduced reuptake, and therefore increased extracellular concentration, of DA in the striatum (Grace, 2000; Saunders et al., 2000). Whereas cocaine inhibits DAT activity, amphetamine also promotes cellular internalization of DAT receptors (Saunders et al., 2000), and stimulates DA release from DA terminals in SNc and VTA (Bernardini, Gu, Viscardi, & German, 1991; Pifl, Sitte, Reither, & Singer, 2000; Saunders et al., 2000). In a seemingly paradoxical way, low dosages of psychostimulants can actually suppress DA cell firing, which perhaps explains how psychostimulants may have therapeutic effects for attention deficit hyperactivity disorder (ADHD). However, this is explainable by the fact that there are inhibitory D2 autoreceptors on the DA terminals providing negative feedback on DA cell firing (Grace, 2000). In fact, when D2 receptors are blocked, the inhibitory effects of D-amphetamine go away, leaving a net excitatory effect for the drug (Shi, Pun, Zhang, Jones, & Bunney, 2000). Thus, when interpreting psychostimulant effects, it is necessary to recognize that low dosages may actually lead to

net inhibitory effects due to presynaptic D2 effects, whereas higher dosages lead to net excitatory effects due to postsynaptic effects (Frank & O'Reilly, 2006). (By contrast, delivery of antipsychotic D2 antagonists, can lead to DA excitation at low dosages, and inhibition at higher dosages.)

Psychostimulants such as cocaine and D-amphetamine have been shown to facilitate reaction time in both primates (Hienz et al., 1994) and humans (Halliday et al., 1994). In rats, subcutaneous amphetamine injections tend to decrease frequency of behaviors like grooming, lying, and standing still, and increase behaviors such as sniffing, snout contact, and slow and fast locomotion (Carr & White, 1987). In the same rats, direct amphetamine injection into nucleus accumbens leads to much of the same changes, whereas fewer effects are observed in dorsal striatal injections.

1.2.3.10 DA and Reinforcement Learning

In addition to effects on response vigor, DA has many effects on reinforcement learning. Given the LTP and LTD effects predicted in Sections 1.2.3.7 and 1.2.3.8, we would predict that extreme hypo-DA would lead to LTD in the striatal Go cells and PFC cells during Hebbian events, whereas hyper-DA would lead to LTP. On the other hand, in striatal NoGo cells, we'd expect hypo-DA to lead to LTP, and hyper-DA to lead to LTD. This suggests a set of hypothesis about the effects of DA reward phasic bursts and punisher phasic dips. Reward bursts during Hebbian events should effectively reward striatal Go and PFC, but punish NoGo synapses. Punisher dips during Hebbian events, on the other hand, should punish striatal Go and PFC, and reward NoGo synapses. In accordance with these predictions, a recent computational model in conjunction with data taken from human medicated and unmedicated Parkinson's patients suggests that

unmedicated Parkinsonism (which would be a chronic hypo-DA condition) correlates with an emphasis on punishment learning at the expense of reward learning, whereas L-dopa-medicated patients (which may be in a hyper-DA condition at times) often have an emphasis in reward learning at the expense of punishment learning (Frank, 2005; Frank et al., 2004).

What particular learning is potentiated or depotentiated would depend on the particular DA target or targets affected. In the case of hypo-DA, such as occurs in Parkinsonism, motor learning would be expected to be impaired in the early stages when the motor (i.e. most dorsal) striatum is most affected. Accordingly, habit-formation in motor tasks is impaired by dopamine depletion of the dorsal striatum (Faure et al., 2005; Robbins et al., 1990) as well as selective lesions of dorsolateral striatum (Yin et al., 2004). Instrumental conditioning is impaired by lesions to the dorsomedial striatum (Yin, Ostlund, Knowlton, & Balleine, 2005) and there is evidence that reversal learning is impaired when DA is depleted there (O'Neill & Brown, 2007). Procedural learning of cognitive tasks is also impaired in PD (Saint-Cyr, Taylor, & Lang, 1988) which would probably correlate with damage to some of the prefrontal corticostriatal pathways. Pavlovian and instrumental learning are both impaired when ventral striatal (NAc) DA is depleted (Parkinson et al., 2002; Smith-Roe & Kelley, 2000).

In determining the likely effects of hyper-DA, the first thing to note is that psychostimulants such as cocaine and amphetamine can act as primary reinforcers for instrumental behaviors, as has been evidenced by experiments where rats learn to lever-press for self-delivery of drugs into medial NAc shell and medial tubercle portions of ventral striatum (Ikemoto, Qin, & Liu, 2005). As a hypothesis to be tested, it may be

proposed that chronic hyper-concentrations of DA are likely to lead to a state of chronic spurious reward reinforcement in the affected pathways. It would be possible, for example, for the organism to accidentally learn the correct S-R mapping during initial training, but it would also be likely for the organism to learn an incorrect first S-R mapping if the organism "tried" the wrong behavior first because, essentially, the organism is under a chronic reward reinforcement state. Moreover, unlearning of an incorrect mapping, i.e. reversal learning, would be difficult during a chronic hyper-DA state. In fact, reversal learning impairment has been demonstrated for both cocaine (Jentsch, Olausson, de La Garza, & Taylor, 2002) and D-amphetamine (Idris, Repeto, Neill, & Large, 2005) delivery in monkeys and rats.

1.2.4 Anterior Cingulate Cortex

A number of computational models of basal ganglia learning treat the basal ganglia as an actor-critic model (Joel, Niv, & Ruppin, 2002). The model in this dissertation is an example of this, possessing an actor pathway, corresponding to a dorsal striatal pathway, and a critic pathway which models, albeit in a much-simplified fashion, the influence of the dopaminergic nuclei. One question that remains regarding actor-critic systems, however, is, How does the actor manage to initially choose the correct behavior, so that it can be "stamped in" by the critic? This dissertation proposes that there is likely to exist an additional cortico-BG pathway (or more than one, perhaps) that is involved with explorative selection of "random" behaviors, a 'babble' pathway. This pathway is engaged when the organism is "stuck" in some sense, and is motivated to "try something, anything" to change its situation. A likely candidate would be a pathway running through the ventral striatum and anterior cingulate cortex.

1.2.4.1 ACC and Willed Behavior

The human anterior cingulate cortex (ACC) consists of BA 24, 25, and 32 (see Figure 1.1), extending through the medial cortex surrounding the corpus callosum (Paus, 2001). As it has reciprocal connections with the (rest of) PFC, efferents to the motor cortex and even the spinal cord, and afferents from numerous limbic and brainstem areas, it is in a crossroads location between neural areas for representing drive/arousal state, cognition, and motor control (Paus, 2001). Paus therefore suggests that it may be in a position to map intentions into actions, i.e., that it mediates "willed control of action".

But what distinguishes *willed* control of action? Norman and Shallice (1986) suggest: "Experientially, a number of different sorts of tasks appear to require deliberate attentional resources. These tasks fit within the following categories:

1. They involve planning or decision making
2. They involve components of troubleshooting
3. They are ill-learned or contain novel sequences of actions
4. They are judged to be dangerous or technically difficult
5. They require the overcoming of a strong habitual response or resisting temptation." (pp. 2-3)

It is intuitively the case that, when a person is initially learning a task, such as driving a car, a great deal of conscious attention is devoted to controlling the sequences of behavior, but as the learner becomes more experienced, less conscious focus is necessary and more of the task performance becomes "automatized". Marvin Minsky has noted that consciousness is chiefly engaged when a person is faced with a non-routine problem that needs to be solved (Minsky, 1986). Conscious will is essentially a flexible problem-

solving method that is engaged when normal, habitual, largely unconscious behavior will not do.

There are, in fact, a number of neurological disorders that seem to demonstrate a selective impairment of willed control, and these involve damage to ACC and connected regions. *Alien hand syndrome* patients have a limb that seems to spontaneously engage in grasping and manipulating behaviors without their consent (Biran, Giovannetti, Buxbaum, & Chatterjee, 2006). Unilateral (contralateral to the alien limb) damage to ACC and neighboring supplementary motor area seems to be responsible, and the likely mechanism is that the "alien hand" is following well-learned automatized manipulation patterns that have become divorced from inhibition by the circuits that mediate willed movement control. There is evidence that a portion of ACC may be involved in discouraging risky behavior such as drug abuse (Fishbein et al., 2005). An fMRI experiment using a counting Stroop task shows that patients with attention-deficit/hyperactivity disorder (ADHD) have bilaterally less activation of ACC (Bush et al., 1999). Reduced activity of ACC has also been associated with mental fatigue in normal human EEG subjects (Lorist, Boksem, & Ridderinkhof, 2005).

However, unilateral damage to ACC or to the connected corticostriatal circuitry can induce a much more serious dysfunction of will, *abulia* (Grunsfeld & Login, 2006; Tekin & Cummings, 2002). Abulia is characterized by "lack of spontaneous and purposeful behavior" with major symptoms including "(1) difficulty in initiating and sustaining purposeful movements; (2) poverty of spontaneous movement; (3) reduced spontaneous speech; (4) increased response-time to queries; (5) passivity; (6) reduced emotional responsiveness and spontaneity; (7) reduced social interaction; and (8) reduced

interest in usual pastimes." (Vijayaraghavan, Krishnamoorthy, Brown, & Trimble, 2002) Bilateral lesions of ACC lead to an even more severe suppression of spontaneous behavior, *akinetic mutism*, which involves "a wakeful state characterized by marked apathy, mutism and lack of motor initiation" (Grunsfeld & Login, 2006). Importantly, akinetic mutism can also be caused by damage to VTA and lateral hypothalamus, and in this instance, symptoms have been improved by dopamine medication (M. P. Alexander, 2001). The abulia case discussed by Grunsfeld and Login (2006) suffered from damage to the right caudate nucleus which is part of the fronto-subcortical circuit including the ACC. Patients given a bilateral anterior cingulotomy for pain tended to manifest abulic symptoms (Cohen et al., 1999).

The above evidence suggests that there is a fronto-subcortical circuit (or set of circuits) involved with the initiation of willed action control (Tekin & Cummings, 2002). This circuit includes ACC and ventral striatum (ventromedial caudate), and requires DAergic activation from VTA. There is evidence in rat studies that effortful behavior may require DA activation of D1 (but not D2) receptors in ACC (Schweimer & Hauber, 2006), as well as the previously mentioned dependence on ventral striatal DA (Sokolowski & Salamone, 1998). Another study in rat medial prefrontal cortex (which includes rat ACC) suggests that D1 and NMDA receptor activity is required for appetitive instrumental (i.e., R-O) learning (Baldwin, Sadeghian, & Kelley, 2002). Generally, it appears that the prelimbic area is involved with learning of instrumental contingencies (R-O learning) (Cardinal, Parkinson, Hall, & Everitt, 2002). While prelimbic area is considered a separate region of rat medial prefrontal cortex from rat

ACC (Uylings, Groenewegen, & Kolb, 2003), in cats, the area corresponds to BA 32 (Room, Russchen, Groenewegen, & Lohman, 1985), which in humans is a part of ACC.

Thus, a cingulostriatal circuit is likely to be involved both with the cognitive control of willed, effortful action, and with the learning of associations/mappings between responses-actions (Rs) and their likely outcomes (Os). Such associations might either drive outcome-prediction (R-O mapping) or motivated response selection (O-R mapping). A prevalent current theory of the ACC's role in cognitive control is that ACC cells engage in *performance monitoring* and their activity influences *performance adjustment* (Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). The performance monitoring is chiefly implemented by ACC whereas the actual mechanisms of adjustment lie in the lateral parts of PFC (dlPFC and vlPFC) (MacDonald III, Cohen, Stenger, & Carter, 2000; Ridderinkhof, van den Wildenberg et al., 2004).

1.2.4.2 ACC and Consequence Monitoring and Prediction

ACC has ample anatomical connections from and to areas involved in emotional appraisal and arousal, including the amygdala (Vogt & Pandya, 1987), the insula which is believed to maintain a holistic estimate of internal homeostatic state of the body (Augustine, 1996; Craig, 2002), and midline thalamic nuclei (Paus, 2001). ACC is also one the most densely dopaminergically (from VTA) innervated regions in primate brain (Allman, Hakeem, Erwin, Nimchinsky, & Hof, 2001). (Interestingly, humans and great apes alone seem to possess a type of neuron known as 'spindle cells' in ACC (Allman et al., 2001); the specialized function of these isn't well understood yet.) The ventral striatum also projects to ACC (Paus, 2001; Voorn et al., 2004). Such connections, along with connections to motor and prefrontal areas, would allow ACC neurons to monitor the

consequences of behaviors as well as the current general state of the organism, and accordingly trigger adjustments in motor behavior that might be more adaptive for the organism. Moreover, parts of ACC connect with the hippocampal regions (Vogt & Pandya, 1987) supporting, perhaps, an ability of ACC to access/retrieve episodic memory records. Episodic memory retrieval might allow predictions of likely outcomes (success or failure, reward or punishment) to be made based on the similarity of the current situation to previous similar situations the organism has been in. Thus, ACC could allow both monitoring and prediction of consequences, and this information could drive adjustments to ongoing behavior.

Consistent with the ACC's connectivity, it has been implicated in a wide variety of consequence monitoring and prediction assessments. Many studies have discovered ACC monitoring of negative outcomes (Ridderinkhof, Ullsperger et al., 2004). ACC activity is correlated with the subjective distress component of pain (Posner & Rothbart, 1998). Reduced reward (Bush et al., 2002; Ito, Stuphorn, Brown, & Schall, 2003; Shima & Tanji, 1998) or actual loss (Gehring & Willoughby, 2002) have been associated with ACC activity. Response errors (the "oops, I pushed the wrong button" type) can lead to the kind of ACC activation believed to be responsible for an *error-related negativity* (ERN) EEG signal (Bernstein, Scheffers, & Coles, 1995; Gehring, Goss, Coles, Meyer, & Donchin, 1993) , and losses also can trigger ERN (Gehring & Willoughby, 2002). Response errors that relate to negative feedback rather than self-monitoring also trigger ACC activity (Amiez, Joseph, & Procyk, 2005; Ridderinkhof, Ullsperger et al., 2004). A popular hypothesis of ACC functionality, based on performance in tasks such as the Stroop task, is that conflict of response tendencies is monitored by ACC activity

(Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999; Carter et al., 1998; MacDonald III et al., 2000; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001). However, not only errors and punishers, but also rewards, either expected or unexpected, may be signaled by ACC activation (Bush et al., 2002; Ito et al., 2003; Niki & Watanabe, 1979).

Regarding prediction, much of error- and conflict-related activity may be interpreted as involving ACC activation in the face of high risk of error (Allman et al., 2001; Brown & Braver, 2005). Brown and Braver's (2005) model suggests that phasic DA punisher dips may train ACC cells to respond to stimuli/conditions that likely correlate with impending errors. ACC activity may increase when the organism needs to engage in error avoidance, for example by aborting the current behavior (Magno, Foxe, Molholm, Robertson, & Garavan, 2006). ACC cells have been found that fire to the degree that the organism expects/anticipates being rewarded (Shidara & Richmond, 2002). Volatility of reward (Behrens, Woolrich, Walton, & Rushworth, 2007) or decision uncertainty (Ridderinkhof, Ullsperger et al., 2004) may trigger ACC activity, and these are essentially also predictions about outcomes. Probably related to this is the fact that some ACC neurons have been found that fire when the organism has to perform explorative behavior in order to find the sequence of behaviors that will lead to reward (Procyk, Tanaka, & Joseph, 2000). ACC related activity has also been discovered under conditions where the needed response or task is rare (Braver, Barch, Gray, Molfese, & Snyder, 2001).

To summarize, ACC seems capable of detecting a wide array of conditions and making predictions regarding consequences of behavior. While the findings above may seem initially contradictory, the heterogeneous nature of ACC and the possibility of

interspersed subpopulations of cells suggest that the area may indeed represent a wide spectrum of (R-O) associations. One possibility offered without follow-up here, is that, similar to the cellular organization of the striatum, there may be separate D1- and D2-responsive cells. The D2 cells, as in Brown and Braver's model, would be 'trained' by phasic dips to represent estimates of error-risk. The D1 cells, on the other hand, trained by phasic bursts, would learn to represent anticipatory estimates of reward-chance. The D1- and D2-trained cells could then drive behavior-selective cells in premotor and prefrontal cortex. The distinct roles of D1 and D2 receptors in ACC need to be more closely investigated.

1.2.4.3 ACC and Response Modification

There appear to be in ACC separate regions involved in cognitive and emotional processing where the most dorsal and posterior components (mostly BA 24) are specialized in cognitive control and the most anterior and ventral parts (BA 32 and 25) are involved in emotional control (Bush, Luu, & Posner, 2000). These parts seem to compete with one another suggesting a mechanism by which major depression (which has been correlated with hyperactivity in BA 25 (Mayberg et al., 2005)) can interfere with cognitive control and cognitive control, on the other hand, can be used to dampen depression (Goldapple et al., 2004). In humans, the emotional (rostral) part of ACC has been shown to be especially active in obsessive-compulsive disorder patients (Fitzgerald et al., 2005), suggesting that overactivity in this area could trigger the kind of hijacking of behavior by ritualistic response that is seen in these patients. The focus in this dissertation, however, is on the cognitive control portion of ACC.

ACC outputs to a number of areas, both cortical and subcortical, that might be deemed output areas. There are reciprocal connections to other parts of PFC, including dlPFC (Paus, 2001) and the frontopolar cortex (BA 10) (Mufson & Pandya, 1984; Petrides & Pandya, 2007). There are also connections to the skeletomotor neurons and motor cortices, and the periaqueductal grey area (PAG) (Paus, 2001), a brainstem area which is involved in motor expression of emotional states (Panksepp, 1998), including vocalizations. As it is involved in thalamocortical loops, ACC outputs to the ventral striatum (G. E. Alexander et al., 1986; Voorn et al., 2004), probably projecting, in part, to both limbic and associative portions of the output BG nuclei (Joel & Weiner, 1994). ACC projects also to the subthalamic nucleus (STN) (Canteras, Shammah-Lagnado, Silva, & Ricardo, 1990) which may provide a means for ACC to inhibit behavior, given STN's role in the basal ganglia action-selection circuit (Frank, 2006; Mink, 1996). These connections may allow ACC to exert a considerable influence on both specific low-level actions and higher-level plans/tasks.

In corroboration with the above anatomical connections, evidence has been found for ACC involvement in executive control. Some evidence suggests that the mechanism by which ACC cognitive control operates is through amplifying/biasing excitation of its target areas, rather than, for example, direct inhibition of opposing behaviors (Egner & Hirsch, 2005). ACC activity has been found in monkey (Shima & Tanji, 1998) and human (Bush et al., 2002) research which selectively correlates with a future decision to switch behaviors, for example, when less reward was received than expected. ACC activity also is involved in response inhibition, for example, in suppression of response in Go/NoGo, or stop-signal tasks (Garavan, Ross, Murphy, Roche, & Stein, 2002; Magno et

al., 2006; Ridderinkhof, van den Wildenberg et al., 2004). ACC activity may need to be engaged, also, for the organism to adapt to and select behaviors according to changing reward contingencies (Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006). Its connection to hippocampus may allow, for example, reestimation of reward probability given recent history. Error-related negativity has been seen to precede more careful, longer RT trials (Gehring et al., 1993). This suggests a role of ACC in raising the threshold of action selection after an error has been made or when there is more risk of error or response conflict (Brown & Braver, 2005; Kerns et al., 2004; Ridderinkhof, Ullsperger et al., 2004). Especially relevant to this dissertation is correlation of some ACC cell activity at periods during tasks when the organism needs to discover a correct response or sequence of responses through trial-and-error (Amiez et al., 2005; Procyk et al., 2000). Explorative behavior (as opposed to exploiting behavior) is also believed, however, to involve frontopolar cortex (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Koechlin & Hyafil, 2007). Given the involvement of ACC in trial-and-error, and its connection to frontopolar cortex (Mufson & Pandya, 1984), it may be hypothesized that the frontopolar cortex could be the output region of an exploratory behavior circuit consisting of ventral striatum, ACC, and FPC.

The model proposed in this dissertation hypothesizes an exploratory behavior circuit such as outlined above which is engaged when an organism is frustrated with the lack of current success at acquiring a reward. It is plausible, given the types of consequence monitoring responses seen already in ACC, that there might be cells there that learn to represent a dislike of not being rewarded and/or a prediction that reward is not forthcoming without intervention. These cells might project to lateral PFC (including

FPC) in such a way as to pseudo-randomly select a plan or behavior, though it's not clear how the noise is injected into the system (perhaps thalamic noise?). Such a circuit would amount to a 'babble' circuit that is engaged when the organism doesn't know how to proceed.

1.3 Potential Contributions of Research

As Section 1.1.5 suggested, computational modeling approaches may be of wide-ranging interest to both the theoretical and applied sciences because they embody and suggest candidate theories of both principles and specific mechanisms of animal cognition. The model proposed here should prove of interest not only to those interested in the study of animal behavior and cognition, but also to clinicians trying to understand the likely impact of pharmacological interventions and the etiology of dysfunctions of behavior and cognition, and also to engineers and computer scientists trying to engineer systems capable of flexible, adaptive behavior in a real-world environment.

1.3.1 Neuroscientific Importance

While there has been a good deal of focus in the neurosciences on studying functional mechanisms for visual perception or spatial localization, study of the mechanisms of executive control (which includes TOBS) is still at an early stage. It is in some ways a difficult object of investigation for the empirical neurosciences because diverse brain areas are involved and these may be a good deal less specialized in their functionality than, for example, V1 and other early stage areas in the visual pathway. Areas such as prefrontal cortex, basal ganglia, the midbrain DA nuclei, and anterior cingulate cortex participate in a number of functional pathways, and the functional

"module" of executive control is densely connected with a host of other brain regions which exert not yet well-understood influences on the control process.

Computational models may serve as provisional theories to guide empirical investigations and to give researchers a preliminary understanding they can use as a point of departure. The proposed model will constitute a candidate explanation of the functional relationships between TOBS/executive control, reinforcement learning, and dopaminergic neuromodulation; and of the division of labor between PFC, basal ganglia, ACC, and the DA nuclei in implementing TOBS. A comprehensive model, even if it is ultimately incorrect regarding many of the particulars of mechanism, may guide researchers by suggesting experiments and testable hypotheses.

The chief scientific benefit of such a model is probably its explanatory potential. TOBS is a critical component of higher mammalian behavior and its investigation is at an early stage. Integrative theories such as (Miller & Cohen, 2001) are useful in providing a verbal and visual overview of how the different areas of brain may interact to produce the complex functionality of executive control. Modeling research such as (O'Reilly & Frank, 2006; Tagamets & Horwitz, 1998) and this dissertation may build on these theories and on recent findings in the empirical neurosciences, in order to propose some specific mechanisms that might be operative. The empirical neurosciences then may investigate which candidate mechanisms seem to be more likely. The various models may then be refined according to later findings, until, ideally, there is a convergence on a single stable explanatory theory that captures the actual mechanisms of executive function.

1.3.2 Clinical Importance

The model proposed in this dissertation, due to the centrality of the brain areas involved and the involvement of the dopaminergic system, may have implications for the understanding of a wide array of clinical dysfunctions and for DA pharmacology. Frontal cortex and anterior cingulate cortex are, as reviewed in Section 1.2, centrally involved with behavior, and activity in both of these areas is regulated by the basal ganglia and the DA nuclei. Though there are many details that may need to be added to the model, it may provide some initial insights into the neurological causes of a variety of disorders of behavior and cognitive initiation and control, especially disorders related to hypo- or hyper-DA receptor activity. In particular, this model may offer insights about the different ways that reinforcement learning and performance of task-oriented behaviors may be disrupted by conditions such as Parkinsonism or psychostimulant intoxication.

Hypo-DA activity (due to DA cell death) is the cause of Parkinsonism (Cools, 2006) , and hypo-DA conditions have been associated with ADHD (Levy & Swanson, 2001), schizophrenia (Grace, 1991; Yang & Chen, 2005), aging (Braver & Barch, 2002), mental fatigue (Lorist et al., 2005), and abulia and akinetic mutism (M. P. Alexander, 2001; Tekin & Cummings, 2002). Hyper-DA conditions have been associated with alcohol and psychostimulant intoxication (Grace, 2000), ADHD (Levy & Swanson, 2001; Zhuang et al., 2001), schizophrenia (Grace, 1991), and obsessive-compulsive disorder and stereotypy (Berridge et al., 2005).

There is here a seeming paradox that some of these disorders are associated with both hypo- and hyper-DA. However, we may begin to resolve this paradox when we consider the differential effects of DA surplus or deficit on different DA targets, a design

step which this model has taken. For example, regarding ADHD, hypo-DA activity in ACC might interfere with cognitive focus by depleting the cognitive monitoring and behavior adjustment circuitry of DA, whereas hyper-DA activity in the ventral striatum might lead to ADHD distractibility symptoms due to an increased bias towards explorative behavior. On the other hand, major depression might be triggered by hypo-DA activation in ACC regions related to cognitive control, or hyper-DA activity in the BA 25 region of ACC whose activity is associated with distress. Ultimately, it is important to understand the varying causes that may lead to some of the same diagnoses and symptoms, such that pharmacological or surgical intervention can be more selectively administered. A major difficulty of pharmacological treatments is that they tend to have a more global effect than desired, activating or blocking receptors in other brain areas besides the preferred targets (though the differentiation of DA receptor types and development of receptor-specific antagonists and agonists has improved selectivity greatly). What models such as the one proposed here suggest is that technology needs to be developed that can selectively influence specific pathways. Means of selectively tweaking DA activity in different target regions could revolutionize clinical treatment of a bewildering array of affective, cognitive, and behavioral disorders.

1.3.3 Technological Importance

While the benefits of a better understanding of human cognition and bases for its dysfunction seem to be the most immediate interest for psychologists and neuroscientists, there are other contributions to be made by modeling human cognitive processes. Artificial intelligence (AI), which may be regarded as both a sub-discipline of computer science (which is often considered an engineering discipline) and cognitive science

(which is an interdisciplinary study of mind) both are deeply concerned with mechanism as well as descriptive theory. Earlier AI approaches have drawn their inspiration from automata theory, symbolic logic, information theory, etc., and were based on mathematical or (artificial) technological metaphors of how animal cognition might function. Periodically, there have been attempts to integrate concepts and metaphors from the neurosciences into AI, but these explorations were done before neuroimaging (aside from EEG) became commonplace and there have been considerable advances in the functional understanding of the brain since the middle of the 20th century. Some AI researchers had made wildly optimistic predictions about the progress that would be made using the old metaphors, and these have not come to pass. While technological and mathematical metaphors may, in principle, arrive at designs for machines capable of human intelligence, progress will be greatly expedited by incorporating the discoveries and theories that come from the study of natural systems which have the benefit of having been designed over a period of millions of years by natural selection. Animal brains are working exemplars of systems capable of conscious perception, and highly complex and adaptive behavior.

The model in this dissertation presents mechanisms which represent advances over what is yet the mainstream in robotics and AI. Reinforcement learning of behaviors is an important advance. Much current robotics and AI work still involves hand-coding of behavior or knowledge databases. This leads to robots that may perform capably in particular limited domains, but are unable to adapt to new conditions in that domain or learn behaviors in other domains. For example, there now exist robots that are pre-programmed with algorithms for vacuuming or mopping a floor (see iRobot's website

at http://www.irobot.com/ for descriptions and demo videos Roomba, Scooba, and other robots), but they perform no other function, nor can they learn any other. It would be better to have a more mobile, legged robot, equipped with manipulators, that could be trained by its owners to vacuum the floor, or shovel snow in the driveway, clean the sidewalks, or any number of other routine, unpleasant tasks. However, such a robot requires a system capable of a much greater level of adaptation and flexible intelligence than what currently exists. Reinforcement learning is one missing component.

Robots that possess a developmental learning period like young animals have, as opposed to being manually coded to perform behaviors, will be capable of far more general and open-ended behavior adapted to more complex and changing environments (Weng et al., 2001). The addition of reinforcement learning mechanisms that allow the learning of behavioral sequences and other mechanisms that allow more than one task representation to be independently maintained, needs to be implemented for a robot capable of learning complex tasks, but the model proposed in this dissertation should at least provide a template for the learning of simple one-step S-R tasks.

1.4 Overview of Dissertation

This chapter has sought to explain the area of investigation of the dissertation and the overall method of inquiry (Section 1.1), to provide a background neuroscience perspective on the modeling task (Section 1.2), and to express likely contributions of the research (Section 1.3). It remains to explain the simulation model in detail (Chapter 2), to present simulation results and compare them with effects seen in the empirical neuroscience literature (Chapter 3), to lay out the theory suggested by the model (Section 4.1), discuss predictions the model makes (Section 4.2), to reflect on the contributions

and limitations of the research (Sections 4.3-4.4), and to suggest future research in continuation of the work of this dissertation (Sections 4.5).

Chapter 2: Simulation Method

2.1 Modeled Environment and Organism

At the heart of this dissertation is a computational simulation model of task learning and performance, implemented as a set of (Mathworks) Matlab scripts running in a standard Windows/PC environment. Figure 2.1 shows the overall functional structure of this model: that of a cognitive agent acting in a simple environment. The overall approach involves simulating, at varying levels of abstraction, the organism's brain, body, and environment, and how these interact during situations where the organism is required to perform particular tasks. The initial goal for the author was to develop a neurocomputational model of task learning and performance that was consistent with the neuroscience literature as outlined in Section 1.2. This involved testing the model on the learning and relearning of S-R task sets, starting from a "naïve" state where the model made no responses to the stimuli, but was rewarded for correct and punished for incorrect responses. Once this model was capable of learning the task, it was then tested under various simulated conditions of dopamine surplus and deficit, under the stages of initial learning, learned performance, and reversal learning, so that the effects of DA manipulations on task learning and performance could be observed. Together, the model architecture and the testing results embody a preliminary theory of TOBS (which will be discussed in Chapter 4).

The organism and environment depicted in Figure 2.1 are extremely abstracted. The organism's "body" is essentially a floating eye or camera, capable of panning left and right and tilting up and down across a nondescript visual (128x128 pixel) "arena" (represented as black in Matlab) upon which a single colored square (Red, Green, Blue,

or Yellow) may be projected. The "invisible trainer" of the environment rewards or punishes this camera organism according to whether it executes a Nod (sinusoidal tilting) or a Shake (sinusoidal panning) after it is presented with a particular color. The organism, however, also has an "ear" which can be presented with brief (simulated) tones corresponding to the twelve notes of a musical chromatic scale: C,Db,D,Eb,E,F,Gb,G,Ab,A,Bb,B. It can be rewarded with simulated food deliveries or punished with simulated shocks, and it maintains an awareness of its internal state of hunger and/or satisfaction. The system is roughly analogous to an immobilized primate facing a computer monitor and being rewarded with fruit juice when it makes a proper saccade response to a stimulus presented on the monitor. The addition of auditory tones and punishers creates a potentially more general and flexible system for testing task learning and unlearning mechanisms.

The "brain" of the camera organism consists of a *cognitive processing* and a *motor control* module. The cognitive processing module, given the current visual, auditory, visceral, and reward/punisher inputs, will select an appropriate *behavior command* for implementation. There are three such commands: Track (the default) in which the camera tries to center its color-sensitive (16x16 pixel) fovea on any stimulus presented in its (64x64 pixel) color-insensitive peripheral field, Nod in which the camera centers itself in the arena and tilts up and down twice, and Shake in which the camera centers itself in the arena and pans left and right twice. Potentially, simultaneous commands may be issued; for example, Nod and Shake may occur together leading to a two-dimensional oscillatory movement. The motor control module implements the mapping between active behavior commands and the immediately appropriate tilt or pan

behavior/s. It is the cognitive processing portion of the model that will mainly be described in Section 2.3, as it is the portion which is implemented with discrete neurocomputational units rather than being simulated through more basic procedural algorithms.

2.2 Task/Trial Structure

The model has, designed into its structure, representations of distinct colors (Red, Green, Blue, Yellow) and distinct tones (C through B on the chromatic scale), and also representation of three behaviors (Track, Nod, Shake). Working memory traces are implemented in a rigid fashion with tone working memory always remembering the last-heard tone and color working memory lasting from the time the color is seen in the organism's fovea until a time shortly before the occurrence of the next color square presentation trial (see Section 2.3.5). However, there are no representations of *tasks* per se wired into the model. Rather, based on working memory traces of tones and colors, the simulated organism must learn what stimuli might represent different tasks and what mappings are appropriate between the combination of task and current stimulus to the correct behavior response. Through rewards and punishers, the model is able to learn such mappings and unlearn them in favor of new mappings.

Prior to testing of the model under hypo- and hyper-DA conditions, the model was tested under a "full" task set consisting of two tasks: BLUE-SELECT and RED-SELECT. The start of these tasks are respectively signaled by the occurrence of a C or an Eb tone. During a multi-trial run, a random C or Eb tone is chosen after 1-5 (randomly selected) trials under a current task. For each trial under a task, either a Blue or a Red square is projected randomly to a place somewhere in the model's 64x64 retina.

By default, the model is always executing a Track behavior command which causes it to foveate on any object presented in its retina. After it has foveated the object so that it can see its color, however, the model is expected (by the "invisible trainer") to make a Nod or Shake response. Under the BLUE-SELECT task, the model is required to Nod in response to Blue squares and Shake in response to others (Red only because only Blue and Red were used in the results documented in this dissertation). Under the RED-SELECT task, the model is required to Nod in response to Red and Shake in response to other colors (Blue). The model is rewarded for a correct response, punished for an opposite response (Nod instead of Shake or vice versa), and neither rewarded or punished for ignoring the square (and simply fixating on it, instead, according to the dictates of the default Track behavior).

Essentially, training of the model consists of presenting a lot of trials to the model until its learned performance is satisfactory. When it has learned the regular "full" task set (using Blue and Red squares and C and Eb tones), it has learned BLUE-SELECT, RED-SELECT, and the cues (in this case, auditory) which signal these. Another way of analyzing what is learned is to say that the model has learned mappings of four (stimulus) *conjunctions* to behaviors, as follows:

- C AND Blue -> Nod
- C AND Red -> Shake
- Eb AND Red -> Nod
- Eb AND Blue -> Shake.

The first two conjunctions contain the learning of the BLUE-SELECT task, and the latter two the learning of the RED-SELECT task. As will be shown, the model is capable of

learning these four conjunctions, although there are learning capacity issues, probably due to the small number of neural units allocated in the model.

The model passes through three simulation phases. In the first, the *training* phase, the model is presented with the four conjunctions of the full task set until it has learned all. In the second, the *maintenance* phase, the model is simply tested for a few runs on a random sequence of conjunctions to see if the model retains its correct responses. Then, finally in a *reversal* phase, the model is trained on the reversal task set, i.e., the following four conjunctions:

- C AND Blue -> Shake
- C AND Red -> Nod
- Eb AND Red -> Shake
- Eb AND Blue -> Nod.

Essentially, the C tone switches from being a signal of BLUE-SELECT to being a signal of RED-SELECT. Likewise, Eb switches from being a signal of RED-SELECT to being a signal of BLUE-SELECT. The model, trained under the initial task set, has to unlearn the old conjunctions and relearn the new ones. As will be seen, it is successful at this, and, although it is not documented in the dissertation, the model is also capable of relearning the old task set after the reversal task set is learned. The results presented in Section 3.1 show the performance of one instance of the model (each instance being a different randomized set of initial model weights of the neural units). Although exhaustive data was only collected and analyzed for one model instance (because of the immense length of time for training and the volume of data generated), performance appeared to be typical, given adequate model training time.

As will be explained in further detail in Section 2.3.14, a simplified task set was used for the hypo- and hyper-DA tests so that data could be collected for several model instances, and statistics could be collected. To summarize, however, training, maintenance, and reversal phases were done for multiple DA conditions, but only two conjunctions were used: one conjunction, C AND Blue -> Nod, in the training and maintenance phases, and the other, C AND Blue -> Shake, in the reversal phase. The likely effects of DA conditions on the performance of the model on the full task set can be inferred from understanding the effects of the same conditions on the 1-conjunction task sets.

Figure 2.2 shows the (overall, correct) performance of the model after it has been trained on the full task set (and therefore is a sneak preview of the results presented in Section 3.1.1) , and also presents a typical testing (as opposed to training) simulation run. In a simulation run (testing or training), there are 1,000 simulation iterations with each iteration representing roughly 100 ms of "neural real-time", so that the total run simulates a period of 100 s. Within a run there are trial blocks of (randomly) 1-5 trials each. At the beginning of each trial block, a random C or Eb tone (500 ms: 5 iterations) is presented which selects the reward/punisher contingencies (e.g. BLUE-SELECT vs. RED-SELECT), followed by a 2.5 s (25 iteration) delay period. Each trial within a block has the following temporal structure: a Blue or a Red square is projected randomly to a place somewhere in the model's 64x64 retina for a period of 3 s (30 iterations), and then a 2.5 s (25 iteration) delay period follows the visual stimulus offset. If a response is made, either a reward or a punisher of 500 ms (5 iterations) is delivered, when the Nod or Shake execution is almost completed. The "invisible trainer" rewards or punishes the model for

responses made between the onset of the colored square and a period 1500 ms (15 iterations) after its offset. No partial trial is allowed to happen near the end of the simulation run; no new stimuli are presented if sufficient time does not remain for a whole trial. Each simulation run typically consists of about 15 trials.

While testing runs were described above, simpler simulation training runs are used for training the model in the training and reversal simulation phases. Each training run maintains the same general structure as the testing runs described above, but only one color and one tone are used in a given run, with the result that the model is only trained on one conjunction during any run. In the training phase, the model is trained, one conjunction at a time; generally the model learns a conjunction in 1 or 2 runs, though sometimes more are required. After all 4 conjunctions have been trained, a test run is done to verify that all conjunctions remain learned. Often, in fact, some conjunctions have been unlearned due to interference effects (described in Section 3.1.1.2), so another iteration of training is done on the failed conjunctions, followed by another test run to see if there are any remaining failing conjunctions. This process iterates until the test run reveals that all conjunctions have been learned. Finally, a special kind of test run is done wherein learning is turned off and any DA manipulations (not applicable in the full task set data) are reset to a normal level. Performance data for the whole training simulation phase is collected from this ending test run which is shown in Figure 2.2. The trained model is then subjected to a maintenance simulation phase which consists of 5 testing runs (of the normal variety) from which performance data is collected. Finally, the model is subjected to a reversal simulation phase whose training process is essentially identical to the training simulation phase, but the reversal task set is used rather than the normal

task set. The special test run from the end of the reversal simulation phase is shown in Figure 2.3. Figures 2.2 and 2.3 illustrate the learned performance of the model and that, save for a few missed response or behavior repetition errors, the learned performance of tasks is high, both under initial and reversal learning conditions.

For a more intuitive view of the model's performance, movies have been made available online that show the dynamic activity during various model simulations. The URLs for these are given in Table 2.1. Movie 2.1 shows the performance of the model at the end of the training simulation phase; the run is identical to that shown in Figure 2.2. Movie 2.2 shows the performance of the model at the end of the reversal simulation phase; this run is identical to that shown in Figure 2.3.

2.3 Model Architecture

It remains to explain the architecture of the brain of the simulated organism. This brain essentially consists of a number of interconnected rectangular layers of neurocomputational mean firing-rate units. The layers correspond to populations of neurons within areas of mammalian brain. Certain input units take on values set by non-unit parts of the model that are involved with stimulus and working memory representation. Other output units drive the simulated behavior of the model. Some of the layers have their activity and synaptic learning modulated by the model's dopamine mechanism. This section will explain the units and layers of the model and how they are collectively able to learn and perform the task sets described in Section 2.2.

2.3.1 Neural Unit Architecture

Figure 2.4 schematically shows the structure of the neural units as well as the dynamic equations. Each unit is a simple mean firing-rate unit that takes a weighted sum

of the output of units, subjects it to a sigmoidal "squashing function", and moves the output value closer to the sigmoided net input sum. The weighted sum for unit j at time t (in simulation iterations) is defined by

(1) $$i_j(t) = \sum_k w_{jk} x_k(t-1)$$

where x_k is the output (ranging from 0 to 1) of unit k and w_{jk} is the weight from unit k to unit j. The squashing function for the units is defined by

(2) $$\theta(x) = \frac{1}{1+\exp(-\beta(x-\tau))}$$

where β is the gain of the sigmoid, setting the steepness of its central slope, and τ is the threshold of the sigmoid, the location along the x axis where the center of the slope is. The shape of this function is shown on the unit schematic in Figure 2.4, and the output range of the function is from 0 to 1 with the value being exactly 0.5 at $x = \tau$. The sigmoidal function, however, is not applied to only the net input, but rather the net input plus some Gaussian noise. The squashed net input may be defined as

(3) $$s(t) = \theta(i_j(t) + N(0,n))$$

where $N(0,n)$ is a random normally distributed variable with standard deviation n.

The output of the unit is defined by

(4) $$o_j(t) = o_j(t-1) + \Delta\big(s(t) - o_j(t-1)\big)$$

where Δ is a growth/decay rate. This means that the change in the output at iteration t is the scaled difference between the squashed net input (plus noise) and the previous output, which means that the output over time will track the sigmoided net input. The output, ranged between 0 and 1, probably may be thought of as representing the mean firing frequency of an individual neuron with 0 representing minimal activity and 1

representing the maximum frequency of firing. Alternatively, the output may represent the proportion of neurons regarded as active within a small population of neurons.

It is the normal dynamic of the units that has so far been described. However, there are a few modifications to the default unit performance which the model utilizes. First, for many of the units, the DA mechanism can drive changes in the weights w. This will be explained in more detail in Section 2.3.13. The sigmoid thresholds τ can be modulated, either by the DA mechanism (see Section 2.3.8), or by the lateral-inhibition mechanism which will be explained in Section 2.3.2. More sparingly, the sigmoid gain β is modified, specifically in the mechanism for the generation of neural noise, which will be explained in more detail in Section 2.3.10.

2.3.2 Layer Structure and Connectivity

The simulation model consists of a number of layers of one or more units, each corresponding with a neuron or population of neurons in a particular area of mammalian brain. Layers consisting of more than one unit are arranged in a rectangular array in the model (typically 4x4 or 8x8). Layers are interconnected through mostly excitatory (but sometimes inhibitory) feedforward connections. The topology of these connections is organized by afferent arbors of essentially three types: *full*, *uniform sparse*, and *one-to-one*. For full arbors, the feedforward connection is fully connected so that each afferent arbor of a destination unit consists of the entire source layer of units. The uniform sparse arbors are like full arbors with some of their connections omitted, those connections being randomly determined by each weight having the same probability of inclusion. In one-to-one arbors, each unit in the destination layer connects only to the equivalent unit in the source layer, e.g. in connection between two 4x4 layers, source unit (2,3) connects

to destination unit (2,3). The initial weights for all arbors are set randomly when a new model instance is created. Each layer has a Gaussian weight distribution from which weights are drawn.

In the results (Chapter 3) and also in the following explanation of the model, aggregate activity is often measured in layers. The type of activity marked as 'sum' on the plots consists of the summation of all of the unit outputs at the current time step. (Thus, an 8x8 layer may have a sum value between 0 and 64.) The type of activity marked as 'tsum' is a thresholded sum whereby each unit's activity is subjected to a hard threshold (e.g. 0.7) and the sum given is the number of units that cleared the threshold.

Some of the layers have their activity and/or their learning modulated by the DA mechanism, as will be described in later sections. Some of the layers have their collective activity modulated by a competitive lateral inhibitory mechanism known as k winners-take-all (kWTA) (O'Reilly & Munakata, 2000). The heuristic used to implement this is that all of the excitatory units that compete with one another are lumped into a common pool. Generally, the pool consists of a single layer of units, but in the model there is an instance (for the Request units) where two layers are pooled together. Essentially, the sigmoided net inputs $s(t)$ (Equation 3) are calculated for each of the units in the pool. In the event that more than k units of the pool clear the threshold of activation (the unadjusted τ), the top $k+1$ of these sums are examined, and the threshold τ for the whole pool is adjusted so that it lies between the kth- and $(k+1)$th-highest sums. For the layers that use kWTA, the gain β is set very high so that the threshold is sharp. Thus, generally only a maximum of k units will become active over threshold during a simulation iteration. This algorithm is a heuristic implementation of what would

be a more complex circuit involving inhibitory interneurons in a global negative feedback circuit. Use of kWTA allows distributed sparse representation 'sizes' to be set in the model so that, for example, a Plan unit representation (see Sections 2.3.11 and 2.3.13) may consist of 5 units, and there may be different sets (hopefully non-overlapping) of 5 units for each stimulus conjunction represented. It also allows lateral inhibition to be more easily implemented in layers where there is only one desired winner.

2.3.3 Functional Overview of Model

Figure 2.5 shows an overview of the functionality of the model. The model consists of four modules. Three of these involve processing of Ss, Rs, and Os, and the final one involves the communication and mapping between these. The Response Processing (R) block is essentially the motor control module from Figure 2.1. This receives behavior commands for Nod, Shake, and Track, and also the black-and-white 64x64 retinal spatial map information that is needed to guide Track direction, and outputs the immediately appropriate pan/tilt action/s. The other three blocks constitute the cognitive processing module of Figure 2.1.

The Stimulus Processing (S) block takes immediate visual retina information, from both the full and the foveal retinal regions; and also auditory tone information; and outputs information used by the other three blocks. It forwards the spatial visual map information to the R block. It also detects whether either the color or the tone is novel/unfamiliar, forwarding this information to the O and Behavior Selection blocks. It sends a signal detecting the presence of a target square to the Behavior Selection block. Finally, it maintains and updates the working memory of the last-seen colors and last-heard tone and forwards these to the Behavior Selection block.

The Outcome Processing (O) block takes stimulus inputs involving primary rewards and punishers, which signify food and shock delivery, respectively. It also receives the novelty signal from the S block, and a "visceral state" signal which increments as the simulated organism grows more "hungry" or otherwise unhappy about not being rewarded. From these signals, it outputs two signals: a signal that measures the "frustration" of the organism, and the DA "activity-oriented motivation" signal which is one of the focuses of this study.

The Behavior Selection block (detailed in Figure 2.6) is the "pseudo-homunculus" that allows the model to map information from the S and O blocks to a behavior selection in the R block. Thus, it might be also considered the "decision" or "choice" module. It takes information about color and tone working memory, (visual) target presence, and novelty from the S block; and the frustration and DA signals from the O block; and generates a behavior command (Nod, Shake, Track) which it forwards to the R block.

2.3.4 Response Processing Module

Figure 2.6 shows mainly the Behavior Selection module, but also shows in more detail the R block. An Exec (execute) unit exists for each of the three behaviors: Track, Nod, and Shake. The Track Exec unit can be thought of as having constant activation, but the activation is inhibited by the Nod and Shake Exec units. The Nod and Shake Exec units are activated by an equivalent pair of Init (initialization) units, which are a part of the Action Gating mechanism, when the appropriate Init cell activity clears a threshold (0.65). Once the Nod or Shake Exec units are activated—and it is possible for both to be activated simultaneously—they stay active until the motor behavior is completed, releasing the Track Exec unit's activity when they complete.

Where the correspondents of the three Exec units might exist in mammalian brain is not entirely clear. The units here are presumed (in Figure 2.6) to correspond to populations of frontal neocortical neurons (probably in PMC or SMA) that have a recurrent excitatory connection that gives them persistent activation once excited, and must then be inhibited by some signal of successful completion of the behavior. How precisely premotor frontal cells implement behavior sequences seems poorly understood at this stage. In this dissertation, they are treated like central pattern generators (Grillner, 2003) that cause the motor apparatus to first center the model's gaze, then perform the oscillatory movements, then finally deactivate when the movements end in the centered position.

It is also possible that Track, Nod, and Shake behaviors, which involve head orientation, could be implemented by brainstem areas, including the interstitial nucleus of Cajal (Klier, Wang, Constantin, & Crawford, 2002) and the superior colliculus which is the locus of the essential circuitry for oculomotor control (Klier, Wang, & Crawford, 2003; Trappenberg, Dorris, Munoz, & Klein, 2001). The basal ganglia has extensive control over both neocortical and brainstem areas (Grillner et al., 2005), so either locus for head orientation control in these tasks is plausible.

2.3.5 Stimulus Processing Module

Central to the simulated model's sensorium is a 64x64 retina (see Figure 2.1). This is implemented as four fields: a 64x64 set of units which correspond to the color-insensitive rod cells, and three foveal sets of 16x16 units which correspond to color-sensitive cone cells: a set each for red, green, and blue (Rolls & Deco, 2002). The 64x64 rods drive the Track behavior through an algorithm that calculates, for the top, bottom,

left, and right 24 pixel-width borders, average activations of the rod map. If the average intensity of a border is over a threshold then, the camera/eye pans and/or tilts in that direction. This automatic tracking mechanism is disabled when Track Exec unit activity is suppressed. The rods field would probably also drive the Stimulus Presence unit, potentially through a thresholded mean algorithm, though it is implemented more simply in the model.

As for the three sets of 16x16 cone color maps, activity from these is passed into a Red vs. Green, Blue vs. Yellow color opponency algorithm (Rolls & Deco, 2002) which outputs its results to a set of four Colors In units (Red, Green, Blue, Yellow). For example, the Red Color In unit takes as its activation value the average (over the whole 16x16 fovea) Red cone activation minus the average Green cone activation. These Colors In units allow the model's immediate detection of colors and are updated on each simulation iteration according to where the camera/eye is directed within the 128x128 pixel arena. Color opponency in animal vision is implemented as early as the lateral geniculate nucleus in the visual processing pathway, and the color-sensitive (Colors In) units in the model may correspond to cells in V4 (Rolls & Deco, 2002). Although only red and blue squares are actually utilized in this dissertation, the model potentially allows representation of colors which could be represented by mixtures of primary color features.

The other major stimulus input for the model is auditory tone. There are 12 Tones In units, each corresponding to a pitch on the musical chromatic scale (C through B). While it is not clear that humans have cells that are receptive to musical pitch per se, the auditory pathway is tonotopically organized, from as early as the cochlea to the primary

cortex (A1) cells (Kolb & Whishaw, 2003). Therefore, tentatively, the Tones In cells could be said to reside in A1. As with the color units, it is actually possible that multiple tone presences could be simultaneously represented. However, only single brief C and Eb tones are utilized for this dissertation.

Both the Colors In and Tones In units have a set of working memory (WM) units associated with them, i.e., 4 Color WM units (Red, Green, Blue, Yellow) and 12 Tone WM units (C through B). Both sets update their state whenever a new color/tone is presented, but their state maintenance in the absence of change is different for the different modalities. The maintenance latency for color working memory is 5 s (50 iterations), and it is essentially indefinite for tones. The rationale for this is that the tone working memory needs to be maintained over multiple trials whereas the color working memory should only be active for the duration of a single trial. How the working memory system learns these respective durations is a question beyond the scope of this dissertation, and the working memory is implemented algorithmically rather than through afferent neural units. Both modalities of working memory are presumed to reside somewhere in PFC in mammalian brain, probably in the ventrolateral portion.

Distinct "what" and "where" pathways have been proposed for both visual (Milner & Goodale, 1998; Ungerleider & Mishkin, 1982; F. A. W. Wilson, Ó Scalaidhe, & Goldman-Rakic, 1993) and auditory (Arnott, Grady, Hevenor, Graham, & Alain, 2005) perception. Clearly, this model is representative of "what" (or object) processing, since the tasks involve recognizing a (color or tone) identity rather than a spatial location. Whereas the "where" (or spatial, or dorsal) visual pathway runs from the retina through the superior colliculus, pulvinar, and posterior parietal cortex (Milner & Goodale, 1998),

to arrive in dlPFC where it participates in spatial working memory (Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998; Smith & Jonides, 1999; F. A. W. Wilson et al., 1993); the "what" (or object, or ventral) visual pathway runs from the retina through LGN, primary (V1) and extrastriate visual areas (V2, V4, etc.), and inferotemporal cortex (Milner & Goodale, 1998), to arrive in vlPFC where it participates in object working memory (Smith & Jonides, 1999; Ungerleider, Courtney, & Haxby, 1998; F. A. W. Wilson et al., 1993). The auditory "where" and "what" perception pathways also are arrayed dorsally and ventrally, respectively, with inferior frontal areas being activated during pitch tasks (Arnott et al., 2005).

There is also a unit in the model for detecting novel stimuli in the S block. This Novelty unit activates for a 500 ms (5 iteration) burst for the first three times that a color or a tone is presented, and ignores the same color or tone thereafter. There is likely to be a mechanism in the medial temporal lobe (hippocampus and parahippocampal regions) for novelty detection (J. E. Lisman & Otmakhova, 2001; Stern, Sherman, Kirchhoff, & Hasselmo, 2001).

Figure 2.6 shows the final stage of units from the S block which are inputs to the Behavior Section module. The activation of these units is set algorithmically rather than through neural activation propagation.

2.3.6 Outcome Processing Module

As Figure 2.6 shows, there are two output units in the O block: the Frustration unit and the DA Signal unit, and these are likely to have neural correspondents somewhere in ACC, and in the midbrain DA cells, respectively. The Frustration unit measures the amount of hunger and/or frustration that the simulated organism is

experiencing as it continues not to be rewarded. According to Figure 2.7, the visceral state of the organism sends an excitatory signal that steadily increases as the simulated organism becomes hungrier. As Figures 2.7 and 2.8 show, the Frustration unit's activity is reset, through inhibition, by the occurrence of (food) rewards. Presumably, if visceral state is the source of the ramping signal, this state may also be reset by temporary satiation. Alternatively, the Frustration unit might have its ramping up driven by, instead of a visceral state signal, more of a top-down expectation-of-reward signal which is active when the organism is in an environmental context where they are expecting to procure rewards. In any case, the presumed correspondent area of the Frustration unit in animal brain is somewhere in the ACC, as there is much evidence for consequence monitoring and prediction activity in this area (see Section 1.2.4.2). As a part of PFC, ACC cells may innervate VTA directly, or they may exert influence on VTA indirectly, through intermediate PFC cells, or through a ventral striatal (NAc) pathway (see Figure 1.7).

Figure 2.7 shows that the DA Signal unit, which corresponds to SNc and VTA cells, is excited by the Frustration unit, and also by Reward and Novelty unit activity; and inhibited by Punish unit activity. Figure 2.8 shows these influences on the DA signal in a sample model run. A base-level DA activity level is maintained tonically, and the afferents superimpose their influence on this. As Frustration unit activity ramps up, it drives up the tonic DA signal until a reward resets the frustration. Additionally, rewards cause massive phasic spikes in the DA signal and punishers massive phasic dips. As will be seen (Section 2.3.13), the extreme phasic levels contribute learning signals (LTP and LTD) to the rest of the model, and the tonic level sets the threshold of the Action and

Babble Gating modules. The likely anatomical connection between the Frustration and DA Signal units was suggested above. There are many potential locations of the Reward signal, including the insula and the lateral hypothalamus. One possibility for the source of the Punish signal might be the serotonergic raphe nuclei which could possibly phasically signal punishment (Daw et al., 2002). It is likely, though, that an excitatory aversive signal could increase activity in the inhibitory, rather than DA-ergic VTA cells (Ungless et al., 2004). The source of this excitatory signal could be ACC or amygdala since both of these areas have excitatory responses to aversive consequences or stimuli. The Novelty unit to DA Signal unit connection may correspond to an indirect pathway from the hippocampus and parahippocampal areas to VTA (J. E. Lisman & Otmakhova, 2001).

2.3.7 Behavior Selection Module

It is the Behavior Selection module, shown in Figure 2.6, which allows S and O block information to influence the R block. Essentially, this module consists of three pathways: a frontal neocortical, dorsal striatal 'actor' pathway which mediates S-R mappings; an anterior cingulate, ventral striatal 'babble' pathway which mediates random O-R mappings triggered either by frustration or novelty; and finally a DA-ergic 'critic' pathway which innervates both of the above pathways, affecting both the threshold of activation and learning.

The 'actor' pathway begins at the Stimulus Presence and Color and Tone WM units. These are fully connected to an 8x8 set of Plan units by learnable weights that initially start out very small. These use a kWTA mechanism such that $k = 5$. This allows representations of size 5 to develop for conjunctions of activity of the S block

afferent units. The Plan units, in turn, are fully connected through more learnable weights to two 4x4 sets of (Nod and Shake) Request units which compete to select Nod or Shake behaviors. Both sets of Request units are lumped into the same kWTA pool which only has one winner. Thus, although there are only two behavior choices, there are multiple Nod and Shake units that may each potentially specialize in responding to different 5-unit Plan patterns, so the representation allowed for, for example, Nod could be a disjunct of multiple conjunctions. This is needed in a situation where the system must learn something like:

(C AND Blue) OR (Eb AND Red) -> Nod.

The Request units drive the basal ganglia Action Gating mechanism which allows or vetoes the actual response of the Nod or Shake Exec units. In mammalian brain, the Plan units may reside in either PFC or in premotor areas like PMC. The Request units are presumed to reside in PMC or SMA, as are the Exec units in the R block. Section 2.3.8 will discuss possible basal ganglia anatomical connectivity to the frontal neocortical areas involved in the requesting and executing the desired behavior.

The 'babble' pathway begins at the Frustration and Novelty units, which both innervate the Babble Request unit in a way such that activation of either triggers Request activity. The Babble Request unit drives a basal ganglia (ventral striatal) Babble Gating mechanism that, in turn, drives the 8x8 set of Babble units. Not shown in Figure 2.6 is the additional factor of "neural noise" that drives the Babble units randomly (see Section 2.3.10) whenever there is a stimulus present. The Babble units, however, require the additional gating of the basal ganglia mechanism (described in Section 2.3.9). There is a single winner in the 8x8 set of Babble units, and the result is that the winner is essentially

random. The Babble units connect through random sparse connections to the Plan and Request units which means that these are randomly activated as a result of a babble being gated. It is through this mechanism that a random choice of a Nod or Shake behavior command is made when the model is prompted through frustration or novelty to "try something new". The anatomical evidence for a babble pathway is reviewed in Section 1.2.4. The Frustration to Babble Request unit projection is probably an intra-ACC connection. The fact that the hippocampal regions project to ACC (Vogt & Pandya, 1987) suggests the likely correspondent of the Novelty to Babble Request unit connection. ACC projects, in large part, to the ventral striatum (Voorn et al., 2004), which would then project, through the BG output, back to ACC. This would correspond to the Babble Gating connections. As for the Babble units, they may correspond to either ACC or FPC units, as discussed in Section 1.2.4.3. ACC connects to both PFC and premotor units (Paus, 2001), and FPC projects to other (more posterior) regions of PFC (Koechlin & Hyafil, 2007); so these pathways may potentially implement the Babble to Plan and Request unit connections.

Finally, the 'critic' pathway spreads out from the DA Signal unit to the Plan and Request unit synapses, and to the Action and Babble Gating modules. The Action Gating module receives its DA-ergic input from the SNc, as the dorsal striatum receives its innervations from that midbrain nucleus. The Babble Gating and Plan and Request units receive their DA-ergic innervations, on the other hand, from VTA. The DA-ergic connections are not implemented as weights, but, rather, the DA Signal level modulates learning in target synapses and/or the thresholds τ of output sigmoidal functions.

2.3.8 Action Gating Module

The 'actor' pathway is gated by a complex basal ganglia dynamic gating mechanism, depicted in Figure 2.9. This, as well as the Babble Gating mechanism described in Section 2.3.9, was designed to implement the braking release mechanism described in Section 1.2.2.2 (see also Figure 1.4), though the specific units and their connectivity are not exactly analogous to the actual BG circuitry. Though action selection is an often-proposed and plausible role of the BG circuitry (Gurney et al., 2001; Redgrave et al., 1999a), in this model, the BG has more of a role of output gating (Brown et al., 2004). Brown and colleagues (2004) in their model suggest a layer specificity in frontal neocortex, where activity in layers II, III, and Va represents, in essence, requests (potentially conflicting) for particular behavior choices; and layer Vb activity represents the actual behavioral output gated. The more superficial layers (II, III, Va) project to the BG, and the BG outputs, in an inhibitory fashion, to the thalamus which projects in excitatory connection to the deep output layer (Vb). (Layer VI provides a top-down enabling signal to both input and output layers in the (Brown et al., 2004) model, but this layer is ignored in this dissertation.) In the dissertation model, the Request units take on the role of the input layer units, and the Init units are analogous to the Layer Vb cells. The Request and Init units are essentially considered to be colocated within the same cortical columns. When Nod or Shake (or both) are requested by Plan unit excitation of the Request units, the Action Gating mechanism decides what behavior (if any) should be allowed to be initiated, and sets the appropriate Nod or Shake Init unit appropriately. If these units are excited over threshold, they trigger Exec activity which cannot be interrupted until the behavior is completed. What behaviors are permitted or rejected by

the gating mechanism can be learned by the Go and NoGo units, respectively. The Go units learn to "open the gate wide", in response to particular working memory contexts, when gating of the behavior during those contexts leads to reward. On the other hand, the NoGo units learn to "shut the gate", in response to particular working memory contexts, when gating of the behavior during those contexts leads to punishment.

A (somewhat archaic and whimsical) military metaphor might be used to explain the performance of the Action Gating mechanism. Imagine a fortified city with two gates, manned by sentries and commanded by a governor who rewards the sentries' performance by providing them with wine. It is important to the governor that certain dignitaries be allowed into the city through specific gates (which correspond to specific behaviors in the model), and that certain undesirables are shut out. The observable characteristics of the people permitted or denied entrance correspond to working memory contexts in the model (in our task, the context of the last-heard tone and the currently present colored square). If a desired dignitary is allowed into the city through the correct gate (corresponding to an appropriate behavior, given the current situation), then the city benefits, and the governor is happy and rewards the sentries temporarily with more wine (a dopamine burst). If a spy or other enemy is allowed into the city and is able to work mischief (corresponding to a wrongly-timed behavior in the model), or if the right dignitary is ushered in through the wrong gate (corresponding to the wrong behavior being executed at the right time), then the governor is unhappy and punishes the sentries by temporarily withholding wine from them (a dopamine dip). The level of the governor's overall happiness with the sentries' performance can be directly measured by the amount of wine he's providing them.

Now, there are three types of sentries in the city, greeters and gate openers (the Go cells), gate shutters (the NoGo cells), and a single captain of the guard (the Conflict unit). The captain of the guard is unable to actually observe and inspect the people that desire entrance to the city, but is aware of the operation of all of the gates, and if the greeter sentries are in the process of opening more than one of the gates at the same time (which is a cause for suspicion in this city), he vetoes their activities. The genial gate openers generally allow more people to enter the city when they have more wine, but are less likely to let people in when they have less. The unfriendly gate shutters, on the other hand have a predisposition to shut everybody out, but when provided with enough wine, desist. In the sentries' hierarchy, the gate shutters are able to overrule the gate openers. Each gate opener or shutter has a limited memory allowing them to recognize only one kind of person (one working memory context), and among the openers or shutters, only one sentry in each is allowed to operate at a time at each gate, the sentry that best recognizes the person to be permitted or denied access. So each sentry specializes in the recognition of a particular class of person, and when they encounter this class, they will let them in or shut them out.

Ordinarily, when wine is at an intermediate level, the gate openers will let in those requesting access by default, though after a significant waiting period. The gate shutters, by contrast, will idle their time away. If an important dignitary arrives, one of the gate openers, recognizing them, will excitedly open the gate for them with less delay than usual. If one of the gate shutters, however, recognizes a particular person as a spy, however, they will insure that the gate is shut. When things are going well and there is more wine flowing, the gate openers are more likely to let people in and the shutters less

likely to shut people out, whereas the reverse is true if times are bad and there is less wine. Each sentry responds in a simple-minded way, generally, reacting to the particular kind of person they see, and not changing their responses. However, when the governor gives the sentries a wine reward, the gate opener who was responsible for letting in the dignitary learns that dignitary's identity and will let them in without delay in the future. On the other hand, a shutter who is not inclined to let in the person in will learn to ignore them in the future. When the governor withholds wine, however, the opposite is true: the opener will learn to ignore the person in the future, and the shutter will learn to quickly shut them out in the future. The result is that the city garrison is capable of learning whom to admit and whom to turn away based on the consequences for the city (in the model, whether the organism is rewarded or punished for executing a particular behavior in a particular context).

Moving from the metaphor to the model, the Request unit activity for both Nod and Shake drives the respective Go E (excitatory) units and also the Conflict unit which is only engaged if both behaviors are requested. The Conflict unit likely corresponds to feedforward inhibitory interneurons—not medium spiny neurons— in the striatum which inhibit the Go E cells when they are sufficiently excited by cortical afferents (Brown et al., 2004). The Go E cells are excited by color and tone working memory units, and also by the neural noise mechanism explained in Section 2.3.10, as well as by the Request units. DA Signal activity modulates both learning of the WM to Go E unit weights, and the output thresholds τ of the Go E units. The operation of the learning mechanism will be detailed in Section 2.3.13. The output threshold effects are shown in Figure 2.10: as DA level increases, the thresholds of the Go E units decrease, making Go E unit

activation easier, and as will be seen in Chapter 3, more prompt. The 4x4 Go E units compete through the kWTA mechanism for a single winner. The gating noise provides a random bias on which unit is the winner during the occurrence of a visual stimulus. Request activity in conjunction with this random bias activity is generally enough to trigger Go E unit activity, but a sufficient decrease of DA can inhibit Go E activation altogether. The Go E units in their activity all excite a single inhibitory Go I unit; this arrangement corresponds to a proposed efferent connection between direct-pathway striatum neurons and GPe inhibitory indirect pathway neurons (Frank, 2006). The direct pathway would actually be a double-inhibitory (i.e. disinhibitory) connection (see Figure 1.4), but it is abstracted in this dissertation as a double-excitatory connection from the Request unit to the Go E units to the Init unit which represents an amalgamation of the GPi/SNr, thalamus, and the neocortical thalamic target. Each Init unit is excited by its corresponding Request unit, and the Go E units, and inhibited by the corresponding NoGo and Go I units. The Init unit's activity depends on, simultaneously, the Request and Go E unit activity. The feedforward excitation/inhibition of the Go-to-Init pathways essentially causes Request activity to phasically excite the Init unit, but disables new Init activity until the behavior request is released so that the Go I units are deactivated.

The 4x4 NoGo units also compete for a single winner through kWTA. They may be excited by learned working memory conditions, or by a conjunction of random gating noise-determined bias and the occurrence of Init unit activity. Thus, for example, any time a Nod is triggered, one of the 4x4 Nod NoGo units is activated by the Nod Init activity, so that DA phasic activity is able to train the response of the NoGo pathway. Both the random noise and the signal from the Init units probably come from a

thalamostriatal pathway (Brown et al., 2004; Mengual, de las Heras, Erro, Lanciego, & Giménez-Amaya, 1999). As shown in Figure 2.10, the NoGo units have their output thresholds affected in the opposite fashion from the Go units: increasing DA increases the thresholds, making it harder to activate NoGo units when the DA level is high.

2.3.9 Babble Gating Module

The gating mechanism for the 'babble' pathway, as shown in Figure 2.11, is simpler than that of the 'actor' pathway, but similarly-organized. Whereas the former pathway is dorsal striatal, the latter is most likely ventral striatal. A Babble Request unit, probably in ACC, excites the 4x4 set of Babble Go E units which compete through kWTA for a single winner. This unit excites an inhibitory Go I unit, and the Request, Go E, and Go I units in their dynamics lead to a phasic activation of the Babble Init unit by Babble Request activation, and then the requirement of disengaging Request unit activity before a new babble can be requested. If the Babble Init unit is driven over threshold (0.5), the Babble Exec unit turns on for a period of 2.5 s (25 iterations), and, while it's on, inhibits the Babble Request unit directly. A refractory period of 5 s (50 iterations) is then required before a Babble Init signal can trigger a new 2.5 s Babble Exec. While the Babble Exec unit is active, it, in addition to babble noise (see Section 2.3.10), drives a single random Babble unit to activation.

The Babble NoGo pathways works in the same fashion as the Nod and Shake NoGo pathways work. Learned working memory contexts may suppress Babble Init activity by triggering NoGo activity, but Babble Init activity also triggers Babble NoGo activity—randomly in one of the 4x4 units—so that the NoGo pathway may be subject to learning through DA Signal bursts or dips.

Figure 2.10's depiction of the threshold modification by DA applies to the Babble Gating, as well as the Action Gating, mechanism: DA makes Babble Go cells more responsive and the NoGo cells less responsive. The DA Signal is, of course, also involved in training the Babble Request to Go E and NoGo pathways, with phasic bursts training the former, and dips the latter.

2.3.10 Neural Noise Generation Mechanism

The Action and Babble Gating mechanisms and the random behavior generation allowed by the 'babble' pathway require a source of randomness, or perhaps, rather, pseudo-randomness. Where that noise source or sources might be located is unclear. However, it is proposed here that the thalamus might be a source, as it receives a wide variety of input related to both external and internal stimuli, and the sheer variability of occurring thalamic states, as a result, might provide a viable source of pseudo-random activation. An illustrative analogy for the algorithm used by the model is that there are layers of n neural units where 1 out of all n of the units is driven to activation by noise input, implementing, in essence, an n-sided die. Sparse random feedforward connections from these "dice" layers to their targets allows randomization of neural activity in these targets.

Figure 2.12 shows the actual implementation of the neural noise source. This may or may not be neurally plausible, but the important result is that there are two sets, 8x8, of units that, upon the appearance of a visual stimulus (a colored square), activate randomly one unit each: the Gated Dice layer and the Babble layer. The Gated Dice activation drives the Action and Babble Gating mechanism noise, projecting to the Go and NoGo units. The Babble units are the source of random Plan and Request activity

during initiation of babble behaviors. A single 8x8 Neural Noise layer generates considerably varying random output activity in all of its units. Activation from this projects in a one-to-one fashion to 8x8 Neural Dice and Babble Dice layers which both have kWTA dynamics set up for a single winner. However, the dynamics of these units is specially modified. Each 8x8 Dice layer has self-excitatory connections for each unit, and a reciprocal connection to a kind of 'shadow' 8x8 inhibitory layer (not shown in Figure 2.12). The gain β of the output units is set extremely low by default, so the kWTA mechanism doesn't yield any clear winners, and the Neural and Babble Dice units basically mirror the activity of the Neural Noise layer. However, when the Dice Set unit is activated, the gain on the Neural and Babble Dice units significantly increases for the duration of Dice Set's activation, so that a kind of working memory dynamic is implemented in the Dice layers, similar to the working memory dynamics the author has used in a previous work (Chadderdon & Sporns, 2006). Essentially, this working memory dynamic, in combination with the kWTA algorithm, freezes the most active unit of the Dice layers into activity until the Dice Set unit is inactivated and the gain returns to its original low level. The Neural and Babble Dice units project, respectively, to the Gated Dice and Babble layers, and do so also in one-to-one projections. These output layers, essentially, take a logical AND of activity from their afferent Dice layers, and the activity of a Dice Gate unit. Thus, the Dice Gate unit's activity signals the Gated Dice and Babble units being set to the values of the Neural Dice and Babble Dice layers, respectively. The presence of a stimulus (perhaps detected by thalamic lateral geniculate nucleus) activates both the Dice Set and Dice Gate units simultaneously. As a result, the

output layers generate no random activity in the absence of a colored square, but each (usually) has a single random unit activated when a visual stimulus appears.

Movie 2.3 shows the noise generation mechanism operating during the simulation run shown in Figure 2.2. Labels for the plots are: (1,1) Neural Noise Units; (1,2) Neural Dice Units; (1,3) Gated Dice Units; (2,2) Neural Dice Set; (2,3) Neural Dice Gate.

2.3.11 Model Implementation of Learned Responses

Figures 2.13 through 2.15 illustrate the dynamics of the model's performance in making learned responses. Figure 2.13 shows an overview of how Nod responses are generated by the model during the same simulation run as that depicted in Figure 2.2 (and Movie 2.1): the test run that occurs after the model has been trained successfully on the full task set. The activity of color and tone working memory can be seen (second and fourth traces). It can be seen, also, that around 3 Plan units are activated during a visual stimulus. For C AND Blue or Eb AND Red, the Nod Request units are activated: generally only one unit winning because of the kWTA mechanism. Nod Request activity triggers Nod Go E activity which, in turn, triggers Nod Go I activity which is inhibitory in nature. The Nod Init unit is activated and, in all but one case, goes over the Nod Exec threshold shown in green, which triggers a Nod behavior. Figure 2.14 shows, for the same simulation run, how the afferents to the Nod Init unit contribute to its activation. The Request and Go E unit activity (shown in black) is excitatory, whereas the Go I and NoGo units' activity (shown in red) is inhibitory. The activation from these sums to form the NodInitSum trace. This last trace is the net input to the Nod Init unit; the actual output of the Nod Init unit follows in the next trace. Interestingly, it can be seen that one of the desired Nods—the last one—fails to execute a Nod behavior because the Nod Init

unit doesn't quite clear the Nod Exec threshold (shown in green in the Nod Init trace). The Shake behavior works in the same fashion as Nod, so figures are omitted for it.

Figure 2.15 shows the crucial spatial structure of learned Plan and Request unit activity. Figure 2.15(a) shows, in one simulation iteration (iteration 45 of the run shown in Figures 2.2, etc.), how the occurrence of a blue square in the context of a C tone affects the Plan and Request layer activity. In the first column of plots, the Visual Presence unit, and the color and tone current stimuli and working memory traces are shown, along with Babble unit activity. Plot (1,2) (first row, second column) shows the weights in fanout from the single Visual Presence unit to the 8x8 Plan layer. Plots (2,2) and (2,3), respectively, show the fanout weights from the Blue WM and Red WM units to the Plan layer. Plots (3,2) and (3,3), respectively, show the fanout weights from the C and Eb tone WM units to the Plan layer. Essentially for this example, the logical AND of Plots (1,2), (2,2), and (3,2) gives the pattern seen in Plot (4,2), the actual Plan unit activation in the condition of Visual Presence AND Blue AND C. Plot (1,3) shows the net input activation of the Plan layer input arbors to each of the Nod Request units; Plot (4,3) shows the same for the Shake Request units. Plot (1,4) shows the net input activation for the Babble layer input arbors to each of the Nod Request units; Plot (4,4) shows the same for the Shake Request units. Plots (2,4) and (3,4) show the Nod and Shake Request unit outputs, respectively. It can be seen that unit (4,2) of the Nod Request layer wins the competition. Figure 2.15(b) shows how the occurrence of a red square in the context of an Eb tone affects the Plan and Request layer activity. It can be seen that a different set of Plan units becomes active, and unit (4,4) of the Shake Request layer wins the behavior request competition. Ideally, each color/tone conjunction is represented by non-

overlapping sets of Plan and Request units. 3-5 units in a Plan representation are sufficient to sustain the correct output that selects the correct Request unit, and to provide some robustness in the representations. As will be explained in more detail in Chapter 3, however, accidental overlap in learning causes behavior unlearning and is, therefore, a source of capacity limitation in the model.

2.3.12 Model Implementation of Babble Responses

Figure 2.16 shows the first half (500 iterations) of run 4 of the training phase for the full task set. (Movie 2.4 also shows this run.) During this run, the model has learned the C AND Blue -> Nod and C AND Red -> Shake conjunctions, but it must now learn the Eb AND Red -> Nod conjunction. The first three times the model is presented with a red square, it does not react, so it is neither rewarded nor punished. The level of the Frustration unit activation, however, rises as the model goes unrewarded. Although the model also responds with Novelty unit activation during the Eb tone deliveries, it is primarily the Frustration unit that drives the Babble Request unit. Once the Request unit is driven to a high enough level, the Babble Go units are finally activated enough to cause the Babble Init unit to activate over the Babble Exec threshold (shown in green in the Babble Init unit trace). The first babble (shown in Figure 2.17(a)) is a Shake which is a wrong guess, for which the model is punished. A second babble fails to trigger either a Nod or a Shake, but the third and last babble (shown in Figure 2.17(b)) triggers the correct Nod behavior, and for the rest of the simulation run, the Babble Request unit is never pushed high again because the model responds correctly to the stimulus and the resulting rewards reset the Frustration activity. So the model ceases babbling once it has learned the conjunction.

Figure 2.17 shows more closely the first and third babble from the simulation run. It can be seen that a single random Babble unit is activated during each babble. A random distribution of 3-6 Plan units is activated as a result of this, and a different behavior is randomly selected (through the Request units).

2.3.13 Model Implementation of Learning

Thus far, the performance of the model in the absence of (or before) learning has been discussed. The afferent weights to the Plan, Request, Go, and NoGo units, however, are subject to learning and unlearning depending on the operation of the DA mechanism. Two prominent features of the model's learning algorithm are that learning is *value-driven*, and *eligibility traces* are used for selecting which synapses are eligible for weight modification. Section 1.2.3.7 discusses value-driven learning and DA's involvement. Figure 2.18 shows graphically the relationship between DA level and LTP (weight increase) and LTD (weight decrease). When the DA level is in the intermediate range, no synapses are modified. On the other hand, when the DA level is in the upper or lower regions, eligible synapses are subject LTP or LTD. The units dominated by D1—the Plan, Request, and Go units—have LTP triggered when the DA level is driven over 0.75 and LTD triggered when the DA level is driven below 0.20. The D2-dominated units, the NoGo units, have LTD in their upper range and LTP in the lower range (for reasons explained in Section 1.2.3.7). Generally, the DA level falls in the intermediate range where no weights change, but phasic (reward or novelty) bursts drive the level into the upper range, and phasic (punisher) dips drive the level into the lower range. The weights are all positive values (never changing signs), generally no greater than 1.0. Additionally, each learning layer of the model has a maximum weight value that caps the

value for each synaptic weight. There is also heterosynaptic competitive weight renormalization (Abbott & Nelson, 2000) for most of the DA-innervated areas which is implemented by a total (afferent) weight maximum. If, after synaptic modification, the total weight runs over this cap, all weights are rescaled to bring the total afferent arbor weight to this maximum.

Eligibility traces are used in a number of models to solve what would otherwise be a difficult credit assignment problem (Barto, Sutton, & Anderson, 1983; Brown et al., 1999; Sutton & Barto, 1990). The difficulty is that Hebbian events (co-occurring pre- and post-synaptic activity at a synapse) are transient affairs, and by the time a reward or a punisher is delivered for the behavior they were responsible for, the co-occurrence of activity is no-longer present. However, if a synaptic memory existed that remembered the Hebbian event over a period of a few seconds, then by the time the reward or punisher is delivered, the eligibility trace, the synaptic memory, would still be active, and the synapses could then be rewarded or punished accordingly. Ca^{2+} channel and voltage-gated activity at specific synapses might allow the necessary synaptic memory (Magee & Johnston, 1995; Takechi, Eilers, & Konnerth, 1998) when the pre- and post-synaptic activity, together, generate enough calcium influx. In the model, a coincidence of pre- and post-synaptic output activity over a threshold (0.7) causes an eligibility trace to be initiated for a period of 3.8 s (38 iterations). During the eligibility trace activation, the synapse is subject to the DA effects shown in Figure 2.18.

Figure 2.19 shows an example of high DA initiating LTP in a single synapse (a), and an example of low DA initiating LTD in another (b). It can be seen in trace 3 of both examples (both of them color working memory-to-Plan synapses) that the eligibility trace

remains on 3.8 s after the co-occurrence of pre- and post-synaptic activity ceases. In the fourth trace, the DA activity is shown with the LTP and LTD thresholds superimposed on them. It can be seen in the final traces that, in the first example the weight increases at the time when the DA level moves over the LTP threshold, and in the second the weight decreases when the DA level drops below the LTD threshold. (The blue lines shown in the Weight trace are the maximum weight values for the synapse.)

The collective behavior of the learning can best be understood by examining how the model forms distributed representations for the conjunctions it learns. Figures 2.20 and 2.21 show an example of the model's initial learning, during the training phase, of the C AND Red -> Shake conjunction. (Movie 2.5 shows the run, and Movie 2.6 shows the superset of the iterations that Figure 2.21 was drawn from.) In the first behavior (and babble), the model guesses the correct behavior output, and the second and future responses (though the latter aren't shown) are correct learned responses. During the babble in iteration 42 (see Figure 2.21(a)), it can be seen that a pattern of (primarily) 5 units is activated in the Plan layer. Plot (3,4) shows that unit (4,4) of the Shake Request layer is activated. The Plan and Request unit activity is stimulated by the Babble unit activity shown in Plot (4,1). It should be noted that there are existing fanout weights during the babble; these are the Plan unit afferent weights for the previously trained conjunction, C AND Blue. During iteration 100, after the reward has been delivered (see Figure 2.21(b)), the pattern previously seen in the Plan units appears in the fanout weights atop the existing pattern for the previous conjunction. The pattern is in the Visual Presence, Red WM, and C WM fanout weights which means that the next

occurrence of Visual Presence AND Red WM AND C WM will cause the corresponding Plan units to reactivate: the same as were active in Figure 2.21(a).

The learning dynamics of the Plan layer is typical in the model. Each Plan unit is innervated with Visual Presence, Color WM, and Tone WM activity. When the model is rewarded for a correct behavior in these simulations and a Plan unit is active, the Visual Presence unit is usually still active, and the color WM unit for the viewed color and the tone WM unit for the tone context are always active. Because of the individual weight maximum, the afferent arbor maximum, and a high (fast) learning rate, the units are able to learn in one reward presentation a conjunction of active afferents that can potentially learn "don't care" conditions. For the Plan units, the maximum individual weight, the arbor maximum, and the learning rate (the increment of the weight during LTP and decrement during LTD) are all set to 0.55, and it is the case that an afferent output driving a single weight of 0.55 is enough drive the unit over the sigmoidal threshold. So, if only, for example, a Blue WM unit were turned on, and there was no tone working memory or Visual Presence unit activity, then the weight from the Blue WM to the Plan unit would adjust to 0.55 on reward, and thereafter the unit would be activated whenever the blue working memory was engaged, but the tone and Visual Presence activity would be treated as "don't care" conditions. However, if, as is typical, the Visual Presence unit is on, and one of the color, and one of the tone working memory units, then all of the three afferent units try to add 0.55 to their weights to the Plan unit during the reward-triggered LTP. However, this leads to an arbor sum of 1.65 which is over the 0.55 maximum arbor threshold. Therefore, the renormalization algorithm operates so that all of the 0.55 weights are scaled down so that their sum is 0.55, i.e., each of the 3 weights is

scaled down to 0.183. The result is that all three afferent units need to be active in the future in order to drive the Plan unit over threshold; there are no "don't cares". Generally speaking, the total afferent activity needs to be over 0.5 for the unit to be activated. A single afferent unit's activation only provides 0.183 and any pair of active units only provides 0.367 in afferent net input, and these are not enough to activate the Plan unit. The renormalization algorithm essentially forces a logical AND to be established for all of the previously active afferent units.

The Nod and Shake units operate according to a similar dynamic to the Plan units. However, they only receive 2 inputs: Color WM, and Tone WM input. Thus, they form conjunctions that can have at most 1 "don't care" condition. Movie 2.7 shows the learning the Shake Go weights during the time frame of the Figure 2.20 simulation.

The Request units learn under a somewhat different dynamic. The maximum arbor weight is set much higher, 1.0. The learning rates, however, and the individual maximum weights are set much lower, 0.2. Thus, at least 3 Plan units connected to learned weights have to be activated in order to make the Request unit activate over threshold. Generally, somewhere between 3 and 5 units are activated during a babble, so the Plan unit representation for a conjunction requires only 3 out of potentially 5 units to be active, and it doesn't matter which 3 units. This provides some robustness in the model to degradations that come from losses incurred during learning of new conjunctions or failed babbles which trigger unlearning in previously well-allocated Plan units.

One of the features of the model is its capability of reversal learning. Figures 2.22 through 2.24 show an example of how a conjunction can be unlearned and relearned.

Specifically, what is shown is from the first simulation run of the reversal phase. The desired conjunction is C AND Blue -> Shake. However, what has been previously learned in the training phase is C AND Blue -> Nod. During the first presentation, the model tries the behavior it learned previously and is punished for it. Thereafter it ceases to make that response, but finally, during the sixth presentation of the blue square, the model babbles and ends up choosing the correct response which it then learns. Figure 2.22 (and Movie 2.8) shows the overall dynamics. Figure 2.23 (and Movie 2.9) shows how the weights are changed during the initial punishment unlearning. The pattern seen in the Plan activation in 2.23(a) is effectively subtracted from the weights of the Visual Presence, Blue WM, and C WM fanouts to the Plan units in Figure 2.23(b), so that the occurrence of the stimulus conjunction will no longer trigger those Plan units in the future. Then in Figure 2.24(a), it can be seen that the correct babble activity seen in the Plan units gets added to the Visual Presence, Blue WM, and C WM fanout weights in Figure 2.24(b) so that the those Plan units will be activated in the future. In addition, though it is not shown, these Plan units will have learned weights to Shake Request unit (4,3) so that the new response is a Shake rather than a Nod. Other learning, not shown, transpires in the Go units that allows the appropriate action gating.

2.3.14 Dopamine Effects Manipulation

So far, the performance of the model under normal DA conditions has been explained. However, this dissertation is also interested in the effects of modifying the dopaminergic activity of the model, specifically producing path-selective conditions of hypo- or hyper-DA concentration, and testing both learning and performance under these conditions. Thus far, learning has been understood as being directly controlled by the

DA Signal unit's activity. In fact, there are six intermediate DA effects variables that are maintained for the DA targets, and these variables are the true input values for the DA learning threshold function shown in Figure 2.18:

- Plan unit DA effect
- Request unit DA effect
- Nod/Shake Go unit DA effect
- Nod/Shake NoGo unit DA effect
- Babble Go unit DA effect
- Babble NoGo unit DA effect.

Agonism and antagonism, respectively, for these DA effect variables, are implemented by adding or subtracting numbers between 0 and 1 to the DA Signal unit activation. The resulting DA effect values are then clipped between 0 and 1. Effectively, the agonist/antagonist effects shift the range of the DA signal at which LTP or LTD occurs in the target units (for the Plan, Request, Go and NoGo units), and shift the activation threshold τ for the target units (Go and NoGo only; see Figure 2.10).

The DA manipulation results (covered in Section 3.2) utilize a different task set for training and testing the model in order to simplify and expedite the training process and conserve data storage space, so that statistics can be collected on the results. Instead of the full, 4-conjunction task set, only one conjunction is learned by the model: C AND Blue -> Nod in the training and maintenance phases, and C AND Blue -> Shake in the reversal phase. 10 different dosages of agonism and antagonism are simulated under 4 different pathway conditions. The dosages are notated as follows:

- x3: hypo-DA 0.3 (subtract 0.3 from DA Signal)
- x2: hypo-DA 0.2 (subtract 0.2)
- x15: hypo-DA 0.15 (subtract 0.15)
- x1: hypo-DA 0.1 (subtract 0.1)
- x05: hypo-DA 0.05 (subtract 0.05)
- norm: normal DA (use the default DA Signal)
- X05: hyper-DA 0.05 (add 0.05 to DA Signal)
- X1: hyper-DA 0.1 (add 0.1)
- X3: hyper-DA 0.3 (add 0.3)
- X5: hyper-DA 0.5 (add 0.5).

The pathway conditions are as follows:

- ns/NS: (hypo-/hyper-DA) Nod/Shake (dorsal striatal) pathway (modify Nod/Shake Go and NoGo units)
- b/B: Babble (ventral striatal) pathway (modify Babble Go and NoGo units)
- pr/PR: Plan/Request (neocortical) pathway (modify Plan and Request units)
- g/G: global hypo-/hyper-DA modification (modify all pathways the same way).

So, there are 37 total DA manipulations: 36 hypo- and hyper-DA conditions, e.g., ns15 would mean hypo-DA in the Nod/Shake pathway with 0.15 as the subtraction performed; and 1 norm condition where DA is unmodified. For each DA manipulation, multiple model instances have data collected for them so that means and standard deviations can be calculated for that data.

Figures 2.25 through 2.28 show examples of how different DA manipulations affect the target area DA effects. Figure 2.25 shows the normal (non-manipulated) DA

effect: the DA effects at all of the targets track exactly the DA Signal unit. Figure 2.26 shows the DA effects for the ns2 (Nod/Shake hypo-DA 0.2) condition: the Nod and Shake Go and NoGo units selectively have their DA effects suppressed by 0.2. Figure 2.27 shows the DA effects for the PR5 (Plan/Request hyper-DA 0.5) condition: the Plan and Request units selectively have their DA effects boosted by 0.5. Finally, Figure 2.28 shows the DA effects for the g2 (global hypo-DA 0.2) condition: all DA targets have their DA effects suppressed by 0.2.

Chapter 3: Results

Two sets of results were collected for the model. Section 3.1 discusses data collected under the full task set regime explained in Section 2.2. Essentially, a single simulation 'subject' is studied which is a particular instance of the model, meaning that the initial weights are randomly set only once and that instance is used for all of the simulation runs. The model was first trained in the initial training simulation phase. This training-phase model was then tested through the maintenance and reversal simulation phases to determine how well the model could maintain or reverse its learning, respectively.

The results described in Section 3.2 involve testing the effects of selective hypo- and hyper-DA level manipulation (see Section 2.3.14) on performance in the training, maintenance, and reversal simulation phases. Multiple 'subjects', i.e., different randomized model instances, were tested under the DA manipulation conditions under the three simulation phases with, again, the training simulation phase being followed, in a bifurcating fashion, by simulation/training in the maintenance and reversal phases. Means and standard deviations were collected for performance measures and measures of aggregate model activity.

3.1 Full Task Set Results

3.1.1 Training Phase

Figure 3.1 and Table 3.1 show the performance of the model over the entirety of the training phase. Figure 2.2 and Movie 2.1 show the performance over the last 13 trials which are of the special testing type (see Section 2.2).

3.1.1.1 Full Task Set Successfully Learned

A quick appraisal of Figures 3.1 and 2.2 shows that the model successfully learns the full task set. In the final 13 trial testing run, 11 trials (84.6%) are correct responses, and the remaining 2 (15.4%) are misses, i.e., failures to respond. After around 190 or so trials (13 simulation runs / 13,000 iterations / 1,300 s. "real-time"), all 4 conjunctions are successfully learned.

Figure 3.2 shows how the model accommodates the 4 conjunctions. Each of the 4 conjunctions has essentially non-overlapping representations in the Plan, Request, and Go units. During the initial babble, random Plan and Request units are excited by the Babble units, and the Request activity, in conjunction with random gating noise activity, generates random Go unit activity. The kWTA lateral inhibitory mechanism causes this activity to be focused on a few units, making the representations sparse. If the babble leads to no behavior being generated, as sometimes happens, then there is no learning. If the babble leads to a correct behavior guess, then the active random representations get "stamped in" by the reward. On the other hand, if the babble leads to a wrong behavior, then the random representations get "stamped out" by the punisher. A learned correct behavior, i.e., a correct response in the absence of a babble, then, leads to reinforcement of its driving representations when it receives a reward.

3.1.1.2 Conjunctions Can Be Overwritten During Training

However, the punishment of incorrect babbles, as well as the learning of new conjunctions, can lead to unlearning of previously learned conjunctions. An example of this is shown in Figures 3.3 and 3.4 and Movie 3.1. In run 2 of the training simulation phase, the model is tasked to learn C AND Red -> Shake, but in the process of doing so,

the conjunction learned in run 1, C AND Blue -> Nod, is unlearned. Figure 3.3 shows that between iterations 410 and 470 there is an incorrect (i.e., Nod instead of Shake) babble, for which the model is punished.

As it turns out, this punishment leads to unlearning. Figure 3.4 shows how. Early in the babble (at iteration 428) (see Figure 3.4(a)), 2 Plan units are triggered which participate in the C AND Blue -> Nod conjunction. The corresponding weights are made eligible for reward or punishment after this iteration. Not shown, but viewable in Movie 3.1, is the fact that as the babble progresses, a different set of Plan units is activated by the active Babble unit. Afferent weights to these Plan units are made eligible, also, but the weights made eligible in iteration 428 are the critical ones that lead to trouble. As Figure 3.4(b) shows, after punishment, these weights are decremented, and it can be seen that only a conjunction of 2 Plan units could be activated by the Visual Presence AND Blue WM AND C WM conjunction. As mentioned in Section 2.3.13, at least 3 units need to be active to trigger Request unit activity, so the previous conjunction has been effectively disrupted. Had the babble resulted in the correct response (Shake), however, the C AND Blue -> Nod conjunction would still have been unlearned because the 2 Plan units shown in Figure 3.4(a) would have been co-opted by the new conjunction.

Thus, the model is subject to capacity limitations due to the relatively large number of Plan units (3 to 5) required out of 64 units (8x8) for each conjunction. At most 21 (64 / 3 rounded) conjunctions could be represented in the Plan layer, and this would only be possible if the random unit selection miraculously caused no collisions of new unit representations with the old ones. In an actual brain, it is likely that there would be a much larger number of units so that the activation by stimulus conjunctions would

be much sparser and far more conjunctions could be learned. Nonetheless, it is likely that analogous interference effects might be found in prefrontal cortical learning.

3.1.1.3 Novelty Sometimes Drives Learning

It is not always the case that learning is triggered by reward delivery. Figure 3.5 shows a curious instance where the model tries a babble (around iteration 100) and fails to generate a behavior, but nonetheless eventually makes the correct response without generating another babble (around iteration 225). Movie 2.1 shows the entire simulation run this is taken from. It is, in fact, the first simulation run of the training simulation phase, when the model is tasked to learn C AND Blue -> Nod.

How does the model learn without being rewarded? Figures 3.6 (and Movie 3.2) and 3.7 show how this is possible. Figure 3.6(a) shows the babble-induced Plan unit activity, and that the Plan weights are initially zero. Figure 3.6(b) shows that the weights corresponding to most of the Plan units active in (a) are learned by iteration 150. Figure 3.7 shows the learning of one of these weights: the weight from the Blue WM unit to Plan unit (6,1). Returning to Figure 3.5, it can be seen that a novelty burst was delivered just before iteration 150. This was triggered by the third occurrence of the blue square (which happens to be the last occurrence which triggers novelty). It is not clear why the Nod Go unit is not initially activated, but it is clear from the fifth and sixth plots in Figure 3.5 that the Plan and Request units of the model have learned the conjunction due to the novelty-induced DA burst. Later, the Nod Go unit finally responds to Nod Request activity, probably due to the decreasing Go threshold as Frustration unit activity drives the DA level up. This causes the Nod behavior to finally be performed, so that the Nod Go units,

as well as the Plan and Request units, are reinforced. Thus, the behavior is finally learned.

3.1.2 Maintenance Phase: Task Set Learning Maintained

Figure 3.8 and Table 3.2 show the performance of the model over the entirety of the maintenance phase. Mostly, the model succeeds in maintaining its learning. Of the 73 trials, 54 (74.0%) are correct responses (or repeats in 2 cases), and the remaining 19 (26.0%) are missed responses. For each of the trials, the reaction time (RT) is measured from the appearance of the square to the onset of the response (Nod or Shake). The average RT over all 73 trials turns out to be 1.12 s (11.2 simulation iterations). This is the typical performance of the model under normal DA concentration conditions.

3.1.3 Reversal Phase: Reversal Task Set Successfully Learned

Figure 3.9 and Table 3.3 show the performance of the model over the entirety of the reversal phase. Figure 2.3 and Movie 2.2 show the performance over the last 13 trials which are of the special testing type (see Section 2.2). As with the training phase, 11 (84.6%) of these trials are correct responses (though one has a repeat response, a double-Shake), and 2 (15.4%) are missed responses. After around 70 or so trials (5 simulation runs / 5,000 iterations / 500 s. "real-time"), all 4 of the reversal conjunction were successfully learned. Figure 3.10 shows the new Plan, Request, and Go unit representations for the 4 conjunctions. Again, they are non-overlapping. They are also notably different configurations than the representations in the training (and maintenance) phase which are shown in Figure 3.2.

3.2 Dopamine Manipulation Results

Two classes of effects are seen in response to hypo- and hyper-DA manipulations: behavioral vigor effects, and learning effects. The main vigor effects to be explained are

- Hypo-DA slowing of RT;
- Hypo-DA suppression of behavior initiation;
- Hyper-DA speeding of RT.

The learning effects to be explained include

- Hypo-DA impairment of acquisition (the training phase);
- Hypo-DA impairment of acquired performance (the maintenance phase);
- Hypo-DA impairment of reversal learning (the reversal phase);
- Hyper-DA impairment of acquisition (the training phase);
- Hyper-DA impairment of reversal learning (the reversal phase).

3.2.1 Hypo-DA RT Slow-down

As would be expected in a Parkinson's patient, hypo-DA leads to a slowed reaction time, but whether that effect is seen depends on which DA pathway is impaired. Figure 3.11 shows the effect of DA manipulations on the average RT of a maintenance simulation phase trial. All RT data is collected from 5 trained (through the training simulation phase) 'subjects', each of which is submitted to 5 runs of trials, and is scored a mean RT by averaging over all trials where there is a response. Nod/Shake pathway manipulation, which corresponds to the dorsal striatal pathway (SNc to dorsal striatum) leads to progressively slower RT as the 'dosage' is increased of the antagonism. No equivalent effect, however, is seen when only the Babble pathway (VTA to ventral striatum) has its DA depleted. At the most extreme conditions (pr3 and pr2), an increase

in RT is seen when the Plan/Request (VTA to frontal cortex) pathway is DA-depleted, but it is an artifact due to the fact, to be discussed in Section 3.2.5, that DA depletion causes unlearning of the task. The effects of global DA depletion (all pathways) are (in general, throughout all of these results) reflective of the superposition of the Nod/Shake, Babble, and Plan/Request effects.

This DA RT-slowing effect is consistent with effects seen in Parkinson's patients and in rat and primate studies where DA antagonists are delivered or DA pathways are lesioned. Figure 3.12 shows results of a meta-analysis of many human Parkinson's patients, both L-dopa medicated and unmedicated, performing RT tasks (Gauntlett-Gilbert & Brown, 1998). Both unmedicated and medicated patient groups have slower RTs relative to the control (non-PD) groups. However, the L-dopa medicated subjects have less severe RT impairment, as can be seen in the plot by the fact that their performance is closer to the diagonal line. Parkinsonism, at least at the early stage, corresponds with hypo-DA conditions in the Nod/Shake (dorsal striatal) pathway (Cools, 2006). The medicated and unmedicated conditions, then, may be viewed as two different levels of DA antagonism in the Nod/Shake pathway. Another study of untreated Parkinson's patients performing a simple reaction task (Muller et al., 1999) also suggests that PD leads to decreases in reaction and movement time.

In an operant conditioning rat study in which the subjects learn to respond to slow vs. fast stimuli (either visual or auditory pulse trains), selecting one of two levers for a food pellet reward (Robbins et al., 1990), delivery of different dosages of a DA antagonist to rats that have learned the task leads to increased reaction times, as can be seen in Figure 3.13(bottom). Both the time it takes for them to poke their noses into the

food magazine to receive their reward (magazine latency) and the time it takes to press the correct lever (latency to correct) are increased with increasing dosage of the DA antagonist. Figure 3.14 shows, for the same study, that during training sessions on this same task, rats show decreasing RT as they learn the task. However, RT is significantly slower, both during the course of the learning and at the end, when the dorsal striatum (CAUD trace) is DA-lesioned through selective injection of the neurotoxin 6-OHDA. On the other hand, consistent with the model results shown in Figure 3.11, DA lesions to the ventral striatum (NAS trace) do not show a significant increase in RT. This is consistent with the idea that the dorsal striatum is selectively implicated in the 'actor' pathway, the pathway that performs learned behaviors, while the ventral striatum is not.

Lesion and drug-delivery studies in primates also show that DA depletion leads to slowing of RT. In a macaque monkey experiment where MPTP was used to selectively lesion nigrostriatal DA neurons, it was seen (following L-dopa treatment necessary to restore some level of motor activity) that RT was increased in a simple reaching for food task (Schultz et al., 1989). Another experiment with rhesus monkeys showed that delivery of either a D1 or D2 antagonist leads to slowed RT on a simple reaction task (Weed & Gold, 1998). D1 antagonism in the model would lead to a raising of the Go unit thresholds and therefore decreased Go unit activity, whereas D2 antagonism would lead to a decrease in the NoGo unit thresholds and therefore increased NoGo unit activity.

Figures 3.15 and 3.16 suggest, in terms of the model, how hypo-DA in the dorsal striatal pathway might lead to slower RTs. The Nod/Shake Request units ramp up in their activity after the presentation of a colored square (see the first trace of Figure 3.15).

Once that activity clears a threshold, the Request unit is able to activate corresponding Go unit activity (as seen in the second trace). Simultaneous Request and Go activity, then, initiates corresponding Init unit activity, as seen in the third trace, and this, in turn, leads to the execution of the behavior. Hypo-, as compared with normal, DA conditions lead to an increase in the threshold for the Go units, shown by the shift of the horizontal blue line to the red. This leads to a longer interval between the stimulus and the Init unit activity. Figure 3.16 shows that hypo-DA leads to an average decrease of overall Go unit activity, an effect that is caused by the raised Go unit thresholds.

3.2.2 Hypo-DA Behavior Suppression

Not only do hypo-DA conditions lead to slowed RT, however, but also a net suppression of behavior: extreme akinesia. Figure 3.17 shows that hypo-DA in the Nod/Shake pathway leads to an increase in the percentage of missed responses, failures to react to the presented stimuli. Correspondingly, Figure 3.13(top) shows, for the rat study involving learning of the lever-pressing task, that the percentage of correct responses for rats that have learned the task decreases with increasing dosages of a DA antagonist (Robbins et al., 1990). In the hypokinesia primate study of Schultz and colleagues (1989), MPTP delivery initially caused extreme akinesia, absence of self-initiated skeletal and eye movements. Several days of L-dopa treatment were required before the monkeys were able to react sufficiently to have RT data collected in the task.

In the model, the source of this behavior suppression can be traced to the lowering of the NoGo thresholds and the raising of the Go thresholds. Figure 3.18 compares maintenance phase runs for one of the 'subjects' under normal (a) vs. hypo-DA (b) conditions. The Nod Plan and Request unit activity is basically the same in response to

the blue squares. However, in the hypo-DA case, the Nod Go E units fire less reliably, and the Nod NoGo threshold is decreased and the NoGo units are activated during stimulus presentations, except periodically.

What causes the hypo-DA model to respond around every third time rather than each time? It is notable that after a response that the model is rewarded for, it always misses the next trial. As the Go E units fail to activate in each of these cases, the likely cause is that the Go threshold is too high. This is a result of the DA level dropping to its lowest level when the Frustration unit is reset. In the second response after the rewarded trial, the Go E units activate somewhat late, but activation of the NoGo units precedes it, being triggered immediately at the onset of the Blue WM unit. During the actual responses, the NoGo unit response tends to be weaker and the Go E units respond more quickly and less sluggishly. Figure 3.16 shows how Nod/Shake hypo-DA conditions lead to decreased Go unit activity, and Figure 3.19 shows that (Nod/Shake) NoGo unit activity is increased by Nod/Shake hypo-DA conditions. Through both influences, behavior in the model is suppressed until the Frustration level drives the DA level high enough to lower the Go thresholds and raise the NoGo thresholds enough to allow the Nod or Shake behavior to be gated.

3.2.3 Hyper-DA RT Speedup

While hypo-DA conditions in the Nod/Shake pathway lead to increases of RT, hyper-DA leads to decreased RT, as can be seen in Figure 3.11. Figure 3.20 shows that, in a simple reaction time experiment involving baboons, moderate dosages (either chronic or acute) of cocaine, an indirect agonist, lead to decreases in RT (Hienz et al.,

1994). In a human subjects study, D-amphetamine, another indirect agonist, speeded RT in two discriminative response tasks (Halliday et al., 1994).

Figure 3.21, mainly, shows how the Nod/Shake hyper-DA condition leads to a decreased RT. Increasing dorsal striatal DA level leads to a decrease in the threshold on the Request units' triggering of a corresponding Go unit response (shown by the movement of the horizontal blue line to the red line). Correspondingly, the interval between the onset of the visual stimuli and the onset of the Go unit, and thus the Init unit, shrinks. Figure 3.16 shows that Nod Go unit activity, in fact, increases under Nod/Shake hyper-DA conditions, but not under Babble hyper-DA conditions. Ventral striatal DA has no significant effect on RT.

3.2.4 Hypo-DA Impairment of Acquisition

As Figure 3.22 shows, hypo-DA has a negative impact on learning of the task during the training simulation phase. The data points represent the percentage of 10 'subjects' that successfully learn the initial (C AND Blue -> Nod) task. The task is deemed learned if, in the last 13 trials of the training simulation phase, the model makes a correct response at least 25% of the time. At extreme hypo-DA conditions, both Nod/Shake and Plan/Request (and global) manipulations lead to extreme likelihood of failure to learn the task, with pr2 and pr3 conditions effectively disabling learning and ns3 driving learning chances down to 10% (1 out of 10 subjects). Babble hypo-DA is progressively, but less, disruptive, with the extreme b3 conditions leading to 4 out of 10 subjects learning the task.

Consistent with the Nod/Shake hypo-DA impairment, there is evidence that DA lesions of the dorsal striatum lead to significant deficits in task learning in rat and primate

studies (Packard & Knowlton, 2002; Robbins et al., 1990; Yin et al., 2004; Yin et al., 2005). Figure 3.23 shows that DA lesions of rat dorsal striatum lead to significantly increased numbers of errors made before the rats acquire a dual lever-pressing task (Robbins et al., 1990). This number of errors is also increased, but far less, for ventral striatal lesions. Figure 3.24 shows that lesions made specifically to the medial portion of the dorsal striatum can lead to impairment of lever-pressing task acquisition (Yin et al., 2005). In another rat study, 6-OHDA-generated lesions to specifically the lateral part of the dorsal striatum have been shown to disrupt stimulus-response habit formation (Faure et al., 2005). There is also evidence that D1 antagonism, as well as NMDA antagonism localized to the rat nucleus accumbens core (part of the ventral striatum) leads to impairment of appetitive task learning (Smith-Roe & Kelley, 2000).

Figures 3.25 through 3.28 suggest how the effects seen in Figure 3.22 are caused by the structure of the model. Figure 3.25 shows that Nod/Shake hypo-DA can lead to failure of the Nod/Shake Go E units to be activated, a consequence of the raised Go unit thresholds when DA level is increased. The model attempts to babble several times, but none of the babbles lead to an actual behavior being emitted. (The final Nod behavior is probably due to novelty learning of the Plan/Request representations early in the run.) The Request units try many times to request a Nod, but the Go units' inactivity prevents them from driving the Nod Init unit over the activation threshold.

Figure 3.26 shows how Plan/Request hypo-DA conditions can impair learning. A specific weight between the Blue WM unit and one of the Plan units is shown. Near the middle of the run, there is a Hebbian event and the eligibility trace becomes active. A reward is delivered during this period and this drives the DA Signal level up above the

LTP threshold. However, the hypo-DA manipulation in the Plan/Request pathway leads to a shifting of the DA effect down so that the reward is unable to drive the effect over the LTP threshold. Thus, the weight fails to learn, despite reward. Generally, the Plan and Request units fail to learn the necessary representations under extreme hypo-DA conditions.

Babble hypo-DA conditions may or may not lead to acquisition failure. In both Figures 3.27 and 3.28, early learning leads to the Plan and Request units learning a representation for C AND Blue. However, too few Plan units learn the representation in Figure 3.27 during what is probably a novelty-driven babble. No further babbles happen in this run, probably due to the raised Babble Go unit threshold caused by the DA depletion. On the other hand, in Figure 3.28 novelty-driven learning leads to enough Plan units being a part of the representation that the model ends up responding to the conjunction appropriately. (See Section 3.1.1.3 for a discussion of how novelty-learning works in the model.)

3.2.5 Hypo-DA Impairment of Acquired Performance

Not only does hypo-DA impair learning of the initial task (C AND Blue -> Nod), but in the most extreme cases, also causes actual unlearning of the previously learned task. Figure 3.29 shows the effects of DA manipulations on 5 'subjects' that, during the maintenance simulation phase, are required to make the same Nod response to C AND Blue. The percent of subjects that have retained the learning is determined, by looking, for each subject, at whether the last 13 trials have greater than 25% correct responses. In the two most extreme Nod/Shake and Plan/Request (and global) cases, loss of learning of

the task is effectively guaranteed, with the exception of ns3, where 1 of the 5 subjects manages to retain the task response.

The rat DA-lesioning study of Robbins and colleagues (1990) provides an example of how DA impairment of the dorsal striatal pathway can impair a learned task. Some of the rats trained on the dual lever-pressing task are given actual dorsal striatal lesions, and the others are given sham lesions. As Figure 3.30 shows, when the sham group is returned to the task, only a few sessions are necessary for the rats to regain their performance. On the other hand, the dorsal striatal DA-lesioned group requires a relearning period comparable in length to the initial training period. In another rat study, one involving the dorsomedial striatum, temporary deactivation of this area by the GABA agonist, musicimol, leads to reduced performance of a learned lever-pressing task, as shown in Figure 3.31 (Yin et al., 2005). Thus, at least a portion of the dorsal striatum seems to be important in the 'actor' performance of learned tasks.

Figures 3.32 and 3.33 show how the loss of the learned task transpires in the model. As shown in Figure 3.32, Nod/Shake hypo-DA leads to unlearning of the Go unit weights shortly after the model is rewarded for a correct behavior. This happens because reward resets the Frustration level which causes the DA Signal activity to fall to its minimum. Under normal DA conditions, this doesn't cause any weight changes, but here the DA level for the Nod/Shake Go units drops below the LTD threshold, and the weights are unlearned. As seen in Figure 3.33, a similar unlearning transpires for Plan/Request units when the Plan/Request pathway is DA-depleted, though the unlearning happens immediately upon the appearance of the stimulus because the Plan units are activated

right away before the Frustration level has had a chance to drive the DA level high enough to avoid unlearning.

3.2.6 Hypo-DA Impairment of Reversal Learning

Hypo-DA impairs learning, not only during the training and maintenance phases, but also during the reversal phase when the model, having been trained successfully on the C AND Blue -> Nod task, is retrained on the C AND Blue -> Shake task. Figure 3.34 shows a similar degradation to the reversal learning as was seen for the initial task acquisition shown in Figure 3.22. The data points represent the percentage of 10 'subjects' that successfully learn the new (C AND Blue -> Shake) task. At extreme hypo-DA conditions, both Nod/Shake and Plan/Request (and global) manipulations lead to high likelihood of failure to learn the reversal task, with pr2, pr3, and ns3 conditions effectively disabling relearning. Babble hypo-DA is progressively, but less, disruptive, with the extreme b3 conditions leading to 6 out of 10 subjects learning the reversal task.

The effect of hypo-DA during the reversal simulation phase is effectively a combination of its effects on the training and maintenance phases. The effects shown in Figures 3.32 and 3.33 would cause the initially learned task to be become unlearned, and the effects shown in Figure 3.25 through 3.27 would also apply, impairing any possible learning of a new task.

3.2.7 Hyper-DA Impairment of Acquisition

Figure 3.22 shows that not only hypo-DA conditions, but also hyper-DA conditions, can impair learning in the model of the initial (C AND Blue -> Nod) task. Neither Nod/Shake (dorsal striatal) or Babble (ventral striatal) hyper-DA conditions lead

to learning failures, but Plan/Request (frontal cortex) hyper-DA leads to difficulty learning the conjunction with only 6 out of 10 subjects succeeding in the PR5 case.

Figure 3.35 shows evidence that rats' learning of lever-pressing tasks may be impaired by D-amphetamine (Idris et al., 2005). The percentage of correct lever-presses is, in fact, decreased for both the initial and reversal tasks.

But how does hyper-DA impair learning? Figures 3.36 and 3.37 suggest that the cause might be perseveration on wrong responses. In Figure 3.36, it can be seen that on the third babble, when the model is trying to learn the C AND Blue -> Nod task, it randomly chooses the wrong behavior (Shake). It continues for most of the rest of the simulation run making that wrong response, despite being repeatedly punished. Figure 3.37 shows why: the Plan DA effect level is shifted up by the hyper-DA effects in the Plan/Request (frontal cortex) pathway. Whereas the punisher dips cause the DA Signal unit level to drop below the LTD threshold, this is not true of the Plan DA effect variable. Therefore, the punishers fail to cause unlearning, although the other condition for plasticity is met, i.e., eligibility trace activity. As Figure 3.38 shows, however, a correct guess during a babble can allow the model to avoid such perseveration.

3.2.8 Hyper-DA Impairment of Reversal Learning

As well as impairing the initial acquisition of the DA manipulation task (C AND Blue -> Nod), perhaps not surprisingly, hyper-DA also impairs reversal learning (i.e., learning of the C AND Blue -> Shake task). Figure 3.34 shows that, as in Figure 3.22, neither Nod/Shake or Babble hyper-DA conditions lead to learning failures, but Plan/Request (frontal cortex) hyper-DA leads to difficulty learning the reversal conjunction. Only 4 out of 10 subjects learn the reversal conjunction in the PR5 case.

As mentioned in Section 3.2.7, Figure 3.35 shows that hyper-DA can affect reversal learning on lever-pressing tasks in rats (Idris et al., 2005). In the model, the reasons for the failure are similar to the reasons for failure during the initial acquisition, but there is an added difficulty caused by the fact that the model begins having acquired the wrong response. Figure 3.39 shows that the model is punished several times for performing the erstwhile correct response, though it recovers during a correct babble near the end of the run. Figure 3.40 shows that, again, the problem is that the Plan DA effect never dips below the LTD threshold, even during punisher dips. Moreover, the model is in a chronic hyper-DA state where the Plan units tend to be in an LTP mode, always ready to learn when a Hebbian event occurs. It is as if the model is stuck in a spurious reward state that is insensitive to the reality of the punishment it receives. What seems to allow the model to recover is that it makes a correct babble that interferes with the incorrect response and the new correct response overwrites the representation driving the wrong behavior.

Chapter 4: Discussion

4.1 Theory Embodied in the Model

The neurocomputational model in this dissertation, and the results collected for its performance on the full task set and under the DA manipulation conditions using the simplified task set, together suggest a preliminary theory of dopamine's functional role in the learning and performance of stimulus-response tasks. This section will elucidate that theory and summarize it and discuss its implications.

As laid out in Section 1.1.4, the research questions this dissertation set out to address were as follows:

1. What is the neural substrate of task-oriented behavior selection (TOBS)?
2. How are TOBS behaviors learned by this substrate?
3. What role does the neurotransmitter dopamine play in the learning and performance of these behaviors?

The next three subsections review theoretical elements for each of these. The italicized statements collectively represent a summary of the theory. Section 4.1.4 then discusses some salient implications of the theory.

4.1.1 Neural Substrate of TOBS

4.1.1.1 'Actor' Pathway

A frontocortical (dorsal striatal / 'actor') pathway is involved in the performance of learned TOBS behaviors.

The green-labeled blocks in Figure 2.6, along with the Action Gating mechanism, constitute the dorsal striatal pathway. This is the pathway that is ultimately responsible for initiating a voluntary behavior, either learned or randomly explored.

The prefrontal cortex (PFC) represents stimulus context (e.g. color and tone working memory).

The Stimulus Presence, Color WM, and Tone WM units maintain these representations. Details are given in Section 2.3.5. Similar types of units can be imagined for other sensory modalities.

PFC (Plan units) and its connection to frontal premotor areas (Request units) mediates a mapping between stimulus context and response.

As explained in Sections 2.3.7 and 2.3.11, the Plan units respond selectively to Stimulus Presence, Color WM, and Tone WM conjunctions. Ultimately, the Plan units self-organize in their responses so that particular stimulus conjunctions lead to distributed representations of 3 to 5 units, such as those shown in Figures 3.2 and 3.10. The Plan layer activity propagates forward to the Nod and Shake Request units.

Frontal premotor units (Request/Init/Exec) are gated by a dorsal striatal (Action Gating) circuit.

Sections 2.3.7 and 2.3.11 explain the operation of the Request units. The Request and Init units are assumed to be a part of the same premotor area cortical columns, and the Action Gating module provides a means of gating or vetoing Request activity, leading to Nod or Shake Init unit activity. Figure 2.9 shows the architecture of the Action Gating module, and Section 2.3.8 describes its complex operation.

Frontal premotor units (PMC/SMA) initiate motor responses.

Section 2.3.4 explains the Nod, Shake, and Track Exec units. The Nod and Shake Exec units are driven by the corresponding Init units (see Figure 2.9). Once the Nod or Shake Exec units are engaged, they do not disengage until the behavior is completed.

4.1.1.2 'Babble' Pathway

A cingulostriatal (ventral striatal / 'babble') pathway may be involved in triggering of random explorative behaviors when an organism is motivated and hasn't received a reward over a long interval.

The red-labeled blocks in Figure 2.6, along with the Babble Gating mechanism, constitute the ventral striatal pathway. This pathway randomly stimulates activation and behaviors in the dorsal striatal 'actor' pathway.

Anterior cingulate cortex (ACC) takes as input reward and visceral state information to monitor hunger/frustration (Frustration unit).

Section 2.3.6 describes the operation of the Frustration unit, Figure 2.7 shows its input connectivity, and Figure 2.8 shows an example of its dynamic. By default, the Frustration unit activity increases, but it is reset by reward deliveries.

ACC (Babble Request units) monitors frustration and novelty in stimuli and triggers random "babble" activation (in Babble units) under high frustration or novelty conditions.

Section 2.3.7 describes the activity of the Babble Request unit. It is driven high by great Novelty or Frustration unit activity. If this activity is gated by the Babble Gating circuit, it randomly triggers Babble unit activity such that one of the 8x8 units is activated.

ACC babble units are gated by a ventral striatal (Babble Gating) circuit.

Section 2.3.9 describes the operation of the ventral striatal Babble Gating module, and Figure 2.11 shows its unit connectivity.

ACC (Babble units) outputs to PFC (Plan) and premotor (Nod/Shake Request) units to trigger explorative behaviors.

As Section 2.3.12 explains, a single Babble unit is randomly selected every time a Babble Exec activation is made. Section 2.3.10 describes the mechanism that generates neural noise which is used, not only in selecting which Babble unit is activated, but also which striatal units are selected in the Action and Babble gating mechanisms. The random selection of the Babble unit, in combination with random feedforward weights to the Plan and Request units, leads to random activation in the Plan and Request layers. Request activation, then, can result in initiation of the babble behavior in the 'actor' pathway.

4.1.1.3 'Critic' Pathway

Reinforcement of explorative behaviors that were triggered by the cingulostriatal pathway allows learning of new TOBS stimulus-response mappings in the frontocortical pathway.

There is a 'critic' pathway extending from the outcome various outcome processing units (see Figure 2.7) through the DA Signal unit to the Plan and Request units and the Go and NoGo units of the Action and Babble Gating modules. Signals from this allow the "stamping in" of rewarded random babble behaviors, as well as the "stamping out" of punished behaviors.

DA-ergic midbrain cells signal reward-learning and punishment-unlearning: substantia nigra pars compacta (SNc) and ventral tegmental area (VTA).

Section 2.3.13 explains the learning algorithm. As reviewed in Section 4.1.3.2, it is the phasic DA signal that is responsible for learning and unlearning.

DA cells set the permissiveness of Action and Babble Gating.

As explained in Section 2.3.8, and as reviewed in Section 4.1.3.1, the tonic level of DA is involved in setting the permissiveness of Go and NoGo unit activity in the Action and Babble Gating modules.

The dorsal striatum (Nod/Shake path) is innervated by SNc.

This pathway, the nigrostriatal pathway, is the most intensively studied DA pathway, probably due to its involvement in Parkinson's disorder.

The ventral striatum (Babble path) and PFC (Plan/Request path) are innervated by VTA.

The mesoaccumbal pathway has been mostly studied in drug addiction and electrical reward self-stimulation research. The mesocortical pathway has been studied in working memory research, as the proper function of PFC working memory depends on DA. The precise role of VTA activity in signaling cortical learning, however, needs more thorough investigation.

4.1.2 Dynamics of TOBS Substrate

The dorsal striatal ('actor') pathway begins in a naïve, unresponsive state.

The Plan and Request afferent weights are initialized to a very small number, so that the 'actor' pathway is initially unresponsive to stimuli.

The ventral striatal ('babble') pathway triggers explorative behaviors in absence of recent reward.

If the model will not respond to stimuli initially, then any behaviors it selects must be explorative in nature. The 'babble' pathway allows random trial of a behavior when the organism is motivated to "try something".

Correct guesses trigger learning along the dorsal striatal pathway, causing learning of stimulus context-to-response mappings.

Assuming a correct babble, and sufficient stimulation of random Plan and Request layer representations, reward leads to "stamping in" of the Plan and Request representations in the context of the stimulus.

Sudden reversal of reward conditions triggers punishment which causes unlearning of old mappings.

If, as during reversal learning conditions, the old response patterns are now punished rather than rewarded, then the old Plan and Request representations are actively "stamped out".

Under lack of reward under reversal conditions, babble behaviors re-emerge which leads to learning of new correct mappings.

When the old responses have been "stamped out", effectively the model is returned to a state where it must relearn the proper stimulus mapping from scratch. This it does, in the usual way, through random babbling, followed by reward for a correct guess.

4.1.3 DA Role in TOBS Learning and Performance

4.1.3.1 Tonic DA's Role

Base-line (tonic) rates of DA activity signal 'activity-oriented motivation'.

This is one of the primary hypotheses of this dissertation. Overall high levels of DA signal occur during times when an organism is highly motivated to act, for whatever reason. Such reasons may include anticipation of rewards or craving or extreme deprivation. Low levels of DA, on the other hand, occur when the organism is less

motivated to act, whether because of essential satiation of needs, or because of withdrawal from its external environment.

Frustration (e.g. from unsatiated hunger) and novelty in the environment encourage action over inaction. Thus, both excite DA release.

Figure 2.8 shows how Frustration and Novelty unit activity both excite the DA Signal level, though the Novelty unit's effect is phasic in nature. It is intuitive that either situations of long-standing non-reward or novelty would encourage an organism to try out new behaviors, explore new ways of interacting with the environment in an attempt to find adaptive (or more-adaptive) responses to the stimuli.

The tonic dopamine level regulates the baseline level of gating allowed by striatal pathways (high-DA decreasing Go unit thresholds and increasing NoGo unit thresholds).

Figure 2.10 shows the influence of DA level on Go and NoGo threshold activity. Essentially, high-DA states potentiate initiation of action, whereas low-DA states suppress initiation of action. Through different pathways, Nod and Shake, and Babble initiations are both regulated by DA level.

The tonic DA effect in dorsal striatal pathway also regulates RT because of slow cortical behavior request activity onset (low-DA increasing RT and high-DA decreasing RT).

Figures 3.15 and 3.21, respectively, show how hypo- and hyper-DA conditions in the 'actor' pathway affect RT. Hypo-DA effects correspond with Parkinsonian akinetic effects on RT, whereas hyper-DA effects correspond with psychostimulant motor effects. Section 3.2.1 describes the hypo-DA RT slow-down, and Section 3.2.3 the hyper-DA RT speedup.

4.1.3.2 Phasic DA's Role

Phasic dopamine signals (superimposed on the tonic signal) regulate learning in the frontal cortex and striatum.

Section 2.3.13 describes the learning algorithm and Figure 2.18 shows the LTP and LTD ranges for the two types of DA target cells.

Phasic bursts, driven by (food) rewards or novel stimuli, trigger LTP in most target areas (frontal cortex and D1-dominated striatal 'Go' cells) and LTD in others (D2-dominated striatal 'NoGo' cells).

Figure 2.18 shows the described ranges. Phasic bursts, in the absence of significant ambient hypo-DA effects, lead to DA levels that rise over the upper learning threshold.

Phasic dips, driven by punishers, trigger LTD in most target areas (frontal cortex and D1-dominated striatal 'Go' cells) and LTP in others (D2-dominated striatal 'NoGo' cells).

Figure 2.18 shows the described ranges. Phasic dips, in the absence of significant ambient hyper-DA effects, lead to DA levels that fall below the lower learning threshold.

4.1.3.3 Synopsis of DA's Role and Behavioral Implications

DA cell activity superimposes 'activity-oriented motivation' (tonic) and learning/unlearning (phasic) signals.

This is a statement of the essential hypothesis of this dissertation. Some implications of the superposition of these two signals will be discussed in Section 4.1.4.1.

Depletion of DA can lead to sluggish behavior initiation, slower RT, failure to learn tasks, and even spurious unlearning of tasks.

These effects were discussed in Sections 3.2.2, 3.2.1, 3.2.4, and 3.2.5, respectively, and shown in Figures 3.17, 3.11, 3.22, 3.29, and 3.34.

Excess DA can lead to faster RT, spurious learning, and failure to unlearn incorrect behaviors.

These effects were discussed in Sections 3.2.3, 3.2.7, and 3.2.8, and shown in Figures 3.11, 3.22, and 3.34.

4.1.4 Further Implications of the Theory

4.1.4.1 Interaction of Tonic and Phasic DA Effects

Because of the DA learning dynamics shown in Figure 2.18 and the simultaneous effect on the thresholds of the Go and NoGo units, as shown in Figure 2.10, there is an interaction between the phasic (learning) and the tonic (vigor) DA effects. As Figure 2.10 shows, phasic bursts lead to temporary threshold dips in the Go units which could allow them to temporarily potentiate new behaviors. On the other hand, the same bursts lead to spikes in the NoGo thresholds which could temporarily disable vetoing power of the NoGo units. Phasic dips have an even more pronounced effect, temporarily raising the Go thresholds and dropping the NoGo thresholds. This could respectively inhibit behavior initiation and potentiate vetoing of behaviors.

Conversely, there is also an effect of the tonic DA level on learning. Hypo-DA in a pathway effectively shifts the DA activity for that pathway with respect to the thresholds shown in Figure 2.18 down. If sufficiently inhibited, phasic DA bursts will fail to trigger learning in the Go, Plan, and Request units and unlearning in the NoGo

units. In extreme hypo-DA conditions, the model can be stuck in a default state where the Go, Plan, and Request units unlearn during Hebbian events and the NoGo units learn. On the other hand, tonic hyper-DA conditions shift the DA activity for a given pathway up with respect to the LTP and LTD thresholds. At extreme enough hyper-DA conditions, the DA level will be stuck by default over the upper threshold, and phasic dips will fail to drive the DA level below the lower threshold. The result impairs punishment unlearning and triggers spurious learning after Hebbian events.

Frank and colleagues (2004) have noted that Parkinson's patients that are unmedicated tend to learn more effectively through punishment, whereas L-dopa-medicated patients tend to learn more effectively through reward. The model in this dissertation could explain this through the tonic/phasic DA interactions. Specifically, hypo-DA (the unmedicated PD condition) could lead to a higher likelihood of phasic dips falling below the LTD threshold of Figure 2.18 (i.e., for Go, Plan, and Request units), whereas hyper-DA (the L-dopa condition) could lead to a higher likelihood of phasic bursts rising over the LTP threshold.

4.1.4.2 Evolution of Activity States and DA's Role in Individual Temperament

Why might it have been adaptive for organisms to develop the kind of tonic/phasic DA control mechanism described in this dissertation? There are distinct circumstances for an organism when more or less behavioral activity is desirable. Generally speaking, too little activity will lead to fewer rewards being gained and, in the most extreme cases, starvation. On the other hand, too much activity will expend more metabolic energy that must be replenished through more food intake, and could lead to more tissue damage. Hyperactivity is likely to lead the organism to engaging in

physically or socially risky behaviors that may jeopardize the organism's life or reproductive success. Having distinct waking and sleeping states addresses some of this issue, as the waking state allows the organism to forage and mate, whereas the sleep state allows the organism to conserve metabolic energy and repair (more rapidly) damaged tissues and replenish depleted neurotransmitter stores.

However, even in the waking state, there are times when more or less activity is beneficial. When the organism is hungry or when cues in the environment suggest that the organism has an opportunity to gain a reward, then it may be worthwhile for the organism to respond more frequently and easily. In fact, it might be said that when the organism has the sense that the opportunity cost of inaction is high, then the organism would be better off in a state of higher base-level motivation (Niv et al., 2006). On the other hand, if the organism is satiated and would gain little by further foraging or mating, then it would probably be preferable for the organism to fall into a more restful, less active state which would conserve energy and keep the organism out of trouble. It would therefore be adaptive for an organism to evolve a global signal for 'activity-oriented motivation'. The centrally located DA nuclei, with their global neuromodulatory influences on frontal neocortical and striatal circuitry involved in behavior, are well-situated to provide such a role.

The DA nuclei are likewise well placed to perform a role as a global learning/unlearning signal. Why might these signals piggyback on each other? It may have been an evolutionary accident. However, it would have been a fortuitous accident because, when the organism is in a high-motivation state, it probably makes sense that the organism should ignore relatively minor punishers and be more sensitive to even

potential rewards. On the other hand, if the organism is in a low-motivation state, it would be less adaptive to take risks and, therefore, it makes sense to ignore minor rewards and be more sensitive to punishers.

With the above in mind, we may imagine that organisms tend to operate in a typical range of motivation states during their waking hours, peaking when they are hungry or in a friendly, potentially rewarding environment; and dipping when they are in a hostile or unrewarding environment or are satiated or fatigued. While an individual's motivation state can be expected to fluctuate within this range, it seems plausible to suppose that some organisms might tend to run "hot" or "cold" in their motivation. The particular bias they have in their range of motivation may be due to simple, genetically-determined differences in the DA receptor proteins, such that some individuals possess more sensitive, higher-affinity DA receptors (the "hotter" temperaments), and others have more sluggish, lower-affinity DA receptors (the "colder" temperaments). Depue and Collins (1999) discuss extraversion (what they term "agency") as being determined by base DA levels. The model in this dissertation could model "extraverted" tendencies through a mild global hyper-DA effect; on the other hand, certain aspects of "introversion" could be modeled through a mild hypo-DA effect in the model. Hyperactivity, sensation- and novelty-seeking, and tendency to more risk-taking and ignoring of punishment consequences would be expected of hyper-DA individuals. On the other hand hypo-DA individuals would tend to be less active (more "phlegmatic"), more sensitive to punishment, and more risk-averse, perhaps. Genetic DA receptor factors may set a baseline temperament of individuals, but it is also possible that the overall level of reward vs. risk in the environment could foster the development of an

individual bias in DA activity, though the potential developmental mechanisms need to be investigated.

Assuming that serotonin generally has an opposing effect to DA (Daw et al., 2002), this may explain the anti-anxiety and soporific effects of serotonin. 5-HT may tend to drive the organism into a lower motivation state which would be less impulsive and more relaxed, though not necessarily one of increased positive affect. (In fact, hypo-DA or hyper-serotonin would probably both lead to states of anhedonia and/or apathy, although 5-HT may reduce stress response through its effect on the hypothalamic-pituitary-adrenal axis (Panksepp, 1998) .)

It is a tempting hypothesis that strong emotions that motivate immediate action are likely to temporarily boost the tonic DA signal. Anticipation, anger, fear, sexual desire, or acute pain or distress should lead to a willingness of the organism to expend more effort to take actions that address the strong emotions. On the other hand, chronic depressive states (e.g. grief or physical exhaustion) may inhibit the DA signal temporarily so that the organism tends to behave in a more reserved and risk-averse fashion.

The above hypotheses potentially suggest means of pharmacological intervention under various affective disorders, or at least provide a preliminary explanation of how the currently prescribed drugs may exert their motivational and affective influences.

4.1.4.3 Pathway Dependence of DA Effects

An important observation this dissertation emphasizes is that the effects of pharmacological (or otherwise) manipulation of DA level will depend on which DA pathways are affected. Teasing apart the distinct roles of the various DA (and other neuromodulator) pathways is a major task that is fruitful to undertake at this stage.

Delivery of globally-acting agonists or antagonists is a blunt instrument of clinical intervention. In the worst case, delivery of a nonselective (D1/D2) antagonist can lead to such global impairment of motor behavior that the therapeutic benefits are outweighed by the side-effects. Development of D2-specific neuroleptics, for example, allowed more selective influence (presumably of the NoGo pathways). It would be beneficial if a technology could be developed that could, for example in Parkinson's patients, stimulate DA receptor activity selectively in the dorsal striatal pathway, at least at the early stages of the disease when only the dorsal striatal pathway is impaired. For patients suffering high-anxiety depression states, it might be useful to selectively DA-antagonize the BA 25 portion of the anterior cingulate cortex. (It might also be useful for depression if serotonergic drugs could be developed that selectively affect 5-HT's influence on the hypothalamic-pituitary-adrenal axis, so the stress response is selectively inhibited.)

How might pathways be selectively agonized or antagonized? One possibility might be that new drugs could be developed that are selective for specific pathways. It is not apparent to the author how this might be done, though it would be theoretically possible if the DA receptor G-proteins had subtle distinctions in the different pathways that were analogous to the differences currently recognized between DA receptor types. If every pathway effectively had its own receptor type, then a drug might, in theory, be developed selective for each pathway.

Given current (or near-future) technology, the only alternative that occurs to the author involves selective surgical implantation of a device that releases or stimulates the release of DA in a particular pathway in order to allow selective agonism. For antagonistic effects, the implant would need to inhibit, in a pathway-specific fashion,

release of the neurotransmitter. Such an implant may be a better alternative than selective lesioning, as its effects could perhaps be reversed or progressively tweaked in the manner that drug dosages may be progressively modified. In the distant future, perhaps localized genetic manipulation of neuronal or glial cells in the pathological areas, or outright grafting of healthy cells, may be an option.

4.1.4.4 Hyper-DA, Perseveration, and Spurious Learning

An interesting implication of the model with respect to psychostimulant effects is the idea, observed in Sections 3.2.7 and 3.2.8, that hyper-DA conditions may lead to a chronic spurious reward state in the organism. This state leads to both spurious learning during an incorrect babble, and a perseveration of wrong responses in the face of punishment. It is plausible that this kind of effect could be seen during psychostimulant intoxication. However, it is not clear what DA pathway is responsible for this effect. The model suggests that the VTA-to-frontal cortical pathway is responsible for the perseveration; however, cocaine, for example, does not have much effect on the mesocortical pathway, due to the lack of DAT reuptake in frontal neocortex. However, some indirect effect due to VTA DA cell excitation, perhaps caused by mesoaccumbal pathway activity, cannot be entirely ruled out.

4.1.4.5 Novelty-Induced Learning and Behavior

Extant work on phasic DA learning effects mentions, but does not place much emphasis on, the potential role of novelty in inducing learning. Novelty is likely, also, to induce spontaneous behaviors. Both effects are seen in this model.

But why should novelty stimulate spontaneous behavior and learning? What is the adaptive value? Generally, when a novel stimulus is encountered by an organism, it

represents a perceived change in the organism's environment. Whether the change is for better or for worse is not immediately apparent, but it is in the organism's adaptive interest to be more alert and to expend more effort to learn more about the stimulus to determine if it is a potential reward or a threat. Engaging non-specific explorative behavior patterns in the presence of salient stimuli encourages the organism to learn more about them (Ikemoto & Panksepp, 1999).

In addition, it might be useful for the stimulus to trigger learning, at least during the first occurrences when it is still considered novel. This would stimulate the formation of a default learned response. This would mean having an optimistic bias initially towards a novel stimulus, perhaps, provided there are no obvious threat cues accompanying it. If the encouraged behavior was the wrong one, or the new stimulus emitted threat cues later, then this default learned behavior could be unlearned, and avoidance behavior could perhaps be learned later, in its place.

Without an optimistic bias towards novel stimuli, organisms may be less likely to capitalize on new advantages entering the environment. There may be another pathway that exists, however, that can learn to map novelty to anxiety and fearful responses. (The amygdala seems a likely candidate area for performing this mapping, and this in turn could trigger activity in the hypothalamic-pituitary-adrenal axis.) Chronically hostile environments might cause the stress responses to novel stimuli to overcome the explorative tendencies encouraged by the DA pathways.

4.2 Model Predictions

Primarily, the model was constructed to explain known effects of hypo- and hyper-DA manipulations: in particular, effects on RT of both hypo- and hyper-DA

conditions; and the impairment of learning caused by hypo-DA conditions in the dorsal striatum, such as occurs in Parkinsonism. Table 4.1 summarizes the results covered in Section 3.2. Each entry in the table represents a potential test that might be performed in animal subjects and the predicted result of that test. Some of the effects shown (for example, the dorsal striatal hypo-DA slowdown of RT) have been clearly demonstrated in the literature, but to the author's knowledge, no systematic study has been performed comparatively manipulating the specific anatomical pathways with DA antagonists and agonists, though Robbins and colleagues (1990) compare two of these pathways for hypo-DA conditions (dorsal vs. ventral striatal). Doubtless, there are other DA pathways not studied or modeled here that it would also be useful to manipulate, such as the VTA-to-amygdala pathway. Likely, each of these pathways has its own set of DA executive control and learning effects.

Some of the salient behavioral effects predicted by the model that bear further investigation include the effects of hypo- and hyper-DA manipulations of the VTA-to-frontal cortex pathway on learning, and the triggering by novel stimuli of both explorative action and learning. The model would suggest that mesocortical DA depletion should disrupt learning of S-R mappings that require neocortex. Some brainstem-basal ganglia pathways might still allow S-R learning, but more flexible tasks requiring frontal neocortical implementation may be much more difficult to learn. Moreover, the model would predict that DA depletion in the PFC and premotor areas may actually cause unlearning/forgetting of existing task mappings. The model would also predict that too much DA activity along the mesocortical pathway would tend to lead to spurious learning and behavioral perseveration, as discussed earlier. Generally, the

effects of DA on frontal cortical learning need to be more closely investigated. As already mentioned (in Section 4.1.4.5), the model predicts that both behavior initiation and learning may be triggered by novel stimuli. The mesocortical pathway, again, would be important in the learning, and the ventral striatal (cingulostriatal) pathway is likely to be important in novelty cueing of explorative behaviors.

Probably the most interesting and useful predictions made by the model relate to specific neural mechanisms that are behind the observed performance and learning effects. One of the key mechanisms proposed in this dissertation is the cingulostriatal 'babble' pathway for producing explorative behavior. How spontaneous, quasi-random behaviors are generated and what circuits are involved is an area that deserves more investigation in animal studies. The model also suggests a mechanism (depicted by Figures 3.15 and 3.21) by which the level of dorsal striatal DA regulates the RT for learned responses. In short, the idea is that the striatal "request" afferent activity takes a substantial period of time to build up. The striatal Go units' thresholds will be set by tonic DA activity and this will cause the latency between the start of "request" activity buildup and basal ganglia gating of a behavior to vary with DA level.

The model also makes some predictions about the cellular mechanisms of learning that deserve closer investigation. The existence of synaptic memory for "Hebbian events", i.e., the specific kind of eligibility trace proposed in this model, should be tested for in animal studies, both in vivo and in vitro. A further interesting feature of the learning algorithm in this model is the heterosynaptic competition that is used in the LTP of the learning synapses. One problem with Hebbian learning without competition is that weights may grow over a period of time so that the unit is hyperactive in its

response and responds in a logical OR fashion to a wide variety of stimuli so that its response is, in a sense, too 'promiscuous' and not selective enough to be useful. Heterosynaptic competition, however, allows neural units to specialize more in their response to particular stimuli or stimulus conjunctions. In this dissertation, the bounded arbor-weight total allows learning of specific stimulus conjunctions with "don't care" conditions. (See Section 2.3.13 for further discussion of this.)

4.3 Research Contributions

This dissertation has attempted to formulate a preliminary theory of the learning and performance of task-oriented behavior selection, and the role played by dopaminergic neuromodulation. The model developed here embodies an explanatory theory (discussed in Section 4.1), and makes some predictions (discussed in Section 4.2) regarding likely behavioral consequences of DA manipulations and also of possible neural mechanisms involved in task learning and performance. Probably the key novel prediction is that there exists at least one corticostriatal pathway—probably cingulo-ventral striatal—that is involved in generation of exploratory behaviors, a pathway that allows the trial component of trial-and-error instrumental learning. The model developed is capable of reversal as well as initial learning of a simple S-R task set. Another key contribution of this research and the model is the emphasis that it places on the importance of separate anatomical DA neuromodulatory pathways. It is an initial attempt to develop a neurocomputational model that recognizes an array of distinct pathway-dependent DA effects and fits them into a larger hypothesis of the function of centralized dopamine release. It is a necessarily incomplete picture at this stage, but the author hopes a possible beginning to develop an integrative behavioral functional theory of one of the most

clinically important of neurotransmitters. Additionally, the model may suggest some components of a general learning architecture for artificial intelligence, an architecture that is grounded in an animal learning and comparative neuroscience model.

4.4 Limitations of the Model

Given the scope of the model proposed in this dissertation, it is inevitable that there is a good deal that is incomplete or that may corroborate uneasily with recent data. This section will discuss limitations of the modeling of the actor, babble, and critic pathways, and finish with a discussion of general issues that apply to all three pathways. Some suggestion will be made of how future research might address some of these issues.

4.4.1 'Actor' Pathway Issues

Due to time limitations, less detail was modeled of the working memory apparatus than was originally planned. One consequence of this is that the effects of hypo- and hyper-DA on working memory maintenance and updating are not modeled, though such effects would be significant. Four ranges of mesocortical DA level were planned for modulating the effects of working memory:

- No maintenance range: the lowest range of PFC DA level. Working memory fails to retain its trace, disappearing after the afferent input vanishes. Much evidence suggests that some level of tonic DA is necessary to stabilize recurrent excitation that allows working memory maintenance (Brunel & Wang, 2001; Durstewitz et al., 1999; Durstewitz et al., 2000; Gao et al., 2001).
- Buffer maintenance range. Above the no maintenance range, working memory remembers only the last-perceived input with new stimuli overwriting the old

traces, but without new inputs the old trace is maintained. (Tanaka, 2002) models such a dynamic at an intermediate level of D1 activation.

- Exclusive maintenance range. Above the buffer maintenance range, working memory traces are resistant to disruption by new or otherwise distracting stimuli. This is the type of maintenance generally modeled in other research, e.g. (Durstewitz et al., 1999).
- Overload disruption range. Above the exclusive maintenance range, the high level of DA disrupts the existing working memory traces. Rat studies have confirmed that too much D1 receptor activation disrupts working memory (Zahrt et al., 1997). This is likely due to a potentiation of the GABAergic interneurons in the cortical columns (Brunel & Wang, 2001; Muly III, Szigeti, & Goldman-Rakic, 1998).

The working memory disruptions caused by sufficient hypo- and hyper-DA in the mesocortical pathway would cause additional failures of maintenance performance, though not necessarily unlearning of the task, once it has been acquired. Naturally, acquisition of the task or reversal learning would be both difficult if working memory were not functioning correctly.

Not only DA modulation, but the actual circuitry of working memory was not modeled for lack of time. Because of the need to maintain multiple working memory traces in certain tasks, there are likely to be multiple "stripes" of PFC working memory whose maintenance and/or update may be independently gated through some type of dynamic gating mechanism involving the associative striatal pathways through the basal ganglia (Frank et al., 2001; O'Reilly & Frank, 2006). The dynamic gating mechanism

presumably would be driven by the same reward/punisher learning as was developed in this model. PFC modality representations would not be determined a priori, as they are in the dissertation model, but would be learned in a self-organizing fashion.

More recent data suggests that the dorsal striatal pathway actually has at least two distinct components: a dorsolateral path involved in habit-learning (Faure et al., 2005; Yin et al., 2004) and a dorsomedial pathway involved in goal-directed learning (Yin et al., 2005). When an instrumental response is initially being learned, the goal-directed learning pathway seems to be involved; the evidence is that performance of the required behavior is contingent on the outcome reward maintaining its hedonic value (Yin et al., 2005). When a food outcome is devalued, either through satiation or through pairing of the food with a nausea-inducing substance, the animal ceases to perform responses that are still under goal-directed control. Fixed-ratio reward schedules tend to yield continued goal-directed performance of the response. Variable-ratio reward schedules, however, or overtraining, often lead to habit formation which is resistant to outcome devaluation. The acquisition and performance of habitual S-R responses, then, is mediated by the dorsolateral habit-learning pathway. By contrast, the model in this dissertation treats all S-R learning like habit-learning. A more complex and realistic model needs to be developed exploring the dorsomedial goal-directed learning pathway and modeling how the goal-directed and habit-learning pathways interact, and also how these two pathways relate to the ventral striatal pathway/s involved in explorative behavior generation.

One notable lacuna in the model is the matter of task switching costs. Typically, the RT for responses increases after a new task is signaled (Monsell, 2003; Rogers & Monsell, 1995; Wylie & Allport, 2000). In the dissertation model, there is no latency for

task-switching, nor any interference effects between tasks. In the full task-set, the switch between the BLUE-SELECT and RED-SELECT tasks entails no RT penalty or switch latency. In order to bring such latencies into a model, one solution might be to require the dynamic gating mechanisms of working memory to be subject to a gating latency, in the same manner that Nod or Shake gating is in this model. In that case, hypo-DA in the working memory dynamic gating striatal pathways could lead to slower update of which working memory "stripes" are maintained, and hyper-DA could speed the switching. The likely importance of DA in switching time is mentioned in (Cools, 2006).

One final potential discrepancy to mention in the 'actor' pathway is the possibility that PFC plasticity may be LTD by default during Hebbian events and normal DA levels. Some of the literature suggests that LTD is the default response when cells are stimulated at high frequencies (Law-Tho et al., 1995). The consequence of this would mean that Plan and Request units would tend to unlearn their representations unless they were rewarded. It may be, however, that background neocortical extracellular DA levels under normal conditions, relieve this default LTD condition (Matsuda et al., 2006). More investigation is needed.

4.4.2 'Babble' Pathway Issues

Although DA affects the Babble Go and NoGo units of the model, there is an additional DA-ergic pathway likely to be involved in the 'babble' pathway: the direct VTA connection to ACC. In a rat experiment, both D1 and NMDA receptors need to be active in medial prefrontal cortex (which includes ACC) in order for appetitive instrumental learning to develop (Baldwin et al., 2002). It may be that the disruptions to acquisition of task sets due to hypo-DA conditions may be more severe than what is seen

in Figure 3.22. It may also be the case that the neural correspondent of the Frustration unit may require learning that is DA-dependent in order to correctly drive babbling.

Another improvement that should be made to the 'babble' pathway is to develop a way of utilizing memory in order to remember the last-tried responses. The primate study of Procyk and colleagues (2000) suggests that the subjects remember previously tried responses and do not tend to erroneously retry them. The hippocampus' connection with the ventral striatum and ACC makes this a plausible source of the memory, though PFC working memory could be another medium for traces of the tried responses.

4.4.3 'Critic' Pathway Issues

Another feature of the model that was planned, but not fully developed, was a more detailed 'critic' pathway. Brown and colleagues (1999) developed a model of an adaptive critic which implements the dynamics of phasic bursts and dips. (The conditions for DA bursts and dips were reviewed in Section 1.2.3.4.) The critic used in the present model is non-adaptive and differs in its dynamics from the phasic activity dynamics seen in animal studies. In the present model, there are unconditionally DA phasic bursts only during rewards and phasic dips only during punishers. In a more realistic critic system, fully predicted rewards would not trigger phasic bursts, but omitted rewards would lead to phasic dips. Moreover, stimulus cues that predict reward would lead also to phasic bursts. Future work on this model should involve integrating a more realistic critic with the rest of the model, one capable of the learning of reward cues (S-O learning). This would be a separate ventral striatal pathway from the 'babble' pathway: another pathway subject to its own set of hypo- and hyper-DA effects. Hypo-DA in this pathway would likely disable reward cue learning and possibly even lead to unlearning of

the cues. Hyper-DA in this pathway would probably lead to spurious cue associations with reward.

4.4.4 General Issues

As was discussed in more detail in Section 1.2.3.5, modeling of the midbrain DA cells as a single compartment is probably an oversimplification. The SNc and VTA compartments are likely to be functionally distinct, and there are also likely to be multiple VTA compartments. How activity in these compartments (and their efferent pathways) is related would be a useful direction to explore. Based on the literature review conducted by this author, it seems a likely hypothesis that there exists a hierarchy in the mesostriatal axis, with the nigrostriatal (SNc-to-dorsal striatum) pathways controlling the narrow responses of the organism to particular contexts, and the increasingly ventromedial portions of striatum innervated by VTA exerting a training, activating, and regulating influence over the more dorsolateral striatal pathways. The habit-learning 'actor' pathways are probably the most dorsolateral, followed by the dorsomedial goal-directed learning pathways. Ventromedial to these would be the proposed 'babble' pathway/s. Ventromedial to these would be ventral striatal (accumbal) pathways involved in selection of goals. Still further ventromedial would be pathways involved in Pavlovian (S-O) learning of the relation of stimulus cues to impending reward, as well as pathways involved in arousal of the more dorsolateral pathways in response to incentive cues. It seems to this author that a proper global understanding of mammalian behavior control requires continued and increasingly detailed investigation into the basal ganglia apparatus and its relationship to the rest of the brain. The royal road to the (volitional) soul does indeed run through the striatum.

The brainstem, diencephalon (thalamus and hypothalamus), and neocortex all funnel information into the striatum, and the output nuclei of the basal ganglia send controlling connections back to the brainstem and (through the thalamus) to the neocortex. This control system learns through reinforcement signals that are themselves under basal ganglia control. Brainstem, diencephalon, and basal ganglia together constitute a flexible learning machine, which, however, lacks the memory mechanisms of the hippocampus, amygdala, and PFC, the complex sensory analytical processing of the posterior neocortex, and the high-level motor control of the frontal neocortex, including PFC. Cerebral cortex provides layers of refinement and flexibility (and possibly sentient awareness) to a core behavior system which is already flexible and adaptive.

One avenue that was considered, but not fully developed, in this dissertation was an exploration of the effects of D1- and D2-specific agonists and antagonists. The roles of D1 and D2 receptors even in the striatum are not fully understood, though, as Section 1.2.3 shows, much has been already elucidated. One complication that prevented the author from modeling D1 and D2 manipulation and collecting data for it is the biphasic nature of D2 receptor influence as DA levels increase (Frank & O'Reilly, 2006). At low DA extracellular levels in the striatum (and consequently at low dosages of D2 agonist delivery), D2 receptor activity primarily consists of the D2 autoreceptors on the DA cell terminals. The autoreceptors inhibit DA release, so the behavioral and learning effects are primarily inhibitory. However, at higher D2 agonist dosages, the postsynaptic D2 cells in the NoGo striatal units become activated, and this effect is disinhibitory. Naturally, this complicates the behavioral effects of striatal D2 agonism and antagonism requiring distinctions be made between low and high dosages.

Another general limitation in the model is that extracellular DA levels are not adequately modeled at the respective DA targets. The 6 DA effect variables are a beginning at modeling this, but what is missing is extrinsic and intrinsic modulation of the extracellular DA levels at the DA cell targets (Dreher & Burnod, 2002; Katz, 1998). It is not only the rate of DA cell firing that determines DA release, but glutamatergic activity around the DA cell terminals. As an example of extrinsic modulation, PFC activity is believed to lead to larger striatal extracellular DA concentrations (Grace, 1991). As a (proposed) example of intrinsic modulation, PFC activity may locally stimulate VTA DA release into PFC (Chadderdon & Sporns, 2006; Dreher & Burnod, 2002); this may provide a local control over working memory "stripe" activation. Thus, in addition to more global DA signaling, local control at DA terminals needs to be accounted for.

Another factor that will ultimately need to be accounted for is the array of influences of other neurotransmitters on activity and plasticity at the DA targets. Serotonin, norepinephrine, and acetylcholine, and various neuropeptides all probably add additional influences to the discussed DA influences. For example, in neocortex, acetylcholine is likely to be important in plasticity, as it has shown to be in bat auditory cortex (Ji & Suga, 2003) and human motor cortex (Kuo, Grosch, Fregni, Paulus, & Nitsche, 2007). How acetylcholine and DA interact in control of cortical plasticity is something that ought to be investigated in the future.

4.5 Future Research

This dissertation has developed a complex neurocomputational model which has attempted to begin to explain the division of labor of cortical, basal ganglia, and midbrain

areas involved in task learning and execution. Due to the limitations mentioned in the previous section, as well as yet unknown factors, it is essentially a work in progress. The most important immediate next steps that should be taken with the model involve incorporating known pathways and neural mechanisms that are currently not modeled.

Most important is incorporating the adaptive critic mechanisms that were modeled in (Brown et al., 1999). This would allow the model to predict rewards and respond according to observed reward- and punisher-related phasic DA dynamics (Schultz, 1998; Ungless et al., 2004). Whereas punishers are currently needed to trigger reversal learning in this model, a properly functioning adaptive critic would cause reward omissions to lead to phasic dips and, therefore, extinction of behaviors. Fully predicted rewards would also cease to trigger phasic bursts. Cues predicting rewards would also trigger phasic bursts which, in turn, could lead to learning of behavior sequences through backwards chaining. Hypo- and hyper-DA effects in the ventral striatal pathway involved in the S-O learning for reward prediction would need to be analyzed.

The next area of improvement in the model would involve modeling DA's effect on working memory as mentioned in Section 4.4.1. First, it might be useful to keep the simple mechanisms of WM currently used, but apply the 4-range modulation mechanism explained in 4.4.1. At a later stage, however, the mechanisms for the dynamic gating of the working memory traces need to be developed. It seems that there could be at least two (associative) striatal pathways (or sets of pathways) involved in working memory: a pathway which provides the base-level of DA to the recurrent excitatory circuits that maintain the working memory traces, and another pathway that is involved in switching on and off the maintenance of the working memory "stripes", perhaps by boosting DA

release at VTA terminals locally in PFC. Much effort will probably need to be devoted to development of a dynamic gating mechanism that works with the rest of the model, though the work would probably build on the mechanisms developed in (O'Reilly & Frank, 2006). Hypo- and hyper-DA's effect on these pathways will need to be investigated.

Modeling of D1 and D2 agonism and antagonism is probably the next most important modification to make. If done correctly, it would allow the model to be useful in making predictions about DA receptor-specific drugs and their effects on TOBS. Effectively modeling the biphasic effects of D2 receptors (mentioned in Section 4.4.4) would be an important step in investigating D1 and D2 agonism/antagonism.

More investigation also needs to be made into the details of the 'babble' pathway. DA modulation of ACC needs to be modeled. More work also needs to be done on the learning dynamics of explorative behavior generation. It is likely that this system needs to learn to respond to particular incentive cues, and may need to later unlearn this (babbling) response once adaptive specific S-R mappings have been learned. The ability of animals to remember last-tried responses suggests that some mechanism needs to also be developed that lets the 'babble' pathway take account of previous tries so that it can avoid repeating them.

Finally, it would be useful to begin to model the distinct dorsal striatal pathways: i.e., the dorsolateral habit-learning and dorsomedial goal-directed learning pathways. In addition to evidence already cited about these pathways, there is also some evidence that DA level may shift the balance between habitual and non-habitual behavior by affecting the relative activity in the striosomal vs. the matrisomal striatal cells (Nelson & Killcross,

2006). High DA may lead to an emphasis of striosomal cell activity which is believed to be more involved in habitual actions. No doubt, more information will continue to arrive on basal ganglia function that will allow more detailed and representative models of TOBS-related striatal pathways to be constructed.

In addition to making continual refinements of the model developed here and the associated comparative neuroscientific theory, it may be useful to apply this theory in other ways. Especially promising may be the possibility of embodying the model in a learning robot. It is hoped that the model developed here—and successive versions that might be developed from it—could constitute the nucleus of a machine learning architecture that would allow the development of robots and virtual automata that are capable of general task-learning.

References

Abbott, L. F., & Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nature Neuroscience, 3*, 1178-1183.

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

Alexander, M. P. (2001). Chronic akinetic mutism after mesencephalic-diencephalic infarction: remediated with dopaminergic medications. *Neurorehabilitation and Neural Repair, 15*(2), 151-156.

Allman, J. M., Hakeem, A., Erwin, J. M., Nimchinsky, E., & Hof, P. (2001). The anterior cingulate cortex: the evolution of an interface between emotion and cognition. *Annals of the New York Academy of Sciences, 935*, 107-117.

Almássy, N., Edelman, G. M., & Sporns, O. (1998). Behavioral constraints in the development of neuronal properties: a cortical model embedded in a real-world device. *Cerebral Cortex, 8* (4), 346-361.

Amiez, C., Joseph, J.-P., & Procyk, E. (2005). Anterior cingulate error-related activity is modulated by predicted reward. *European Journal of Neuroscience, 21*, 3447-3452.

Arbib, M. A., Erdi, P., & Szentagothai, J. (1998). *Neural Organization*. Cambridge, MA: MIT Press.

Arnott, S. R., Grady, C. L., Hevenor, S. J., Graham, S., & Alain, C. (2005). The functional organization of auditory working memory as revealed by fMRI. *Journal of Cognitive Neuroscience, 17*(5), 819-831.

Ashby, F. G., & Casale, M. B. (2003). A model of dopamine modulated cortical activation. *Neural Networks, 16*, 973-984.

Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Research Reviews, 22*, 229-244.

Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829-839.

Baldwin, A. E., Sadeghian, K., & Kelley, A. E. (2002). Appetitive instrumental learning requires coincident activation of NMDA and dopamine D1 receptors within the medial prefrontal cortex. *Journal of Neuroscience, 22*(3), 1063-1071.

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics, 13*, 834-846.

Baxter, M. G., & Murray, E. A. (2002). The amygdala and reward. *Nature Reviews Neuroscience, 3*, 563-573.

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

Bergson, C., Mrzljak, L., Smiley, J. F., Pappy, M., Levenson, R., & Goldman-Rakic, P. S. (1995). Regional, cellular, and subcellular variations in the distribution of D1 and D5 dopamine receptors in primate brain. *Journal of Neuroscience, 15*, 7821-7836.

Bernardini, G. L., Gu, X., Viscardi, E., & German, D. C. (1991). Amphetamine-induced and spontaneous release of dopamine from A9 and A10 cell dendrites: an in vitro

electrophysiological study in the mouse. *Journal of Neural Transmission, 84*(3), 183-193.

Bernstein, P. S., Scheffers, M. K., & Coles, M. G. H. (1995). "Where did I go wrong?" A psychophysiological analysis of error detection. *Journal of Experimental Psychology, 21*(5), 1312-1322.

Berridge, K. C., Aldridge, J. W., Houchard, K. R., & Zhuang, X. (2005). Sequential super-stereotypy of an instinctive fixed action pattern in hyper-dopaminergic mutant mice: a model of obsessive compulsive disorder and Tourette's. *BMC Biology, 3*, 4.

Biran, I., Giovannetti, T., Buxbaum, L., & Chatterjee, A. (2006). The alien-hand syndrome: what makes the alien hand alien? *Cognitive Neuropsychology, 23*(4), 563-582.

Blaha, C. D., & Winn, P. (1993). Modulation of dopamine efflux in the striatum following cholinergic stimulation of the substantial nigra in intact and pedunculopontine tegmental nucleus-lesioned rats. *Journal of Neuroscience, 13*(3), 1035-1044.

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*, 179-181.

Braver, T. S., & Barch, D. M. (2002). A theory of cognitive control, aging cognition, and neuromodulation. *Neuroscience and Behavioral Reviews, 26*, 809-817.

Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: effects of frequency, inhibition and errors. *Cerebral Cortex, 11*, 825-836.

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science, 307*, 1118-1121.

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

Brown, J. W., Bullock, D., & Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks, 17*, 471-510.

Brunel, N., & Wang, X.-J. (2001). Effects of neuromodulation in a cortical network model of working memory dominated by recurrent inhibition. *Journal of Computational Neuroscience, 11*, 63-85.

Bunney, B. S., Chiodo, L. A., & Grace, A. A. (1991). Midbrain dopamine system electrophysiological functioning: a review and new hypothesis. *Synapse, 9*, 79-94.

Bush, G., Frazier, J. A., Rauch, S. L., Seidman, L. J., Whalen, P. J., Jenike, M. A., et al. (1999). Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the counting Stroop. *Biological Psychiatry, 45*, 1542-1552.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences, 4*(6), 215-222.

Bush, G., Vogt, B. A., Holmes, J., Dale, A. M., Greve, D., Jenike, M. A., et al. (2002). Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proceedings from the National Academy of Sciences, 99*(1), 523-528.

Caillé, I., Dumartin, B., & Bloch, B. (1996). Ultrastructural localization of D1 dopamine receptor immunoreactivity in rat striatonigral neurons and its relation with dopaminergic innervation. *Brain Research, 730*, 17-31.

Canteras, N. S., Shammah-Lagnado, S. J., Silva, B. A., & Ricardo, J. A. (1990). Afferent connections of the subthalamic nucleus: a combined retrograde and anterograde horseradish peroxidase study in the rat. *Brain Research, 513*, 43-59.

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and Biobehavioral Reviews, 26*, 321-352.

Carr, G. D., & White, N. M. (1987). Effects of systemic and intracranial amphetamine injections on behavior in the open field: a detailed analysis. *Pharmacology Biochemistry and Behavior, 27*, 113-122.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280*, 747-749.

Chadderdon, G. L. (2008). Assessing machine volition: an ordinal scale for rating artificial and natural systems. *Adaptive Behavior, 16*(4), 246-263.

Chadderdon, G. L., & Sporns, O. (2006). A large-scale neurocomputational model of task-oriented behavior selection and working memory in prefrontal cortex. *Journal of Cognitive Neuroscience, 18*(2), 242-257.

Chivileva, O. G., & Gorbachevskaya, A. I. (2008). Projections of striatopallidal structures to the pedunculopontine nucleus of the tegmentum of the midbrain in dogs. *Neuroscience and Behavioral Physiology, 38*(1), 37-41.

Churchland, P. S., & Sejnowski, T. J. (1992). *The Computational Brain*. Cambridge, MA: MIT Press.

Cohen, R. A., Kaplan, R. F., Zuffante, P., Moser, D. J., Jenkins, M. A., Salloway, S., et al. (1999). Alteration of intention and self-initiated action associated with bilateral anterior cingulotomy. *Journal of Neuropsychiatry and Clinical Neurosciences, 11*(4), 444-453.

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

Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science, 279*, 1347-1351.

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience, 3*, 655-666.

Cunnington, R., Bradshaw, J. L., & Iansek, R. (1996). The role of the supplementary motor area in the control of voluntary movement. *Human Movement Science, 15*(5), 627-647.

Daw, N. D., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks, 15*, 603-616.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*, 876-879.

Day, M., Wang, Z., Ding, J., An, X., Ingham, C. A., Shering, A. F., et al. (2006). Selective elimination of glutamatergic synapses on striatopallidal neurons in Parkinson disease models. *Nature Neuroscience, 9*(2), 251-259.

Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience*. Cambridge, MA: MIT Press.

Deiber, M.-P., Passingham, R. E., Colebatch, J. G., Friston, K. J., Nixon, P. D., & Frackowiak, R. S. J. (2004). Cortical areas and the selection of movement: a study with positron emission tomography. *Experimental Brain Research, 84*(2), 393-402.

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences, 22*(3), 491-517; discussion 518-469.

Dickinson, A., Smith, J., & Mirenowicz, J. (2000). Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. *Behavioral Neuroscience, 114*(3), 468-483.

Dreher, J.-C., & Burnod, Y. (2002). An integrative theory of the phasic and tonic modes of dopamine modulation in the prefrontal cortex. *Neural Networks, 15*, 583-602.

Durstewitz, D., Kelc, M., & Güntürkün, O. (1999). A neurocomputational theory of the dopaminergic modulation of working memory functions. *Journal of Neuroscience, 19*, 2807-2822.

Durstewitz, D., & Seamans, J. K. (2002). The computational role of dopamine D1 receptors in working memory. *Neural Networks, 15*, 561-572.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *Journal of Neurophysiology, 83*, 1733-1750.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience, 8*(12), 1784-1790.

Faure, A., Haberland, U., Condé, F., & El Massioui, N. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *Journal of Neuroscience, 25*(11), 2771-2780.

Fields, H. L., Hjelmstad, G. O., Margolis, E. B., & Nicola, S. M. (2007). Ventral tegmental area neurons in learned appetitive behavior and positive reinforcement. *Annual Review in Neuroscience, 30*, 289-316.

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science, 299*, 1898-1902.

Fishbein, D. H., Eldreth, D. L., Hyde, C., Matochik, J. A., London, E. D., Contoreggi, C., et al. (2005). Risky decision making and the anterior cingulate cortex in abstinent drug abusers and nonusers. *Cognitive Brain Research, 23*, 119-136.

Fitzgerald, K. D., Welsh, R. C., Gehring, W. J., Abelson, J. L., Himle, J. A., Liberzon, I., et al. (2005). Error-related hyperactivity of the anterior cingulate cortex in obsessive-compulsive disorder. *Biological Psychiatry, 57*, 287-294.

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

Frank, M. J. (2006). Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks, 19*, 1120-1136.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective, and Behavioral Neuroscience, 1*, 137-160.

Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience, 120*(3), 497-517.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning. *Science, 306*, 1940-1943.

Funahashi, S. (2001). Neuronal mechanisms of executive control by the prefrontal cortex. *Neuroscience Research, 39*, 147-165.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology, 61*, 331-349.

Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: neural correlates of transient memory. *Journal of Neurophysiology, 36*, 61-78.

Gao, W.-J., Krimer, L. S., & Goldman-Rakic, P. S. (2001). Presynaptic regulation of recurrent excitation by D1 receptors in prefrontal circuits. *Proceedings of the National Academy of Sciences, 98*, 295-300.

Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., & Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. *NeuroImage, 17*, 1820-1829.

Gardner, E. L., & Ashby, C. R. (2000). Heterogeneity of the mesotelencephalic dopamine fibers: physiology and pharmacology. *Neuroscience and Behavioral Reviews, 24*, 115-118.

Gauntlett-Gilbert, J., & Brown, V. J. (1998). Reaction time deficits and Parkinson's disease. *Neuroscience and Biobehavioral Reviews, 22*(6), 865-881.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2002). *Cognitive Neuroscience: The Biology of the Mind* (2nd ed.). New York, NY: W.W. Norton & Company.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385-390.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295*, 2279-2282.

Gerfen, C. R. (1992). The neostriatal mosaic: multiple levels of compartmental organization in the basal ganglia. *Annual Review in Neuroscience, 15*, 285-320.

Gerfen, C. R. (2006). Indirect-pathway neurons lose their spines in Parkinson's disease. *Nature Neuroscience, 9*(2), 157-158.

Goldapple, K., Segal, Z., Garson, C., Lau, M., Bieling, P., Kennedy, S., et al. (2004). Modulation of cortical-limbic pathways in major depression: treatment-specific effects of cognitive behavior therapy. *Archives of General Psychiatry, 61*, 34-41.

Goldman-Rakic, P. S., Muly III, E. C., & Williams, G. V. (2000). D1 receptors in prefrontal cells and circuits. *Brain Research Reviews, 31*, 295-301.

Grace, A. A. (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience, 41*, 1-24.

Grace, A. A. (2000). The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving. *Addiction, 95*, S119-S128.

Grace, A. A. (2002). Dopamine. In K. L. Davis, D. Charney, J. T. Coyle & C. Nemeroff (Eds.), *Neuropsychopharmacology: The Fifth Generation of Progress* (pp. 119-132): Lippincott Williams & Wilkins.

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory, 70*, 119-136.

Greenberg, N. (2003). Sociality, stress, and the corpus striatum of the green anolis lizard. *Physiology and Behavior, 79*, 429-440.

Greengard, P. (2001). The neurobiology of slow synaptic transmission. *Science, 294*, 1024-1030.

Grillner, S. (2003). The motor infrastructure: from ion channels to neuronal networks. *Nature Reviews Neuroscience, 4*, 573-586.

Grillner, S., Hellgren, J., Ménard, A., Saitoh, K., & Wikström, M. A. (2005). Mechanisms for selection of basic motor programs--roles for the striatum and pallidum. *Trends in Neurosciences, 28*(7), 364-370.

Grunsfeld, A. A., & Login, I. S. (2006). Abulia following penetrating brain injury during endoscopic sinus surgery with disruption of the anterior cingulate circuit: case report. *BioMed Central Neurology, 6*, 4.

Gulledge, A. T., & Jaffe, D. B. (1998). Dopamine decreases the excitability of layer V pyramidal cells in the rat prefrontal cortex. *Journal of Neuroscience, 18*, 9139-9151.

Gurden, H., Takita, M., & Jay, T. M. (2000). Essential role of D1 but not D2 receptors in the NMDA receptor-dependent long-term potentiation at hippocampal-prefrontal cortex synapses in vivo. *Journal of Neuroscience, 20*, RC106.

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

Halliday, R., Naylor, H., Brandeis, D., Callaway, E., Yano, L., & Herzig, K. (1994). The effect of D-amphetamine, clonidine, and yohimbine on human information processing. *Psychophysiology, 31*, 331-337.

Harris, G. C., Wimmer, M., & Aston-Jones, G. (2005). A role for lateral hypothalamic orexin neurons in reward seeking. *Nature, 437*, 556-559.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York, NY: Wiley.

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

Hernández-López, S., Tkatch, T., Perez-Garci, E., Galarraga, E., Bargas, J., Hamm, H., et al. (2000). D2 dopamine receptors in striatal medium spiny neurons reduce L-type Ca2+ currents and excitability via a novel PLCβ1-IP3-calcineurin-signaling cascade. *Journal of Neuroscience, 20*(24), 8987-8995.

Hernandez, P. J., Sadeghian, K., & Kelley, A. E. (2002). Early consolidation of instrumental learning requires protein synthesis in the nucleus accumbens. *Nature Neuroscience, 5*(12), 1327-1331.

Hienz, R. D., Spear, D. J., & Bowers, D. A. (1994). Effects of cocaine on simple reaction times and sensory thresholds in baboons. *Journal of the Experimental Analysis of Behavior, 61*, 231-246.

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience, 1*(4), 304-309.

Huang, Y.-Y., Simpson, E., Kellendonk, C., & Kandel, E. R. (2004). Genetic evidence for the bidirectional modulation of synaptic plasticity in the prefrontal cortex by D1 receptors. *Proceedings from the National Academy of Sciences, 101*(9), 3236-3241.

Hyman, S. E., & Malenka, R. C. (2001). Addiction and the brain: the neurobiology of compulsion and its persistence. *Nature Reviews Neuroscience, 2*, 695-703.

Idris, N. F., Repeto, P., Neill, J. C., & Large, C. H. (2005). Investigation of the effects of lamotrigine and clozapine in improving reversal-learning impairments induced by acute phencyclidine and D-amphetamine in the rat. *Psychopharmacology, 179*, 336-348.

Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Research Reviews, 31*, 6-41.

Ikemoto, S., Qin, M., & Liu, Z.-H. (2005). The functional divide for primary reinforcement of D-amphetamine lies between the medial and lateral ventral striatum: is the division of the accumbens core, shell, and olfactory tubercle valid? *Journal of Neuroscience, 25*(20), 5061-5065.

Ito, S., Stuphorn, V., Brown, J. W., & Schall, J. D. (2003). Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science, 302*, 120-122.

Jentsch, J. D., Olausson, P., de La Garza, R., & Taylor, J. R. (2002). Impairments of reversal learning and response perseveration after repeated, intermittent cocaine administrations to monkeys. *Neuropsychopharmacology, 26*(2), 183-190.

Ji, W., & Suga, N. (2003). Development of reorganization of the auditory cortex caused by fear conditioning: effect of atropine. *Journal of Neurophysiology, 90*, 1904-1909.

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

Joel, D., & Weiner, I. (1994). The organization of the basal ganglia-thalamocortical circuits: open interconnected rather than closed segregated. *Neuroscience, 63*(2), 363-379.

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

Katz, P. S. (1998). Comparison of extrinsic and intrinsic neuromodulation in two central pattern generator circuits in invertebrates. *Experimental Physiology, 83*, 281-292.

Kaufer, D. I., & Lewis, D. A. (1998). Frontal lobe anatomy and cortical connectivity. In B. L. Miller & J. L. Cummings (Eds.), *The Human Frontal Lobes: Function and Disorders* (pp. 27-48). New York, NY: Guilford Press.

Keefe, K. A., Zigmond, M. J., & Abercrombie, E. D. (1992). Extracellular dopamine in striatum: influence of nerve impulse activity in medial forebrain bundle and local glutamatergic input. *Neuroscience, 47*(2), 325-332.

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

Kerns, J. G., Cohen, J. D., MacDonald III, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*, 1023-1026.

Kirsch, I., Lynn, S. J., Vigorito, M., & Miller, R. R. (2004). The role of cognition in classical and operant conditioning. *Journal of Clinical Psychology, 60*(4), 369-392.

Klier, E. M., Wang, H., Constantin, A. G., & Crawford, J. D. (2002). Midbrain control of three-dimensional head orientation. *Science, 295*, 1314-1316.

Klier, E. M., Wang, H., & Crawford, J. D. (2003). Three-dimensional eye-head coordination is implemented downstream from the superior colliculus. *Journal of Neurophysiology, 89*, 2839-2853.

Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science, 318*, 594-598.

Kolb, B., & Whishaw, I. Q. (2003). *Fundamentals of human neuropsychology* (Fifth ed.). New York, NY: Worth Publishers.

Kuo, M.-F., Grosch, J., Fregni, F., Paulus, W., & Nitsche, M. A. (2007). Focusing effect of acetylcholine on neuroplasticity in the human motor cortex. *Journal of Neuroscience, 27*(52), 14442-14447.

Law-Tho, D., Desce, J. M., & Crepel, F. (1995). Dopamine favours the emergence of long-term depression versus long-term potentiation in slices of rat prefrontal cortex. *Neuroscience Letters, 188*, 125-128.

Le Moal, M., & Simon, H. (1991). Mesocorticolimbic dopaminergic network: functional and regulatory roles. *Physiological Reviews, 71*(1), 155-234.

Lee, C. R., Abercrombie, E. D., & Tepper, J. M. (2004). Pallidal control of substantia nigra dopaminergic neuron firing pattern and its relation to extracellular neostriatal dopamine levels. *Neuroscience, 129*, 481-489.

Levy, F., & Swanson, J. M. (2001). Timing, space and ADHD: the dopamine theory revisited. *Australian and New Zealand Journal of Psychiatry, 35*, 504-511.

Lewis, B. L., & O'Donnell, P. O. (2000). Ventral tegmental area afferents to the prefrontal cortex maintain membrane potential 'up' states in pyramidal neurons via D1 dopamine receptors. *Cerebral Cortex, 10*, 1168-1175.

Lewis, S. J. G., Dove, A., Robbins, T. W., Barker, R. A., & Owen, A. M. (2003). Cognitive impairments in early Parkinson's disease are accompanied by reductions in activity in frontostriatal neural circuitry. *Journal of Neuroscience, 23*(15), 6351-6356.

Lidow, M. S., Goldman-Rakic, P. S., Gallager, D. W., & Rakic, P. (1991). Distribution of dopaminergic receptors in the primate cerebral cortex: quantitative autoradiographic analysis using [3H]raclopride, [3H]spiperone and [3H]SCH23390. *Neuroscience, 40*(3), 657-671.

Lisman, J. (1989). A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory. *Proceedings from the National Academy of Sciences, 86*, 9574-9578.

Lisman, J. E., & Otmakhova, N. A. (2001). Storage, recall, and novelty detection of sequences by the hippocampus: elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. *Hippocampus, 11*, 551-568.

Ljungberg, T., Apicella, P., & Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. *Journal of Neurophysiology, 67*(1), 145-163.

Lorist, M. M., Boksem, M. A. S., & Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research, 24*, 199-205.

Luria, A. R. (1973). *The Working Brain*. Harmondsworth, England: Penguin.

MacDonald III, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288*, 1835-1838.

MacLean, P. D. (1990). *The Triune Brain in Evolution: Role in Paleocerebral Functions*. New York, NY: Plenum Press.

Magee, J. C., & Johnston, D. (1995). Synaptic activation of voltage-gated channels in the dendrites of hippocampal pyramidal neurons. *Science, 268*, 301-304.

Magno, E., Foxe, J. J., Molholm, S., Robertson, I. H., & Garavan, H. (2006). The anterior cingulate and error avoidance. *Journal of Neuroscience, 26*(18), 4769-4773.

Mahon, S., Deniau, J.-M., & Charpier, S. (2004). Corticostriatal plasticity: life after the depression. *Trends in Neurosciences, 27*(8), 460-467.

Malenka, R. C. (2002). Synaptic plasticity. In K. L. Davis, D. Charney, J. T. Coyle & C. Nemeroff (Eds.), *Neuropsychopharmacology: The Fifth Generation of Progress* (pp. 147-157): Lippincott Williams & Wilkins.

Marder, E., & Thirumalai, V. (2002). Cellular, synaptic and network effects of neuromodulation. *Neural Networks, 15*, 479-493.

Matsuda, Y., Marzo, A., & Otani, S. (2006). The presence of background dopamine signal converts long-term synaptic depression to potentiation in rat prefrontal cortex. *Journal of Neuroscience, 26*(18), 4803-4810.

Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., et al. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron, 45*, 651-660.

Mena-Segovia, J., Bolam, J. P., & Magill, P. J. (2004). Pedunculopontine nucleus and basal ganglia: distant relatives or part of the same family? *Trends in Neurosciences, 27*(10), 585-588.

Mengual, E., de las Heras, S., Erro, E., Lanciego, J. L., & Giménez-Amaya, J. M. (1999). Thalamic interaction between the input and the output systems of the basal ganglia. *Journal of Chemical Neuroanatomy, 16*(3), 185-197.

Middleton, F. A., & Strick, P. L. (2000). Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. *Brain and Cognition, 42*, 183-200.

Middleton, F. A., & Strick, P. L. (2001). A revised neuroanatomy of frontal-subcortical circuits. In J. L. Cummings & D. G. Lichter (Eds.), *Frontal-Subcortical Circuits in Psychiatric and Neurologic Disorders*. New York, NY: Guilford Press.

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience, 1*, 59-65.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review in Neuroscience, 24*, 167-202.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of working memory in prefrontal cortex of the macaque. *Journal of Neuroscience, 16*, 5154-5167.

Milner, A. D., & Goodale, M. A. (1998). *The Visual Brain in Action*. Oxford: Oxford University Press.

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

Minsky, M. (1986). *The Society of Mind*. New York, NY: Simon and Schuster.

Mitz, A. R., Godschalk, M., & Wise, S. P. (1991). Learning-dependent neuronal activity in the premotor cortex: activity during the acquisition of conditional motor associations. *Journal of Neuroscience, 11*(6), 1855-1872.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134-140.

Mufson, E. J., & Pandya, D. N. (1984). Some observations on the course and composition of the cingulum bundle in the rhesus monkey. *Journal of Comparative Neurology, 225*, 31-43.

Muller, T., Eising, E., Kuhn, W., Buttner, T., Coenen, H.-H., & Przuntek, H. (1999). Delayed motor response correlates with striatal degeneration in Parkinson's disease. *Acta Neurologica Scandinavica, 100*(4), 227-230.

Muly III, E. C., Szigeti, K., & Goldman-Rakic, P. S. (1998). D1 receptor in interneurons of macaque prefrontal cortex: distribution and subcellular localization. *Journal of Neuroscience, 18*(24), 10553-10565.

Nelson, A., & Killcross, S. (2006). Amphetamine exposure enhances habit formation. *Journal of Neuroscience, 26*(14), 3805-3812.

Niki, H., & Watanabe, M. (1979). Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Research, 171*, 213-224.

Niv, Y., Daw, N. D., & Dayan, P. (2005). How fast to work: response vigor, motivation and tonic dopamine. In Y. Weiss, B. Schölkopf & J. Platt (Eds.), *Advances in Neural Information Processing Systems* (Vol. 18, pp. 1019-1026). Cambridge, MA: MIT Press.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2006). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology, 191*(3), 507-520.

Norman, D. A., & Shallice, T. (1986). Attention to action: willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and Self-regulation* (Vol. 4, pp. 1-18). New York, NY: Plenum Press.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304*, 452-454.

O'Neill, M., & Brown, V. J. (2007). The effect of striatal dopamine depletion and the adenosine A2A antagonist KW-6002 on reversal learning in rats. *Neurobiology of Learning and Memory, 88*, 75-81.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation, 18*, 283-328.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.

Oades, R. D., & Halliday, G. M. (1987). Ventral tegmental (A10) system: neurobiology. 1. Anatomy and connectivity. *Brain Research Reviews, 12*, 117-165.

Ohbayashi, M., Ohki, K., & Miyashita, Y. (2003). Conversion of working memory to motor sequence in the monkey premotor cortex. *Science, 301*, 233-236.

Otani, S., Auclair, N., Desce, J. M., Roisin, M.-P., & Crépel, F. (1999). Dopamine receptors and groups I and II mGluRs cooperate for long-term depression

induction in rat prefrontal cortex through converging postsynaptic activation of MAP kinases. *Journal of Neuroscience, 19*(22), 9788-9802.

Otani, S., Blond, O., Desce, J. M., & Crepel, F. (1998). Dopamine facilitates long-term depression of glutamatergic transmission in rat prefrontal cortex. *Neuroscience, 85*(3), 669-676.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review in Neuroscience, 25*, 563-593.

Panksepp, J. (1986). The neurochemistry of behavior. *Annual Review in Psychology, 37*, 77-107.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotion*. New York, NY: Oxford University Press.

Parkinson, J. A., Dalley, J. W., Cardinal, R. N., Bamford, A., Fehnert, B., Lachenal, G., et al. (2002). Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. *Behavioural Brain Research, 137*, 149-163.

Passingham, R. E. (1993). *The Frontal Lobes and Voluntary Action*. Oxford: Oxford University Press.

Paus, T. (2001). Primate anterior cingulate cortex: where motor control, drive and cognition interface. *Nature Reviews Neuroscience, 2*(6), 417-424.

Petrides, M., & Pandya, D. N. (2007). Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. *Journal of Neuroscience, 27*(43), 11573-11586.

Pifl, C., Sitte, H. H., Reither, H., & Singer, E. A. (2000). The mechanism of the releasing action of amphetamine. Uptake, superfusion, and electrophysiological studies on transporter-transfected cells. *Pure and Applied Chemistry, 72*(6), 1045-1050.

Posner, M. I., & Rothbart, M. K. (1998). Attention, self-regulation and consciousness. *Philosophical Transactions of the Royal Society of London B Biological Science, 353*, 1915-1927.

Procyk, E., Tanaka, Y. L., & Joseph, J. P. (2000). Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nature Neuroscience, 3*(5), 502-508.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999a). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience, 89*, 1009-1023.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999b). Is the short-latency dopamine response too short to signal reward error? *Trends in Neurosciences, 22*(4), 146-151.

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist, 43*(3), 151-160.

Rescorla, R. A. (1991). Associative relations in instrumental learning: the eighteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology, 43B*(1), 1-23.

Reynolds, J. N. J., Hyland, B. I., & Wickens, J. R. (2001). A cellular mechanism of reward-related learning. *Nature, 413*, 67-70.

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

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science, 306*, 443-447.

Ridderinkhof, K. R., van den Wildenberg, W. P. M., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition, 56*, 129-140.

Robbins, T. W., Giardini, V., Jones, G. H., Reading, P., & Sahakian, B. J. (1990). Effects of dopamine depletion from the caudate-putamen and nucleus accumbens septi on the acquisition and performance of a conditional discrimination task. *Behavioural Brain Research, 38*, 243-261.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207-231.

Rolls, E. T. (1999). The functions of the orbitofrontal cortex. *Neurocase, 5*, 301-312.

Rolls, E. T., & Deco, G. (2002). *Computational Neuroscience of Vision*. New York, NY: Oxford University Press.

Room, P., Russchen, F. T., Groenewegen, H. J., & Lohman, A. H. M. (1985). Efferent connections of the prelimbic (area 32) and the infralimbic (area 25) cortices: an anterograde tracing study in the cat. *Journal of Comparative Neurology, 242*, 40-55.

Rougier, N. P., Noelle, D. C., Braver, T. S., Cohen, J. D., & O'Reilly, R. C. (2005). Prefrontal cortex and flexible cognitive control: rules without symbols. *Proceedings from the National Academy of Sciences, 102*(20), 7338-7343.

Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1988). Procedural learning and neostriatal dysfunction in man. *Brain, 111*, 941-959.

Saunders, C., Ferrer, J. V., Shi, L., Chen, J., Merrill, G., Lamb, M. E., et al. (2000). Amphetamine-induced loss of human dopamine transporter activity: an internalization-dependent and cocaine-sensitive mechanism. *Proceedings from the National Academy of Sciences, 97*(12), 6850-6855.

Sawaguchi, T., & Goldman-Rakic, P. S. (1994). The role of D1-dopamine receptor in working memory: local injections of dopamine antagonists into the prefrontal cortex of rhesus monkeys performing an oculomotor delayed-response task. *Journal of Neurophysiology, 71*(2), 515-528.

Schieber, M. H. (2001). Constraints on somatotopic organization in the primary motor cortex. *Journal of Neurophysiology, 86*, 2125-2143.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80*, 1-27.

Schultz, W. (2007). Multiple dopamine functions at different time courses. *Annual Review in Neuroscience, 30*, 259-288.

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

Schultz, W., Studer, A., Romo, R., Sundström, E., Jonsson, G., & Scarnati, E. (1989). Deficits in reaction times and movement times as correlates of hypokinesia in monkeys with MPTP-induced striatal dopamine depletion. *Journal of Neurophysiology, 61*(3), 651-668.

Schweimer, J., & Hauber, W. (2006). Dopamine D1 receptors in the anterior cingulate cortex regulate effort-based decision making. *Learning and Memory, 13*, 777-782.

Seamans, J. K., Durstewitz, D., Christie, B. R., Stevens, C. F., & Sejnowski, T. J. (2001). Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons. *Proceedings from the National Academy of Sciences, 98*(1), 301-306.

Seamans, J. K., & Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in Neurobiology, 74*, 1-57.

Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science, 249*, 892-895.

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

Shi, W.-X., Pun, C.-L., Zhang, X.-X., Jones, M. D., & Bunney, B. S. (2000). Dual effects of D-amphetamine on dopamine neurons mediated by dopamine and nondopamine receptors. *Journal of Neuroscience, 20*(9), 3504-3511.

Shidara, M., & Richmond, B. J. (2002). Anterior cingulate: single neuronal signals related to degree of reward expectancy. *Science, 296*, 1709-1711.

Shima, K., & Tanji, J. (1998). Role for cingulate motor area cells in voluntary movement selection based on reward. *Science, 282*, 1335-1338.

Smiley, J. F., Levey, A. I., Ciliax, B. J., & Goldman-Rakic, P. S. (1994). D1 dopamine receptor immunoreactivity in human and monkey cerebral cortex: predominant

and extrasynaptic localization in dendritic spines. *Proceedings from the National Academy of Sciences, 91*, 5720-5724.

Smith-Roe, S. L., & Kelley, A. E. (2000). Coincident activation of NMDA and dopamine D1 receptors within the nucleus accumbens core is required for appetitive instrumental learning. *Journal of Neuroscience, 20*(20), 7737-7742.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*, 1657-1661.

Sokolowski, J. D., & Salamone, J. D. (1998). The role of accumbens dopamine in lever pressing and response allocation: effects of 6-OHDA injected into core and dorsomedial shell. *Pharmacology Biochemistry and Behavior, 59*(3), 557-566.

Sporns, O., & Alexander, W. H. (2002). Neuromodulation and plasticity in an autonomous robot. *Neural Networks, 15*, 761-774.

Stern, C. E., Sherman, S. J., Kirchhoff, B. A., & Hasselmo, M. E. (2001). Medial temporal and prefrontal contributions to working memory tasks with novel and familiar stimuli. *Hippocampus, 11*, 337-346.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 12*, 643-662.

Suri, R. E. (2002). TD models of reward predictive responses in dopamine neurons. *Neural Networks, 15*, 523-533.

Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & J. Moore (Eds.), *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (pp. 497-537). Cambridge, MA: MIT Press.

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

Tagamets, M. A., & Horwitz, B. (1998). Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. *Cerebral Cortex, 8*, 310-320.

Takahata, R., & Moghaddam, B. (1998). Glutamatergic regulation of basal and stimulus-activated dopamine release in the prefrontal cortex. *Journal of Neurochemistry, 71*, 1443-1449.

Takechi, H., Eilers, J., & Konnerth, A. (1998). A new class of synaptic response involving calcium release in dendritic spines. *Nature, 396*, 757-760.

Tanaka, S. (2002). Dopamine controls fundamental cognitive operations of multi-target spatial working memory. *Neural Networks, 15*, 573-582.

Tekin, S., & Cummings, J. L. (2002). Frontal-subcortical neuronal circuits and clinical neuropsychiatry: an update. *Journal of Psychosomatic Research, 53*, 647-654.

Thorndike, E. L. (1911). *Animal Intelligence*. Darien, CT: Hafner.

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

Trappenberg, T. P., Dorris, M. C., Munoz, D. P., & Klein, R. M. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience, 13*(2), 256-271.

Ungerleider, L. G., Courtney, S. M., & Haxby, J. V. (1998). A neural system for human visual working memory. *Proceedings of the National Academy of Sciences, 95*, 883-890.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Engle, M. A. Goodale & R. J. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-586). Cambridge, MA: MIT Press.

Ungless, M. A., Magill, P. J., & Bolam, J. P. (2004). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science, 303*, 2040-2042.

Uylings, H. B. M., Groenewegen, H. J., & Kolb, B. (2003). Do rats have a prefrontal cortex? *Behavioural Brain Research, 146*, 3-17.

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *NeuroImage, 14*, 1302-1308.

Vijayaraghavan, L., Krishnamoorthy, E. S., Brown, R. G., & Trimble, M. R. (2002). Abulia: a Delphi survey of British neurologists and psychiatrists. *Movement Disorders, 17*(5), 1052-1057.

Vogt, B. A., & Pandya, D. N. (1987). Cingulate cortex of the rhesus monkey: II. Cortical afferents. *Journal of Comparative Neurology, 262*, 271-289.

Voorn, P., Vanderschuren, L. J. M. J., Groenewegen, H. J., Robbins, T. W., & Pennartz, C. M. A. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends in Neurosciences, 27*(28), 468-474.

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

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature, 411*, 953-956.

Weed, M. R., & Gold, L. H. (1998). The effects of dopaminergic agents on reaction time in rhesus monkeys. *Psychopharmacology, 137*, 33-42.

Weng, J., McClelland, J. L., Pentland, A., Sporns, O., Stockman, I., Sur, M., et al. (2001). Autonomous mental development by robots and animals. *Science, 291*, 599-600.

Wichmann, T., & DeLong, M. R. (1996). Functional and pathophysiological models of the basal ganglia. *Current Opinion in Neurobiology, 6*, 751-758.

Wilson, C. J., & Kawaguchi, Y. (1996). The origins of two-state spontaneous membrane potential fluctuations of neostriatal spiny neurons. *Journal of Neuroscience, 16*(7), 2397-2410.

Wilson, F. A. W., Ó Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science, 260*, 1955-1958.

Wise, S. P. (1985). The primate premotor cortex: past, present, and preparatory. *Annual Review in Neuroscience, 8*, 1-19.

Woods, S. P., & Tröster, A. I. (2003). Prodromal frontal/executive dysfunction predicts incident dementia in Parkinson's disease. *Journal of the International Neuropsychological Society, 9*, 17-24.

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs". *Psychological Research, 63*, 212-233.

Yang, C. R., & Chen, L. (2005). Targeting prefrontal cortical dopamine D1 and N-methyl-D-aspartate receptor interactions in schizophrenia treatment. *The Neuroscientist, 11*, 452-470.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience, 19*, 181-189.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience, 22*, 513-523.

Zahrt, J., Taylor, J. R., Mathew, R. G., & Arnsten, A. F. T. (1997). Supranormal stimulation of D1 dopamine receptors in the rodent prefrontal cortex impairs spatial working memory performance. *Journal of Neuroscience, 17*, 8528-8535.

Zhuang, X., Oosting, R. S., Jones, S. R., Gainetdinov, R. R., Miller, G. W., Caron, M. G., et al. (2001). Hyperactivity and impaired response inhibition in hyperdopaminergic mice. *Proceedings from the National Academy of Sciences, 98*(4), 1982-1987.

Table 1.1 Abbreviations used in the dissertation.

5-HT	Serotonin
ACC	Anterior cingulate cortex
ADHD	Attention deficit hyperactivity disorder
AI	Artificial intelligence
BA	Brodmann area
BG	Basal ganglia
CR	Conditioned response
CS	Conditioned stimulus
DA	Dopamine
DAT	Dopamine transporter
ERN	Error-related negativity
FEF	Frontal eye fields (lateral BA 8)
FPC	Frontopolar cortex (BA 10)
GPi	Globus pallidus internal segment
GPe	Globus pallidus external segment
IT	Inferotemporal cortex
kWTA	k winners-take-all
LDT	Laterodorsal tegmental nucleus
LTD	Long-term depression
LTP	Long-term potentiation
M1	Primary motor cortex
NAc	Nucleus accumbens
NE	Norephinephrine (noradrenaline)
O	Outcome
OFC	Orbitofrontal cortex
PD	Parkinson's disease
PFC	Prefrontal cortex (dl = dorsolateral, etc.)
PMC	Premotor cortex (lateral BA 6)
PPC	Posterior parietal cortex
PPN	Pendunculopontine (tegmental) nucleus
R	Response
RT	Reaction (response) time
S	Stimulus
S^D	Discriminative stimulus
SEF	Supplementary eye fields (medial BA 8)
SMA	Supplementary motor area (medial BA 6)
SNc	Substantia nigra pars compacta
SNr	Substantia nigra pars reticulata
STN	Subthalamic nucleus
TOBS	Task-oriented behavior selection
UR	Unconditioned response
US	Unconditioned stimulus
VTA	Ventral tegmental area

Table 2.1 Simulation movies available online.

Movie	URL/s
2.1 Full task set training phase performance	http://gchadder3.com/dissmovies/mov2s1.wmv (compressed: ~ 700 K) http://gchadder3.com/dissmovies/mov2s1best.wmv (uncompressed: ~ 4.3 M)
2.2 Full task set reversal phase performance	http://gchadder3.com/dissmovies/mov2s2.wmv
2.3 Neural noise generation mechanism performance	http://gchadder3.com/dissmovies/mov2s3.wmv
2.4 Full task set babble activation during Eb/Red -> Nod learning	http://gchadder3.com/dissmovies/mov2s4.wmv
2.5 Example of conjunction learning: C/Red -> Shake	http://gchadder3.com/dissmovies/mov2s5.wmv
2.6 C/Red -> Shake learning: Plan unit weight modification	http://gchadder3.com/dissmovies/mov2s6.wmv
2.7 C/Red -> Shake learning: Shake Go unit weight modification	http://gchadder3.com/dissmovies/mov2s7.wmv
2.8 Example of conjunction relearning: C/Blue -> (Nod to Shake)	http://gchadder3.com/dissmovies/mov2s8.wmv
2.9 C/Blue -> (Nod to Shake) relearning: Plan unit weight modification	http://gchadder3.com/dissmovies/mov2s9.wmv
3.1 Example of accidental conjunction overwriting	http://gchadder3.com/dissmovies/mov3s1.wmv
3.2 Example of novelty driven learning: Plan unit weight modification	http://gchadder3.com/dissmovies/mov3s2.wmv

Table 3.1 Full task set training simulation phase run summary.

Run #	Iteration #'s	Trial #'s	Events
1	1-1000	1-15	Training on C/Blue->Nod; learned successfully
2	1001-2000	16-30	Training on C/Red->Shake; not learned
3	2001-3000	31-44	Training on C/Red->Shake; learned successfully
4	3001-4000	45-59	Training on Eb/Red->Nod; learned successfully
5	4001-5000	60-74	Training on Eb/Blue->Shake; learned successfully
6	5001-6000	75-88	Testing on all conjunctions; C/Blue->Nod, C/Red ->Shake not working
7	6001-7000	89-103	Training on C/Blue->Nod; not learned
8	7001-8000	104-118	Training on C/Blue->Nod; learned successfully
9	8001-9000	119-133	Training on C/Red->Shake; not learned
10	9001-10000	134-148	Training on C/Red->Shake; learned successfully
11	10001-11000	149-163	Testing on all conjunctions; Eb/Red->Nod not working
12	11001-12000	164-178	Training on Eb/Red->Nod; not learned
13	12001-13000	179-192	Training on Eb/Red->Nod; learned successfully
14	13001-14000	193-206	Testing on all conjunctions; all working
15	14001-15000	207-221	Testing on all conjunctions; all working

Table 3.2 Full task set maintenance simulation phase run summary.

Run #	Iteration #'s	Trial #'s	Events
1	1-1000	1-15	12 Hits, 3 Misses
2	1001-2000	16-29	14 Hits
3	2001-3000	30-44	10 Hits, 5 Misses
4	3001-4000	45-59	9 Hits, 6 Misses
5	4001-5000	60-73	9 Hits, 5 Misses

Table 3.3 Full task set reversal simulation phase run summary.

Run #	Iteration #'s	Trial #'s	Events
1	1-1000	1-15	Training on C/Blue->Shake; learned successfully
2	1001-2000	16-29	Training on C/Red->Nod; learned successfully
3	2001-3000	30-44	Training on Eb/Red->Shake; learned successfully
4	3001-4000	45-57	Training on Eb/Blue->Nod; not learned
5	4001-5000	58-72	Training on Eb/Blue->Nod; learned successfully
6	5001-6000	73-87	Testing on all conjunctions; all working
7	6001-7000	88-102	Testing on all conjunctions; all working

Table 4.1 Summary of DA manipulation results discussed in Section 3.2. Action speed (in an inverse way) was measured by RT (see Figure 3.11). Behavior initiation was measured (inversely) by the percent of missed responses (see Figure 3.17). Task acquisition was measured by the percent of 'subjects' that learned the initial C AND Blue -> Nod task (see Figure 3.22). Task learning maintenance was measured by the percent of 'subjects' that maintained learning of the initial task (see Figure 3.29). Task reversal learning was measured by the percent of 'subjects' that learned the new C AND Blue -> Shake task (see Figure 3.34). Each table entry represents a test that could be performed on animal subjects, and results the model would predict. ds = dorsal striatal hypo-DA; vs = ventral striatal hypo-DA; nc = neocortical hypo-DA; g = global hypo-DA; DS = dorsal striatal hyper-DA; VS = ventral striatal hyper-DA; NC = neocortical hyper-DA; G = global hyper-DA.

DA Manipulation Effect	**ds**	**vs**	**nc**	**g**	**DS**	**VS**	**NC**	**G**
Action Speed	↓	0	$\downarrow_1$	↓	↑	0	0	↑
Behavior Initiation	↓	0	$\downarrow_1$	↓	0	0	0	0
Task Acquisition	↓	$\downarrow_2$	X_1	X_1	0	0	↓	↓
Task Learning Maintenance	$\downarrow_1$	0	X_1	X_1	0	0	0	0
Task Reversal Learning	$\downarrow_1$	$\downarrow_2$	X_1	X_1	0	0	↓	↓

↑ = increase of effect
0 = no change of effect
↓ = decrease of effect
X = complete disruption
1 = only effective at the most extreme 2 hypo-DA conditions
2 = effect is relatively minor

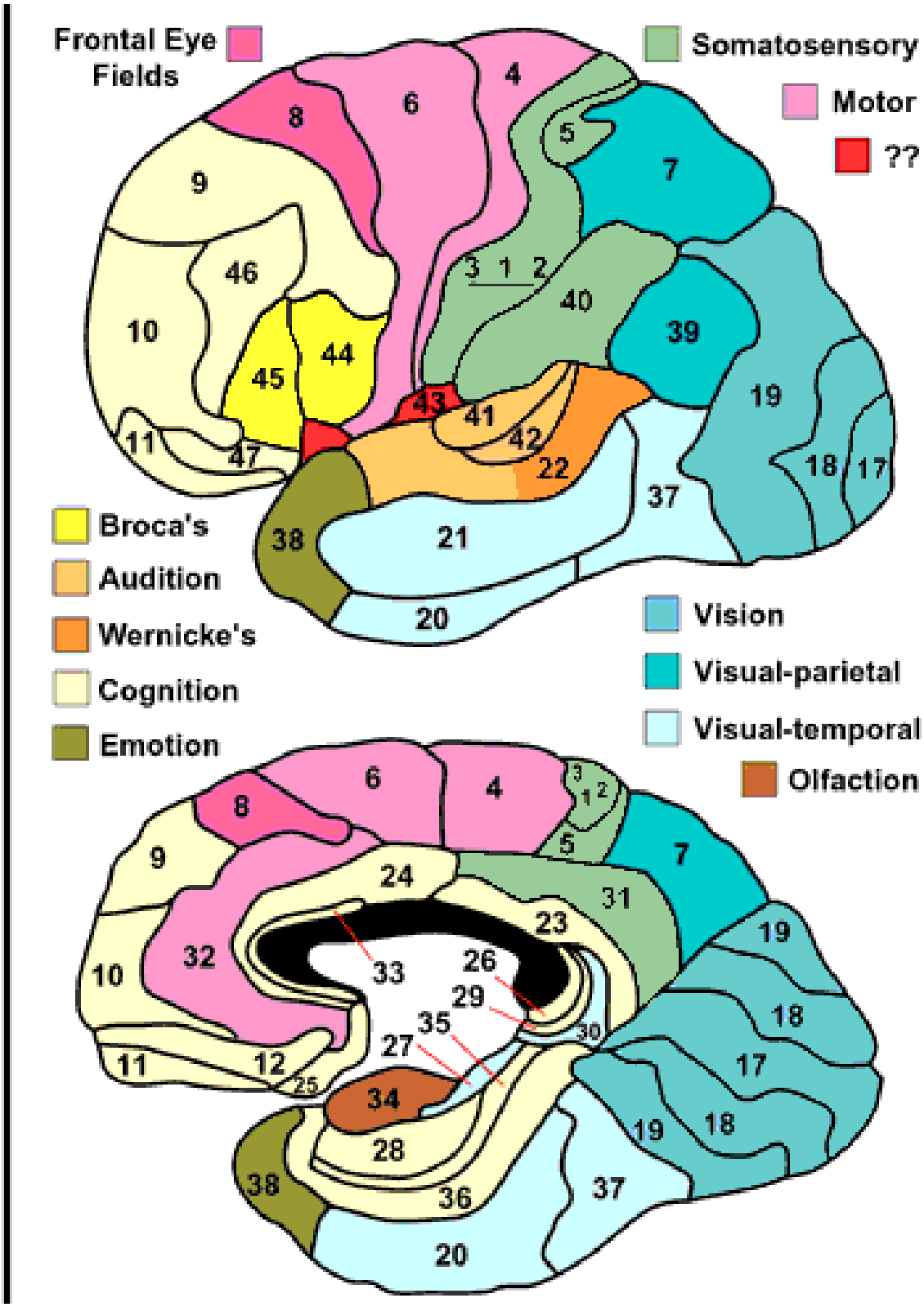


Figure 1.1 Human Brodmann areas. Figure downloaded from http://spot.colorado.edu/~dubin/talks/brodmann/brodmann.html.

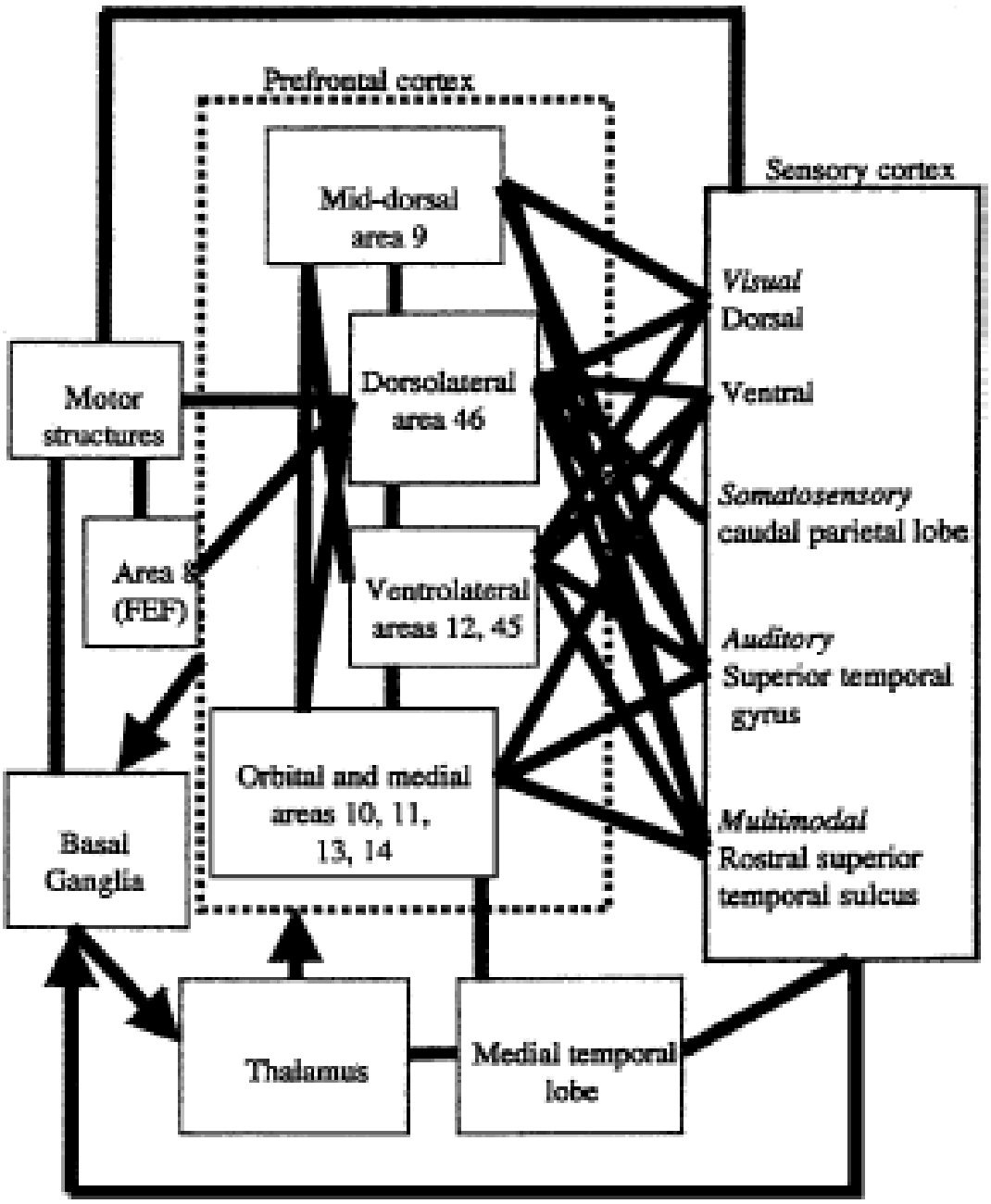


Figure 1.2 Prefrontal cortex and its afferent and efferent connections. Taken from Figure 1 of (Miller & Cohen, 2001).

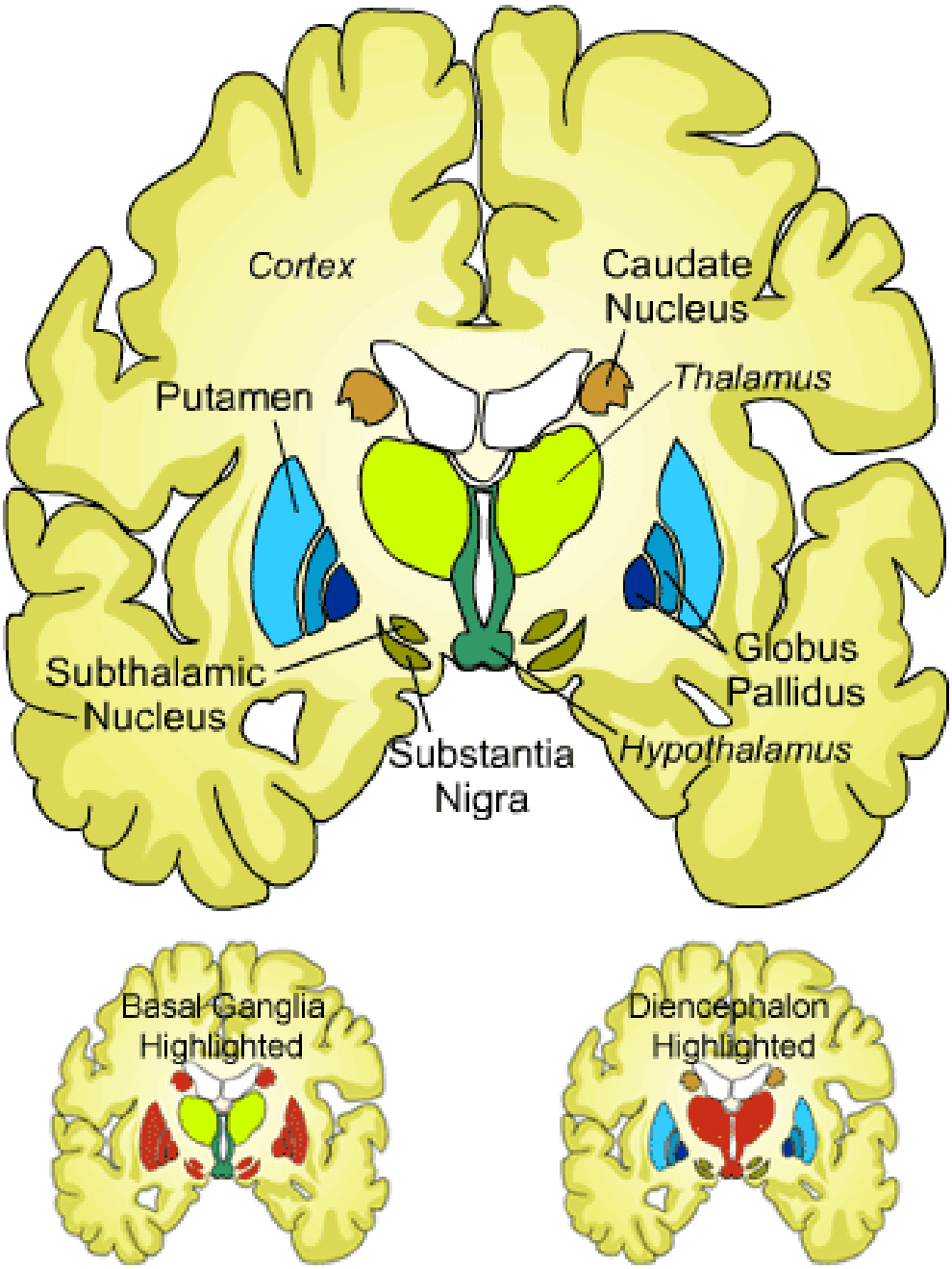


Figure 1.3 Location and components of the basal ganglia. Adapted from www.stanford.edu/.../braintut/f_ab18bslgang.gif. Areas with non-italicized labels are considered components of the basal ganglia. The caudate nucleus and putamen together form the striatum.

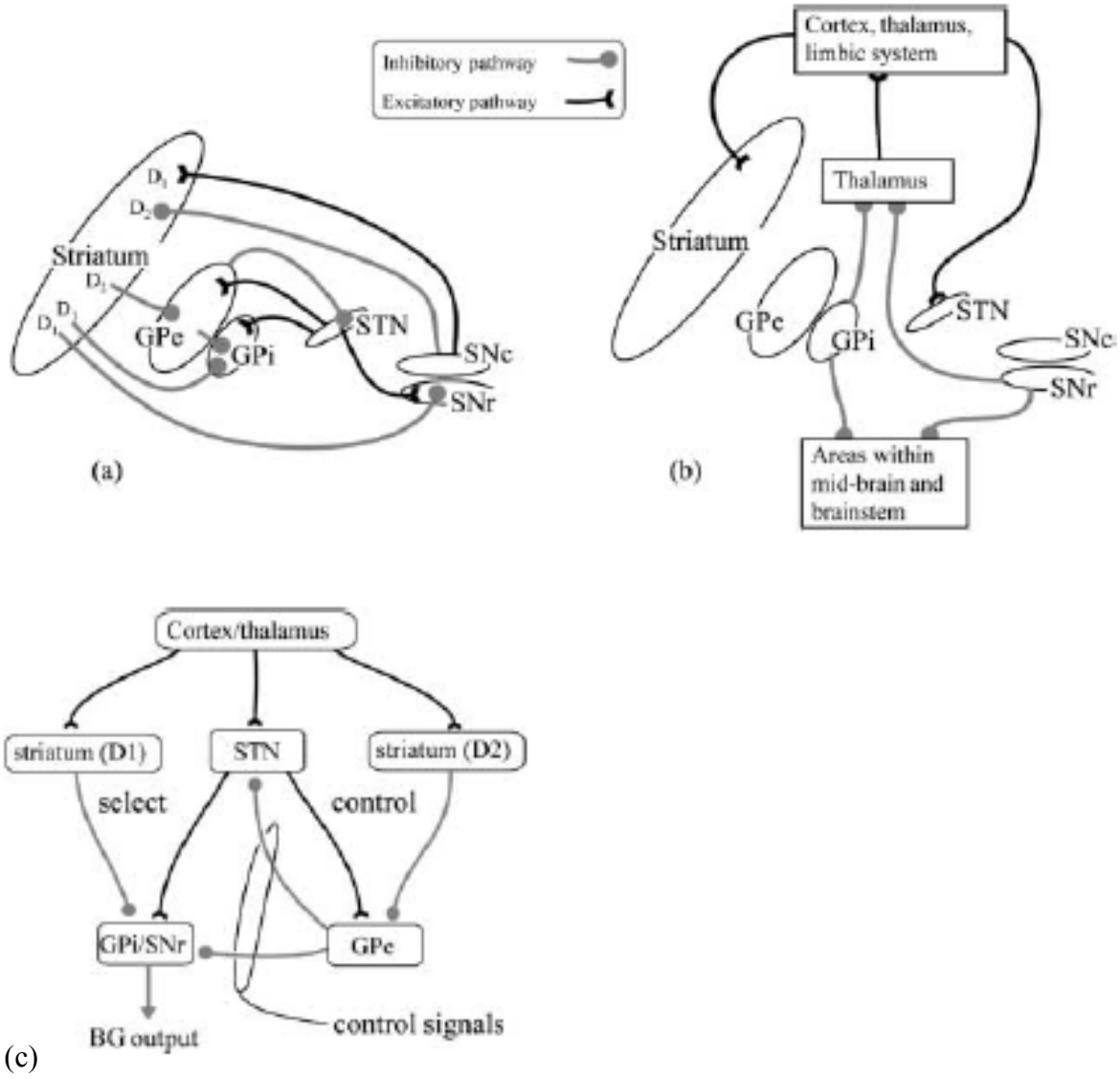


Figure 1.4 Basal ganglia anatomical circuit. Adapted from Figures 1 and 5 of (Gurney, Prescott, & Redgrave, 2001). (a) Internal BG pathways. (b) External BG pathways. (c) Functional architecture. The D1 striatal pathway is a 'Go' pathway because it inhibits the BG output which is itself inhibitory. The D2 striatal pathway (striatum-GPe-GPi/SNr) is a 'NoGo' pathway because it disinhibits the BG output leading to inhibited thalamic, etc., activity. The main STN pathway (STN-GPi/SNr) also has an inhibitory effect on behavior.

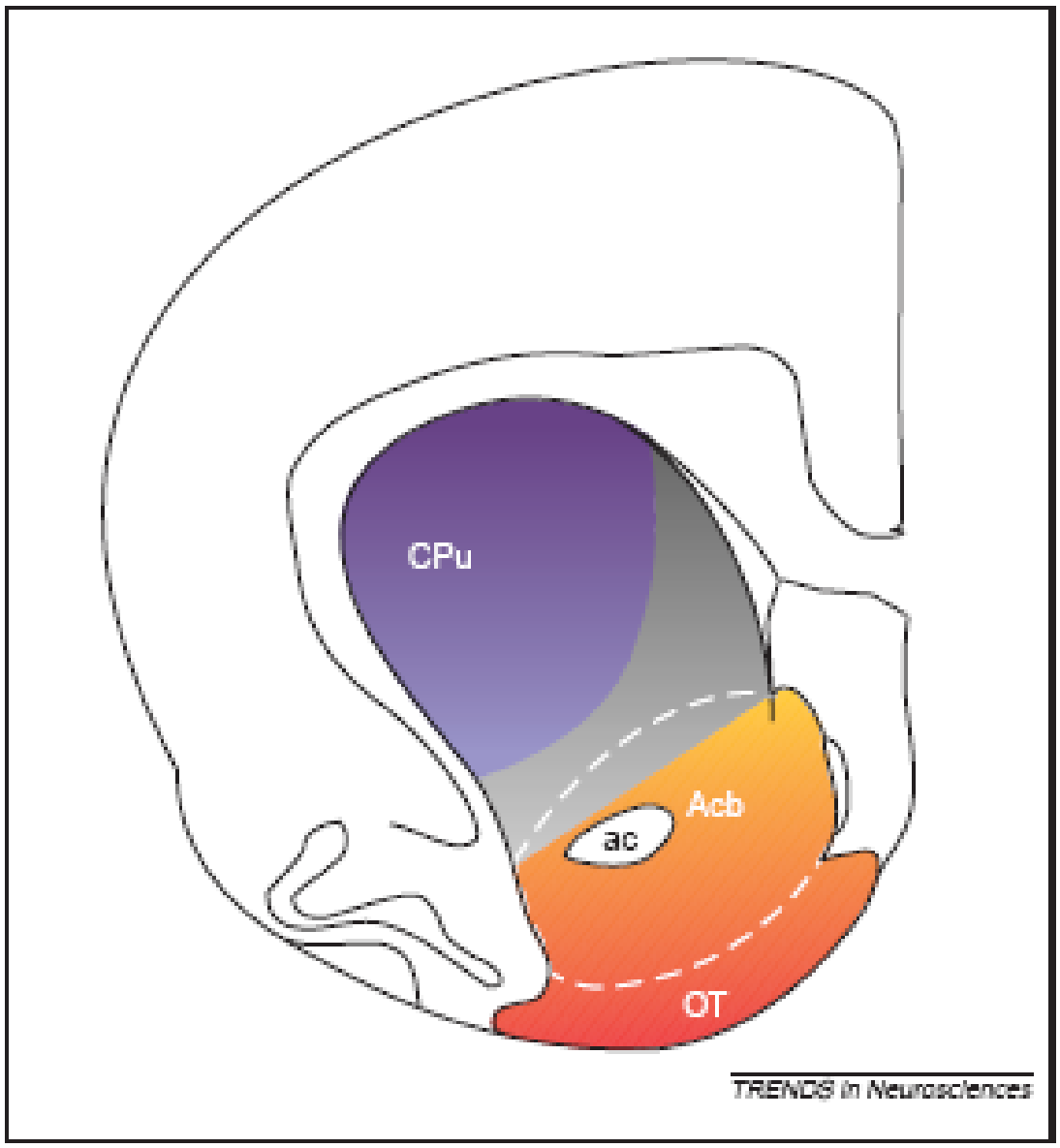


Figure 1.5 Typical anatomical division of the striatum for a rat. Taken from Figure 1 of (Voorn, Vanderschuren, Groenewegen, Robbins, & Pennartz, 2001). The boundary between the dorsal and ventral striatum is typically drawn at the upper dotted line. CPu = caudate-putamen; Acb = nucleus accumbens; OT = olfactory tubercle; ac = anterior commisure.

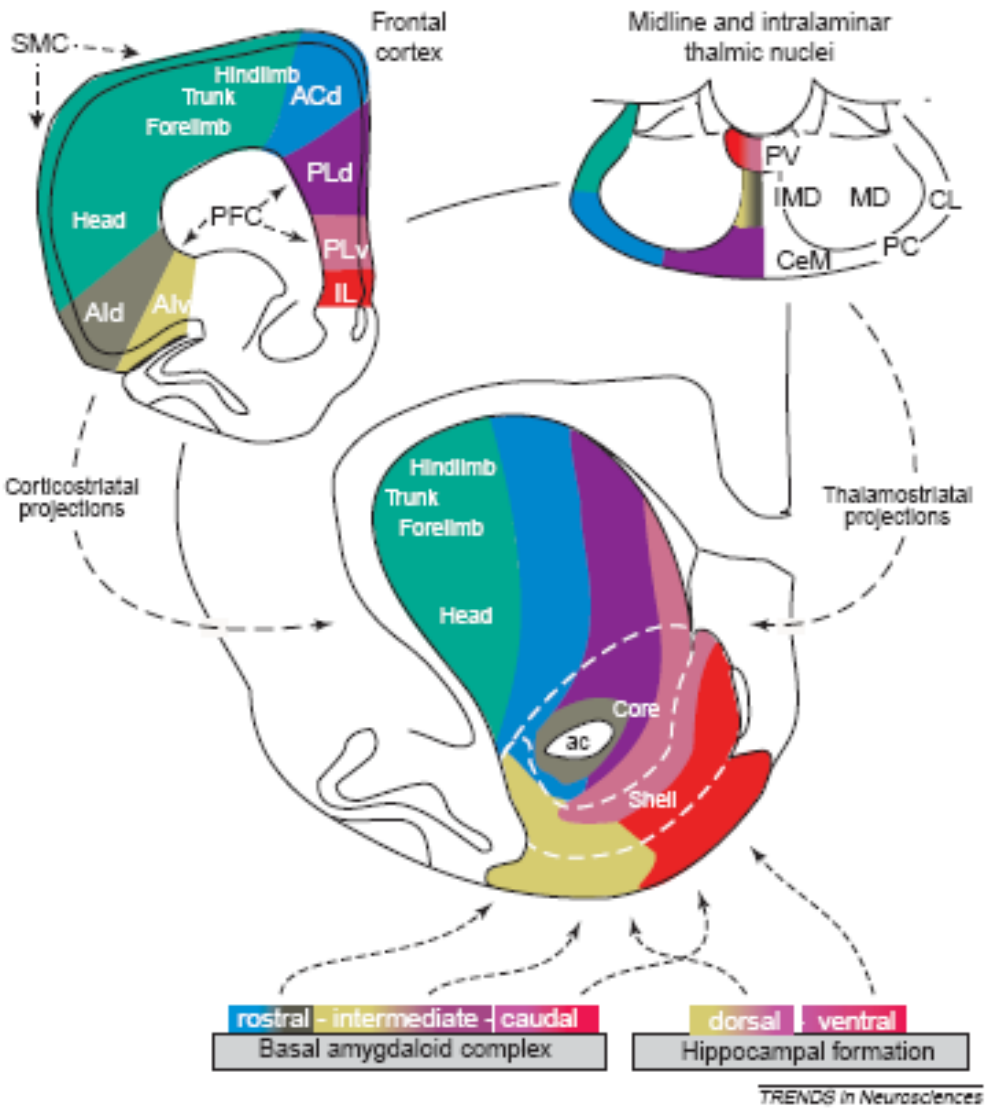


Figure 1.6 Anatomical map of striatum and associated cortical and subcortical connections in a rat. Taken from Figure 3 of (Voorn, Vanderschuren, Groenewegen, Robbins, & Pennartz, 2001). A gradient from dorsolateral to ventromedial according to connectivity is noticeable. Motor areas tend to be innervated by more dorsolateral regions of striatum, whereas associative and limbic regions are innervated more by ventromedial regions. ACd = dorsal anterior cingulate cortex; IL = infralimbic cortex; PLd = dorsal prelimbic cortex; PLv = ventral prelimbic cortex; SMC = sensorimotor cortex. Note that ACd, PLd, PLv, and IL are regions of ACC/medial PFC, and are innervated by more ventral striatal regions.

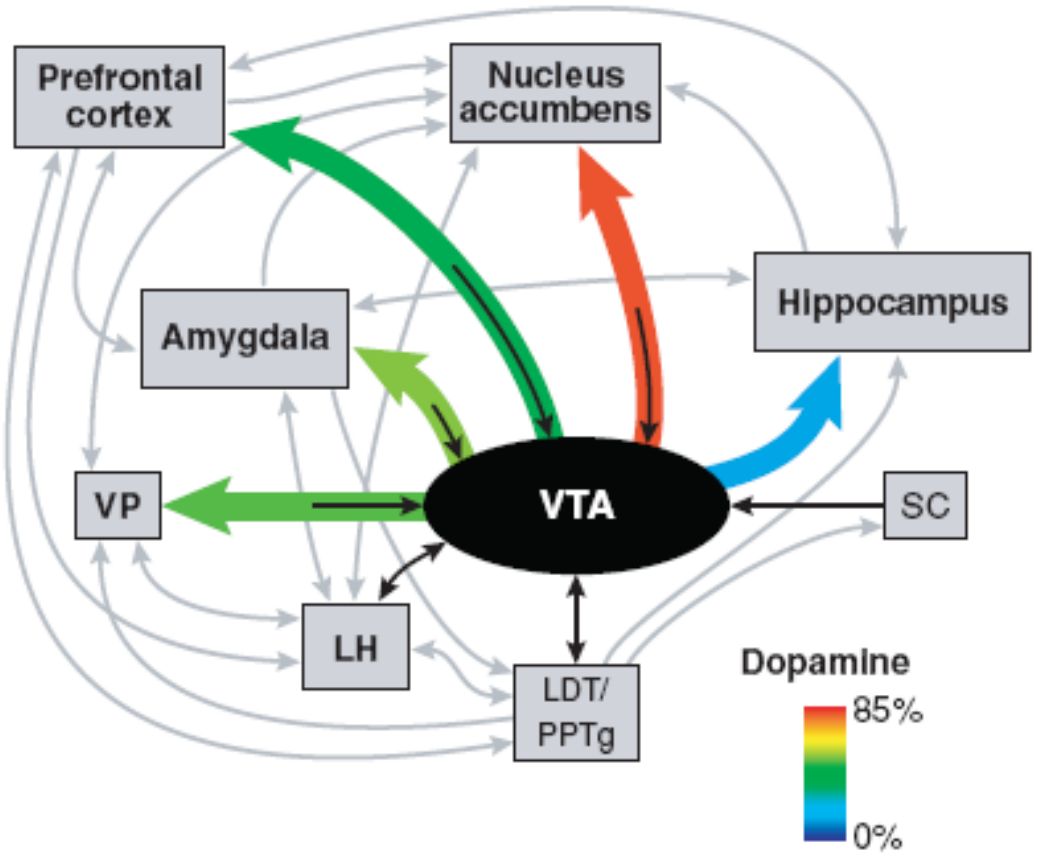


Figure 1.7 Major afferent and efferent pathways of the ventral tegmental area (VTA). Taken from Figure 1 of (Fields et al., 2007). Colored arrows show the percentage of DAergic neurons in the efferent pathways. (VTA also has GABA- and glutamatergic influences.) LDT = laterodorsal tegmental nucleus; LH = lateral hypothalamus; PPTg = pedunculopontine tegmental nucleus; SC = superior colliculus; VP = ventral pallidum.

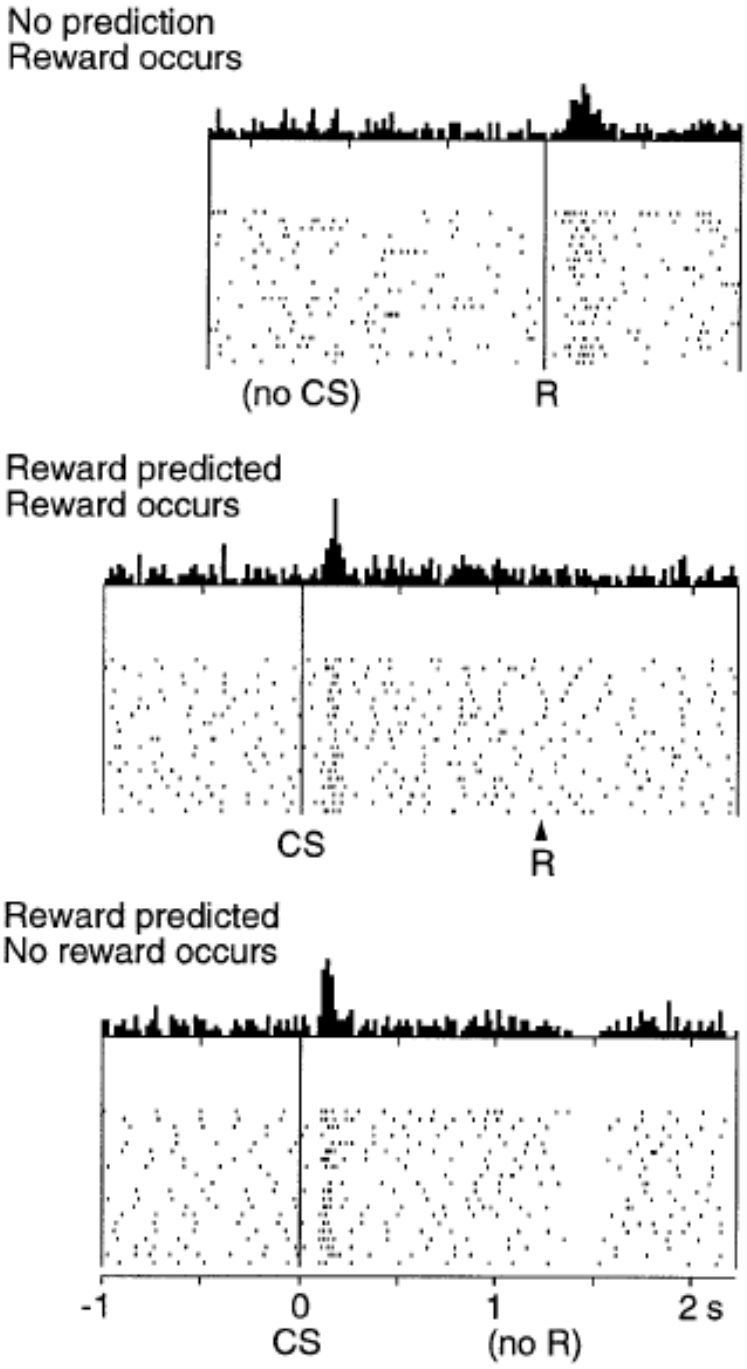


Figure 1.8 Response of DA neurons to reward conditions. Taken from Figure 2 of (Schultz, 1998). (top) DA cells react to unexpected, unpredicted rewards at the time of reward. (middle) When a reward is reliably predicted by a conditioned stimulus, the DA cells burst during the CS, but not during the reward it predicts. (bottom) However, when a reward is predicted, but not delivered, there is a burst at the CS, but a dip at the time that the reward was supposed to be delivered.

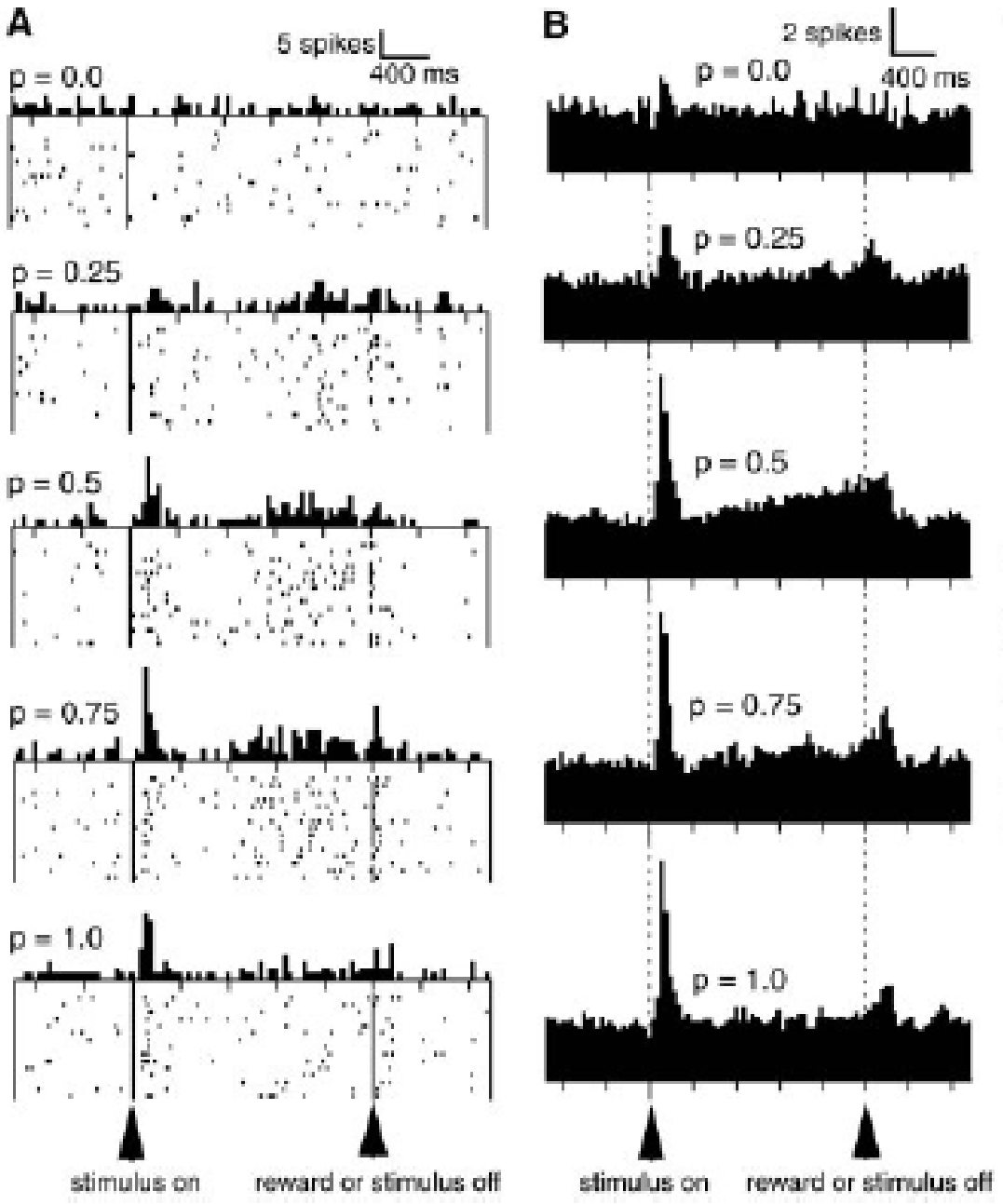


Figure 1.9 Uncertainty-related DA cell firing in response to reward-predictive CSs. Taken from Figure 2 of (Fiorillo et al., 2003). (a) Single cell rasters and histograms of firing related to the probability (p) that a reward is delivered (in Pavolvian fashion) after a cue. E.g. p = 0.75 means a reward is delivered ¾ of the time. (b) Cell population histograms observed under different probability conditions.

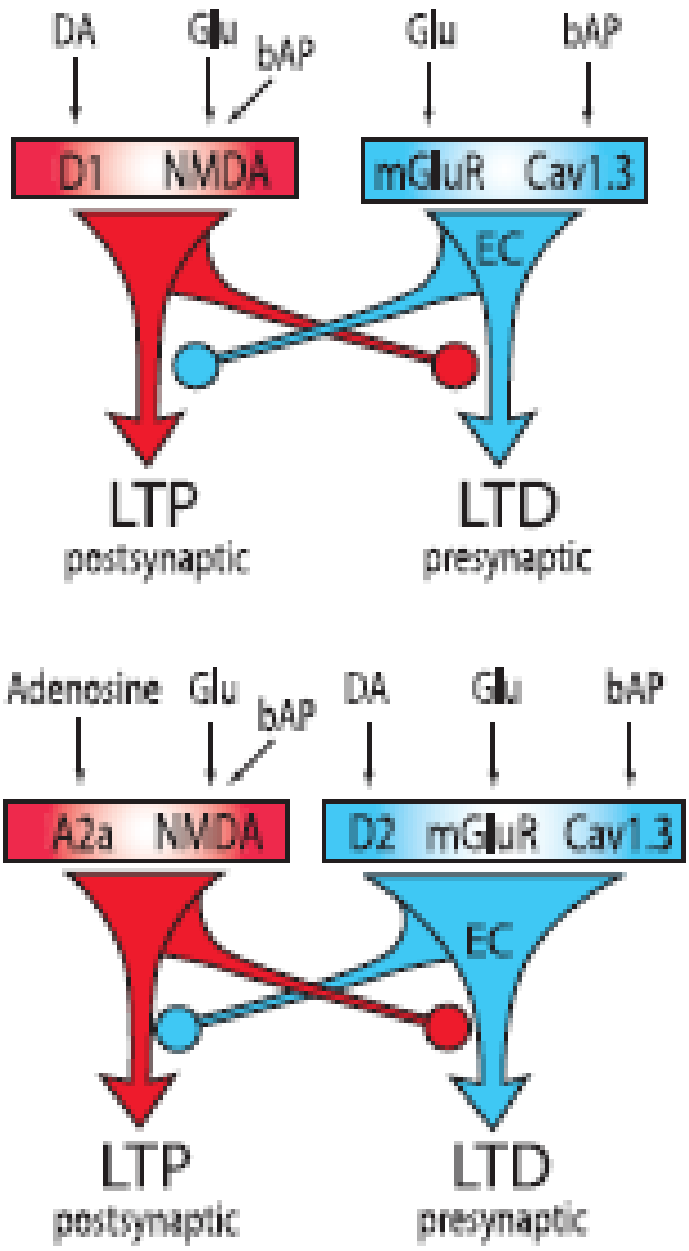


Figure 1.10 Summary of neurotransmitter control of corticostriatal long-term potentiation and depression (LTP and LTD) in Go and NoGo striatal units. Taken from Figures 1 and 2 of (Shen et al., 2008). (top) Go unit (direct pathway) plasticity control. Low DA would be expected to disrupt LTP through hypo-D1 activation. (bottom) NoGo unit (indirect pathway) plasticity control. Low DA would be expected to disrupt LTD through hypo-D2 activation. A2a = adenosine 2A receptor; bAP = backpropogating action potentials; Cav 1.3 = a type of L-type calcium channel; EC = endocannabinoid; Glu = glutamate; mGluR = metabotropic glutamate receptor.

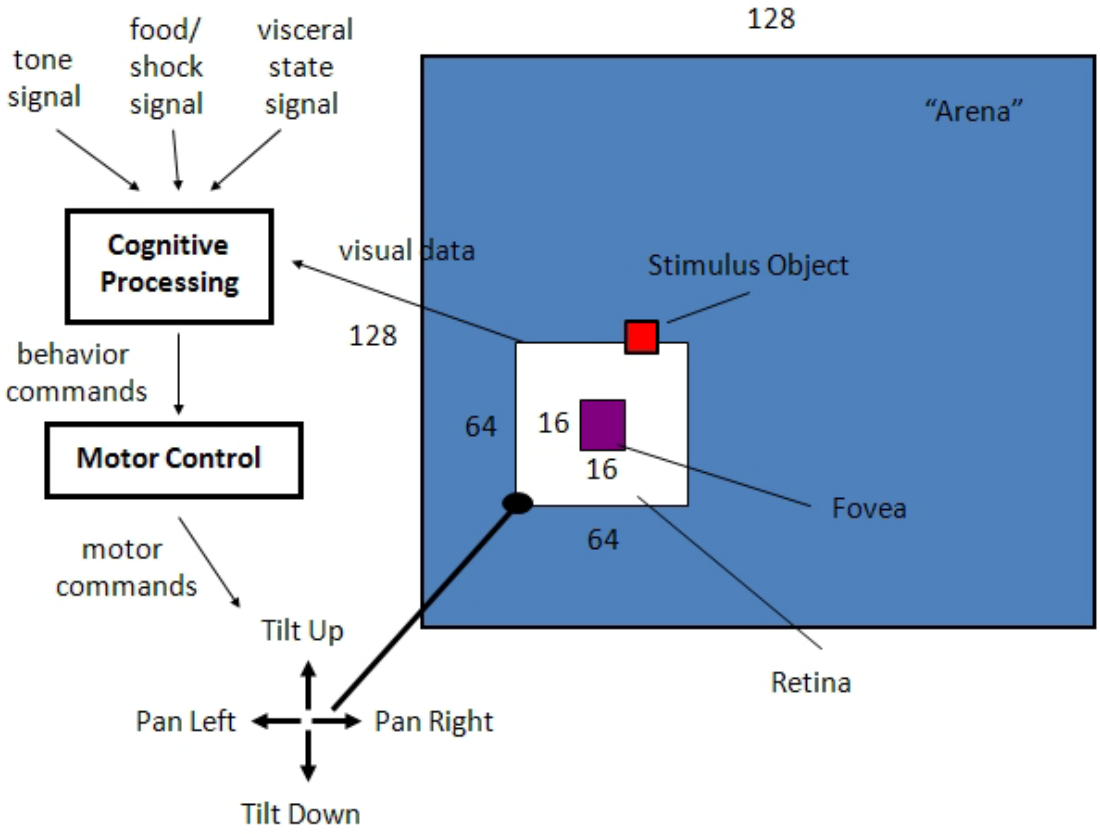


Figure 2.1 Modeled organism and environment. The organism's "body" is a pan-tilt camera moving through a visual (128x128 pixel) "arena". Three behavior commands driving the organism's *motor control* module, Track, Nod, and Shake, allow the organism to track colored stimuli, and make accepting and rejecting responses, respectively, for which the organism may be rewarded for correct and punished for incorrect answers. The *cognitive processing* module takes visual retina, auditory tone, visceral state (hunger/frustration), and reward (food) and punisher (shock) signals and selects the appropriate behavior command. By default, the organism tracks colored squares, placing its (16x16 pixel) color-sensitive fovea over the object. Based on the viewed color of the stimulus and the last remembered auditory tone, the organism is supposed to either Nod or Shake. The modeled organism is capable of learning and relearning mappings of color and tone context to response behavior through operant conditioning using both rewards and punishers.

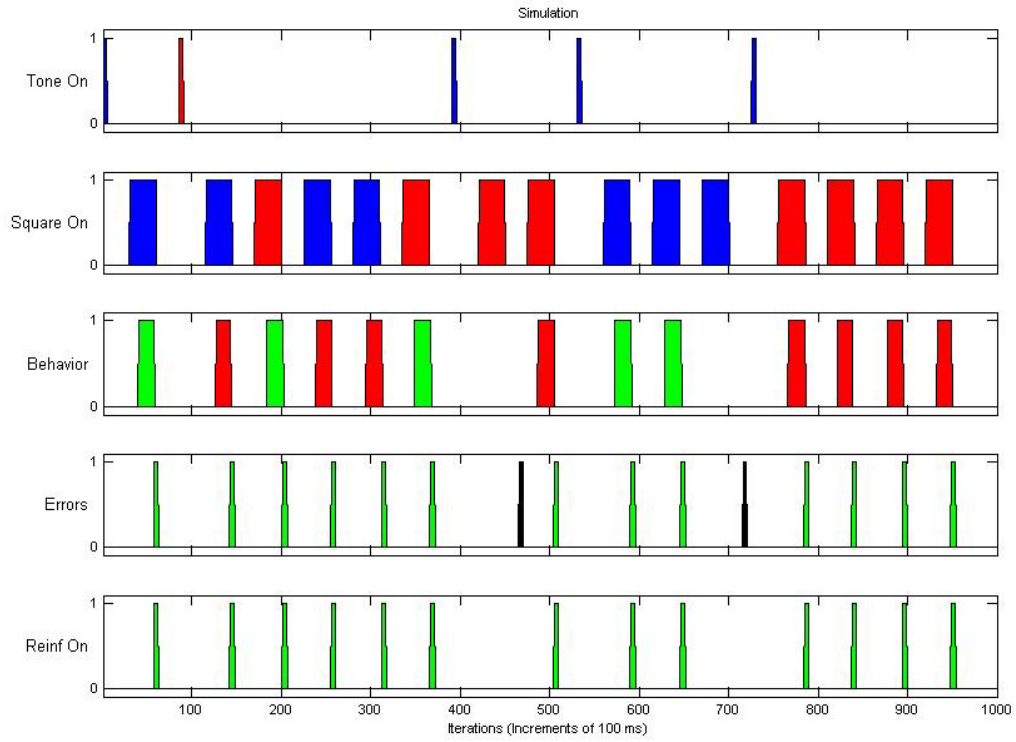


Figure 2.2 Testing stimulation run showing trained full task performance. Tone (Blue=C, Red=Eb) and colored square presentation (Blue, Red), model behavior (Green=Nod, Red=Shake), errors committed (Green=Correct, Black=Miss), and reinforcement delivery (Green=Reward) are shown over all iterations of the 1,000 iteration run. All conjunctions of the task are shown to have been learned (C AND Blue->Nod, C AND Red->Shake, Eb AND Red->Nod, Eb AND Blue->Shake) although there are a couple of missed responses.

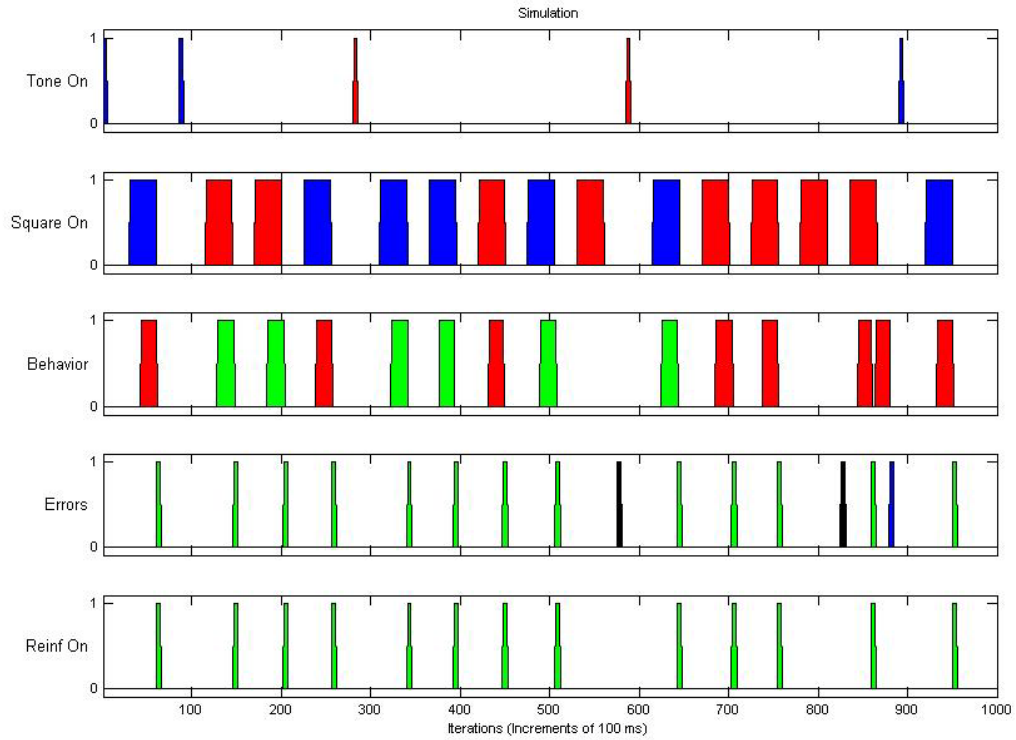
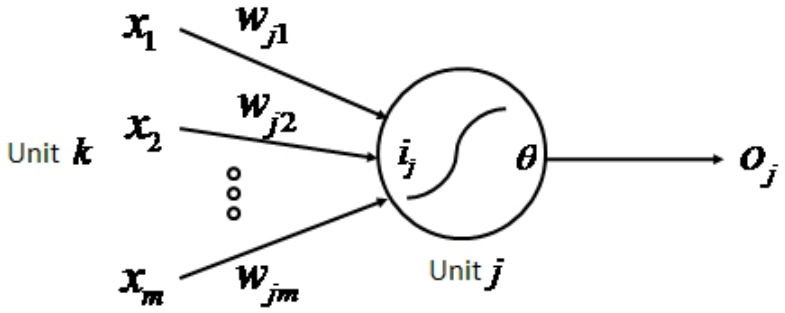


Figure 2.3 Testing stimulation run showing reversal full task performance. Tone (Blue=C, Red=Eb) and colored square presentation (Blue, Red), model behavior (Green=Nod, Red=Shake), errors committed (Green=Correct, Black=Miss, Blue = Repeat), and reinforcement delivery (Green=Reward) are shown over all iterations of the 1,000 iteration run. All conjunctions of the reversal task are shown to have been learned (C AND Blue->Shake, C AND Red->Nod, Eb AND Red->Shake, Eb AND Blue->Nod) although there are a couple of missed responses and a repeated Shake.



$$i_j(t) = \sum_k w_{jk} x_k(t-1) \qquad \theta(x) = \frac{1}{1 + \exp(-\beta(x - \tau))}$$

$$o_j(t) = o_j(t-1) + \Delta[\theta(i_j(t) + N(0,n)) - o_j(t-1)]$$

Figure 2.4 Neural unit architecture. Units are mean firing-rate population units with output values $o_j(t)$ ranging between 0 and 1. Each unit j takes a weighted sum $i_j(t)$ of the outputs of the previous time step. The current output $o_j(t)$ is set to the previous output plus the Δ-weighted difference between the sigmoid-squashed net input $i_j(t)$ plus Gaussian noise of standard deviation n, and the previous output $o_j(t-1)$. The sigmoidal squashing function $\theta(x)$ has a range between 0 and 1, a gain determined by β and a threshold (with respect to x) determined by τ. Over time, effectively the output of the unit tracks the sigmoided net input.

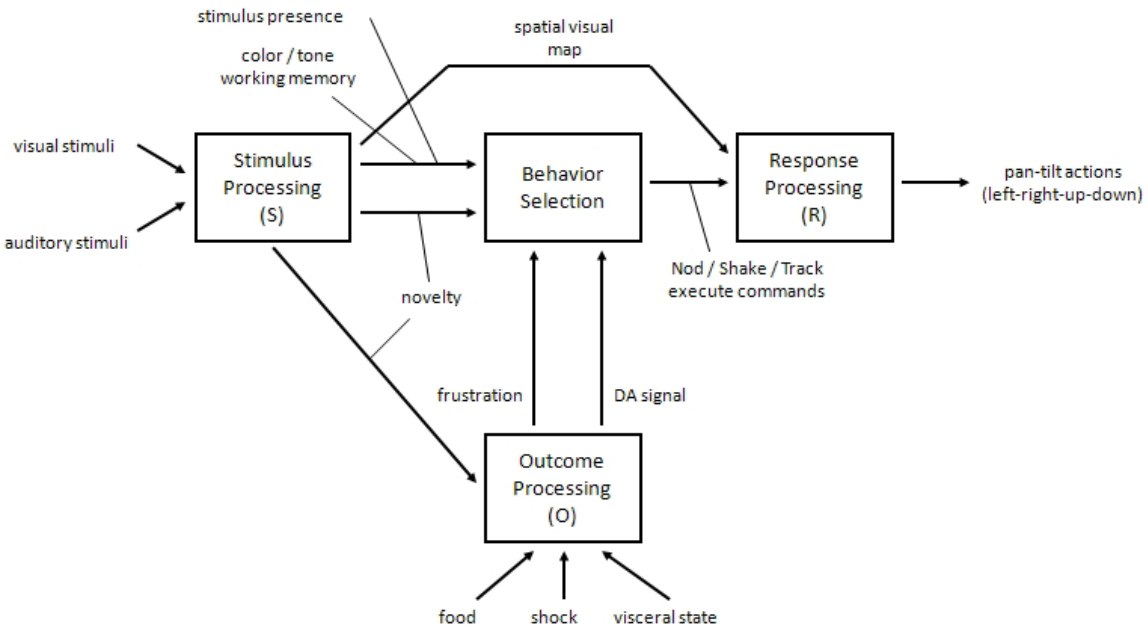


Figure 2.5 Functional overview of model. The Response Processing (R) block is essentially the Motor Control module from Figure 2.1, receiving behavior commands (Nod, Shake, Track), and current spatial visual map information, and outputting the proper immediate pan/tilt action/s. The other blocks constitute the Cognitive Processing module of Figure 2.1. The Stimulus Processing (S) block takes immediate visual and auditory stimuli, and outputs a signal for the presence of a target stimulus, spatial visual map information for guiding tracking behaviors, a signal for novelty detection, and working memory for colors and tones. The Outcome Processing (O) block takes stimulus inputs involving primary rewards and punishers, novelty, and visceral state, and outputs an analog frustration signal and a dopamine signal. The Behavior Selection block takes stimulus and outcome information from the S and O blocks, respectively, and outputs behavior commands to the R block.

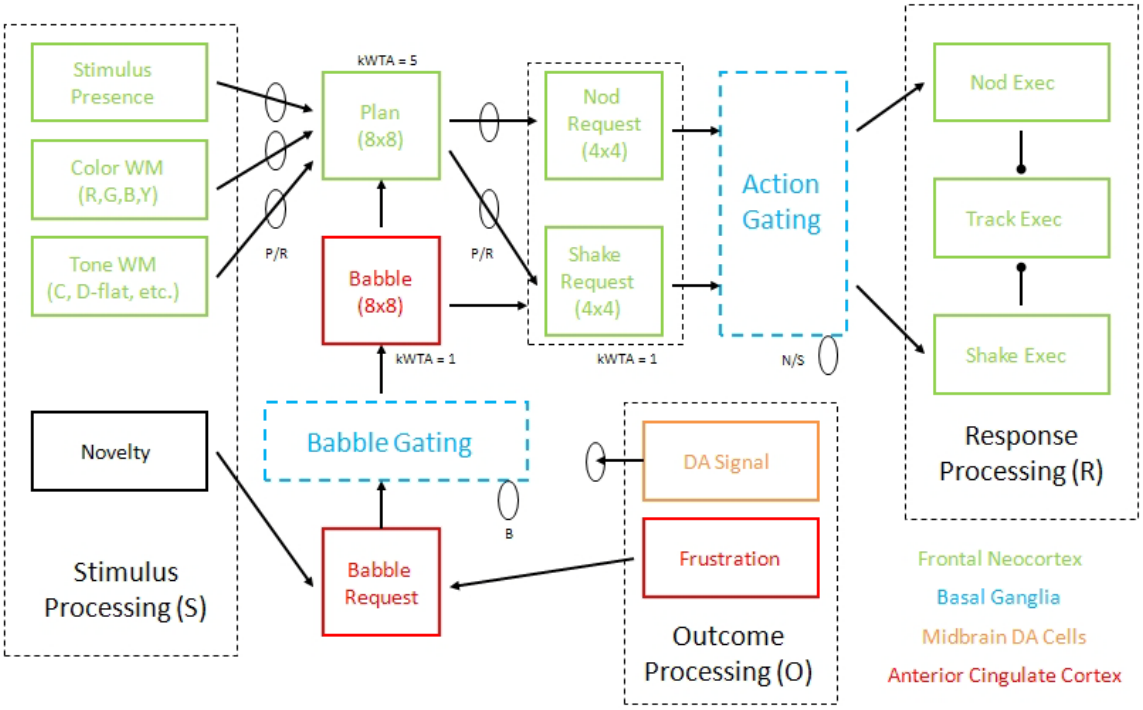


Figure 2.6 Layer architecture and anatomical correspondences of Behavior Selection and surrounding modules. Plan units take (visual) stimulus presence and color and tone working memory input from the S block and output to the Request units. Request units issue a request for Nod and/or Shake behavior/s via the basal ganglia Action Gating module and this gates which behavior is made active in the R block. All areas mentioned so far except the Action Gating module are mediated by frontal neocortical areas. ACC-mediated frustration activity and/or (probably medial temporal area) novelty detection trigger an ACC-mediated request, projecting through the basal ganglia Babble Gating module, to activate an ACC Babble unit randomly. The Babble activity, in turn, activates Plan and Request unit activity downstream which leads to random behavior selection. The midbrain DA cells control learning in the Plan and Request units and the Gating modules, and also activity levels in the Gating modules.

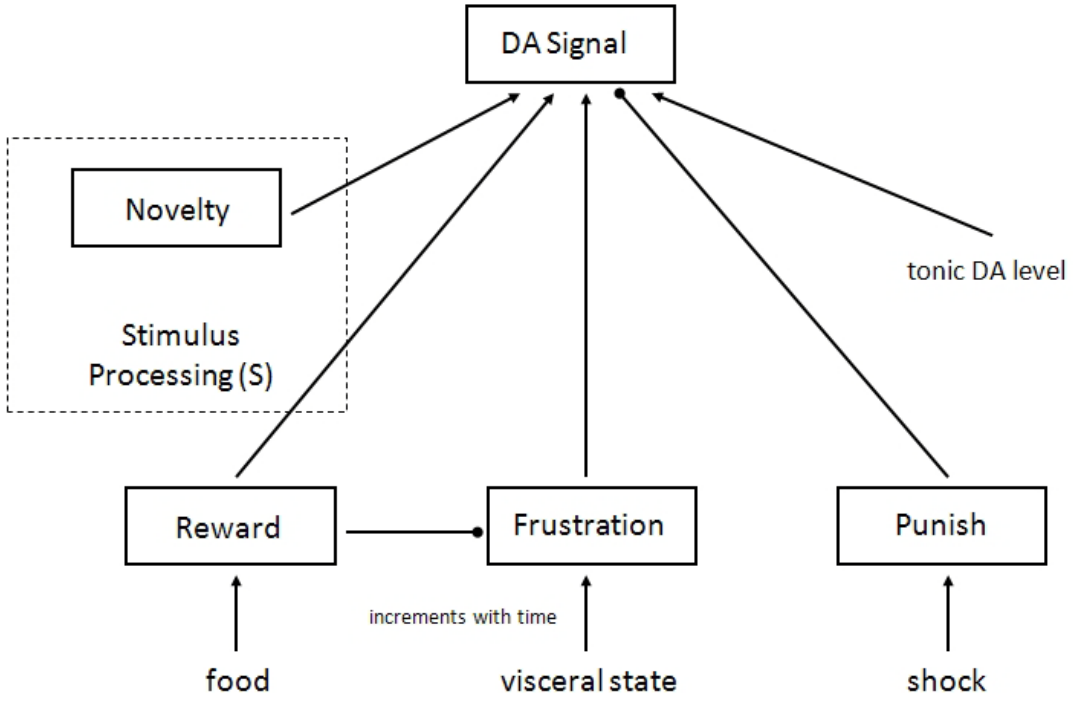


Figure 2.7 Details of outcome processing module. A Frustration unit increments with increasing hunger signaled by the visceral state, but this ramping-up is reset by reward delivery. The DA "activity-oriented motivation" signal, driven by a tonic bias signal, is excited by the Frustration unit, and phasically, by novelty and by (food) rewards. (Shock) punishers, on the other hand, phasically inhibit the DA signal. The Frustration and DA signals are the outputs of the O block.

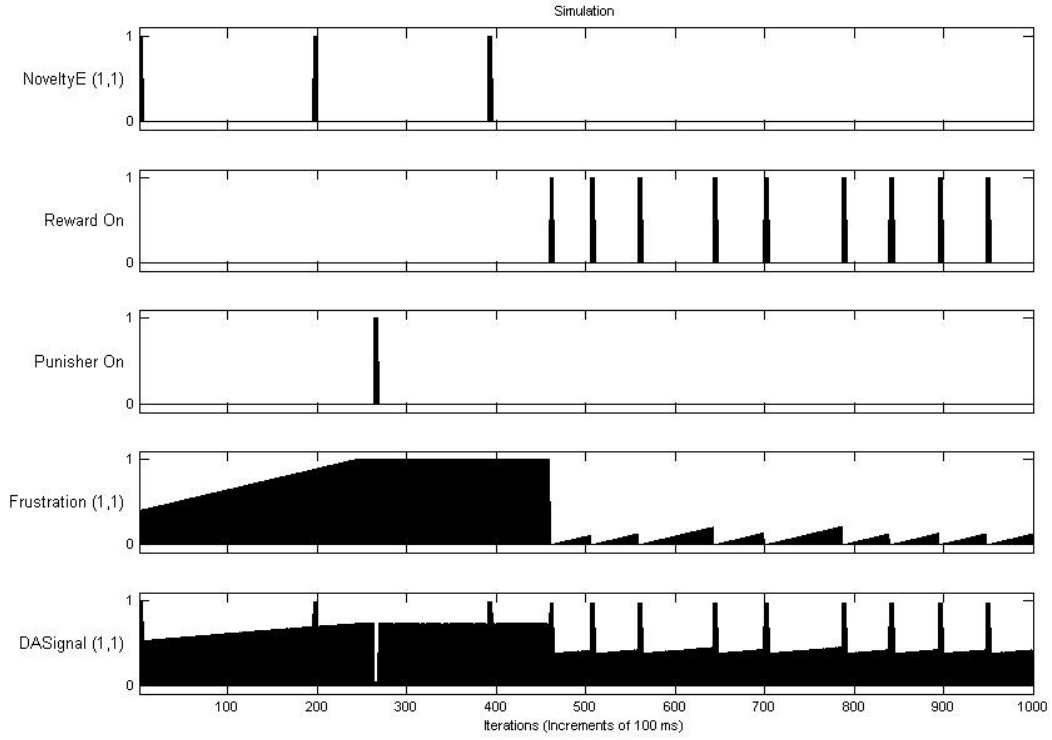


Figure 2.8 Frustration and DA signal unit dynamics. Frustration activity ramps up by default over time (due to increasing hunger or desire to be rewarded) and is reset by reward deliveries. DA signal activity is driven by the Frustration unit activity. It is also highly phasically excited by rewards or occurrences of novelty, and phasically inhibited by punishers. Activity is modified with respect to a basal level of DA signal activity.

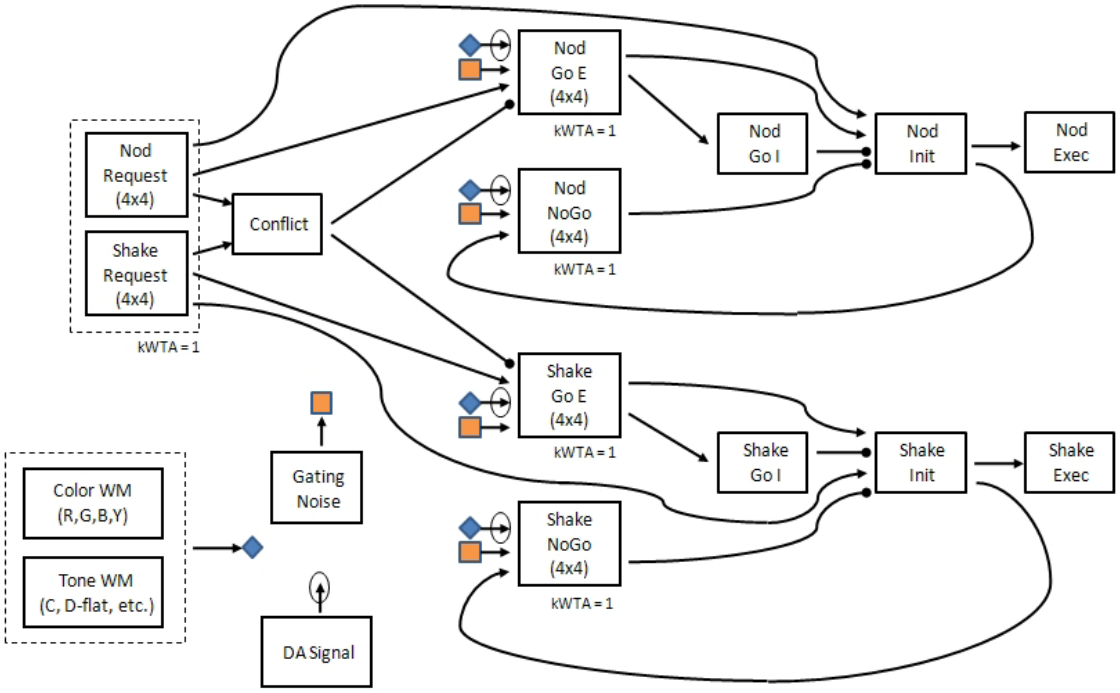


Figure 2.9 Action gating module. Activity from Nod and Shake Request units is gated in such a way that one or none of the corresponding Init units are pushed over the threshold needed to activate an Exec unit (which then would trigger the actual behavior). Both Go (behavior-gating) and NoGo (behavior-vetoing) units are innervated by color and tone working memory (through learnable weights) and by gating noise that is applied during a visual stimulus' presence. Go units are excited by Request activity, except when both Nod and Shake Request are activated triggering inhibitory Conflict unit activity. Init unit activity requires both Request and GoE unit activity, and the absence of NoGo activity. Init activity, in addition to triggering Exec activity, also triggers NoGo activity. Go and NoGo unit activity have their thresholds modified by DA Signal activity with high DA lowering Go unit thresholds and raising NoGo thresholds, and low DA having the opposite effect. Go units learn to fire in response to working memory activity associated with rewards, whereas NoGo units learn to fire in response to working memory activity associated with punishers. Essentially, the system is a dynamic reinforcement learning gating mechanism that learns to admit or reject behavior according to the consequences of admitting the behavior.

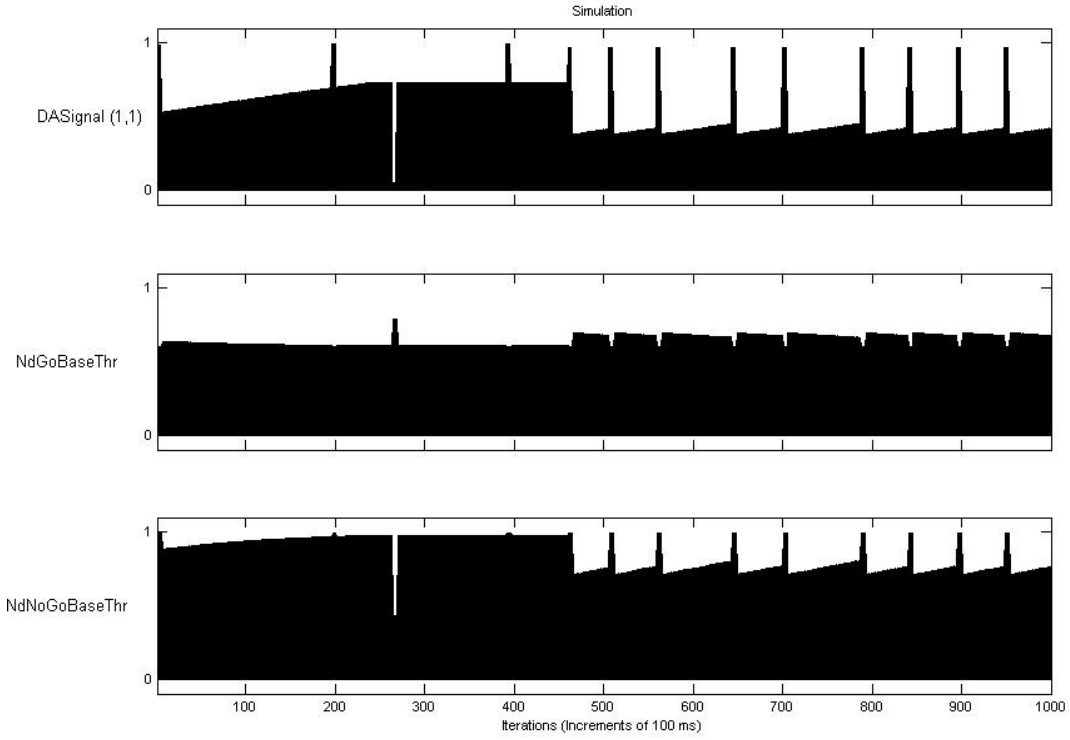


Figure 2.10 Effects of DA signal activity on Go and NoGo τ thresholds. Rising DA levels lower the thresholds of Go units and raise the thresholds of NoGo units. Phasic DA spikes (triggered by novelty or rewards) lead to dips in the minimum threshold for Go units, and spikes in the threshold for the NoGo units. Phasic DA dips (triggered by punishers) lead to significant threshold spikes for Go units, and deep dips for NoGo units. Generally, DA levels encourage Go unit activity, but discourage NoGo activity.

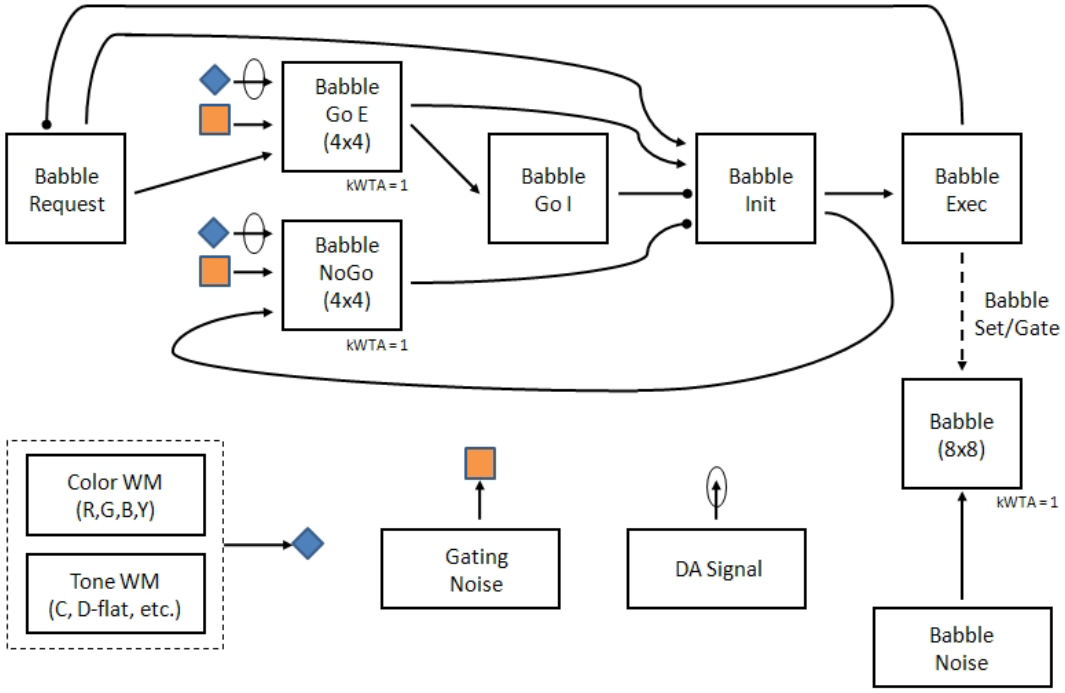


Figure 2.11 Babble gating module. The mechanism is very similar to that of the Action Gating module (shown in Figure 2.9). Babble Request activity is gated by the Go and NoGo units to allow or veto Init unit activity which may then trigger Babble Exec unit activity. Babble Exec activity, in conjunction with random Babble Noise activity, triggers a single random Babble unit be activated. Go units learn to fire in response to working memory activity associated with rewards, whereas NoGo units learn to fire in response to working memory activity associated with punishers.

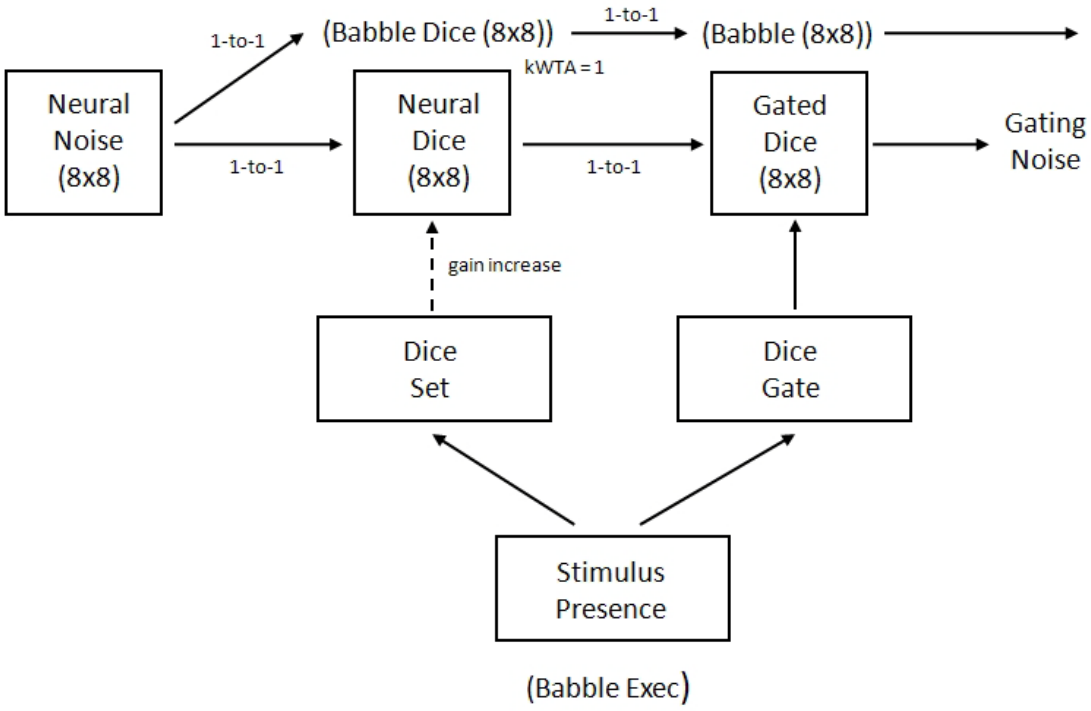


Figure 2.12 Neural noise generation mechanism. The Neural Noise units all generate noisy outputs which, in a one-to-one mapping, excite Neural Dice and Babble Dice units. The Dice units, by default, have low output-function gain which causes their output to mirror the Neural Noise units when the Dice Set unit is off, but causes the kWTA mechanism to lead to only one unit being active when the Dice Set unit is on, through an increase in the output gain. Gated Dice and Babble units are activated by a combination of the one-to-one activation from the Dice layers and the activity of the Dice Gate unit. The Stimulus Presence unit activates both the Dice Set and Dice Gate units simultaneously when a colored square is present. The net result is that the Gated Dice and Babble units are all off when no visual stimulus is present, but when one is presented, each of these areas generally has a single random unit among the 8x8 active. This random activation is a source of randomness for the Action and Babble Gating mechanisms and for Plan and Request unit activity during behavior babbles.

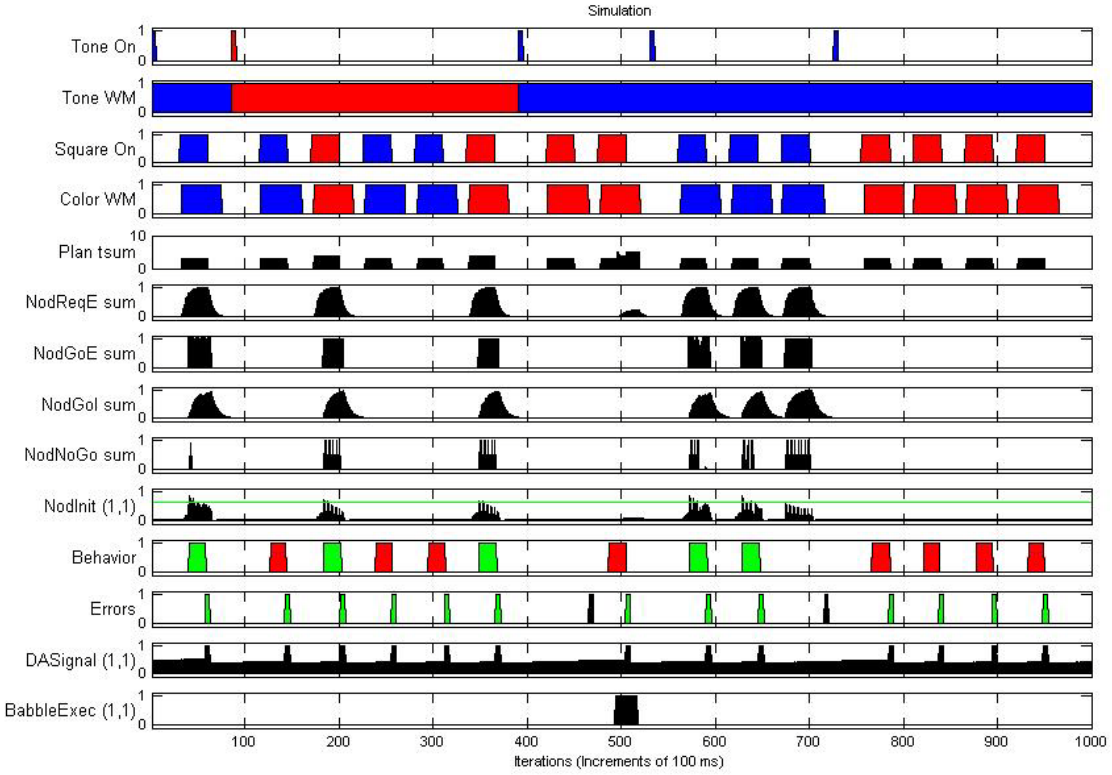


Figure 2.13 Trained full task Nod activation. The simulation run is the same as that of Figure 2.2. The color and tone (Blue=C, Red=Eb) working memory activity are shown, and these (along with the Stimulus Presence unit, not shown) drive Plan unit activity so that around 3 Plan units are made active. During C/Blue or Eb/Red conjunctions, Nod Request units are activated, which in turn activate Nod Go E units. In all cases but one here, the Nod Init unit is driven briefly over threshold, which leads to a Nod behavior. The Nod NoGo layer is also activated during responses which would permit learning in the NoGo pathway.

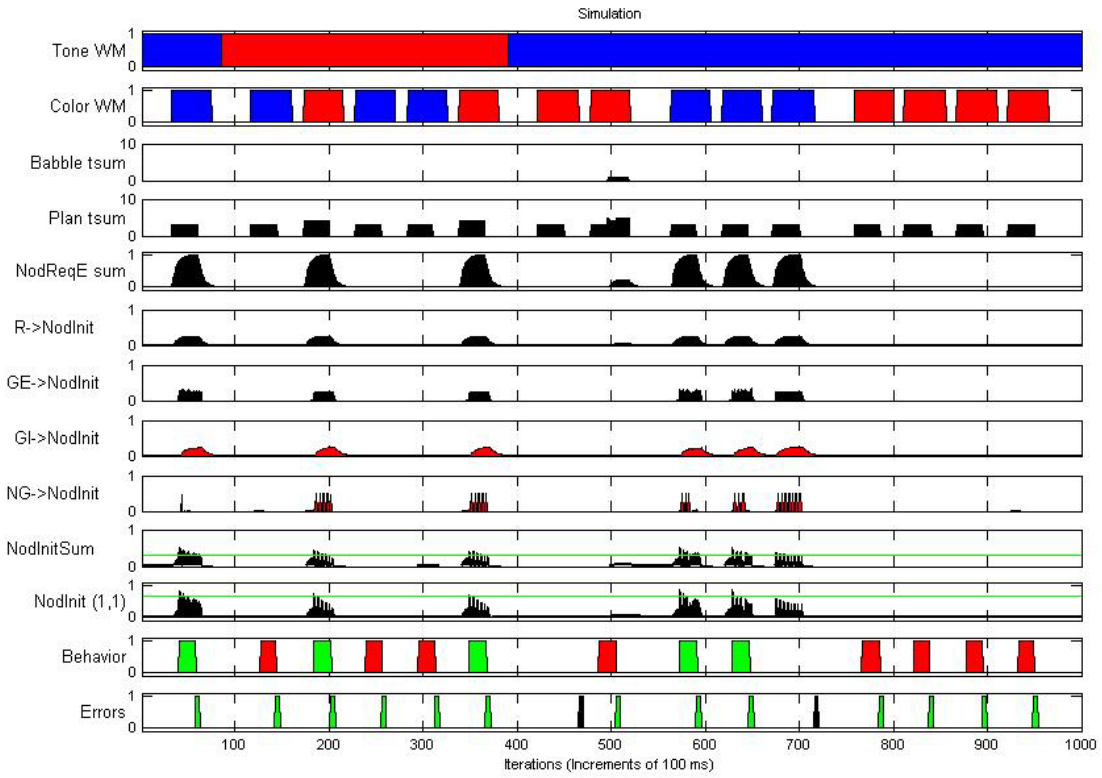
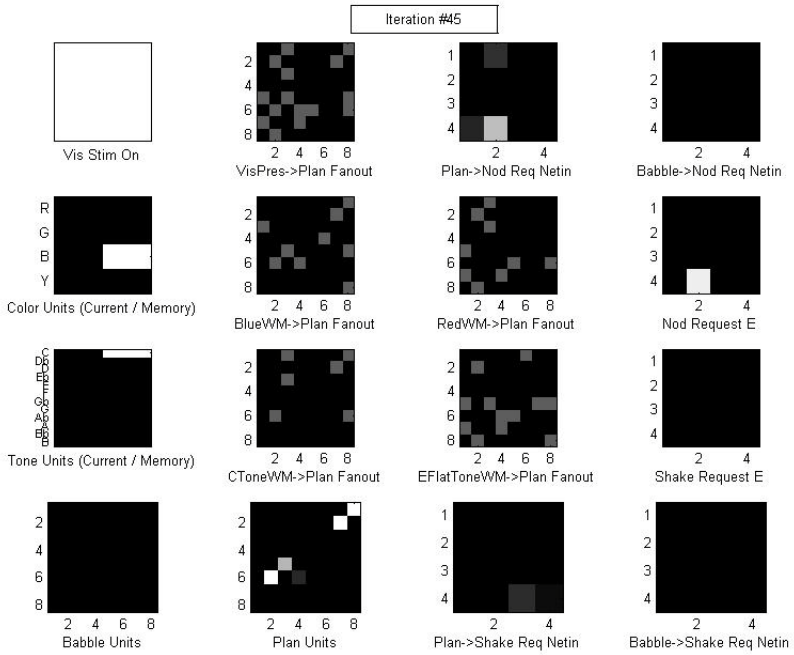
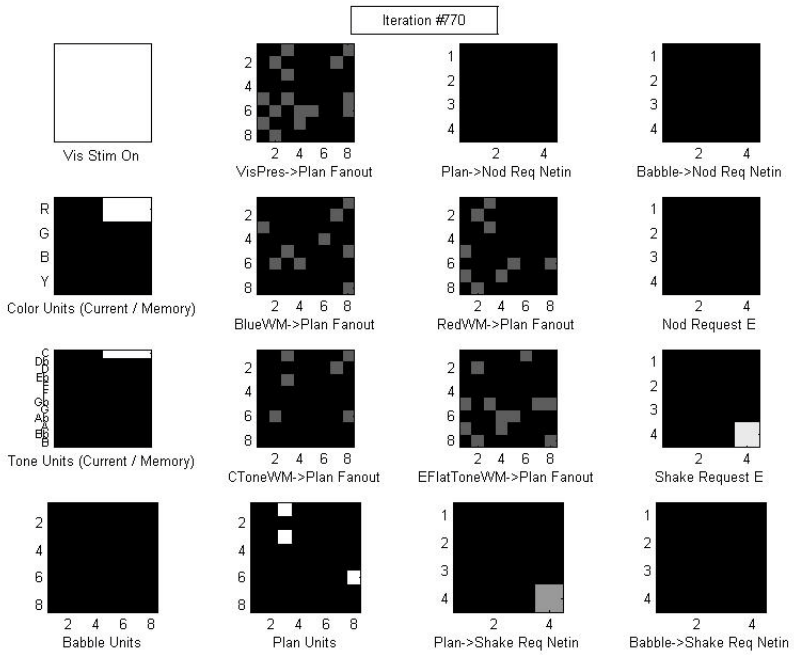


Figure 2.14 Trained full task Nod Init activation. The simulation run is the same as that of Figure 2.2. As in Figure 2.13, a Nod Request unit is activated during C/Blue and Eb/Red conjunctions. Starting with the sixth trace, the contribution of the different afferents to the Nod Init unit (Request, Go E, Go I, and NoGo units) are shown (Black=excitatory, Red=inhibitory). The total sum of Nod Init unit input activation is shown next, followed by the Nod Init output activity with the Nod Exec threshold shown in that trace in Green.



(a)



(b)

Figure 2.15 Plan/Request unit activation in response to working memory conjunctions. (a) C AND Blue. (b) C AND Red. See text for explanation.

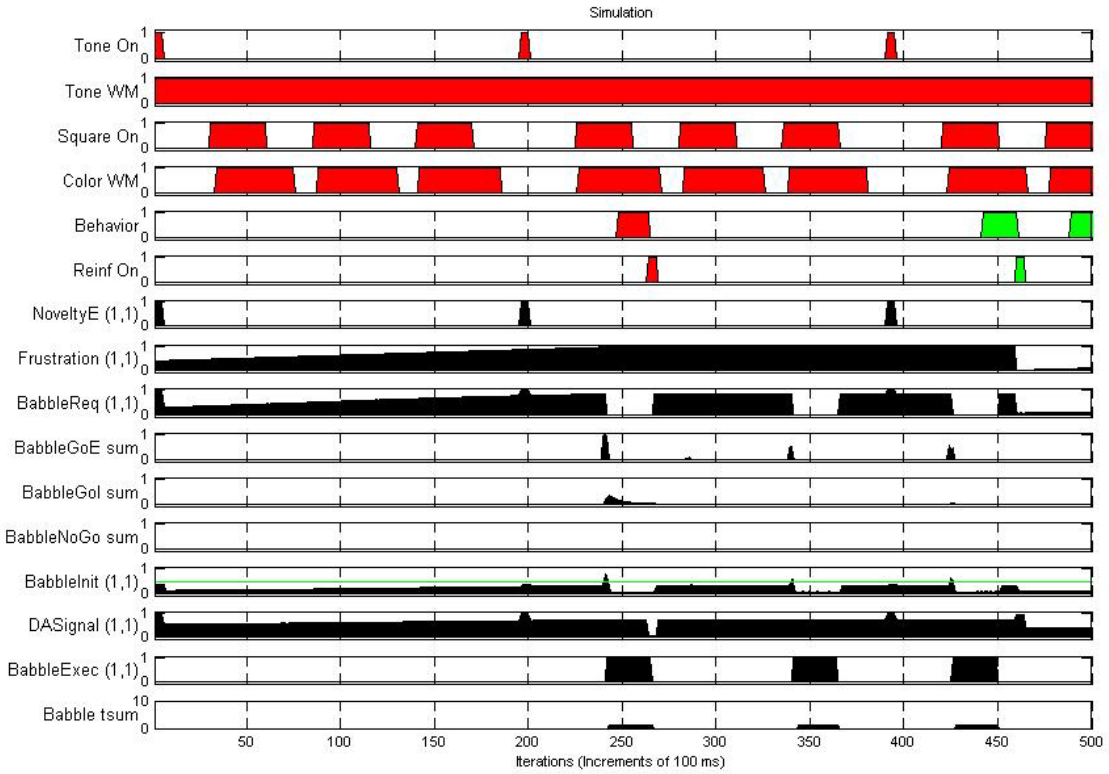
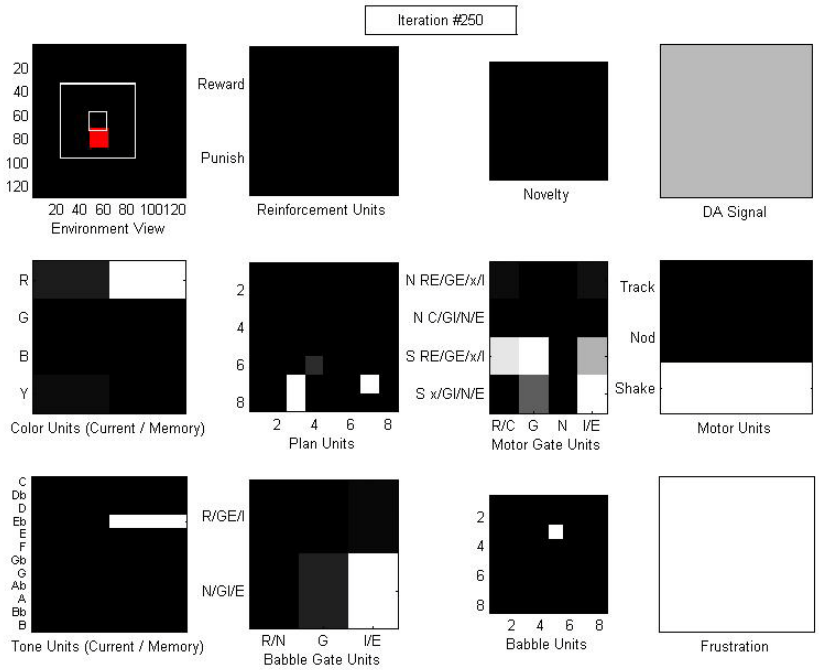
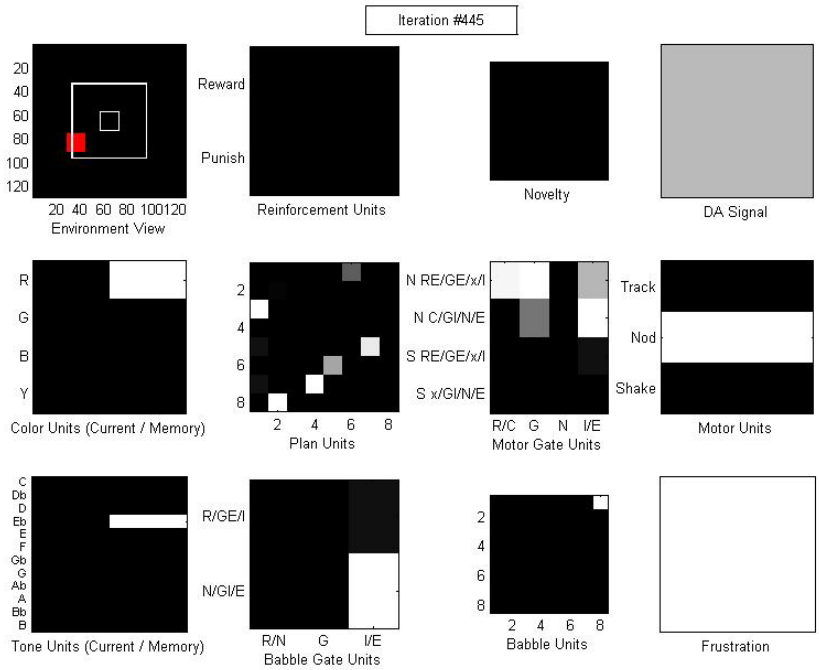


Figure 2.16 Full task Babble activation. The plot shows the first half of run 4 of the full task training phase, in which the system is learning the Eb/Red conjunction for the first time. During the three Eb tone presentations, the Novelty unit is engaged. Frustration is gradually increasing as the model is not rewarded for its inaction. Babble Request activity sums Novelty and Frustration activity, except during babbles. The Babble Go unit is not activated by the Request unit until the Frustration level rises to a certain level. Only then does the Babble Init unit activate over the Babble Exec threshold (shown in green) and trigger a babble. Three babbles are necessary before the correct behavior is learned. The first babble chooses the wrong behavior (Shake), the second fails to trigger either behavior, and the last triggers the correct Nod.



(a)



(b)

Figure 2.17 Babble activation. The same run as in Figure 2.16 is shown. (a) The first (incorrect) babble at Iteration 250. (b) The last (correct) babble at iteration 445.

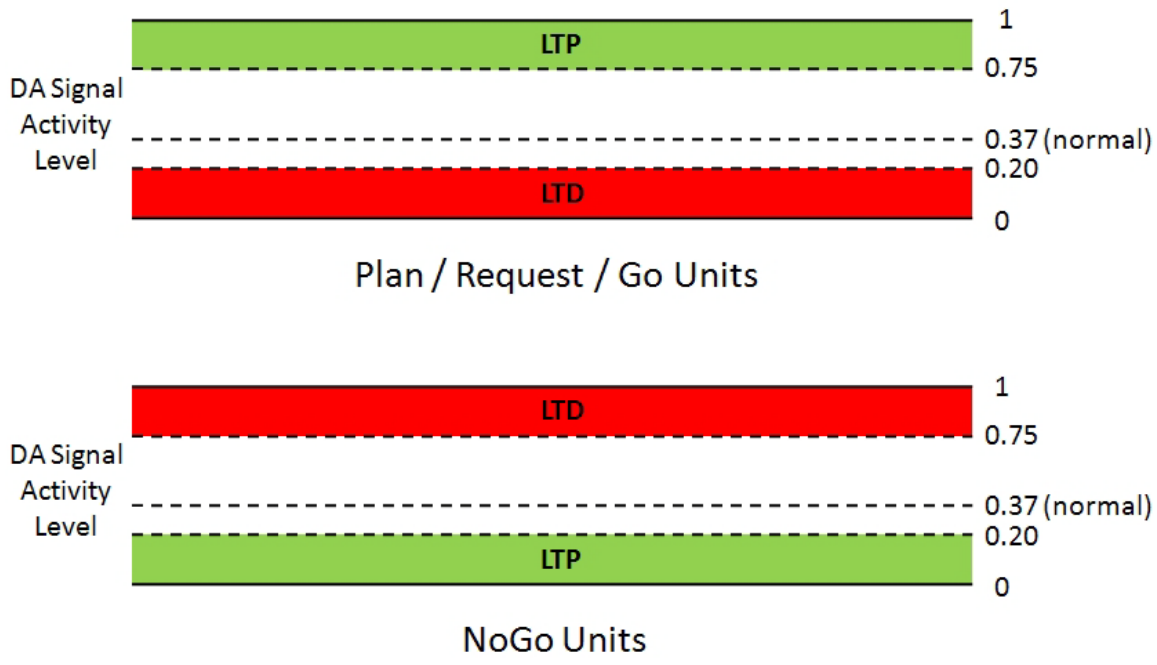
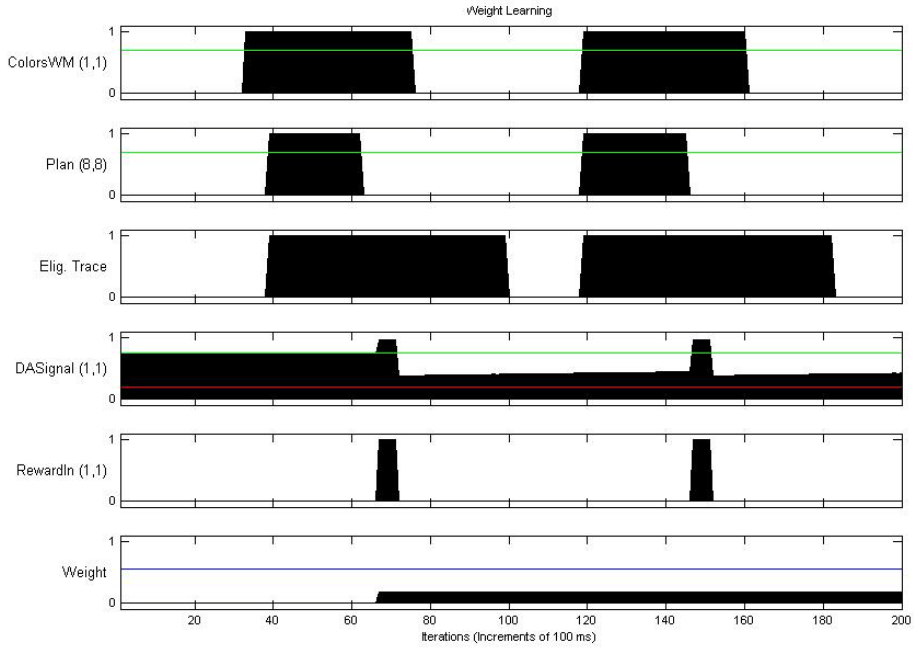
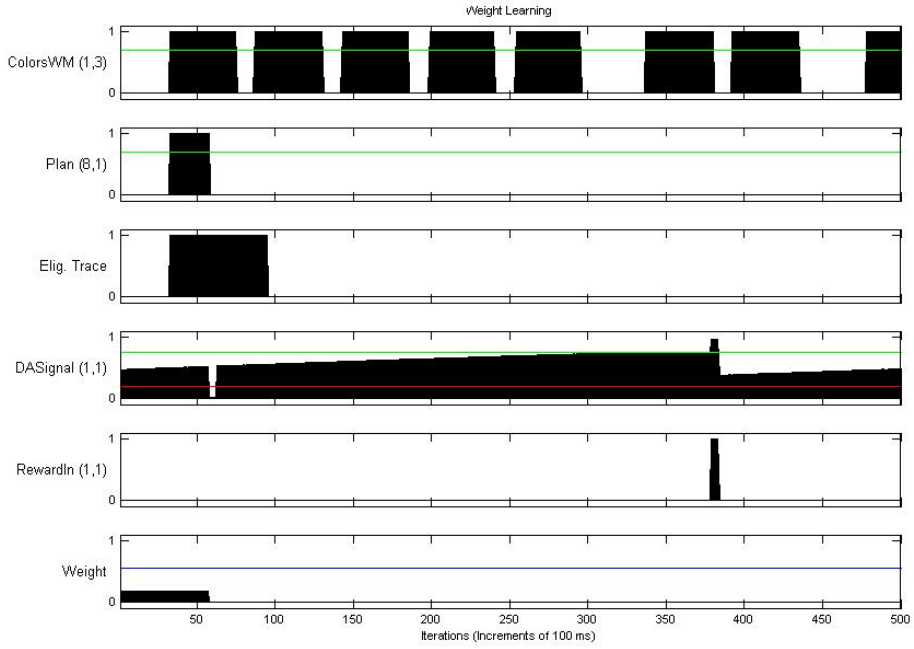


Figure 2.18 DA learning thresholds. (top) Thresholds for the Plan, Request, and Go units. If the DA level rises over the 0.75 activation threshold, LTP is triggered in the target weights during synapse eligibility. If the DA level falls below the 0.20 activation threshold, LTD is triggered in eligible target weights. (bottom) Thresholds for the NoGo units. These are identical with those of the other units, except the LTP and LTD regions are reversed.



(a)



(b)

Figure 2.19 LTP and LTD examples. (a) High-DA LTP example. (b) Low-DA LTD example.

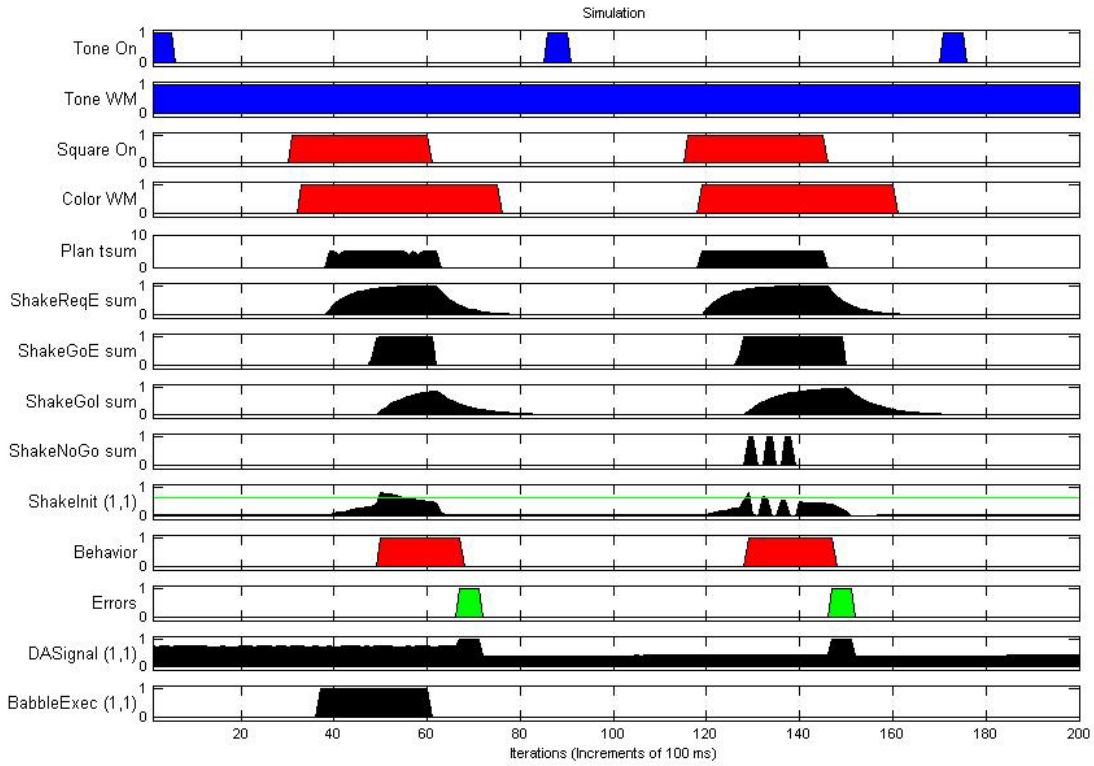
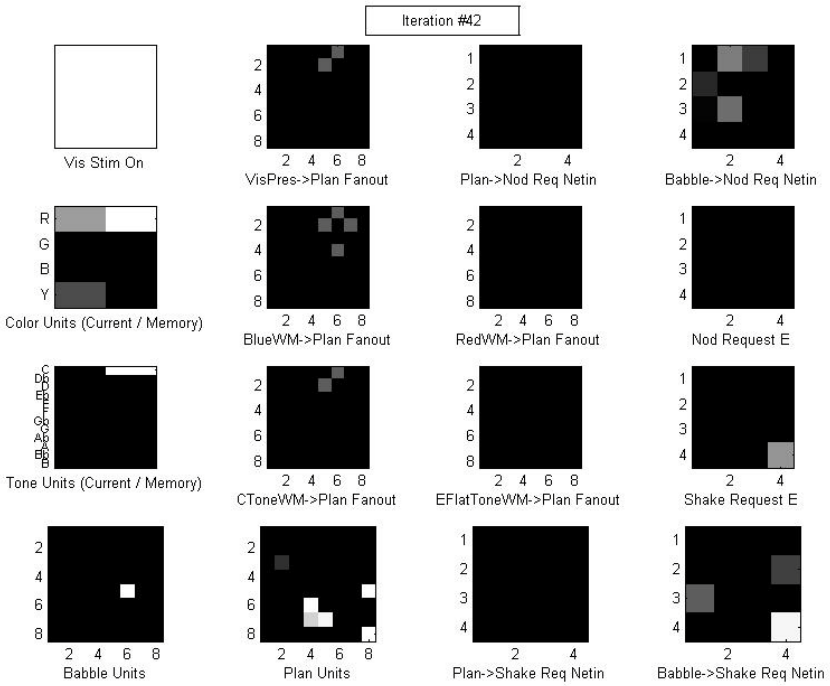
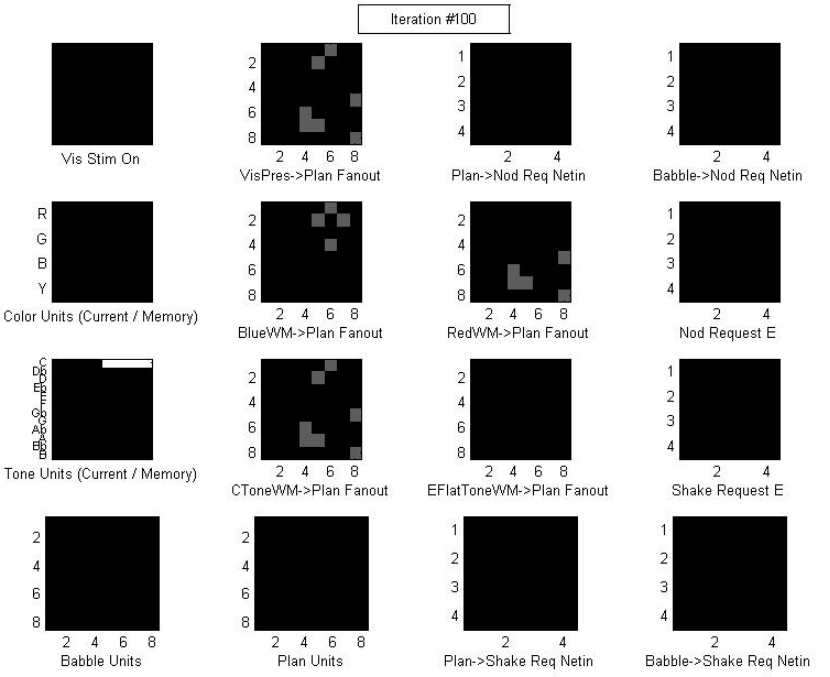


Figure 2.20 Example of conjunction learning. C AND Red -> Shake is successfully learned during the first red square presentation, and performed correctly on the second.



(a)



(b)

Figure 2.21 Weight change during conjunction learning. Simulation run is the same as Figure 2.20. (a) Plan weights during correct babble. (b) Plan weights after reward.

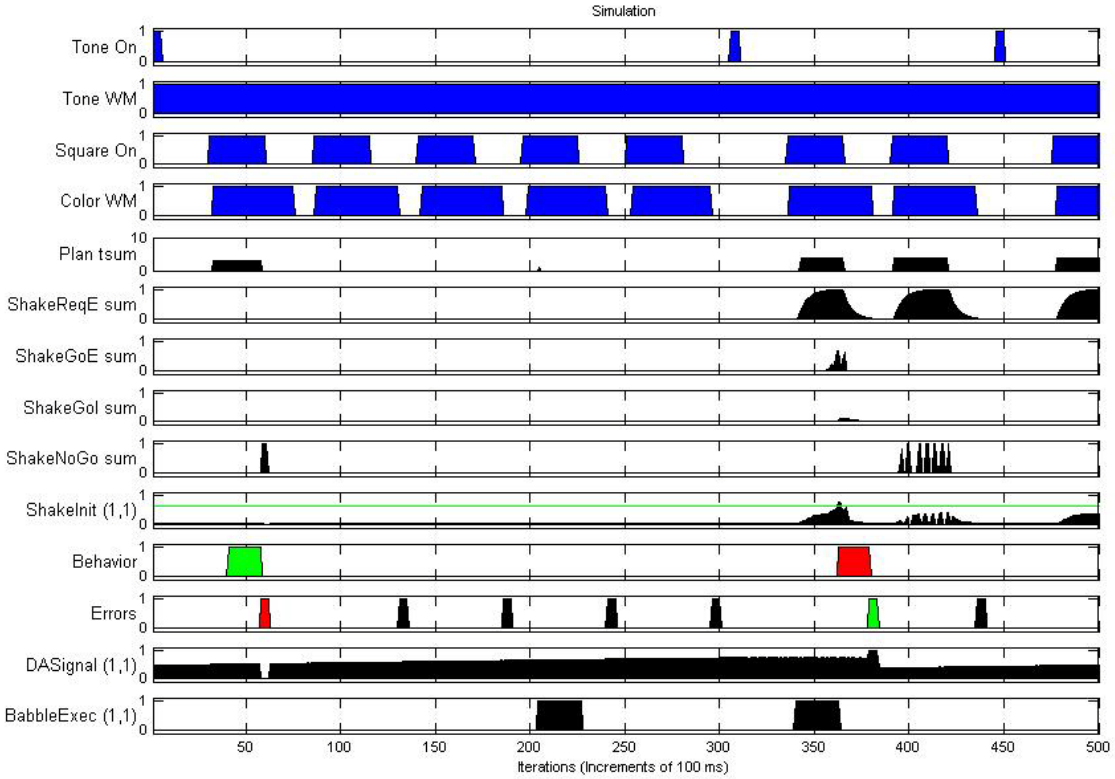
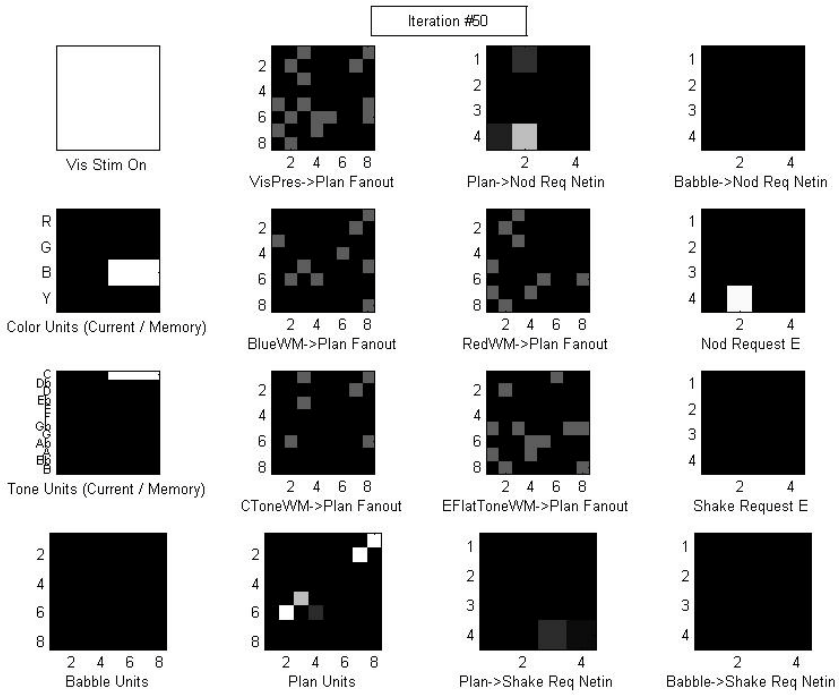
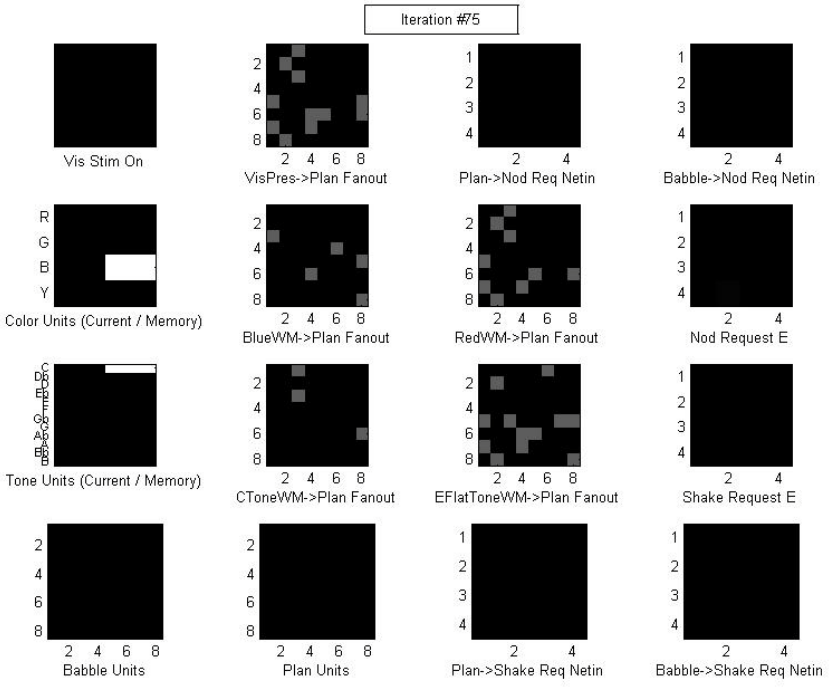


Figure 2.22 Example of conjunction unlearning and relearning. At the beginning of run 1 of the reversal simulation phase, C AND Blue -> Nod is first unlearned through punishment on the first presentation. On the sixth square presentation (and the second babble), the correct (new) behavior, Shake, is learned so that the new conjunction is C AND Blue -> Shake.

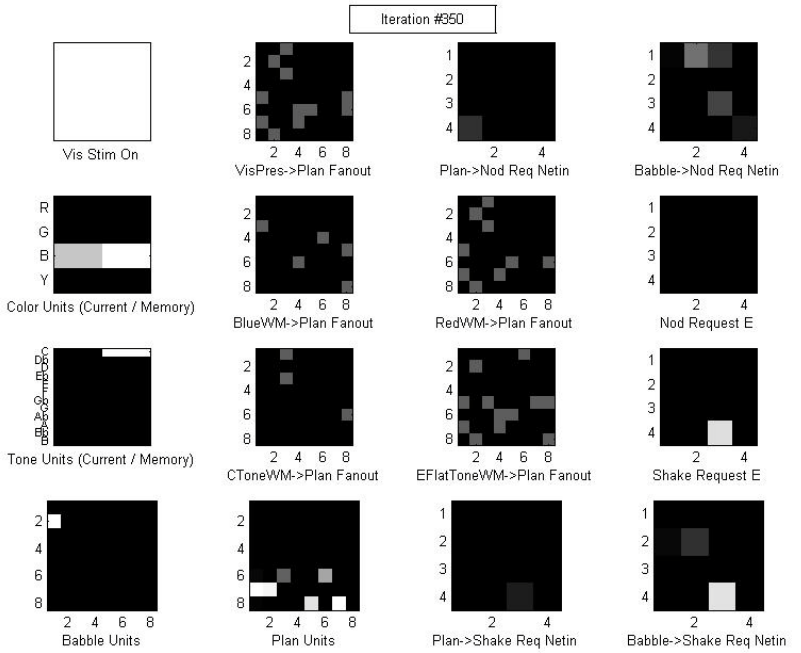


(a)

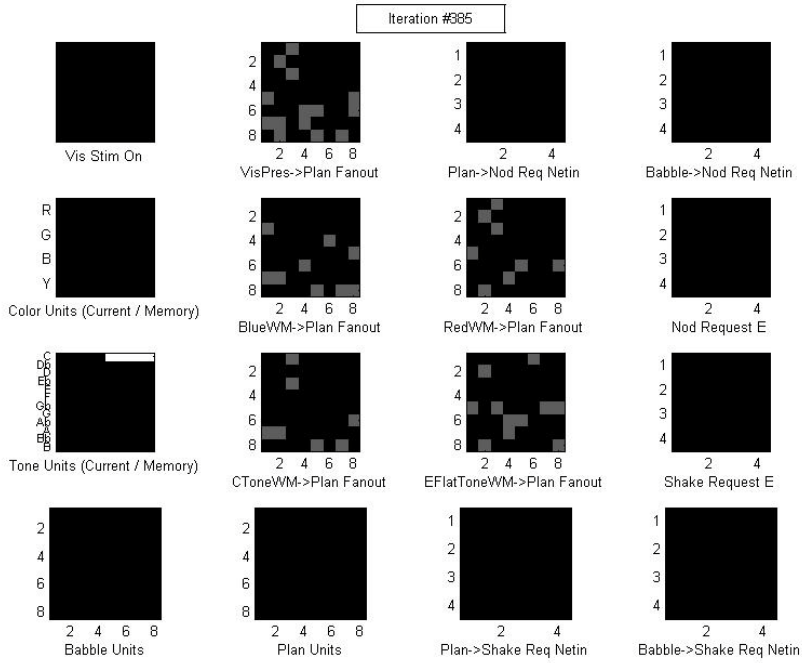


(b)

Figure 2.23 Plan weight change during conjunction unlearning. (a) Plan weights during first response. (b) Plan weights after punishment.



(a)



(b)

Figure 2.24 Plan weight change during conjunction relearning. (a) Plan weights during the correct response babble. (b) Plan weights after reward.

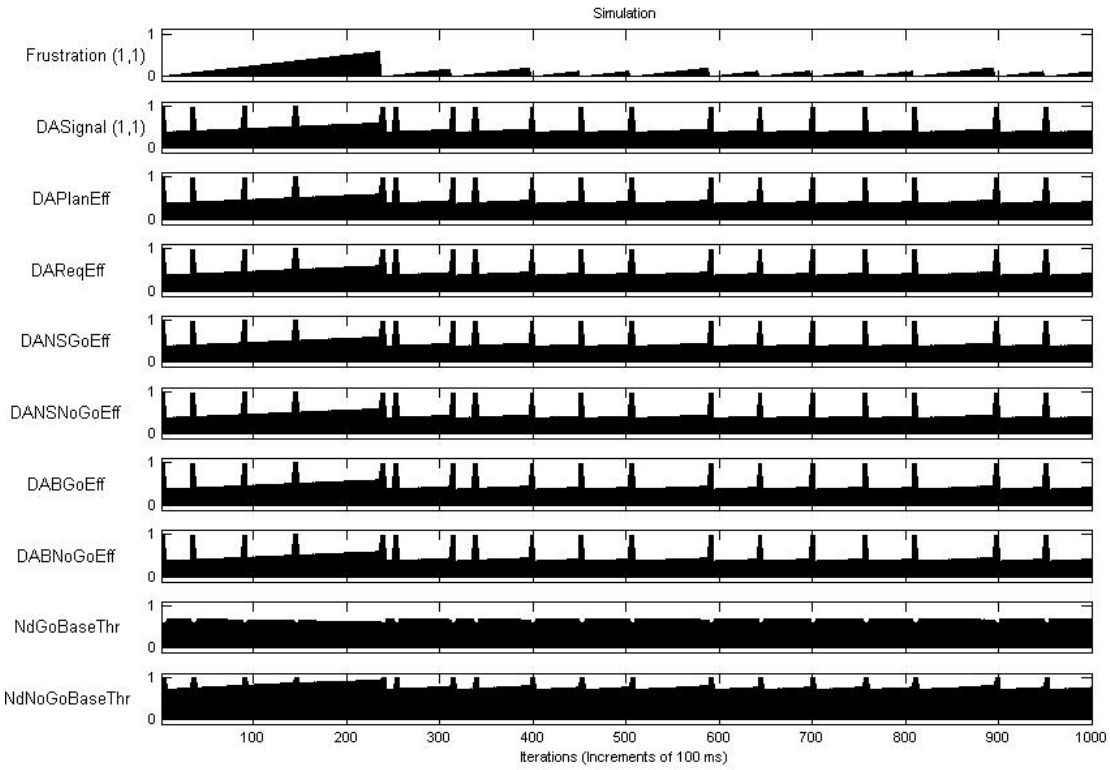


Figure 2.25 Normal DA effects. All DA effect variables are identical to the DA Signal unit output.

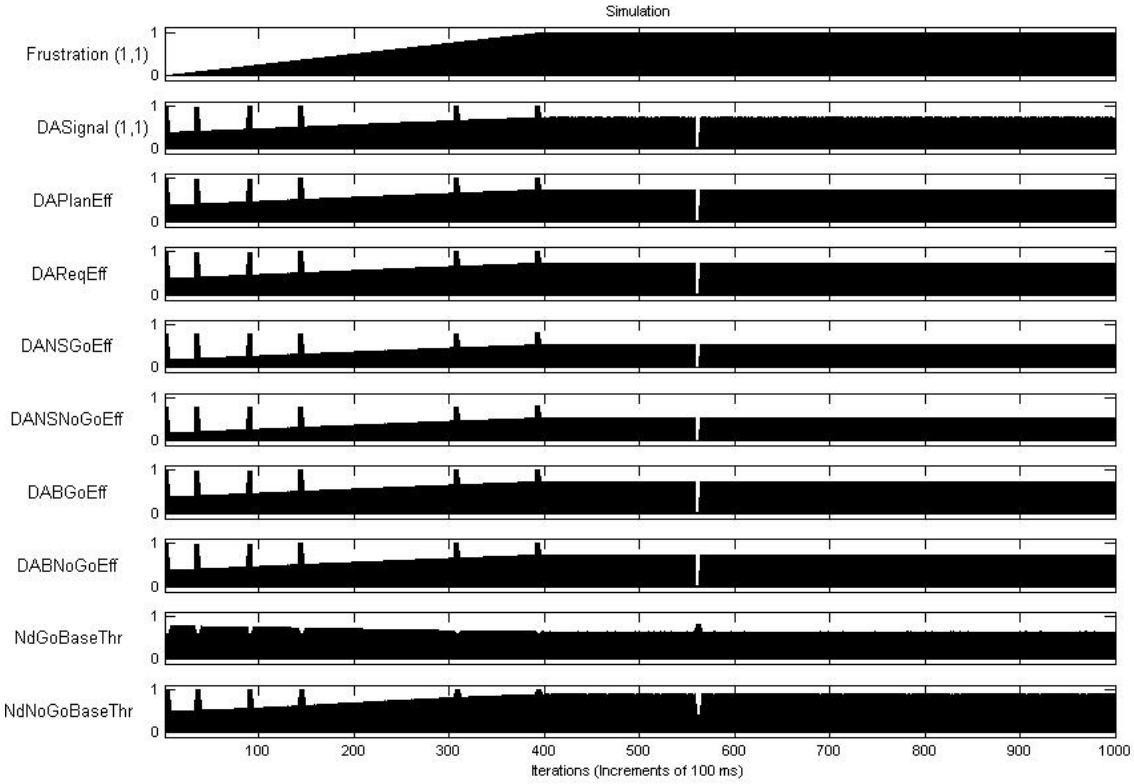


Figure 2.26 Nod/Shake hypo-DA effects. Effects of the ns2 DA manipulation condition. Only the Nod and Shake Go and NoGo units have their DA level effects suppressed.

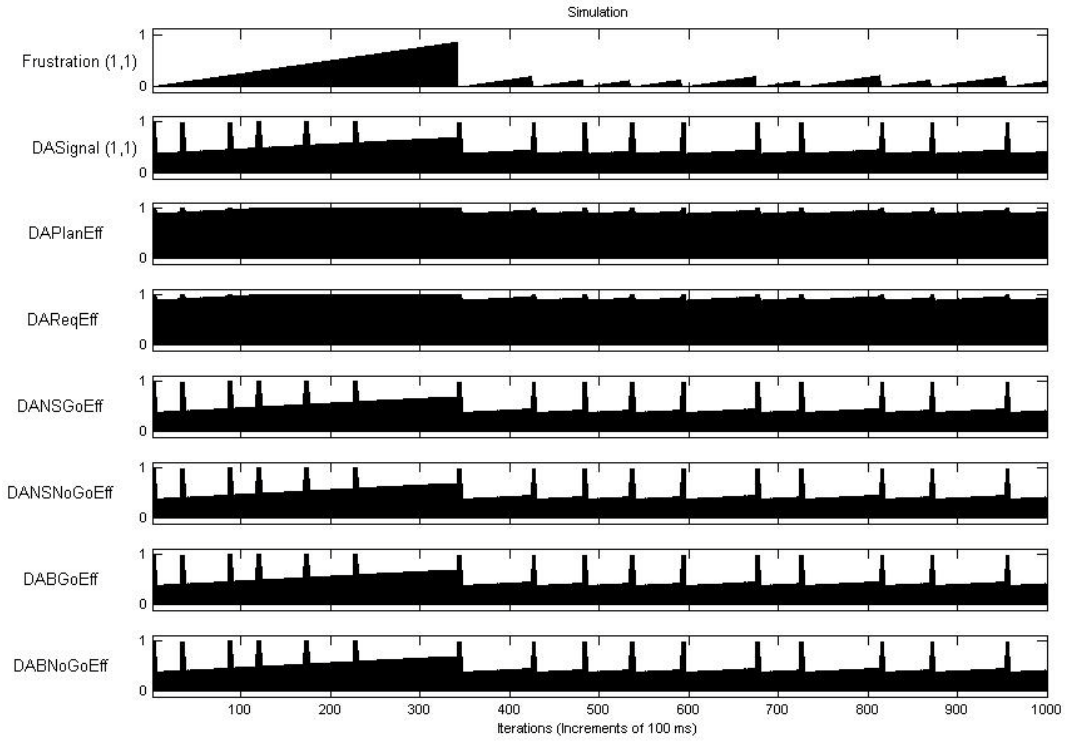


Figure 2.27 Plan/Request hyper-DA effects. Effects of the PR5 DA manipulation condition. Only the Plan and Request units have their DA level effects boosted.

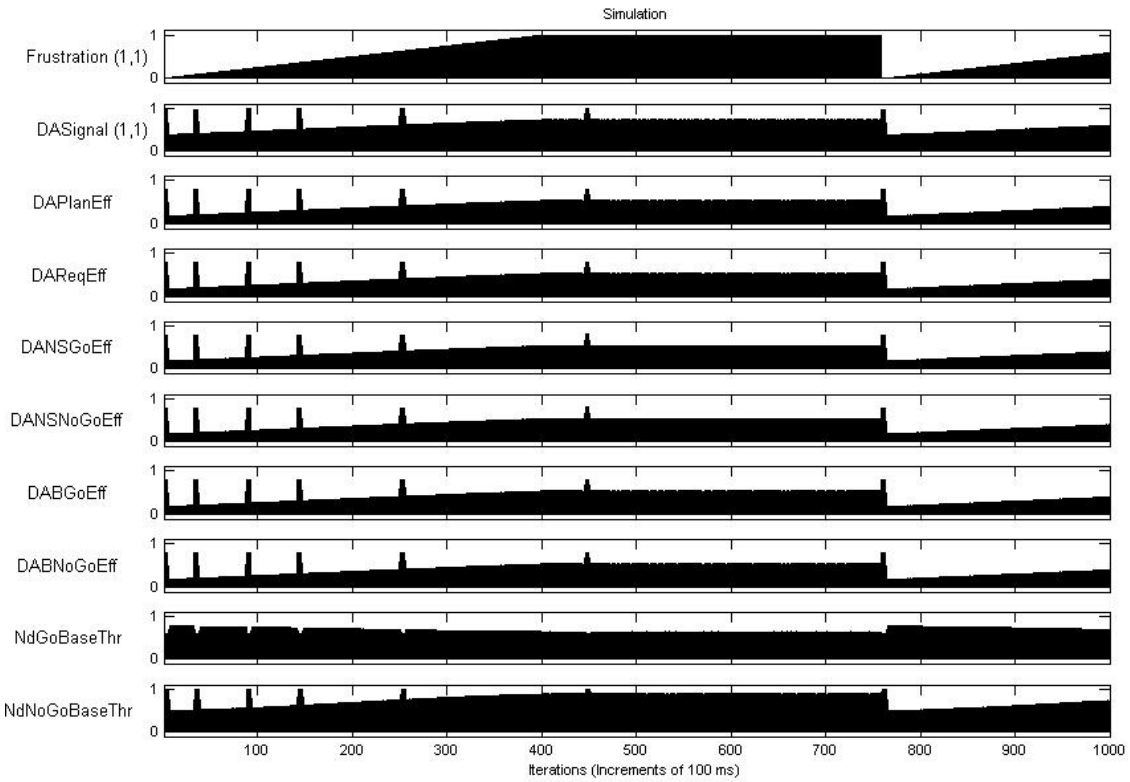


Figure 2.28 Global hypo-DA effects. Effects of the g2 DA manipulation condition. All of the DA-innervated targets have their DA effects suppressed.

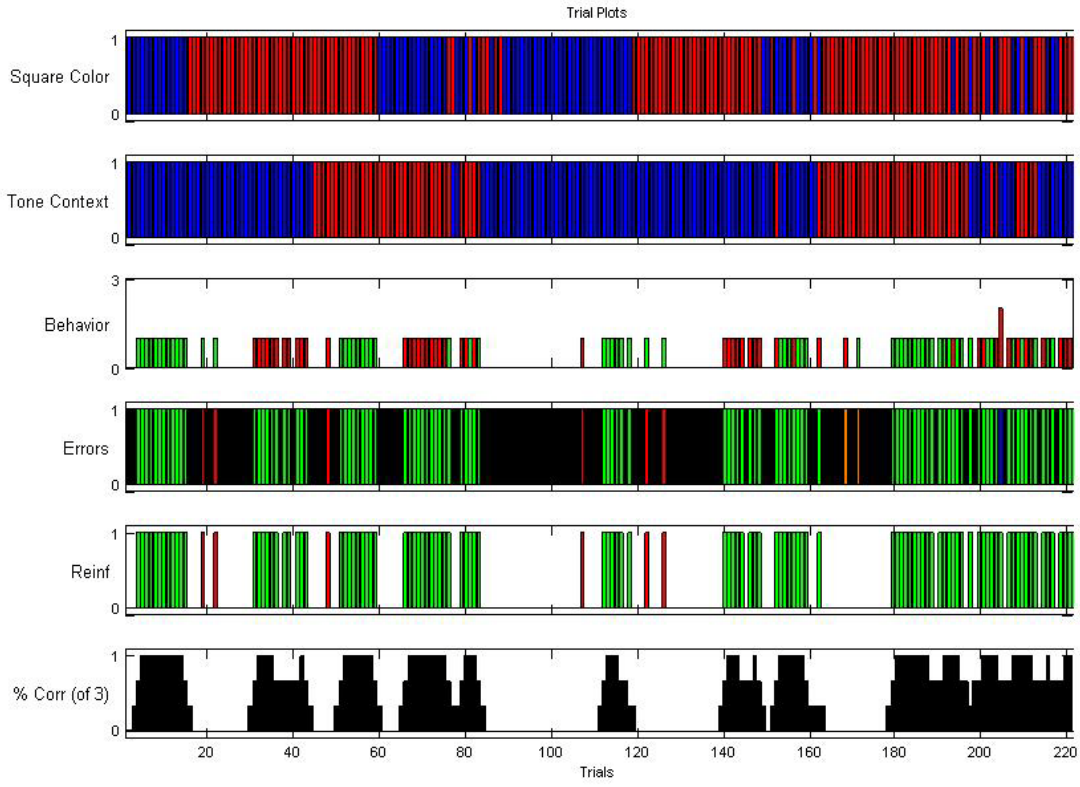


Figure 3.1 Full task set training simulation phase performance. Performance is shown for 221 trials (see Table 3.1 for a summary description). The square color (Blue, Red), tone context (Blue=C, Red=Eb), model behavior (Green=Nod, Red=Shake), errors committed (Green=Correct, Red=Wrong, Black=Miss, Orange=Late, Blue=Repeat), and reinforcement delivery (Green=Reward, Red=Punisher) are shown. Percentage correct over a sliding window of 3 trials is shown in the last trace. In the final 13 trials, 11 (84.6%) are correct and 2 (15.4%) are misses. Thus, training on the full task set is successful.

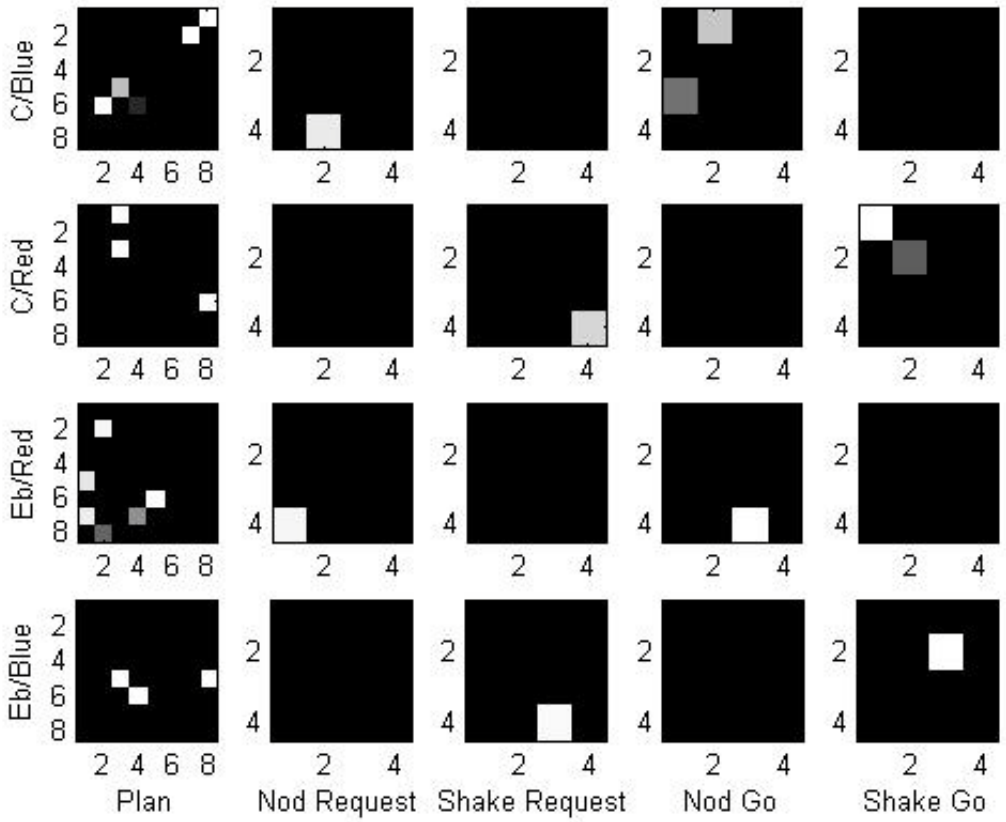


Figure 3.2 Plan, Request, and Go unit representations for training simulation phase. All 4 conjunction representations are shown. Mostly, they are non-overlapping.

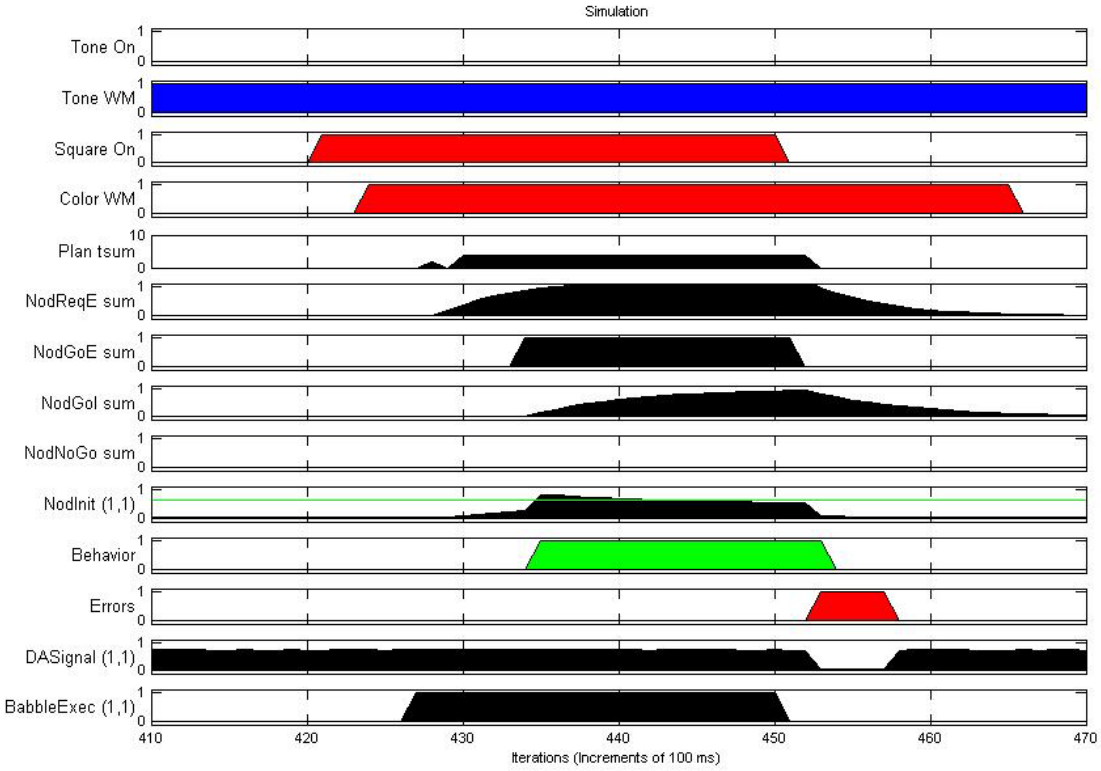
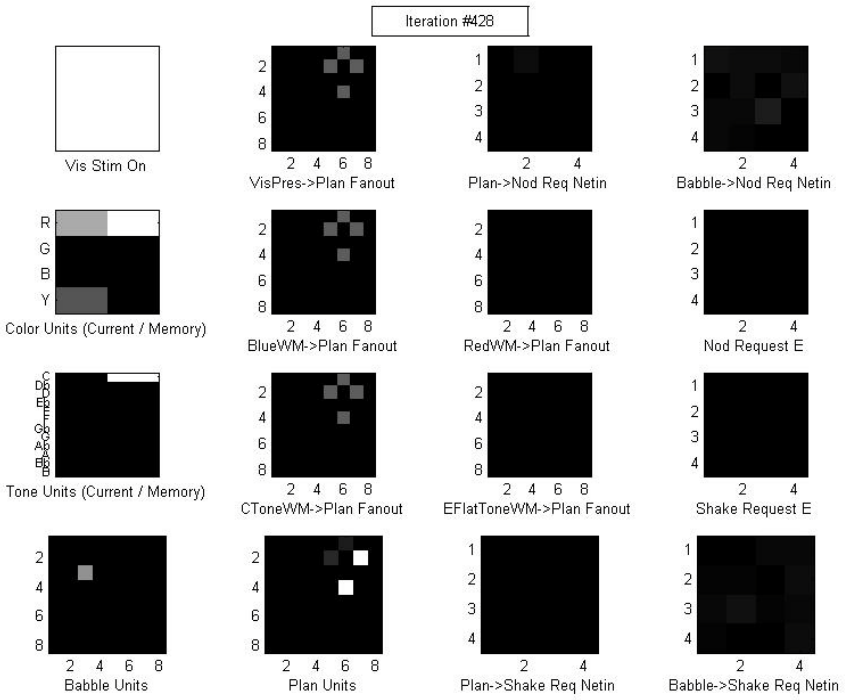
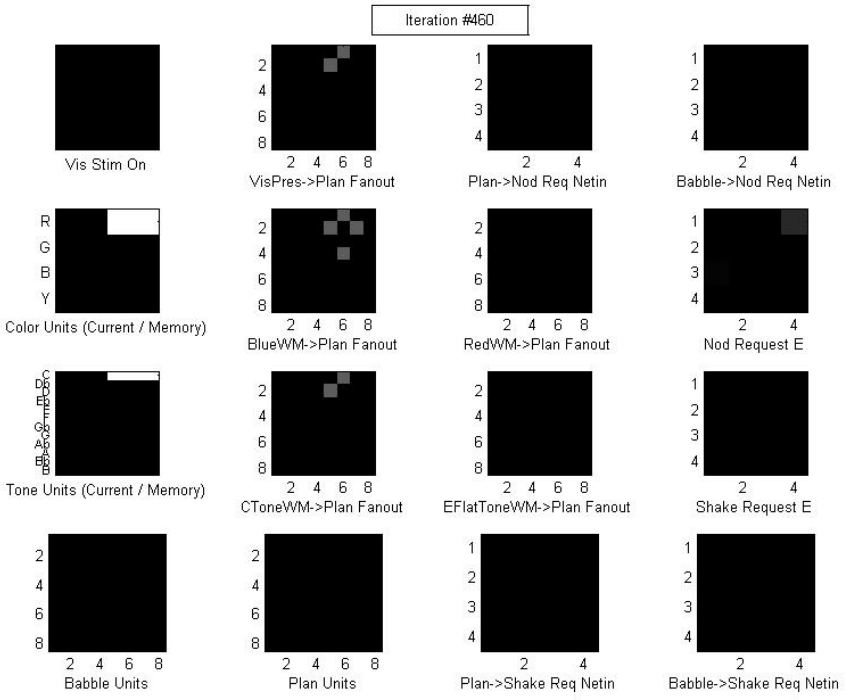


Figure 3.3 Example of an incorrect babble. This is taken from run 2 of the training phase. The desired conjunction is C AND Red -> Shake, but the babble randomly generates a Nod, instead, and the model is punished for the error. As it turns out, this leads to unlearning of the C AND Blue -> Nod conjunction, as is shown in Figure 3.4.



(a)



(b)

Figure 3.4 Weight change during babble-induced conjunction overwriting. (a) Plan weights during babble. (b) Plan weights after punishment.

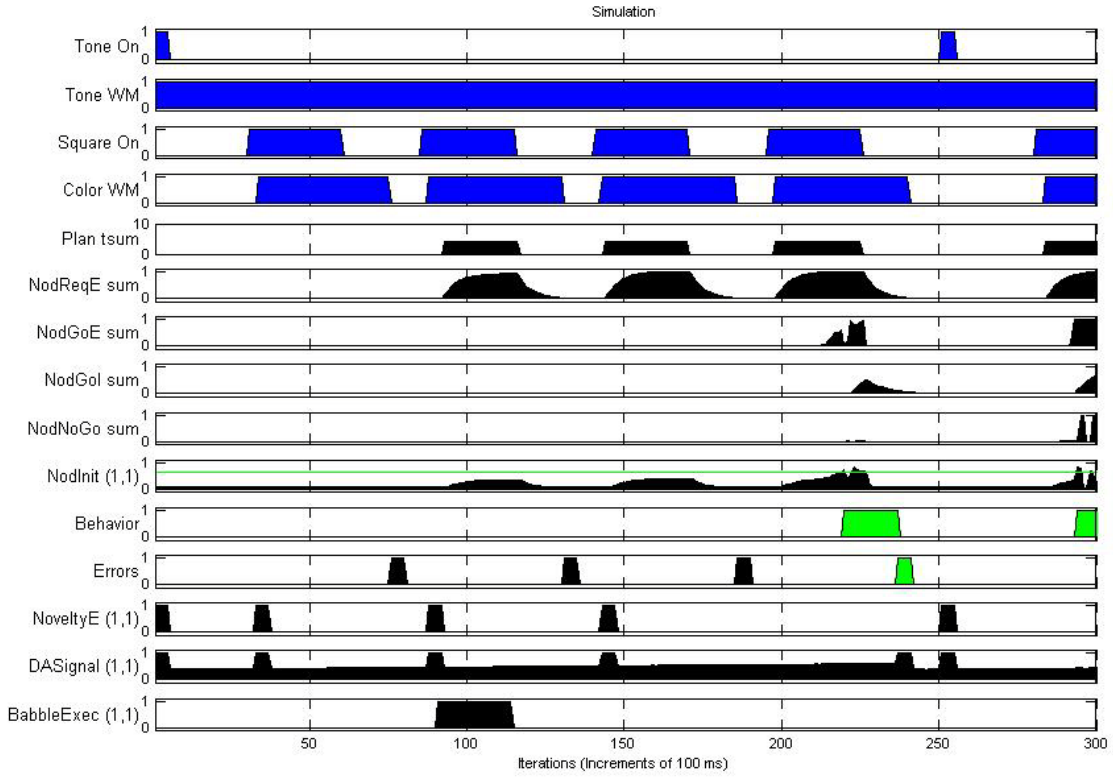
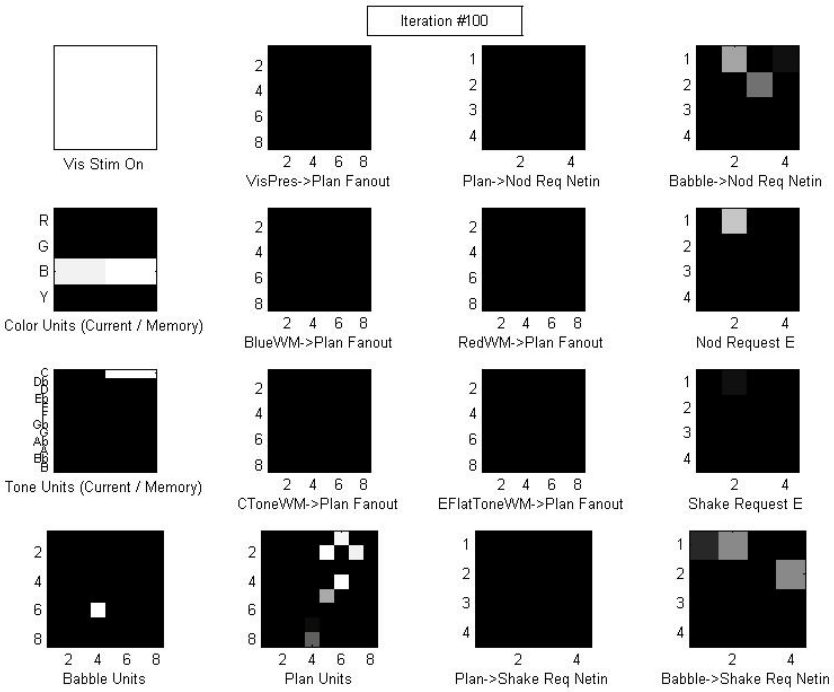
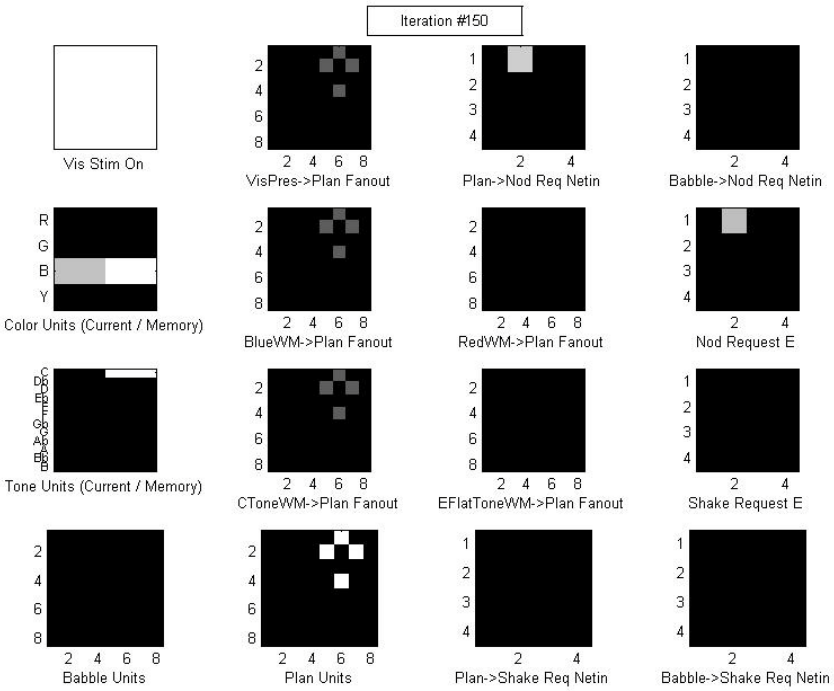


Figure 3.5 Example of novelty learning. The C AND Blue -> Nod conjunction is learned in the first run of the training phase. A babble activates Plan and Nod Request units, but Nod Go units fail to be engaged, so no actual Nod is generated. However, as it turns out, the Plan and Request unit representations are "stamped in" by the third novelty triggered in response to the blue square (shortly before iteration 150). Eventually, this, in combination with increasing DA signal level triggered by frustration, leads to Nod Go activity and a resulting Nod, and the behavior is learned in earnest.



(a)



(b)

Figure 3.6 Weight change during novelty-induced conjunction learning. (a) Plan weights during babble. (b) Plan weights after Blue-induced novelty.

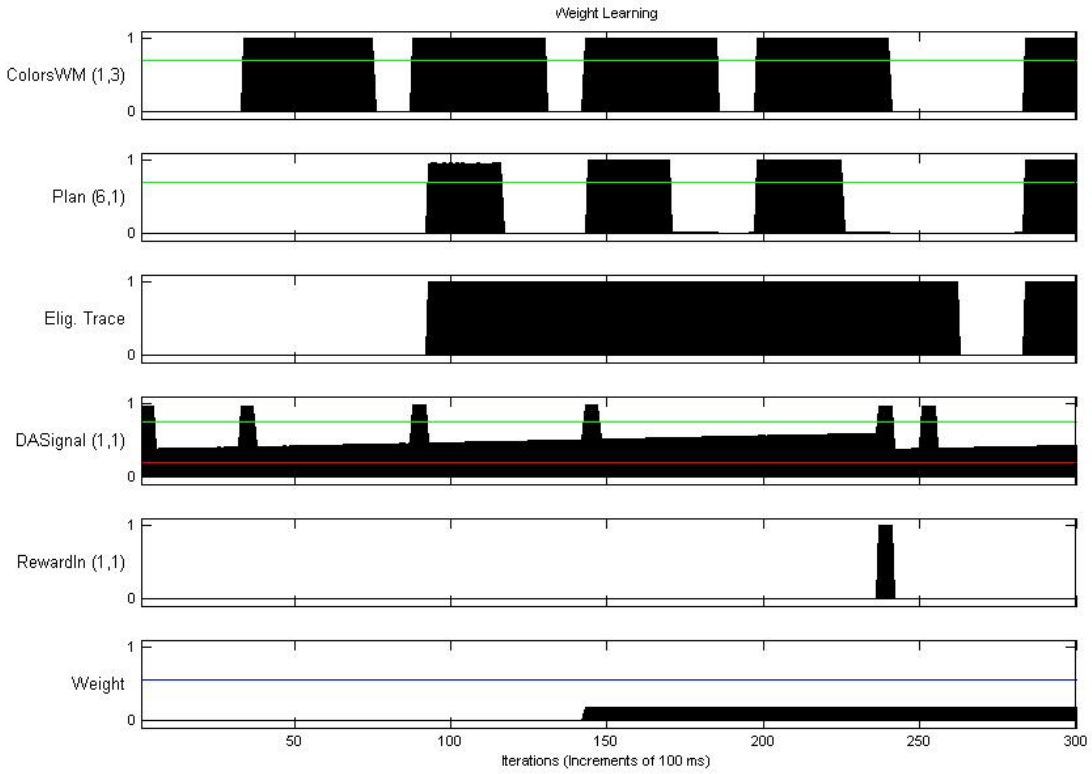


Figure 3.7 Individual weight change during novelty-induced conjunction learning. The weight between the Blue WM unit and Plan unit (6,1) is shown. It can be seen that it is not a reward that triggers the learning. Therefore, as can be verified from Figure 3.5, it is a novelty-induced burst that triggers the learning.

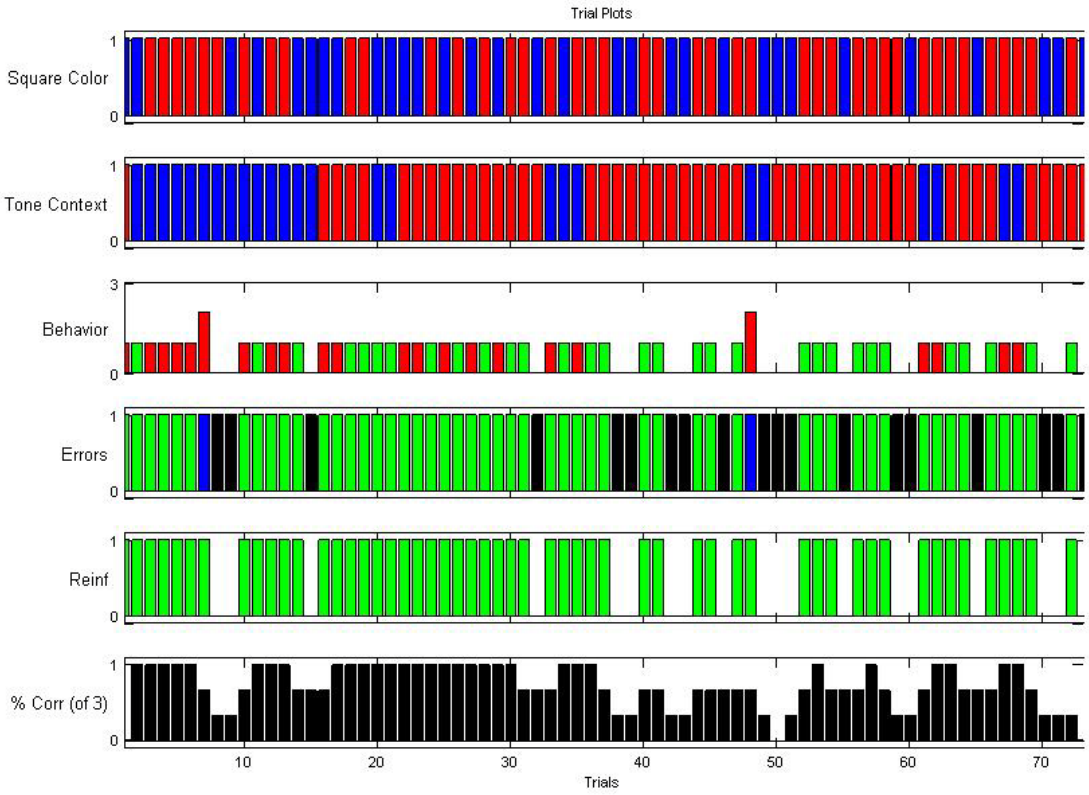


Figure 3.8 Full task set maintenance simulation phase performance. Performance is shown for 73 trials (see Table 3.2 for a summary description). The square color (Blue, Red), tone context (Blue=C, Red=Eb), model behavior (Green=Nod, Red=Shake), errors committed (Green=Correct, Black=Miss, Blue=Repeat), and reinforcement delivery (Green=Reward) are shown. Percentage correct over a sliding window of 3 trials is shown in the last trace. In all 73 trials, 54 (74.0%) are correct and 19 (26.0%) are misses. Thus, maintenance of the full task set is mostly successful.

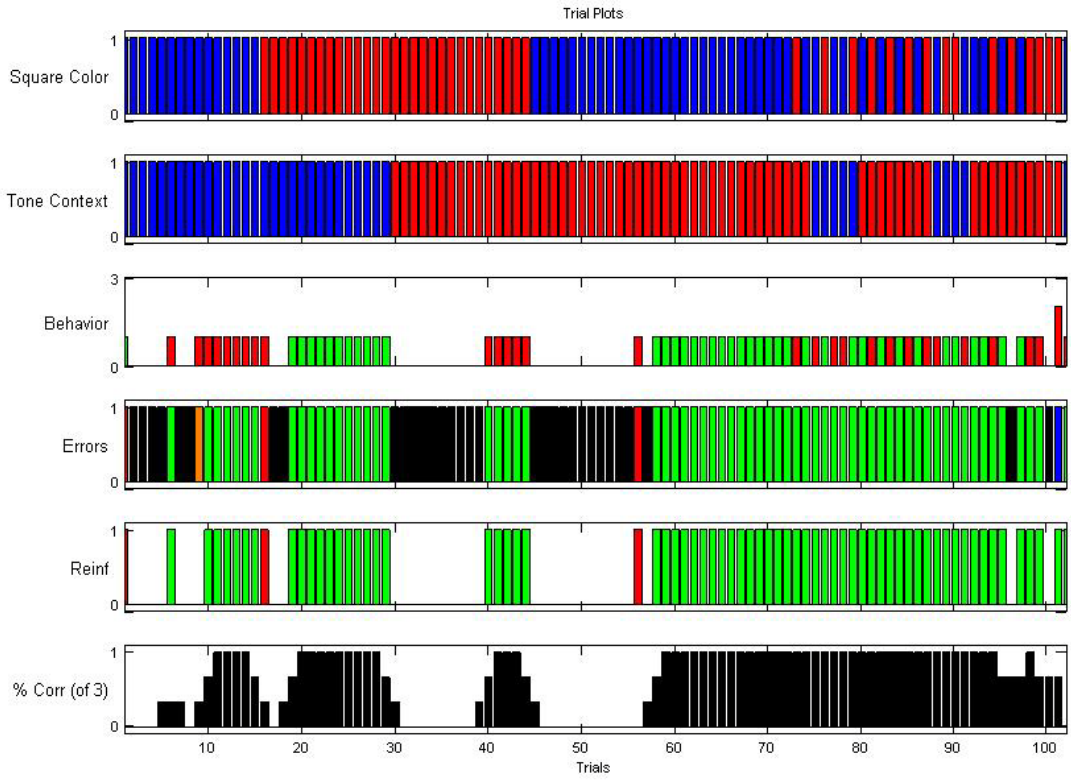


Figure 3.9 Full task set reversal simulation phase performance. Performance is shown for 102 trials (see Table 3.3 for a summary description). The square color (Blue, Red), tone context (Blue=C, Red=Eb), model behavior (Green=Nod, Red=Shake), errors committed (Green=Correct, Red=Wrong, Black=Miss, Orange=Late, Blue=Repeat), and reinforcement delivery (Green=Reward, Red=Punisher) are shown. Percentage correct over a sliding window of 3 trials is shown in the last trace. In the final 13 trials, 11 (84.6%) are correct and 2 (15.4%) are misses. Thus, reversal training on the full task set is successful.

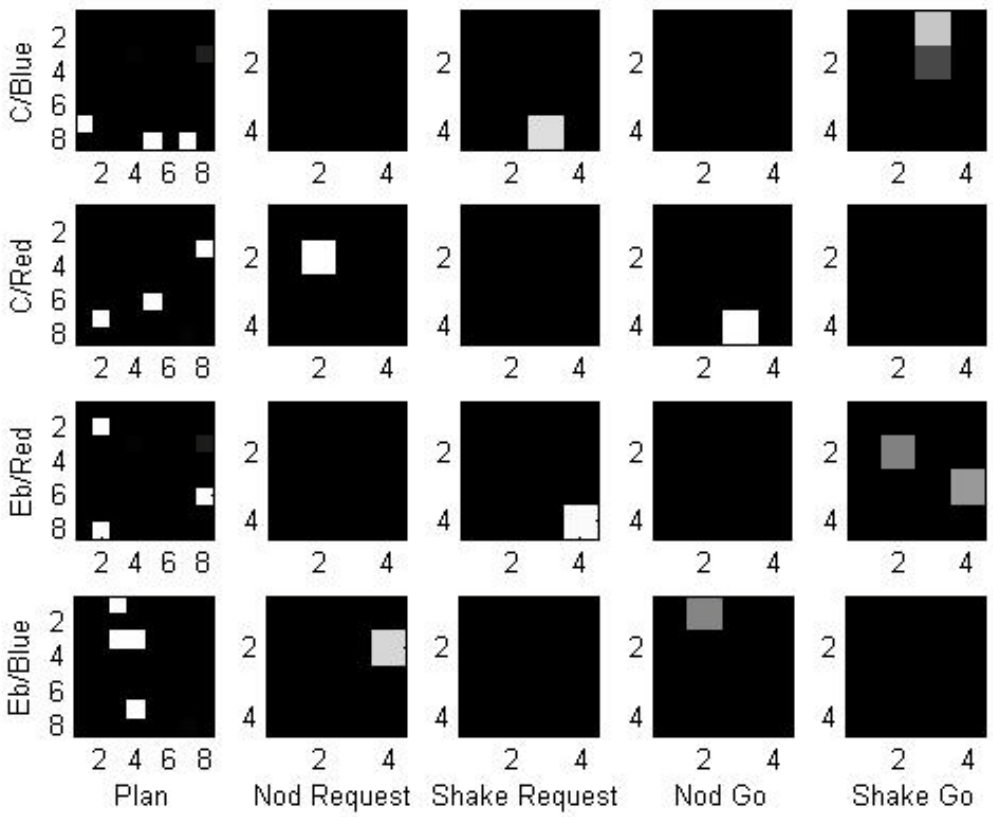


Figure 3.10 Plan, Request, and Go unit representations for reversal simulation phase. All 4 conjunction representations are shown. Mostly, they are non-overlapping. Note also that they are different from those in Figure 3.2.

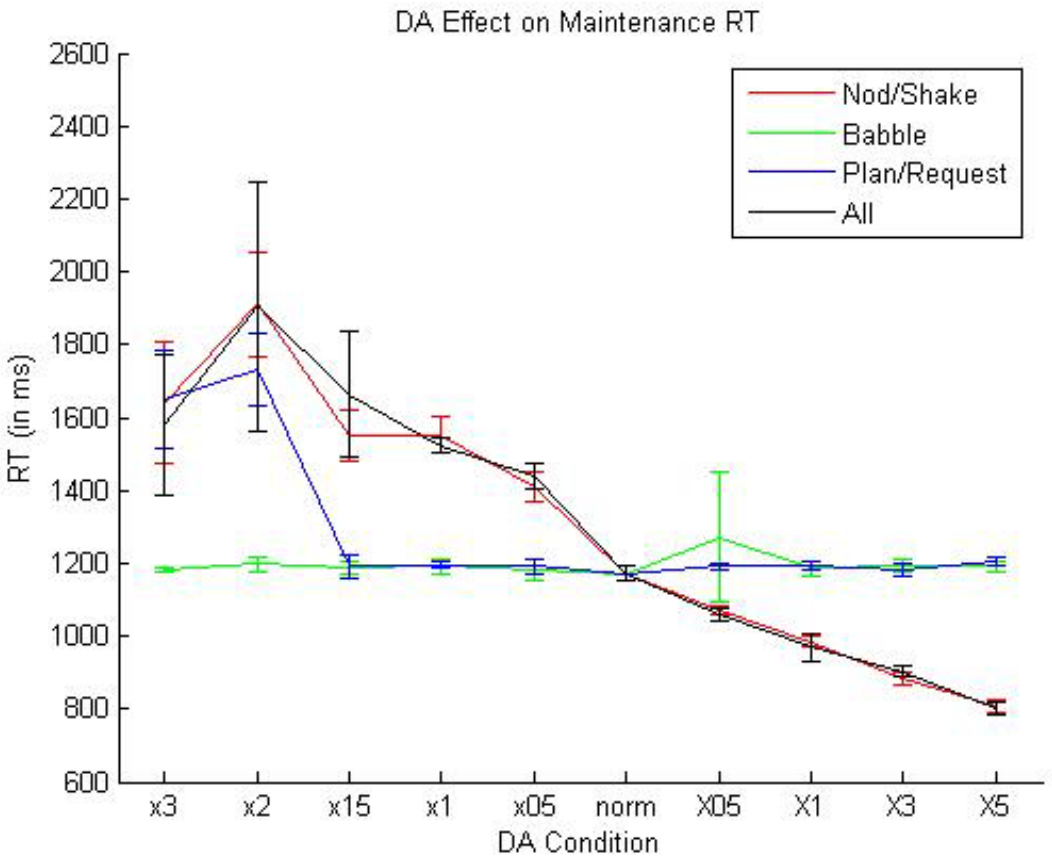


Figure 3.11 DA manipulation effects during the maintenance simulation phase on RT. DA conditions are described in Section 2.3.14. Reaction time (RT) is measured in ms. and reflects the latency between the appearance of a colored square and the onset of a Nod or Shake response. Each point is the mean RT for 5 simulation 'subjects' (i.e., model random weight initializations) over all trials of 5 runs under the particular DA manipulation. Error bars reflect ± 1 standard deviation from the RT means. Hypo-DA effects (left of norm) are as follows. Nod/Shake or global hypo-DA leads to increased (slowed) reaction time. Plan/Request hypo-DA leads to no significant effect, except at the most extreme hypo-DA conditions, where RT is highly slowed (probably reflecting the fact that the model is babbling to respond rather than executing learned responses). Babble hypo-DA has no effect. Hyper-DA effects (right of norm) are as follows. Nod/Shake or global hyper-DA leads to progressively decreasing (speeded) RT. Plan/Request and Babble hyper-DA, on the other hand, has no effect on RT.

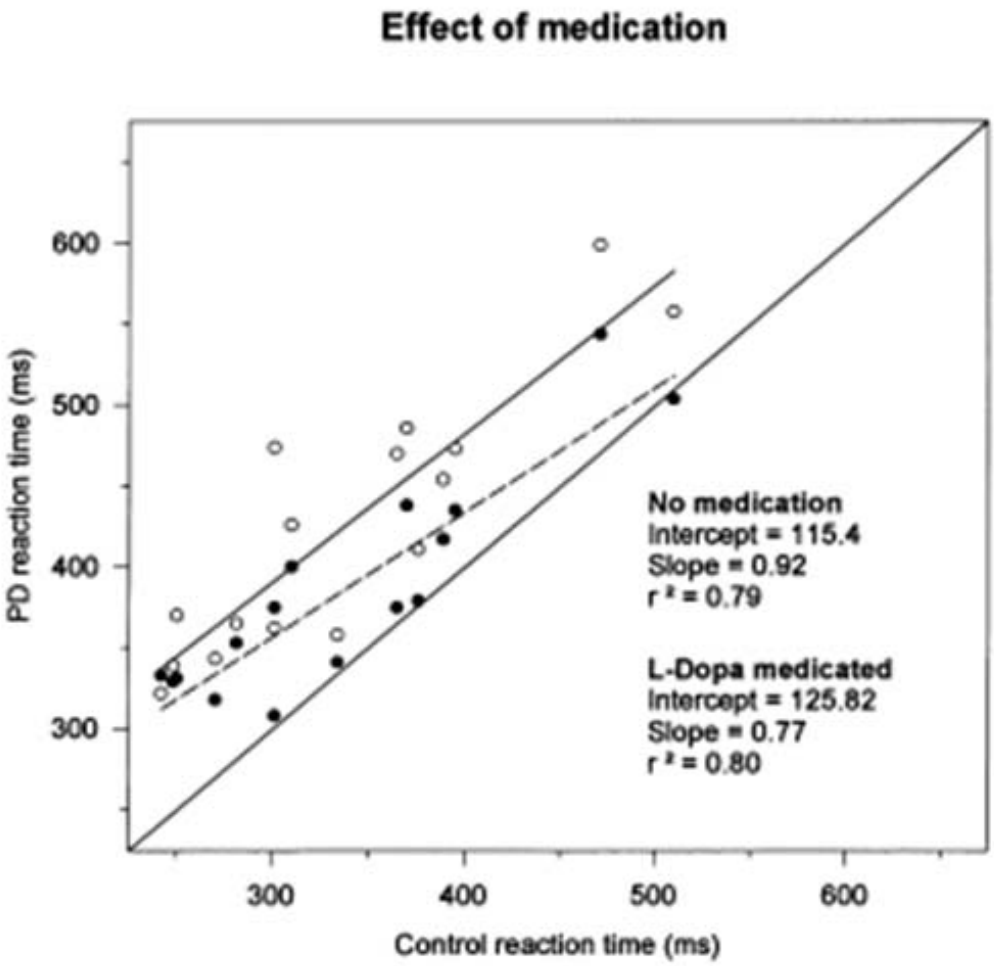


Figure 3.12 Meta-analysis of various Parkinson's studies comparing control and Parkinsonian reaction times. Taken from Figure 3 of (Gauntlett-Gilbert & Brown, 1998). Each point is a study. Unfilled points are mean data for unmedicated subjects, and filled points are mean data for medicated subjects. Medicated performance is closer to normal (control RT) suggesting less severe hypo-DA conditions have less-impeded RT.

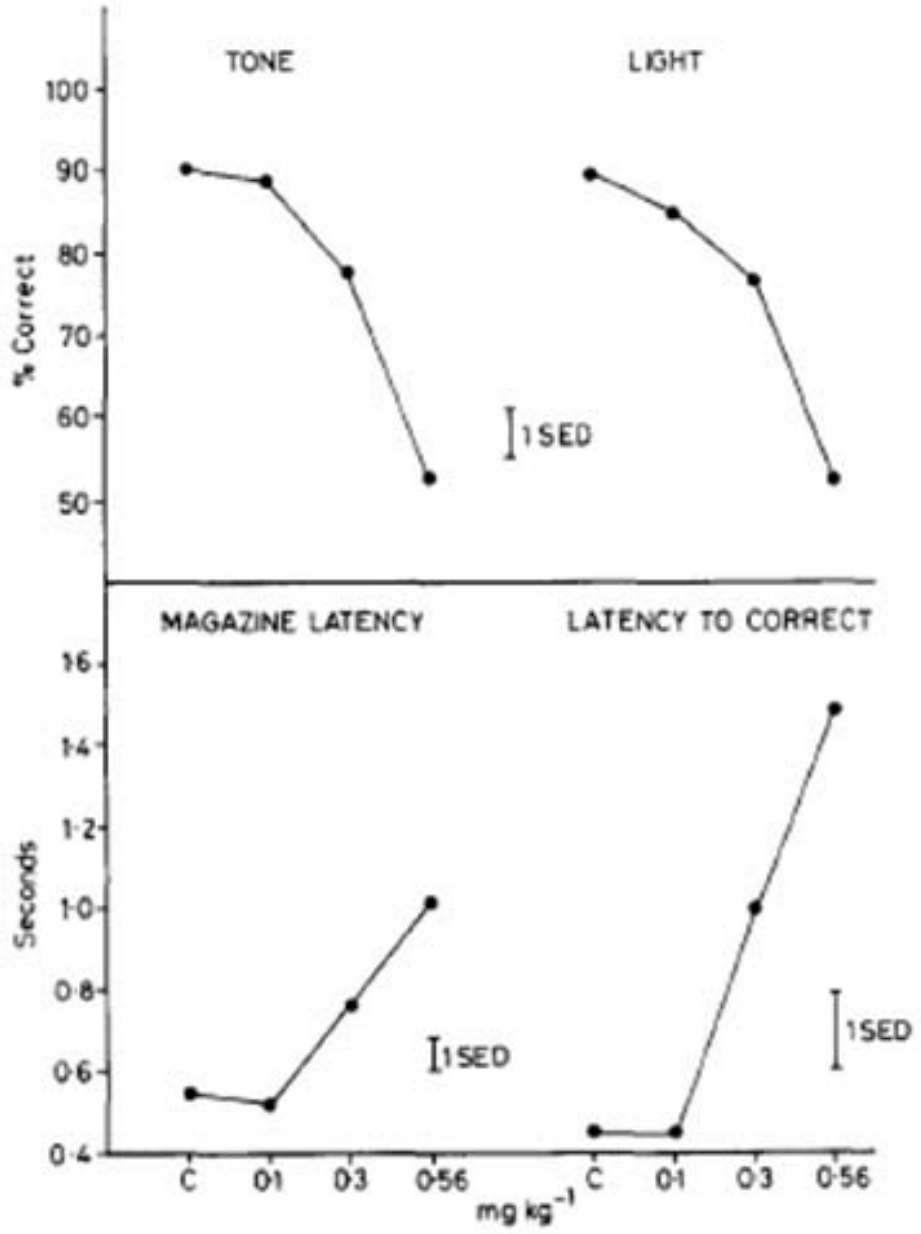


Figure 3.13 Evidence that increasing hypo-DA leads to decreasing responses and increased RT. Figure taken from Figure 6 of (Robbins et al., 1990). (top) Dose-dependent effects of a DA antagonist (α-flupenthixol) on the percentage of correct responses made by rats on a successfully learned discriminative lever-pressing task. As dosage increases, the percent of correct responses decreases for both visual and auditory cues. This suggests suppression of the learned behavior. (bottom) Increasing DA antagonist dosage leads to increasing response latency, both in magazine nose-pokes and the lever presses.

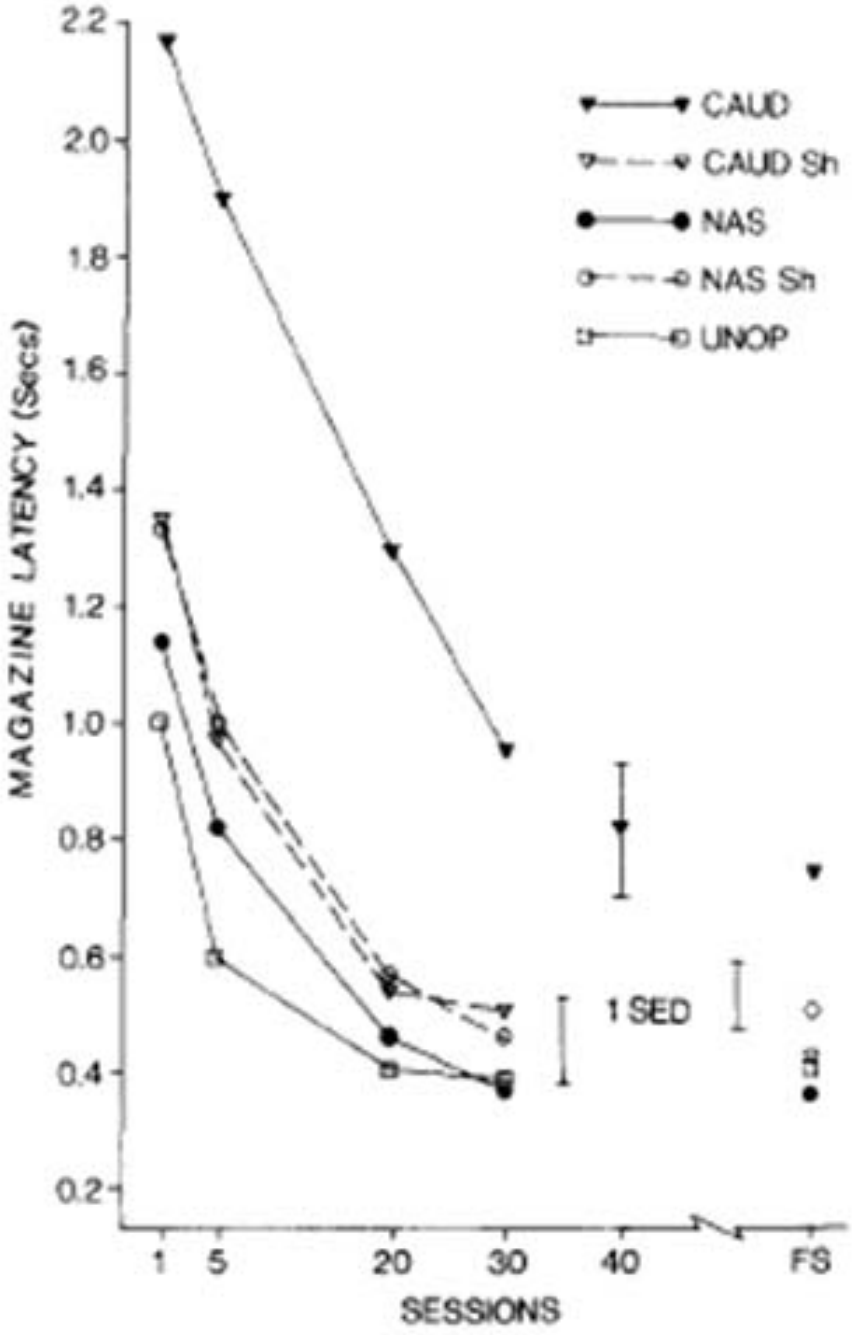


Figure 3.14 Effects of target-specific DA lesions on RT. Figure taken from Figure 3 of (Robbins et al., 1990). Over the course of training on a lever task, rats show a decreased RT. However, this RT is slower for rats that have their dorsal striatum lesioned (CAUD trace), but not the ventral striatum (NAS trace). CAUD = caudate/putamen (dorsal striatum); CAUD Sh = CAUD sham lesion; NAS = nucleus accumbens septi (ventral striatum); NAS Sh = NAS sham lesion; UNOP = unoperated-on.

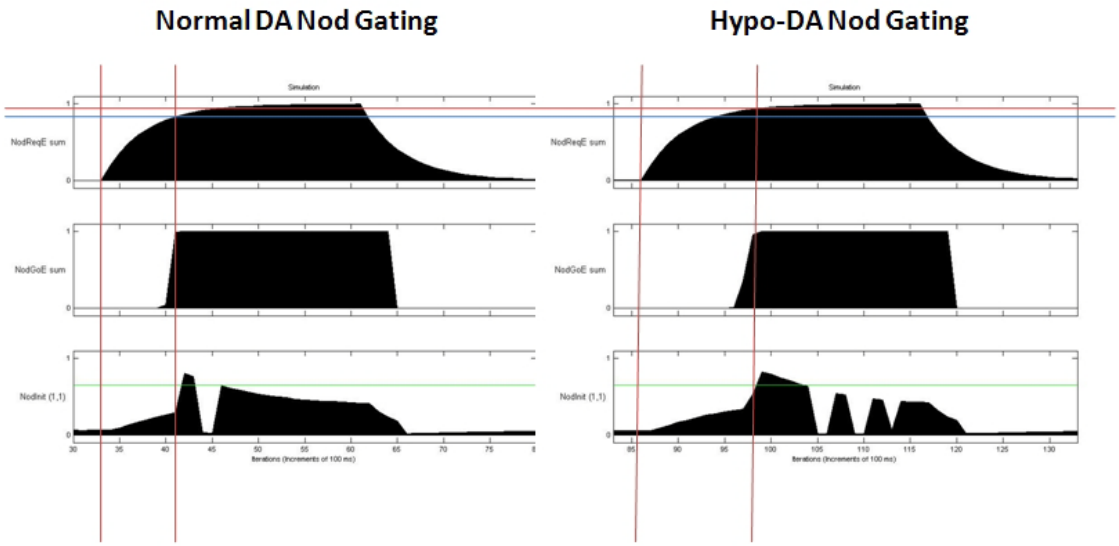


Figure 3.15 Nod/Shake hypo-DA effect on Request-to-Go delay. Hypo-DA in the Nod/Shake pathway leads to raising of thresholds of the Nod/Shake Go units (blue-to-red horizontal line transition). Activity of the Request units takes time to build up in response to the colored square stimuli. Latency (demarcated by the vertical red lines) between the onset of the stimulus and the firing of the Go units enabling the Init unit activity that leads to behavior is longer under hypo-DA because Request activity buildup takes longer to cross the threshold.

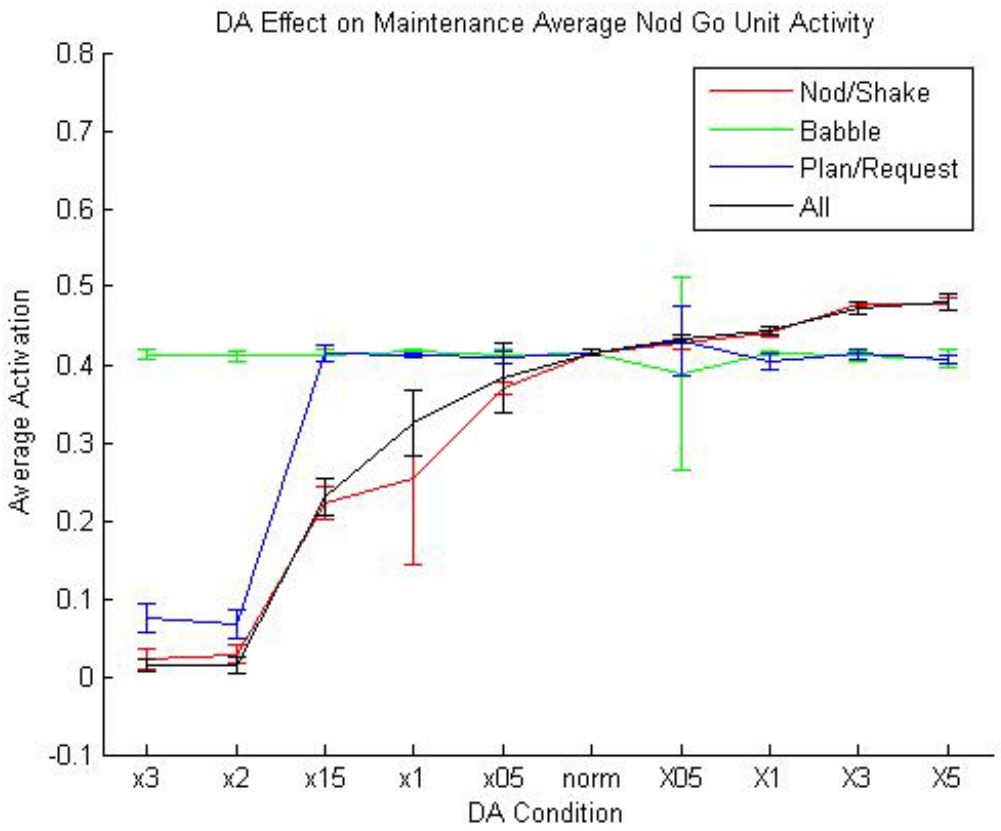


Figure 3.16 DA manipulation effects during the maintenance simulation phase on the average Nod Go unit activity. Each point is the mean of the summed activation of the Nod Go units during all of the trials over all 5 runs for the 5 simulation 'subjects' under the particular DA manipulation. Error bars reflect ± 1 standard deviation from the mean summed activations. Hypo-DA effects are more pronounced than hyper-DA effects. Nod/Shake or global hypo-DA leads to decreasing Nod Go unit activation (because of the threshold-modifying effects of DA). Plan/Request hypo-DA leads to an effect only at the most extreme hypo-DA conditions, but this is probably due to unlearning effects. Babble hypo-DA has no effect. On the other hand, Nod/Shake or global hyper-DA leads to increased Nod Go activity. Babble and Plan/Request hyper-DA, however, have no effect.

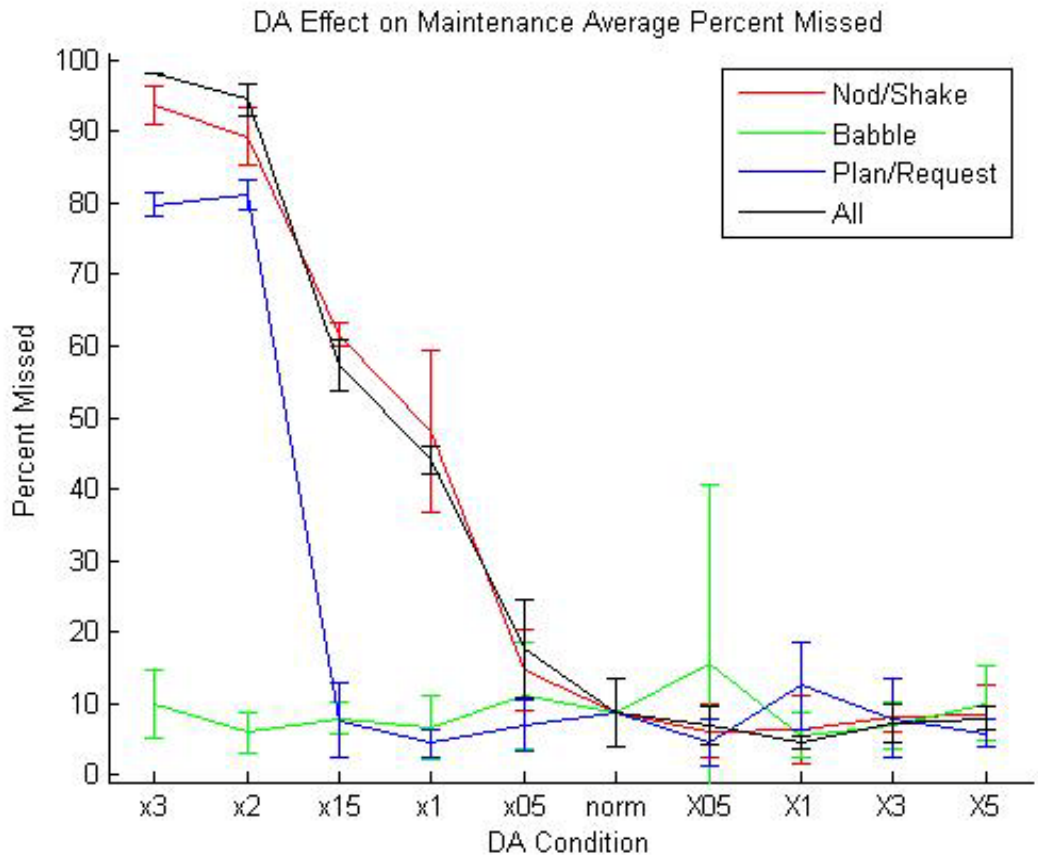
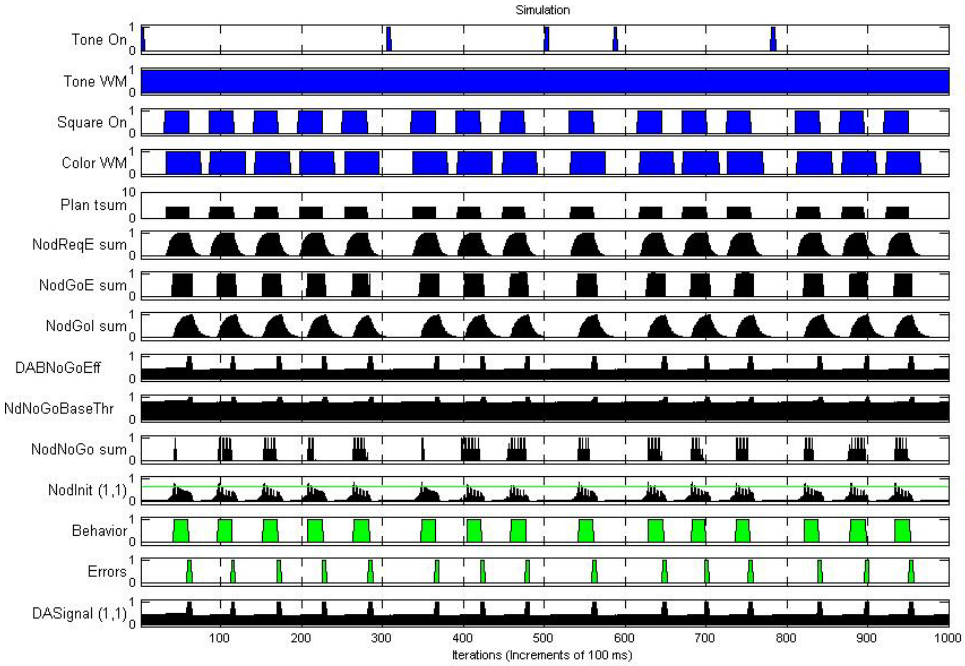
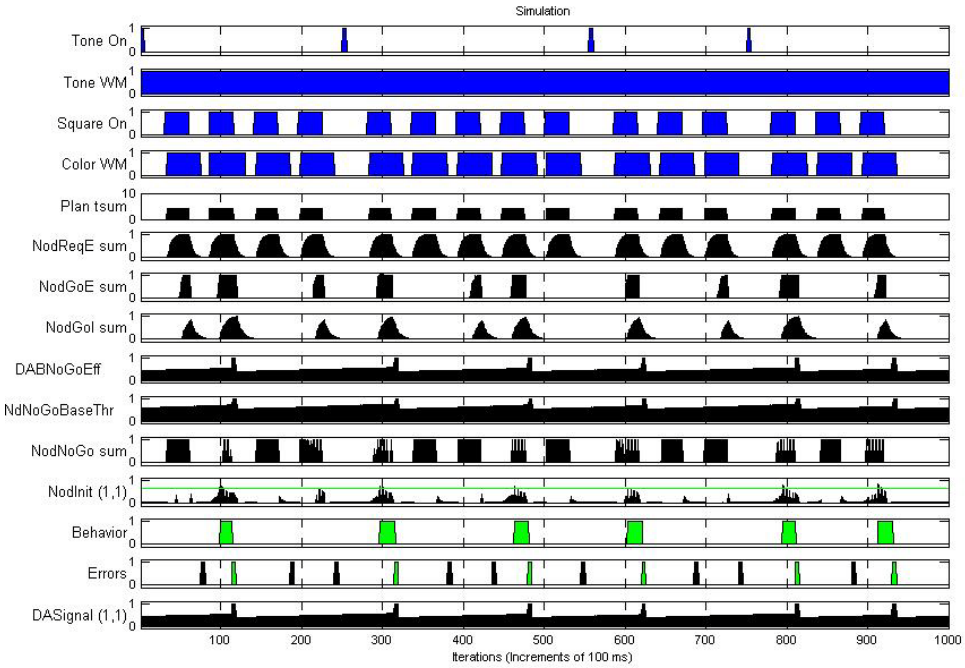


Figure 3.17 DA manipulation effects during the maintenance simulation phase on percent of missed-response trials. Each point is the mean of the percents of trials which lead to missed responses over all 5 runs for the 5 simulation 'subjects' under the particular DA manipulation. Error bars reflect ± 1 standard deviation from the percent of missed-response means. Generally, there are hypo-, but not hyper-, DA effects. Nod/Shake or global hypo-DA leads to increasing percentages of missed responses. Plan/Request hypo-DA leads to no significant effect except at the most extreme hypo-DA conditions, probably due to the conjunction unlearning caused by these extremes. Babble hypo-DA has no effect.



(a)



(b)

Figure 3.18 Effect of hypo-DA on behavior initiation. (a) Normal DA C AND Blue -> Nod response. (b) Hypo-DA C AND Blue -> Nod response.

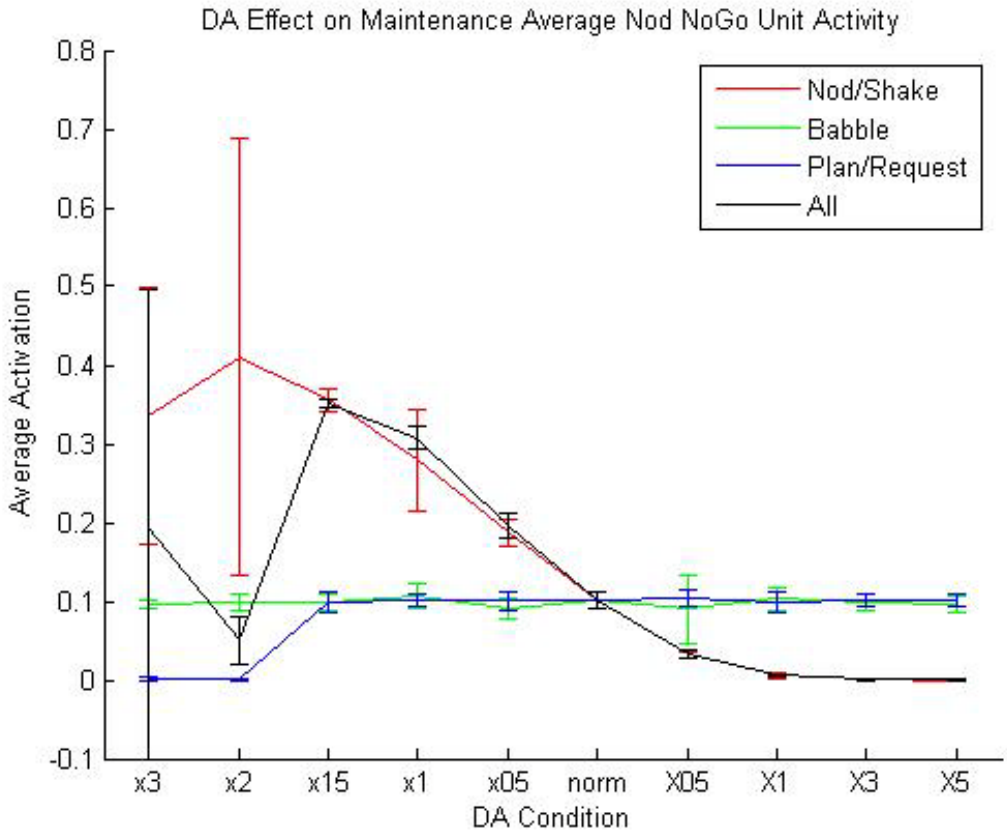


Figure 3.19 DA manipulation effects during the maintenance simulation phase on the average Nod NoGo unit activity. Each point is the mean of the summed activation of the Nod NoGo units during all of the trials over all 5 runs for the 5 simulation 'subjects' under the particular DA manipulation. Error bars reflect ± 1 standard deviation from the mean summed activations. Nod/Shake or global hypo-DA leads to increasing Nod NoGo unit activation (because of the threshold-modifying effects of DA). Plan/Request hypo-DA leads to an effect (in this case a decrease) only at the most extreme hypo-DA conditions, but this is probably due to unlearning effects. The trends are also disrupted for Nod/Shake and global hypo-DA at the extremes. Babble hypo-DA has no effect. Nod/Shake or global hyper-DA leads to decreased Nod NoGo activity: to the point, even, of disabling these units. Babble and Plan/Request hyper-DA, however, have no effect.

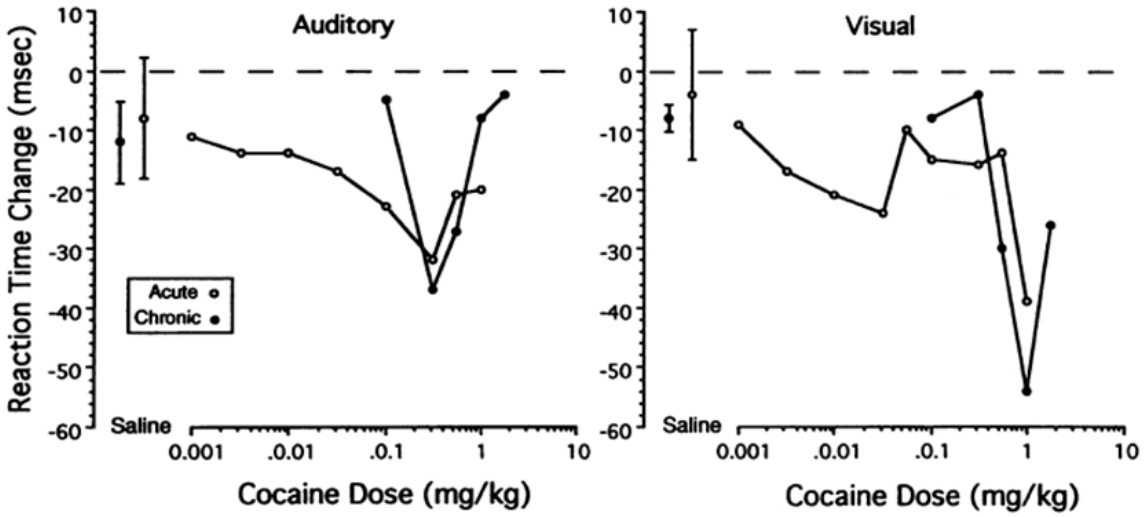


Figure 3.20 Effects of cocaine administration on RT. Figure taken from Figure 9 of (Hienz et al., 1994). Differences in RTs in simple reaction to auditory (left) and visual (right) stimuli in baboons during varying doses (both acute and chronic) of cocaine. Saline error bars are shown for comparison. Generally, moderate dosages of cocaine shorten RT.

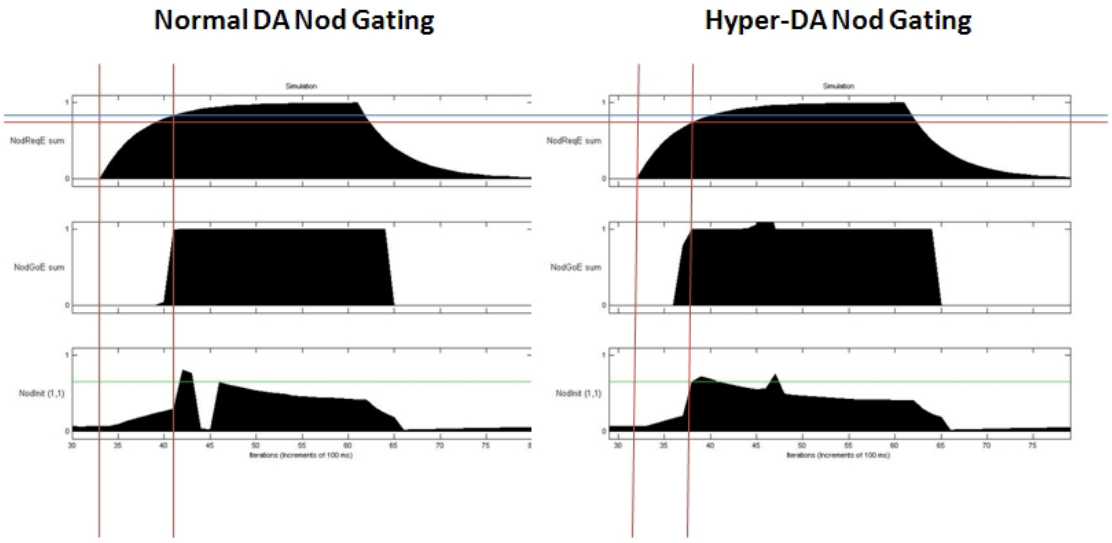


Figure 3.21 Nod/Shake hyper-DA effect on Request-to-Go delay. Hyper-DA in the Nod/Shake pathway leads to lowering of thresholds of the Nod/Shake Go units (blue-to-red horizontal line transition). Activity of the Request units takes time to build up in response to the colored square stimuli. Latency (demarcated by the vertical red lines) between the onset of the stimulus and the firing of the Go units enabling the Init unit activity that leads to behavior is shorter under hyper-DA because Request activity buildup takes less time to cross the threshold.

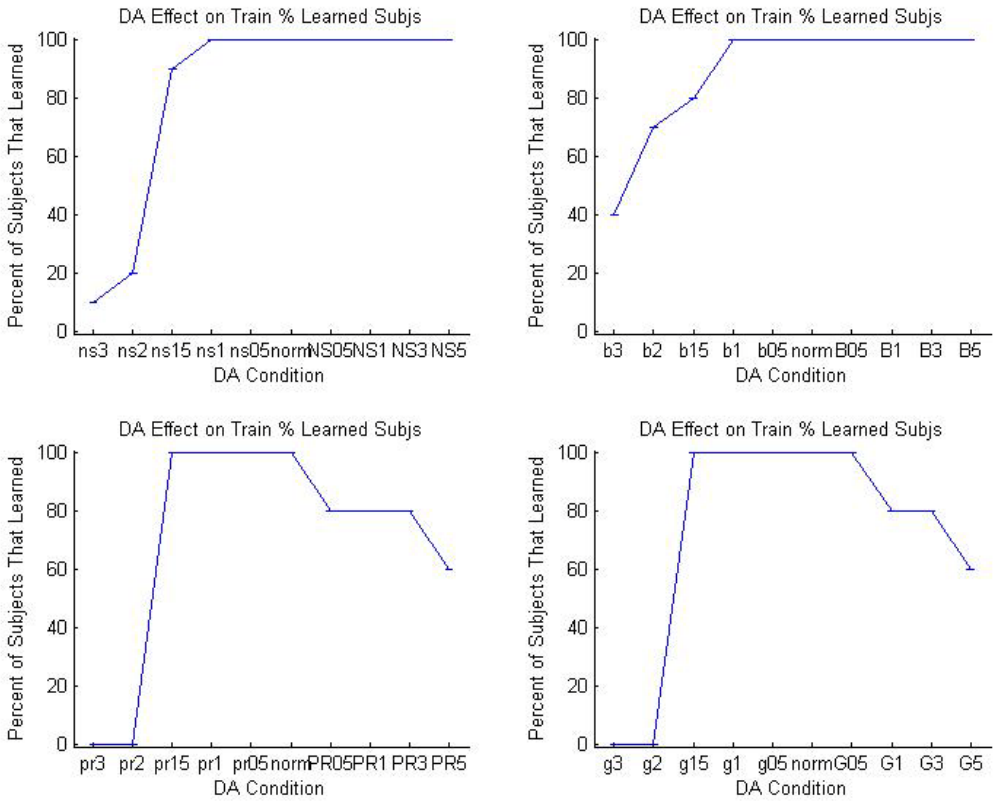


Figure 3.22 DA manipulation effects during the training simulation phase on percent of subjects that learned the task successfully. For each subject, the last 13 trials of the test run at the end of the simulation phase are analyzed and the percentage of correct responses is calculated, and if more that 25% of the trials are correct, the subject has acquired the task. Each point, then, is the percent (out of 10 subjects) that learned the task. Nod/Shake hypo-DA leads to increasing failure to learn the task, with only 1 of the 10 subjects learning it under the ns3 condition. Babble hypo-DA effects are less severe with 4 of 10 subjects learning the task in the worst case. Plan/Request (and global) hypo-DA only has an effect at the two most extreme cases, but effectively disables learning. On the other hand, hyper-DA has a more limited impact, with Nod/Shake and Babble hyper-DA having no effect. However, Plan/Request (and global) hyper-DA decrease the chance of task learning with 6 of 10 subjects learning the task in the most extreme case.

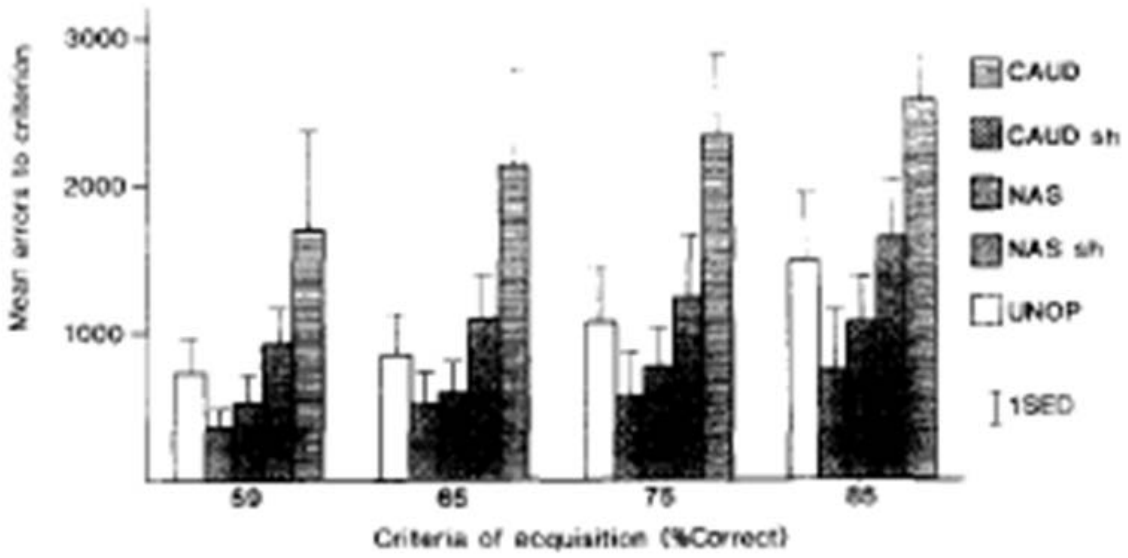


Figure 3.23 Effects of DA lesions on rat learning of discriminative lever-pressing task. Taken from Figure 1 of (Robbins et al., 1990). Dorsal striatal DA lesions of the dorsal striatum (CAUD) cause a significant increase in the number of error trials before the rats acquired criterion performance on the lever-pressing task. CAUD (last bar) = caudate/putamen (dorsal striatum); CAUD Sh (third bar) = CAUD sham lesion; NAS (fourth bar) = nucleus accumbens septi (ventral striatum); NAS Sh (second bar) = NAS sham lesion; UNOP (first bar) = unoperated-on.

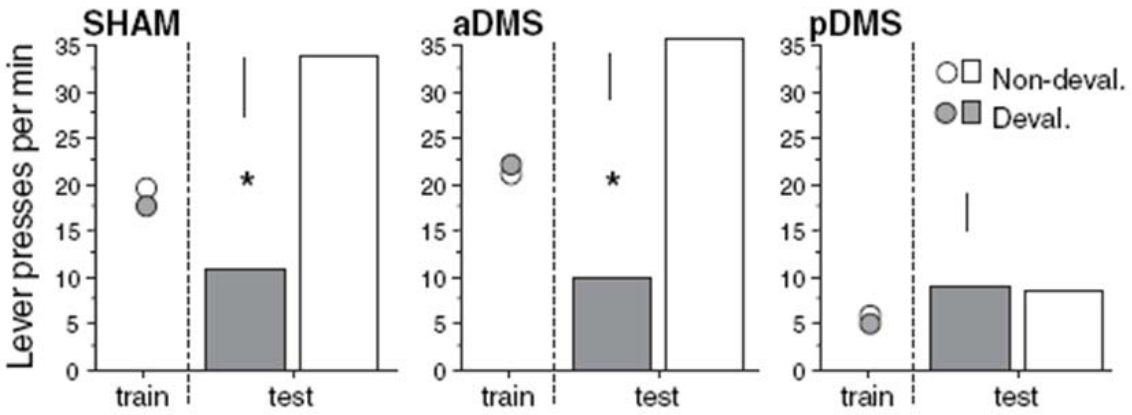


Figure 3.24 Effect of dorsal striatal lesion on task acquisition. Taken from Figure 2 of (Yin et al., 2005). In the learning of a dual lever-pressing task, rats with their posterior dorsomedial striatum (pDMS) lesioned performed far less reliably after training for both levers (see train bars) than the sham-lesioned (SHAM) or anterior DMS (aDMS) lesioned groups. (The test bars show performance on both levers after one has the food reward associated with it devalued.) Non-deval. = non-devalued lever; Deval. = devalued lever.

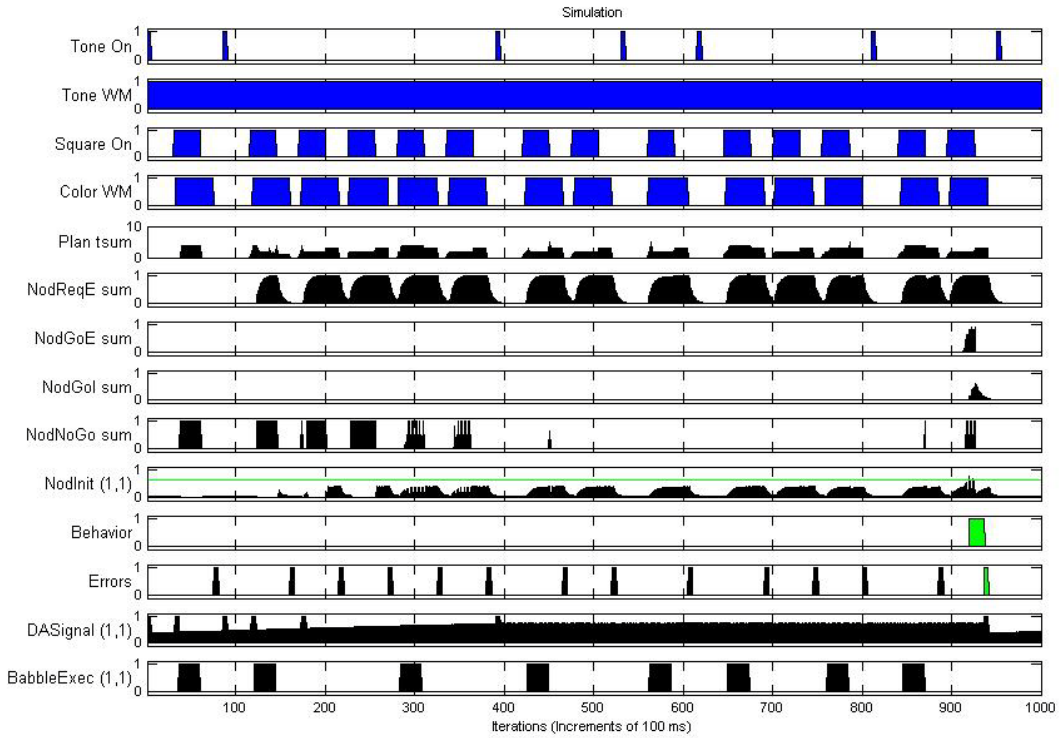


Figure 3.25 Nod/Shake hypo-DA impairment in the model. Extreme hypo-DA in the Nod/Shake pathway leads to raising of thresholds of Nod/Shake Go units, leading to difficulty executing behaviors. It can be seen that the model tries to babble (last trace) several times, and (sixth trace) is constantly executing a Nod Request (probably due to novelty learning). However, in most cases, the Nod Go units fail to activate. The result is that Nod/Shake hypo-DA could suppress babble-generated behaviors, leading to a failure to actually try a response for which it might be rewarded.

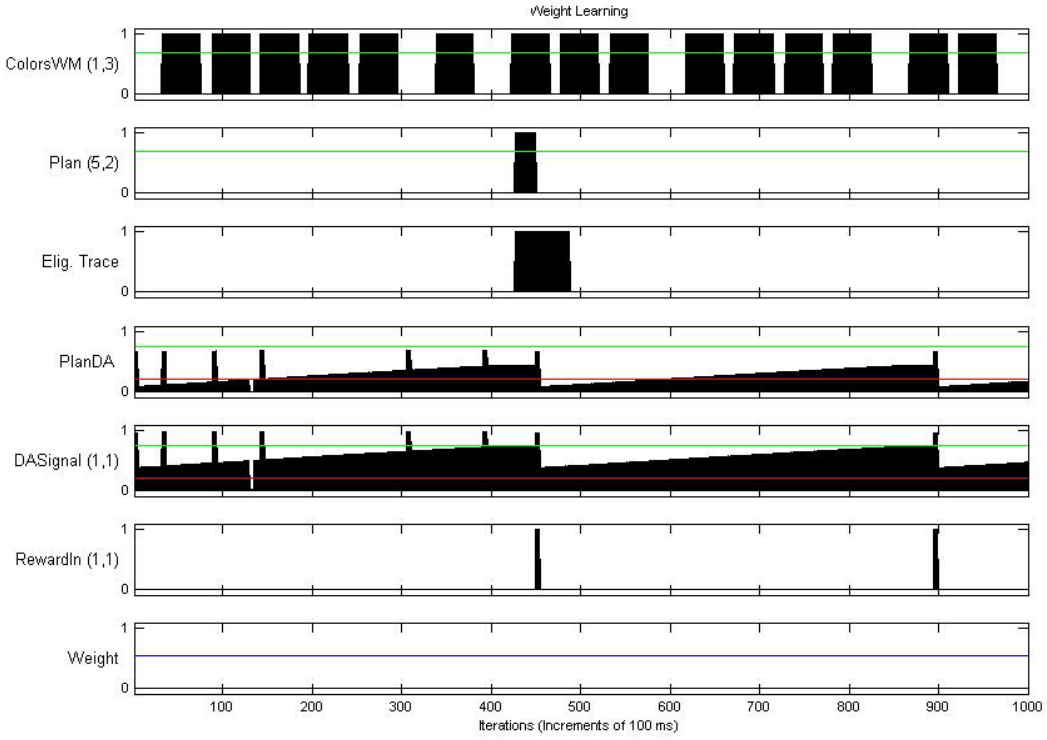


Figure 3.26 Plan/Request hypo-DA impairment in the model. Extreme hypo-DA in the Plan/Request pathway disables necessary LTP in the Plan units. It can be seen that when the eligibility trace is on and a reward is delivered, the DA Signal unit is active over the normal LTP threshold. However, because of the Plan/Request DA antagonism, the Plan DA effect is beneath the LTP threshold, so the Color WM-to-Plan unit weight fails to be incremented.

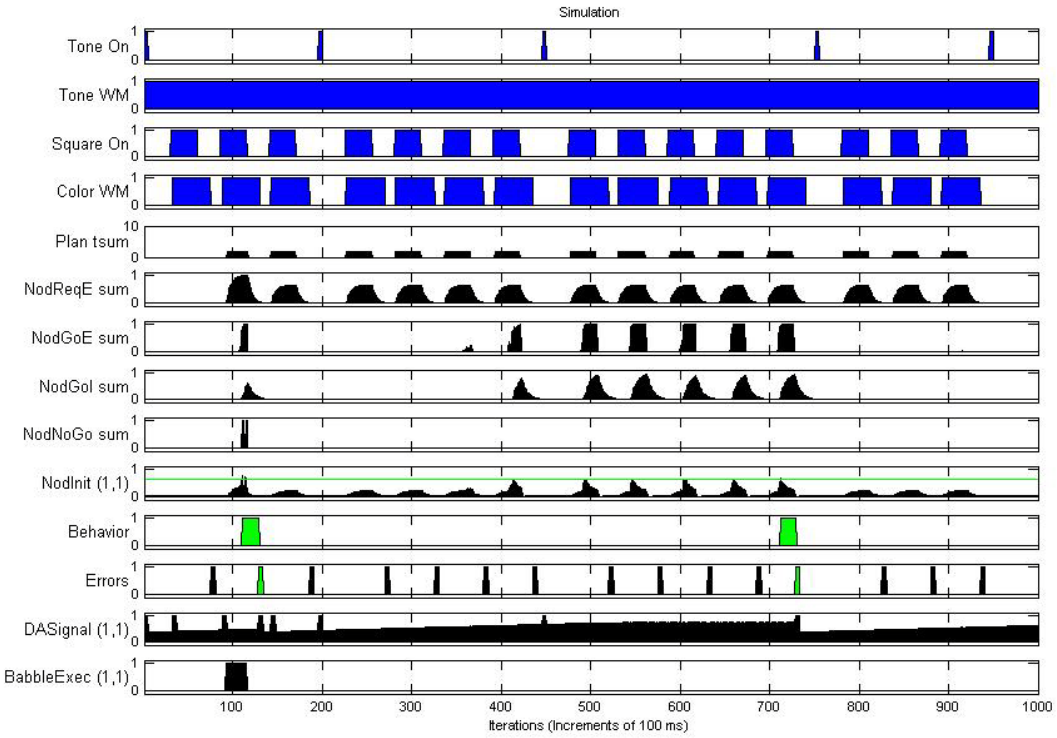


Figure 3.27 Babble hypo-DA impairment in the model. Extreme hypo-DA in the Babble pathway can lead to failure to babble enough to learn the task. The model only tries 1 babble, despite the infrequency at which is rewarded.

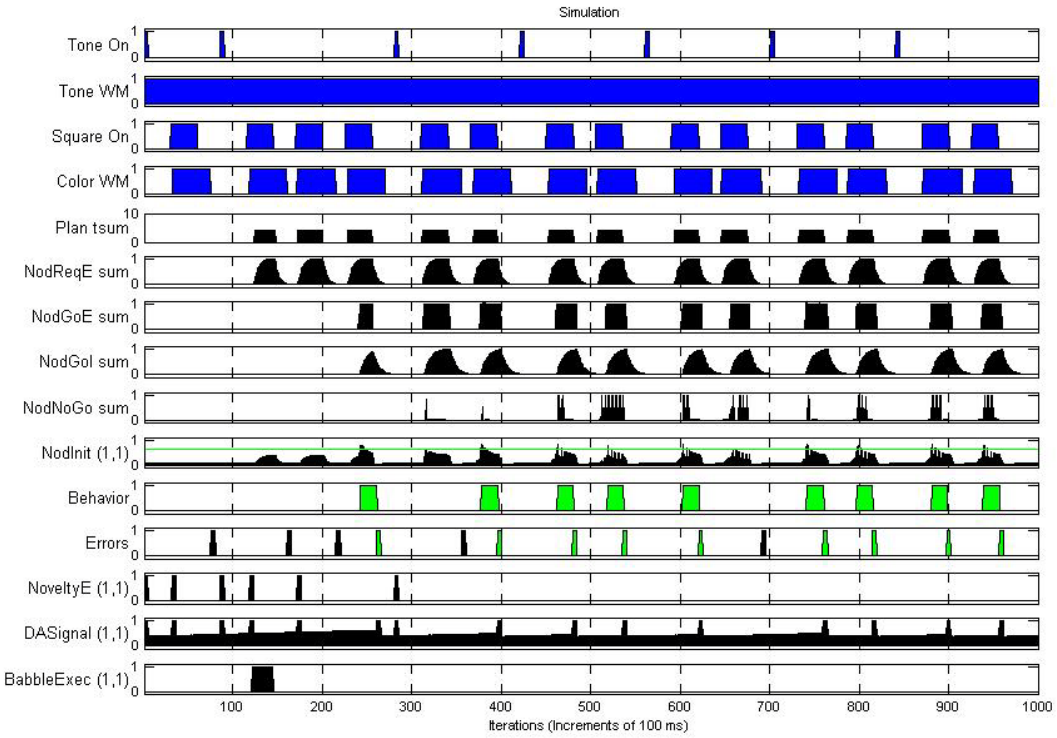


Figure 3.28 Novelty learning offsetting of Babble hypo-DA impairment in the model. Babbles are easily triggered by novelty in the model, so a successful novelty-triggered babble can result in learning of the task despite considerable Babble pathway DA depletion. This probably accounts for the relatively mild learning impairment seen for hypo-DA Babble pathway conditions in Figure 3.22.

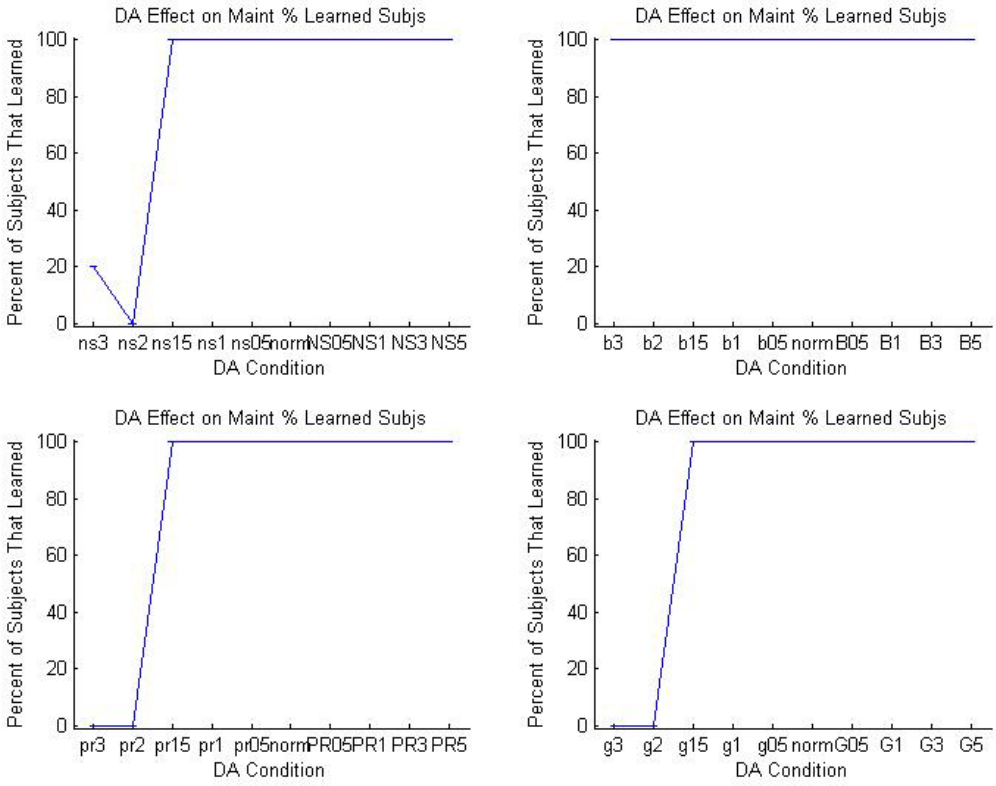


Figure 3.29 DA manipulation effects during the maintenance simulation phase on percent of subjects that retained learning of the task successfully. For each subject, the last 13 trials of the run at the end of the simulation phase are analyzed and the percentage of correct responses is calculated, and if more that 25% of the trials are correct, the subject has retained learning of the task. Each point, then, is the percent (out of 5 subjects) that maintain learning. There are no hyper-DA effects, but, at the two most extreme cases, hypo-DA conditions for the Nod/Shake, Plan/Request, and global cases cause catastrophic unlearning of the task. Babble hypo-DA has no effect, however.

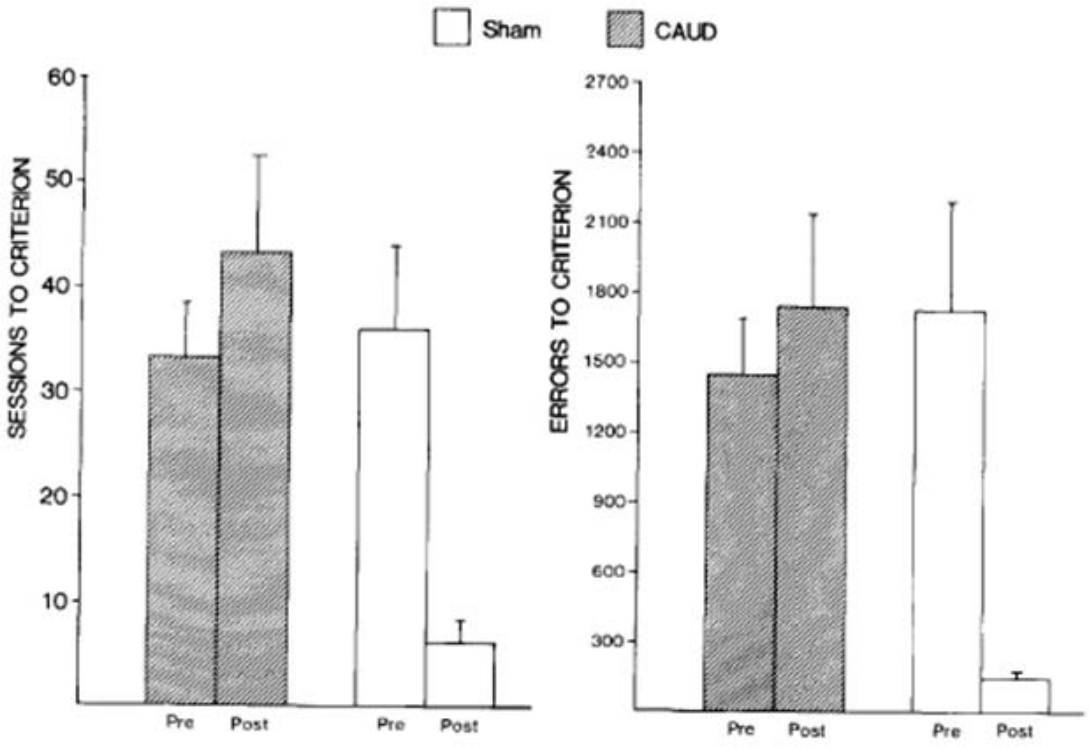


Figure 3.30 Evidence that dorsal striatal DA lesioning leads to impairment of task performance. Taken from Figure 4 of (Robbins et al., 1990). In both control (Sham) and lesion (CAUD) groups, before the surgery (Pre) the number of sessions and errors required before performance criterion is reached is high. After the surgery (Post), the control rats do not require a long relearning phase for the lever-pressing task, but the dorsal striatal-lesioned rats do, suggesting that the task was unlearned due to the DA lesioning.

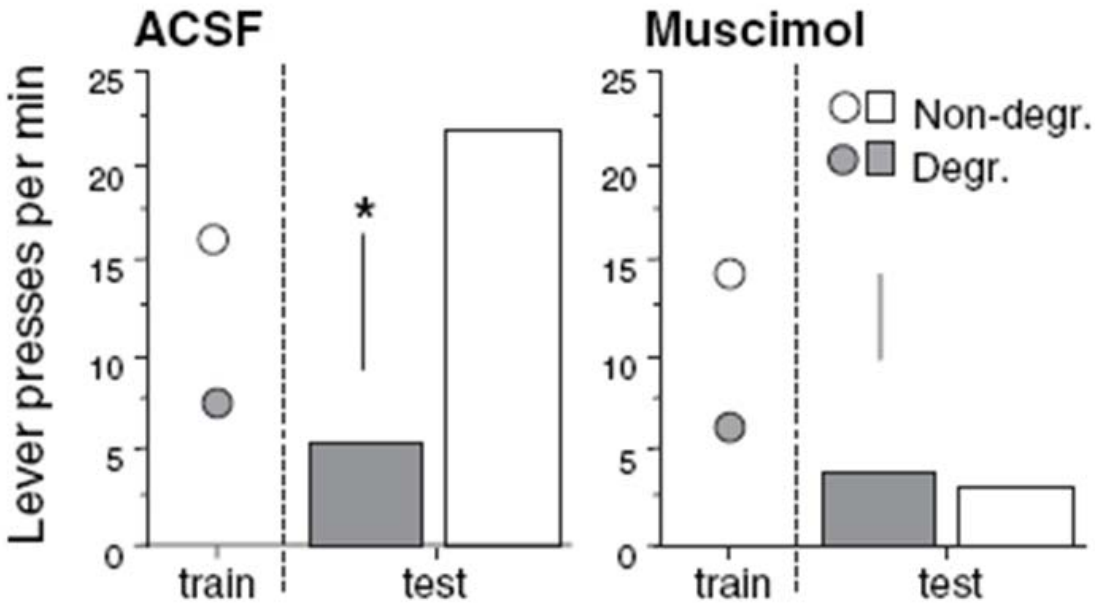


Figure 3.31 Evidence that posterior dorsomedial striatal deactivation leads to reduced lever-pressing task performance. Taken from Figure 8 of (Yin et al., 2005). Both rats that have a sham injunction (ACSF) and those injected with the GABA agonist muscimol exhibit a decreased response lever-pressing response for a lever associated with devalued rewards, but the still-valued reward lever performance is only impaired in the muscimol group. This reinforces the idea that part of the dorsal striatum is involved in learned responses, that the 'actor' pathway passes through the dorsal striatum. Non-deval. = non-devalued lever; Deval. = devalued lever.

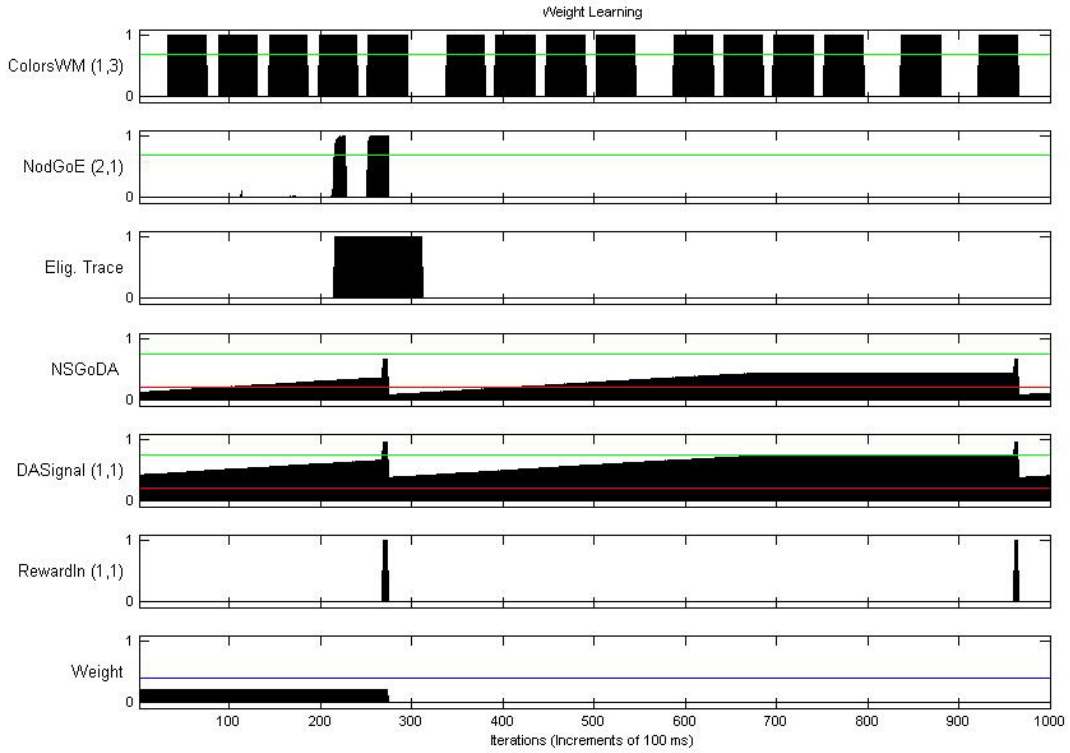


Figure 3.32 Nod/Shake hypo-DA impairment of initial learning in the model. As soon as the model is rewarded for responding correctly, the DA Signal activity drops to its minimum (due to a reset of the Frustration unit). Because the Nod/Shake DA effect level is shifted down, however, the LTD threshold is crossed, and the weights to the Go units unlearn their connections at this point, disrupting the learned task.

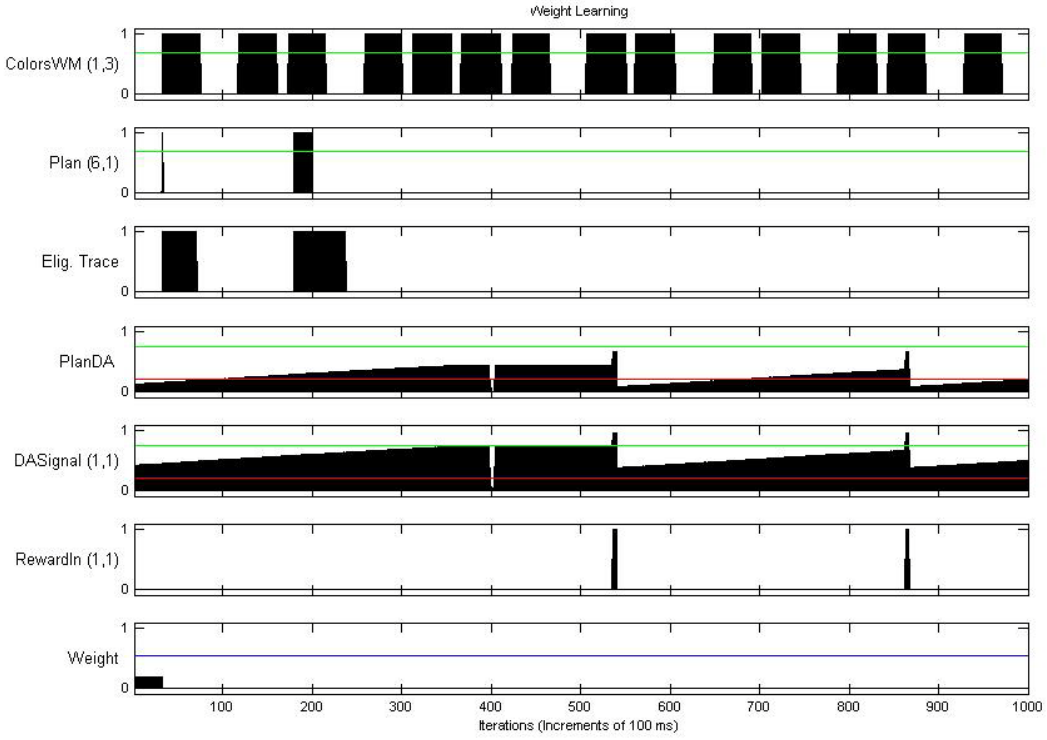


Figure 3.33 Plan/Request hypo-DA impairment of initial learning in the model. The run begins with minimal DA Signal level (due to a start at minimal Frustration activity), and the model rapidly makes a correct response before the DA Signal has a chance to ramp up much. Because the Plan/Request DA effect level is shifted down, however, the LTD threshold is crossed, and the weights to the Plan units unlearn their connections at this point, disrupting the learning of the task.

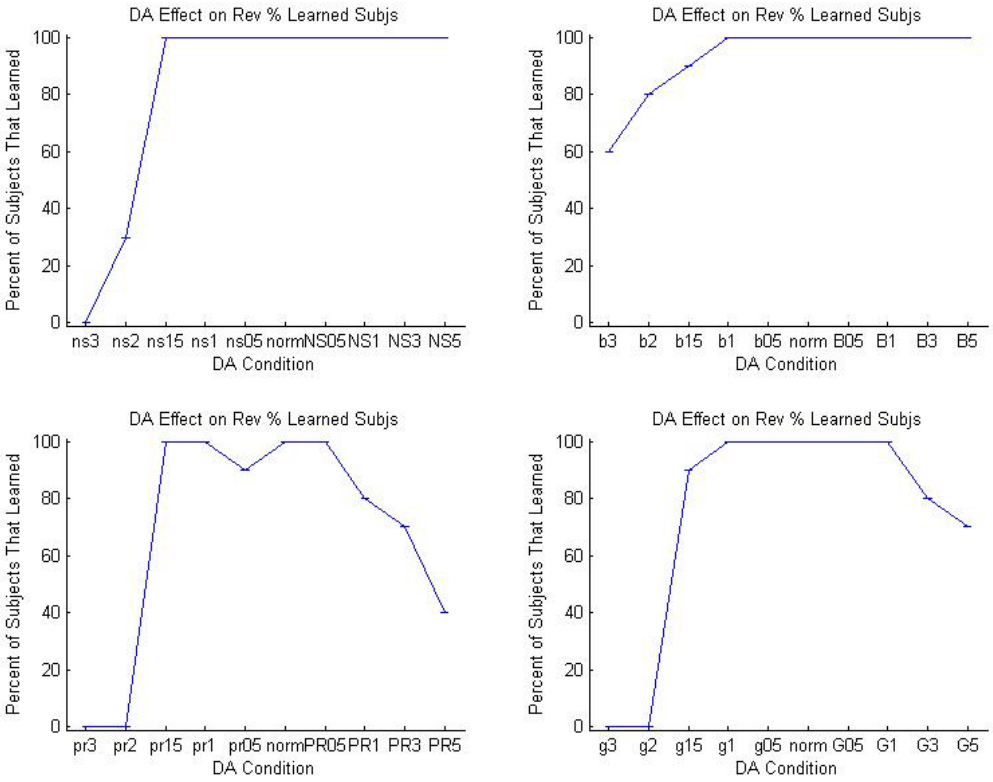


Figure 3.34 DA manipulation effects during the reversal simulation phase on percent of subjects that relearned the task successfully. For each subject, the last 13 trials of the test run at the end of the simulation phase are analyzed and the percentage of correct responses is calculated, and if more that 25% of the trials are correct, the subject has relearned the task. Each point, then, is the percent (out of 10 subjects) that learned the reversal task successfully. Nod/Shake hypo-DA leads to increasing failure to learn the reversal task, with none of the 10 subjects learning it under the ns3 condition. Babble hypo-DA effects are less severe with 6 of 10 subjects learning the task in the worst case. Plan/Request (and global) hypo-DA mostly has an effect at the two most extreme cases, but effectively disables reversal learning. On the other hand, hyper-DA conditions have a more limited impact, with Nod/Shake and Babble hyper-DA having no effect. However, Plan/Request (and global) hyper-DA decreases the chance of task learning. In the Plan/Request case, only 4 of 10 subjects learn the task in the most extreme case, PR5.

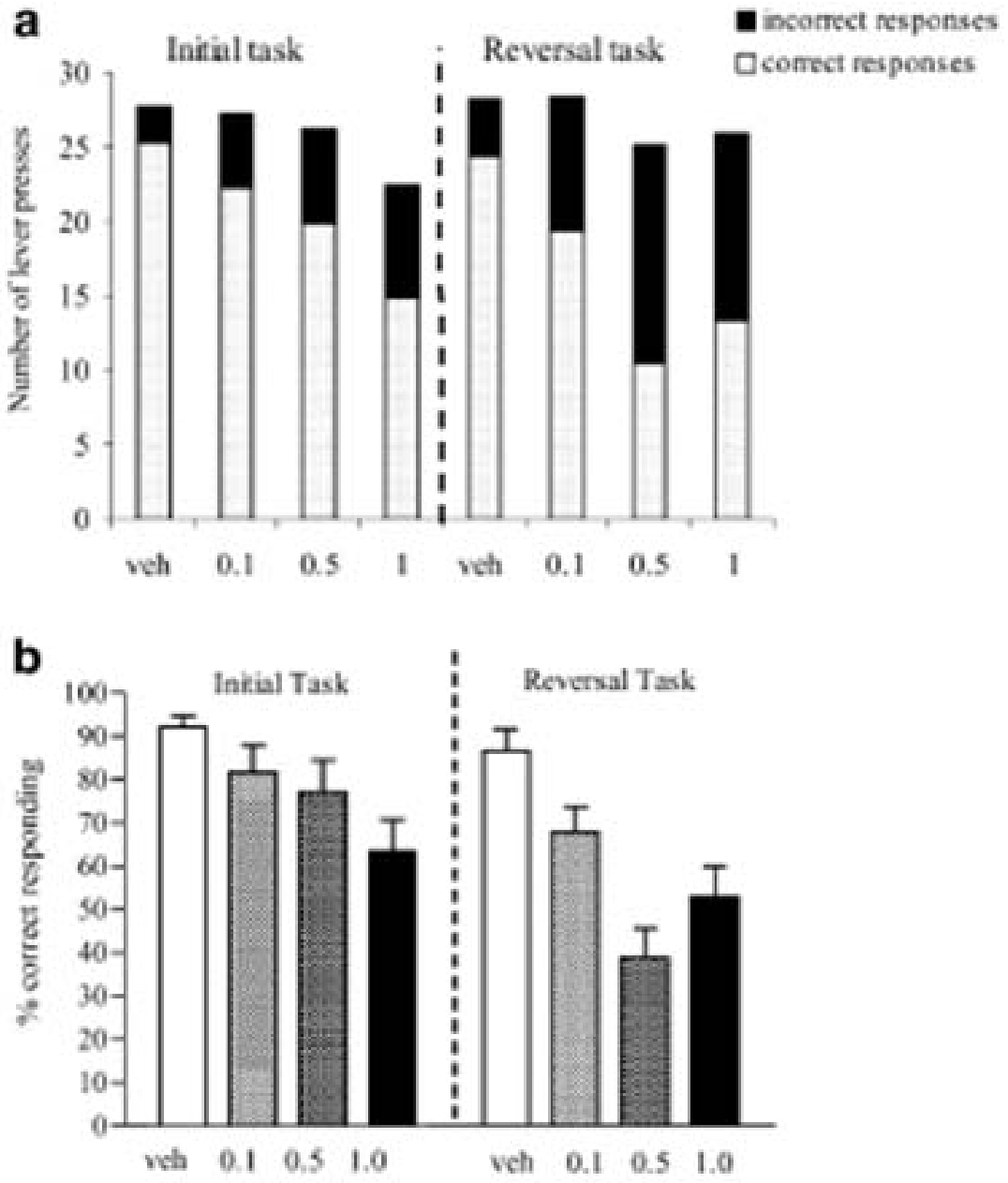


Figure 3.35 Evidence that D-amphetamine impairs both initial and reversal learning. Taken from Figure 5 of (Idris et al., 2005). (a) The number of correct and incorrect responses of rats' lever-pressing at different dosages of D-amphetamine, an indirect DA agonist, in an initial and reversal visual task (responding to the presence or absence of an LED). (b) The percent of correct responses under different amphetamine dosages. Generally, increasing D-amphetamine dosage leads to decreasing correct performance on either the initial or reversal tasks, supporting the role of hyper-DA conditions in impairing task learning and relearning.

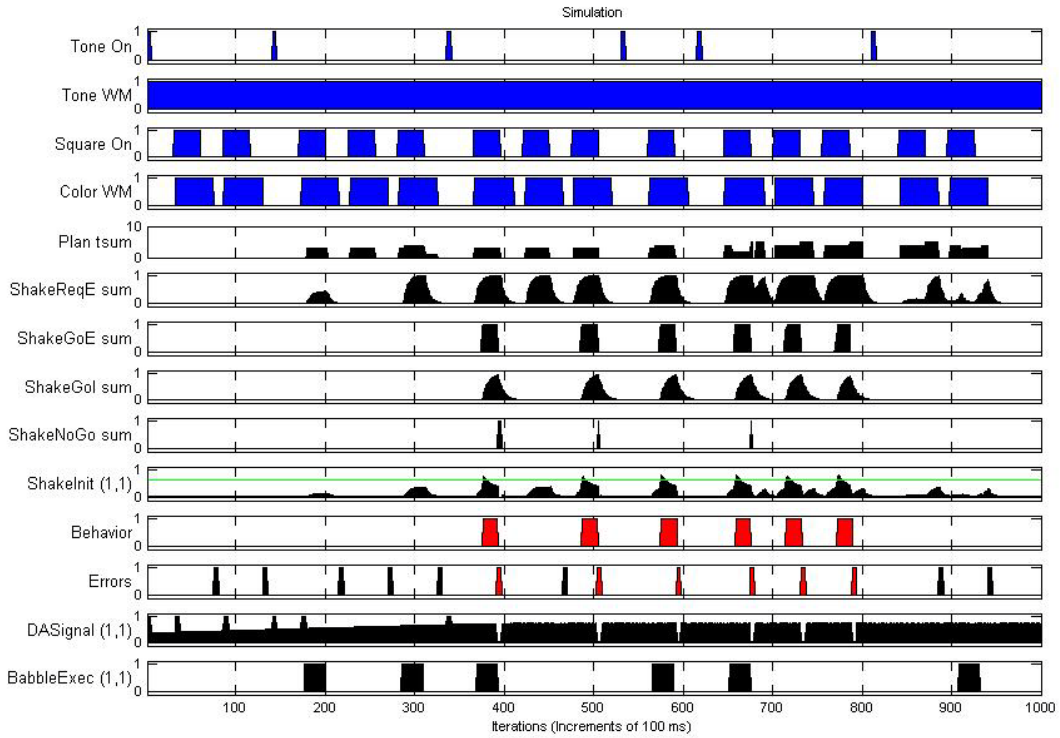


Figure 3.36 Plan/Request hyper-DA impairment of task acquisition. The model, on the third babble, makes a wrong choice and is punished for it. However, it continues to perseverate on the wrong behavior despite repeated punishment.

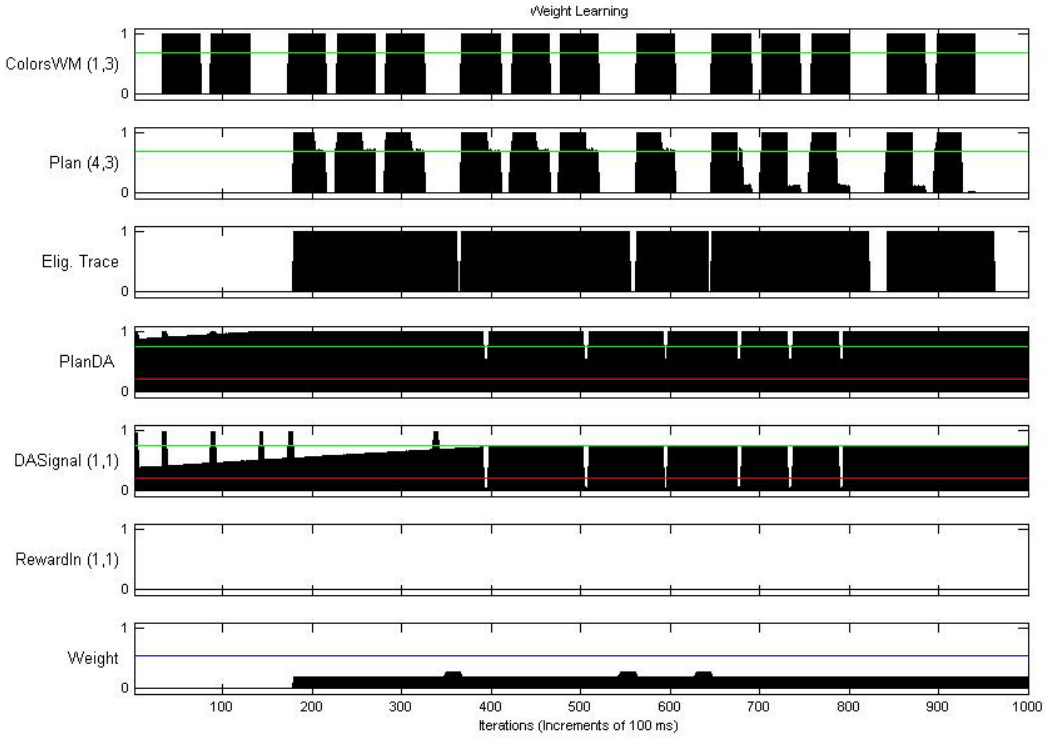


Figure 3.37 Plan/Request hyper-DA generation of perseveration on wrong behaviors. The dynamics of a single Blue WM to Plan unit weight are shown during the simulation run in Figure 3.36. It can be seen that, if the DA conditions shown in the DA Signal unit were to prevail, then LTD would be generated during the punishers. However, the hyper-DA conditions cause the Plan DA effect to be shifted higher so that LTD fails to be engaged during punisher DA dips, leading to a failure to unlearn wrong behaviors. Also of note is the fact that the Plan DA effect level is, by default, chronically over the LTP threshold, as if it were in a state of constant reward.

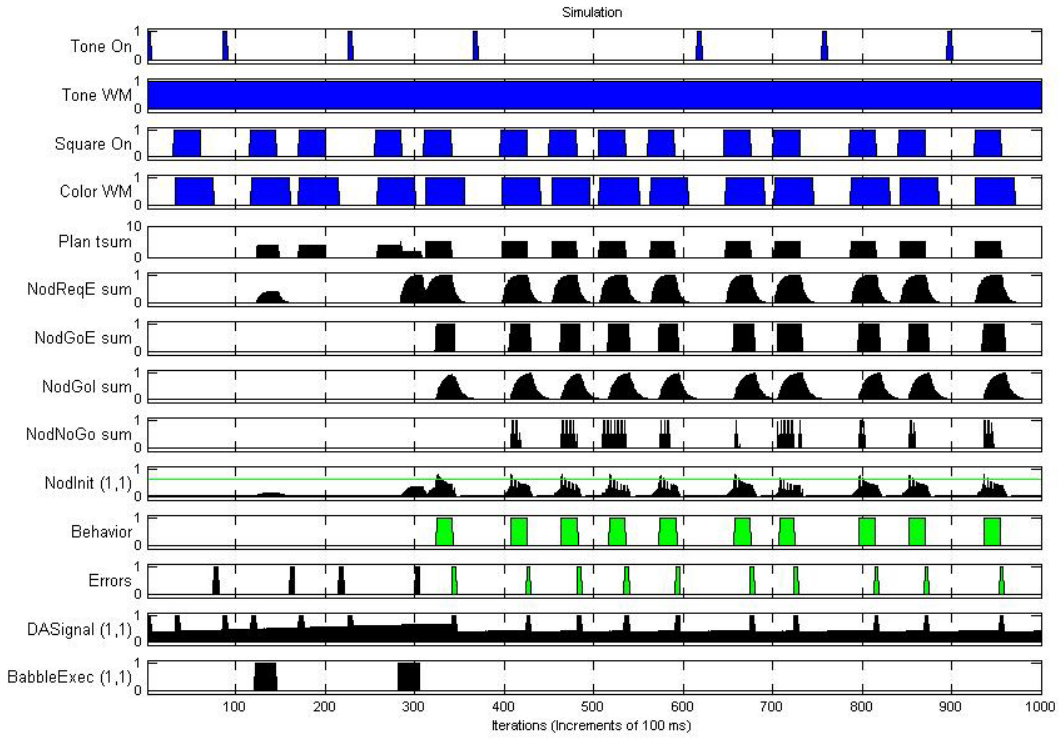


Figure 3.38 Lucky first guess overcoming Plan/Request hyper-DA effects. On the second babble, the model guesses the correct response, thereby avoiding the problem of wrong-response perseveration.

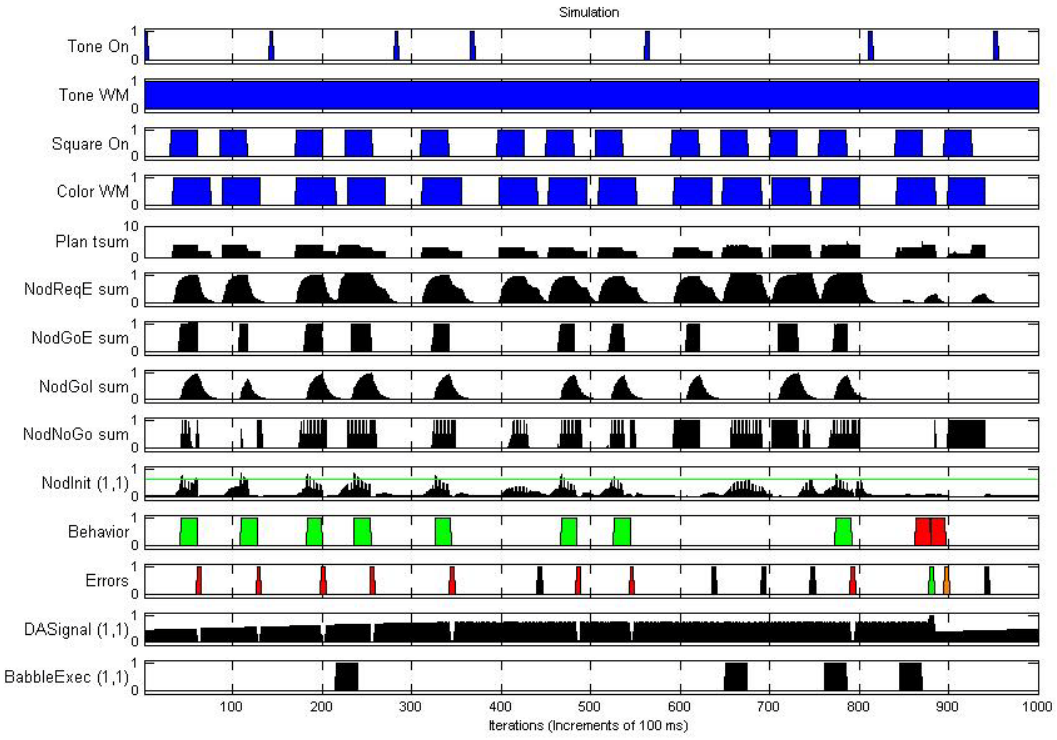


Figure 3.39 Plan/Request hyper-DA impairment of task reversal learning. The model continues to make the initially learned, but now incorrect, response for some time, despite repeated punishment, until a babble selects the right behavior.

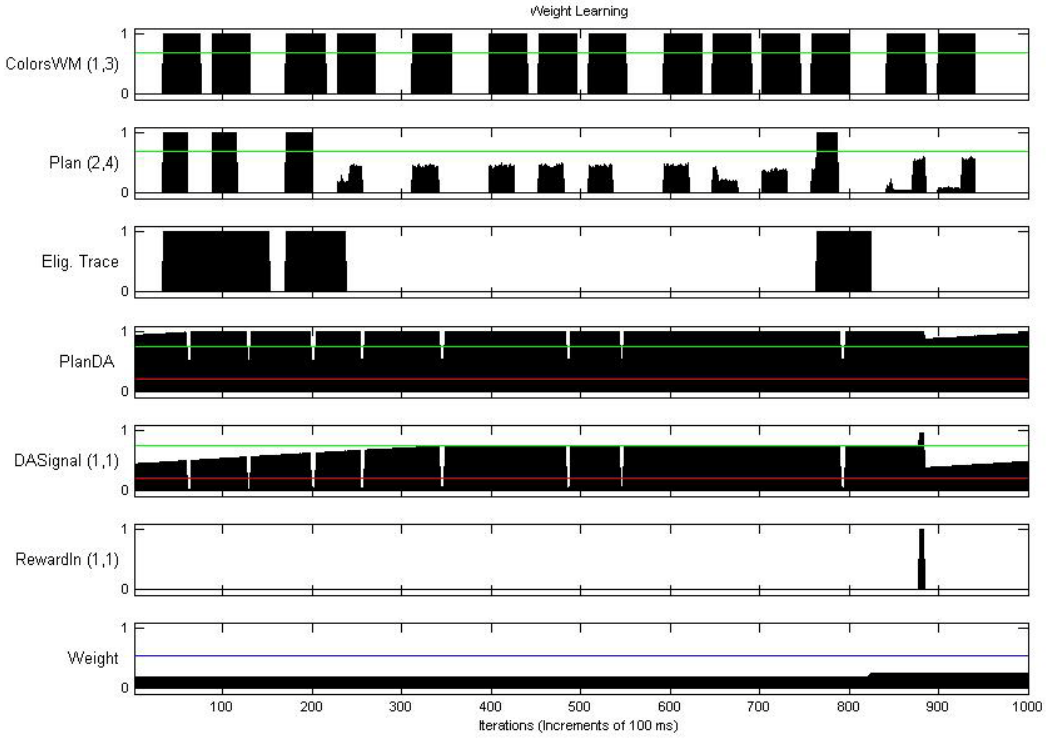


Figure 3.40 Plan/Request hyper-DA generation of perseveration on wrong behaviors during reversal learning. A weight is shown for the run in Figure 3.39 from the Blue WM to a Plan unit. As with the hyper-DA initial acquisition effect shown in Figure 3.37, the Plan DA hyper-DA effect leads to the model never undergoing LTD during the punisher phasic dips. Rather, the model is chronically in an LTP state.

CURRICULUM VITAE

George L. Chadderdon

Institutional Address:
Department of Psychological and Brain Sciences
1101 E. 10th Street
Indiana University
Bloomington, IN 47405
gchadder@indiana.edu

Home Address:
1083 Macomb Road
Galesburg, IL 61401
(812) 322-6947
gchadder3@gmail.com

Education:

Indiana University, Bloomington
Ph.D., Psychology and Cognitive Science with additional Certificate in Modeling in Cognitive Science, March 2009
Specialized Coursework: Cognitive neuroscience, computational neuroscience, neuropsychology, philosophical foundations of cognitive science, dynamical systems, mathematical psychology, partial differential equations

University of California, San Diego
M.S., Computer Science, June 1995.
Specialized Coursework: Artificial intelligence, neurocomputation, pattern recognition, speech recognition, intelligent systems

University of Illinois, Urbana-Champaign
B.S., Computer Engineering, May 1993.

Illinois Mathematics and Science Academy
Charter Class, June of 1989.

Research Interests:

Computational cognitive neuroscience: executive / volitional processing, attention, affective processing, learning, embodied cognition, spatiotemporal pattern recognition, language processing

Research Experience:

Indiana University, Bloomington
Advisor: Dr. Olaf Sporns
In doctoral dissertation, *A Neurocomputational Model of the Functional Role of Dopamine in Stimulus-Response Task Learning and Performance*, developed an integrative theory and corresponding model of neural substrate and dopaminergic mechanisms for reinforcement learning of simple stimulus-response tasks.

Developed a neurocomputational model (using Matlab) to investigate tonic dopaminergic mechanisms of task-oriented behavior selection and working memory in prefrontal cortex. Coauthored publication with Dr. Sporns for Journal of Cognitive Neuroscience. (May 2002-Feburary 2006)

Sony Electronics, San Diego, CA
Research and development of hidden-Markov model-based speaker-independent natural language recognition for consumer electronics products. (1999)

ORINCON Corporation, San Diego, CA
Research and development of algorithms for phoneme segmentation, signal detection and classification, target tracking, and threat evaluation for defense and commercial applications. Methodologies: neural networks, evolutionary programming, fuzzy logic, hidden-Markov models, Kalman filtering. (1994-1999)

University of California, San Diego
Developed hidden-Markov model audiovisual speech recognition architecture. (Paper coauthored with Dr. Movellan.) (1994)

Awards:

Indiana University Cognitive Science Program Summer Research Fellowships (5/2005-9/2005, 5/2006-9/2006), National Institute of Health training grant fellowship in Modeling in Cognition (9/2002-8/2003, 9/2004-8/2005), Toastmasters International CTM, Edmund James Scholar (1989/1990, 1990/1991), Dean's List (1989/1990, Fall 1990, Spring 1993).

Publications (Journal Papers):

Chadderdon, G. L., & Sporns, O. (2006). A large-scale neurocomputational model of task-oriented behavior selection and working memory in prefrontal cortex. *Journal of Cognitive Neuroscience, 18(2)*, 242-257.

Chadderdon, G. L. (2008). Assessing machine volition: an ordinal scale for rating artificial and natural systems. *Adaptive Behavior*, 16(4), 246-263.

Publications (Conference Papers):

Chadderdon, G., & Movellan, J. R. (1995). Testing for channel independence in bimodal speech recognition. In *Proceedings of the 2nd Joint Symposium on Neural Computation, University of California, San Diego and California Institute of Technology* (pp. 84-90). San Diego, CA: University of California, San Diego.

Brotherton, T. W., & Chadderdon, G. (1998). Automated rule extraction for engine health monitoring. In V. W. Porto, N. Saravanan, D. Waagen & A. E. Eiben (Eds.), *Evolutionary Programming VII: Proceedings of the 7th International Conference, EP98, San Diego, California, March 1998* (pp. 725-734). New York, NY: Springer.

Brotherton, T., Johnson, T., & Chadderdon, G. (1998). Classification and novelty detection using linear models and a class-dependent elliptical basis function neural network. In *Proceedings of International Joint Conference on Neural Networks, Anchorage, Alaska, May 1998.*

Brotherton, T., Chadderdon, G., & Graybill, P. (1999). Automated rule extraction for engine vibration analysis. In *Proceedings of the IEEE Aerospace Conference, Aspen, Colorado, March 1999.*

Shea, P., Owen, M., & Chadderdon, G. (1999). Fuzzy control in the deployable autonomous distributed system. In *Proceedings of SPIE: Signal Processing, Sensor Fusion, and Target Recognition VIII, Orlando, Florida, 1999.*

Publications (Abstracts):

Sporns, O., Bulwinkle, D., Chadderdon, G., & Alexander, W. H. (2003). Neuro-robotic models of learning and addiction. *NIH Symposium (Biomedical Information Science and Technology Initiative) Digital Biology: The Emerging Paradigm, Bethesda, Maryland, November 2003.*

Chadderdon, G., & Sporns, O. (2003). A large-scale network model of working memory and neuromodulation. *Society for Neuroscience Abstract.* (presented as poster in New Orleans)

Chadderdon, G., & Sporns, O. (2005). A large-scale neurocomputational model of task-oriented behavior selection and working memory in prefrontal cortex. *Society for Neuroscience Abstract.* (presented as poster in Washington, D.C.)

Courses Taught:

P153 (Introduction to Psychology Laboratory I): Fall semester 2005 at Indiana University: A psychology undergraduate introductory research methods class, taught from my own lesson plan.

Activities:

Interests include: creative writing, music (guitar and composition), philosophy, ballroom dance.

Updated February 28, 2009