Simultaneous Activation of Multiple Memory Systems during Learning: Insights from Electrophysiology and Modeling

By

Catherine Ann Thorn

S.M. Electrical Engineering
Massachusetts Institute of Technology, 2004

B.S. Electrical Engineering
Georgia Institute of Technology, 2002

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2010

Signature of Author:_
Department of Electrical Engineering and Computer Science
May 21, 2010

Certified by: _____
Ann M. Graybiel
Institute Professor and Professor in Brain and Cognitive Sciences
Thesis Supervisor

Accepted By: _____
Terry P. Orlando
Professor of Electrical Engineering
Chair, Committee on Graduate Students

Simultaneous Activation of Multiple Memory Systems during Learning: Insights from Electrophysiology and Modeling

By

Catherine Ann Thorn

Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Electrical Engineering and Computer Science

ABSTRACT

Parallel cortico-basal ganglia loops are thought to give rise to a diverse set of limbic, associative and motor functions, but little is known about how these loops operate and how their neural activities evolve during learning. To address these issues, single-unit activity was recorded simultaneously in dorsolateral (sensorimotor) and dorsomedial (associative) regions of the striatum as rats learned two versions of a conditional T-maze task. The results demonstrate that contrasting patterns of activity developed in these regions during task performance, and evolved with different training-related dynamics.

Oscillatory activity is thought to enable memory storage and replay, and may encourage the efficient transmission of information between brain regions. In a second set of experiments, local field potentials (LFPs) were recorded simultaneously from the dorsal striatum and the CA1 field of the hippocampus, as rats engaged in spontaneous and instructed behaviors in the T-maze. Two major findings are reported. First, striatal LFPs showed prominent theta-band rhythms that were strongly modulated during behavior. Second, striatal and hippocampal theta rhythms were modulated differently during T-maze performance, and in rats that successfully learned the task, became highly coherent during the choice period.

To formalize the hypothesized contributions of dorsolateral and dorsomedial striatum during T-maze learning, a computational model was developed. This model localizes a model-free reinforcement learning (RL) system to the sensorimotor cortico-basal ganglia loop and localizes a model-based RL system to a network of structures including the associative cortico-basal ganglia loop and the hippocampus. Two models of dorsomedial striatal function were investigated, both of which can account for the patterns of activation observed during T-maze training. The two models make differing predictions regarding activation of the dorsomedial striatum following lesions of the model-free system, depending on whether it serves a direct role in action selection through participation in a model-based planning system or whether it participates in arbitrating between the model-free and model-based controllers.

Combined, the work presented in this thesis shows that a large network of forebrain structures is engaged during procedural learning. The results suggest that coordination across regions may be required for successful learning and/or task performance, and that the different regions may contribute to behavioral performance by performing distinct RL computations.

Thesis Supervisor: Ann M. Graybiel
Title: Institute Professor and Professor in Brain and Cognitive Sciences

Acknowledgments

This work would not have been possible without a number of people whose contributions deserve more recognition than this page can provide. First among these of course is Ann, who supported the experiments presented in the following pages and in a multitude of other ways contributed to my development as a scientist. The insights of my committee members, Steve Massaquoi and Leslie Kaelbling, were crucial in the completion of the computational work presented in Chapter 4, and thus to my development as an engineer. A special thanks is due to Eric Grimson for his additional participation in the eleventh hour.

A number of members of the Graybiel Lab family were an indispensable part of this work. Bill DeCoteau, Terra Barnes and Hu Dan went out of their way to teach me the techniques required for these experiments. Other members of the rodent group, including Terra, Hu Dan, Ledia Hernandez, Mark Howe and Hisham Atallah made sure that rats were fed, equipment was working, and experiments kept running rain or shine, sickness or health, during conferences, holidays and the occasional weekend. Without their help, I'd likely still be collecting data. A number of scientific discussions with my officemates Terra, Theresa Derochers, and Joey Feingold, with my co-authors, especially Mark and Hisham, and with everyone else in the Graybiel Lab, but especially Kyle Smith and Patrick Tierney, contributed to the shaping of the ideas presented here. I am particularly indebted to Hisham, Kyle, Ledia, Patrick, Leif Gibb, Dan Gibson, and Yasuo Kubota for their help in proofreading a draft of this document.

Many people provided invaluable technical support and without them the work could not have been done. The veterinary expertise of Bob Marini and the routine animal care by the DCM technicians maintained the health and happiness of the rats. Henry Hall helped with the photography, with maze building and with other construction odds and ends. Pat Harlan and Christine Keller-McGandy performed all the histology. Brandy Baker, Emily Romano, and Gila Fakterman helped assemble the figures in Chapters 2 and 3 for publication. Dan Gibson provided analysis tools, troubleshooting of software and electronics, and patiently endured countless inquiries about data analysis techniques. Henry, Clark Brayton, Jules Bodden, and Yasuo Kubota helped make sure that rules were followed, that deadlines were met, and that the details of lab life were attended to, enabling the rest of us to focus on the experiments.

The work presented in Chapter 2 was funded by NIH/NIMH grant MH60379, ONR grant N000140410208, the Stanley H. and Sheila G. Sydney Fund, European Union grant 201716 and a fellowship from the Friends of the McGovern Institute for Brain Research. The work presented in Chapter 3 was supported by NIH/NIMH grants MH60379 and MH071744.

Finally, without the support of my friends and family over the years I would not have made it this far. Thank you for enduring the process with me.

TABLE OF CONTENTS

INTRODUCTION

The basal ganglia are important for normal behavior and have been implicated in a number of diseases. Best known and longest-studied are the motor deficits caused by neurodegenerative diseases that differentially target basal ganglia nuclei; these include Parkinson's disease and Huntington's disease. More recently, the cognitive deficits in these diseases have gained increasing attention, and the basal ganglia have been implicated in non-motor diseases including obsessive-compulsive disorder and Tourette syndrome. Anatomical connectivity, clinical observations, behavioral and electrophysiological studies are beginning to converge on a role for the basal ganglia in motor and cognitive "action" selection, and this background is summarized in Chapter 1.

The largest input structure of the basal ganglia is the striatum, which has been roughly divided into motor, associative and limbic regions, corresponding to dorsolateral, dorsomedial and ventral striatum in the rat. It has been hypothesized that each of these regions and their associated cortico-basal ganglia loops plays a different functional role in behavioral control, and each may direct behavior during different time periods during procedural learning and skilled performance. It is unknown, however, how the neural activity in each of these regions gives rise to behavior and how these activities may be modulated across learning. Chapter 2 describes the results of experiments in which neural activity was recorded simultaneously from dorsolateral and dorsomedial striatal regions during learning and skilled performance on a T-maze task. The results demonstrate that markedly different patterns of activity develop in each region during training, supporting their distinct functional roles. They further demonstrate that the patterned activity in each region develops with different time courses during learning, though both can be strongly active simultaneously for most of training. A novel scheme is proposed whereby the dorsomedial striatal loop modulates access to behavioral control by the dorsolateral loop, likely through competition of the two activities at downstream targets.

The striatum is one of a number of learning and memory systems in the brain, and the procedural/motor learning supported by the striatum is often contrasted with the episodic/spatial memory supported by the hippocampus. Chapter 3 describes the results of recording experiments in which local field potentials were simultaneously recorded in dorsomedial striatum and hippocampus during learning on a T-maze task. These results show strong oscillations in the theta-band in the striatum during task performance – a result that is at odds with much work suggesting that basal ganglia oscillations appear only during pathological states, and suggests that low frequency rhythms are characteristic of healthy behavioral states as well. Additionally, these results show dynamic modulation of the coherence between striatal and hippocampal theta-band oscillations during task performance, with the strongest coherence expressed during the "decision period" of the task. This pattern of striatal-hippocampal coherence is expressed only in animals that learn the task, and is evident even before good performance is reached, suggesting that cross-structure communication is necessary for learning on the T-maze.

Recently, with the discovery of reward prediction error signalling by the dopamine-containing neurons of the midbrain, attention has focused on reinforcement learning theory and how it may relate to learning mechanisms implemented by the brain. The striatum is intimately interconnected with the dopamine neurons, which likely provide a teaching signal during procedural learning. A number of computational modeling studies have formalized how state- or action-value functions may be computed by the striatum and how the computation of such values may contribute to the key functions of the basal ganglia in movement, procedural learning and habit formation. In Chapter 4, the basic concepts of reinforcement learning (RL) are reviewed, together with their applications to

basal ganglia research. Extending RL concepts to the experimental results presented in Chapter 2, two RL based conceptualizations of striatal function are suggested that can account for the observed patterns of dorsolateral and dorsomedial activation. The first of these proposes that the dorsomedial striatum may be directly engaged in action selection through the computation of action-values according to a model-based planning scheme. The second of these proposes instead that the dorsomedial striatum may be involved in the arbitration between competing model-based and model-free controllers. Future experiments are proposed that may differentiate between the two possibilities.

Combined, the work presented in this thesis shows that a large network of forebrain structures, including learning and memory systems in the dorsolateral striatum, the dorsomedial striatum and the hippocampus, are differentially active during normal procedural learning. These results suggest that coordination across these widely separated and functionally distinct regions may be required for successful learning and/or task performance, and suggest ways in which these different regions may contribute to reinforcement-based learning.

1. Anatomical and physiological evidence for the role of the basal ganglia in motor and non-motor functions

It has long been known that the motor dysfunction seen in Parkinson's disease and Huntington's disease are the result of degeneration of different nuclei within the basal ganglia. More recently, the cognitive effects of basal ganglia dysfunction in diseases such as Tourette syndrome and obsessive-compulsive disorder are gaining more attention. In this Chapter, we summarize background evidence from clinical observations, anatomical studies, and behavioral and electrophysiological experiments, which combined point to a role for the basal ganglia in the selection and evaluation of motor and cognitive actions.

1.1. Basal ganglia subnuclei

The basal ganglia are subcortical nuclei. Current definition of the basal ganglia includes four nuclei and their component subdivisions: the striatum (or "neostriatum", consisting of the caudate nucleus and the putamen), the subthalamic nucleus, the globus pallidus (internal and external segments) and the substantia nigra (pars reticulata and pars compacta subdivisions). Below, we briefly describe the global connectivity of these structures, as they have been defined in primates, including humans. These are illustrated in **Figure 1.1**. Rodents have an analogous set of structures, with overlapping nomenclature, which are mentioned as appropriate.

1.1.1. Striatum

The striatum is the largest input structure of the basal ganglia. It receives topographically organized excitatory input from most areas of cortex and associated thalamic regions onto medium-sized spiny neurons (or "medium spiny neurons," MSNs). The medium spiny neurons then send inhibitory output to the globus pallidus, internal and external segments, as well as the substantia nigra, both pars reticulata and pars compacta segments. Additional modulatory input to the striatum comes from the dopamine-containing neurons of the substantia nigra pars compacta. Finally, numerous cell types intrinsic to the striatum also modulate the firing of the medium spiny input/output neurons. The neurochemical structure of the striatum, including its intrinsic neuron types, is discussed further in Section 1.5.

In primates, the striatum refers to the combination of two structures: the caudate nucleus and the putamen. These are separated by a bundle of descending cortical fibers called the internal capsule. In the rodent, this descending fiber bundle is not as prominent, thus the caudate and putamen are not distinguishable. The single combined structure is referred to as the caudoputamen, or often simply the "striatum."

1.1.2. Subthalamic Nucleus

The subthalamic nucleus (STN) is the other input structure of the basal ganglia and like the striatum, receives excitatory input from most regions of the cortex and nuclei of the thalamus. Unlike the striatum, the STN also receives prominent input from the globus pallidus external segment, as well as the pedunculopontine nucleus of the brainstem. STN sends excitatory projections to the globus

pallidus, internal and external segments, and substantia nigra pars reticulata, though STN projections to these regions tend to be more diffuse than those from striatum. Dopamine neurons of the substantia nigra pars compacta also provide modulatory input to the STN.

1.1.3. Globus Pallidus

The internal segment of the globus pallidus (globus pallidus interna, GPi), along with the substantia nigra pars reticulata (SNr), is the output structure of the basal ganglia. In rodents, the analogous structure is called the entopeduncular nucleus. GPi receives diffuse excitatory input from STN and targeted input from the striatum such that the cortico-striatal topography is preserved through the GPi. Additional inhibitory input comes from the globus pallidus external segment (GPe). Most neurons in GPi send inhibitory projections to the thalamus, which sends projections back to the cortex, thus completing a cortico-basal ganglia-thalamocortical loop. This loop architecture and its implications for cortico-basal ganglia function are discussed further in Sections 1.2 and 1.3. The same neurons that send output to the thalamus branch and send projections to brainstem nuclei including the pedunculopontine nucleus (or midbrain extrapyramidal area) which connects basal ganglia output to the reticulospinal motor system, involved in coordinating automatic movements such as posture and walking, mediating autonomic and pain functions, and facilitating/inhibiting voluntary movements.

The external segment of the globus pallidus (globus pallidus externa, GPe), receives input from and sends output to other basal ganglia nuclei. It receives inhibitory input from the striatum and STN, and sends inhibitory projections back to STN as well as to the output nuclei, GPe and SNr.

1.1.4. Substantia Nigra

The substantia nigra pars reticulata (SNr) is similar to GPi in its connections, receiving input from striatum and STN as well as GPe and sending outputs to thalamic and brainstem nuclei. Because of their similarities in connectivity as well as anatomical and chemical features, SNr and GPi are often considered together as a single basal ganglia output structure.

The substantia nigra pars compacta (SNc) is one of several closely connected midbrain nuclei that contain dopamine neurons. The dopamine neurons of the SNc project predominately to the striatum, and to a lesser extent the other basal ganglia nuclei (GPe, GPi, STN, SNr). The effects of the dopaminergic inputs to the striatum are discussed in more detail in Section 1.4

1.1.5. Summary

In this section, the basic connectivity of the basal ganglia nuclei was summarized. The basal ganglia consist of four component nuclei, each of which can be further subdivided based on anatomical appearance, connection patterns and/or chemical makeup. The input structures are the striatum, including caudate nucleus and putamen, and the subthalamic nucleus. The output structures are the globus pallidus internal segment and the substantia nigra pars reticulata, which are often considered as one structure. The external segment of the globus pallidus is a structure entirely internal to the basal ganglia – it receives input exclusively from and sends projections exclusively to other basal ganglia nuclei. The following section describes the topographical connections from cortex, through the basal ganglia, to thalamic targets and the functional consequences of this parallel loop

architecture. Section 1.3 then describes the direct and indirect pathways through the basal ganglia that arise from this pattern of connectivity and their role in action selection.

1.2. Parallel architecture of cortico-basal ganglia-thalamic loops

Almost all regions of the cortex project to the striatum. These projections are organized topographically – specific cortical regions project to specific striatal regions – and this topography is preserved throughout cortico-striatal, striato-pallidal, and pallido-thalamic projections. This pattern of projections has led to the recognition of multiple cortico-basal ganglia-thalamic loops, and the hypothesis that these loops are, for the most part, anatomically and functionally segregated. The similar layout of these multiple circuits makes it likely that the computations performed at each level are similar, but performed on the different types of information being transmitted through the individual loops. In this section, further detail is provided on the three commonly considered loops (motor, limbic and associative), their functions and modes of interaction. It should be noted that although these three major divisions are generally well accepted, there are no precise boundaries between loops and their divisions can therefore be somewhat arbitrary. Additionally, the motor, limbic and associative loops can each be further subdivided. Nonetheless, it will be useful to consider these three broad functional categories.

1.2.1. Motor

The involvement of the basal ganglia in movement has been well known since the clinical observation that Parkinson's disease, which causes debilitating motor symptoms, was linked to degeneration of the dopamine-containing neurons of the substantia nigra pars compacta. Subsequent investigations have made the motor loop the most studied and best characterized of the three loops. Somatosensory and motor cortical areas, including the arcuate premotor and supplementary motor areas (which, like motor cortex, also project to the spinal cord), send projections to dorsolateral portions of caudate and putamen. These cortico-striatal projections are further organized such that hand, trunk and limb representations in motor and somatosensory cortices converge onto hand, trunk, and limb regions, respectively, in the striatum. In this way, somatotopic maps are preserved in each structure, and it has been suggested that even further subdivision may be possible. Projection neurons in the striatal motor region form synapses with neurons in the ventrolateral portion of both internal and external segments of the globus pallidus, preserving somatotopy at the next level of processing in the basal ganglia. The ventrolateral GPi then sends projections to the ventrolateral nucleus of the thalamus, which sends projections back to motor, premotor and supplementary motor areas, completing the motor loop. In the rodent, the analogous loop begins in the motor and somatosensory cortex, which sends projections to the dorsolateral portion of the caudoputamen, which then projects to lateral pallidal structures, ventrolateral thalamus and back to sensorimotor cortical regions.

In their now-classic paper, Alexander, DeLong and Strick (1986) point out that striatal stimulation can result in movement, and pallidal neurons are activated in response to active or passive movements. However, striatal and pallidal activation occurs after cortical discharge and generally during or after movement initiation. Alexander et. al. suggest that the basal ganglia motor circuitry may be involved in movement preparation, direction and amplitude modulation, but not directly in movement production per se. More recent studies have similarly concluded that the basal ganglia are probably not directly involved in initiating and controlling movements, but more likely play a role in

selecting or inhibiting movements, sequencing movements, evaluating movements, and developing procedural (motor) habits. The role of the striatum in these functions through activation of direct and indirect pathways through the basal ganglia is further discussed in Section 1.4, and the results from recent lesion and recording studies are summarized in Section 1.6.

1.2.2. Associative

The associative loop connects regions of prefrontal cortex to ventral anterior and mediodorsal thalamic nuclei through ventral and medial caudate regions and medial pallidal areas. It is thought that interconnected prefrontal areas may send projections to the same general region of striatum, but these projections interdigitate rather than converge onto the same single units within the striatum (Selemon and Goldman-Rakic, 1985). Like the motor loop, the associative loop can be further subdivided. Uylings et. al. (2003) have suggested that prefrontal cortex in primates and rodents can be subdivided into at least three broad regions: orbitofrontal, anterior cingulate (or medial prefrontal), and dorsolateral prefrontal areas. Alexander et al. defined the orbitofrontal and dorsolateral loops as associative, while the anterior cingulate loop was defined as limbic. However, the anterior cingulate is known to have associative as well as limbic functions, and the anterior cingulate and medial prefrontal cortex send projections to associative regions of the caudate and putamen in addition to the ventral striatum/nucleus accumbens (limbic striatum).

The anterior cingulate projections to striatum are of particular interest with respect to the experimental results presented in Chapter 2. This region of cortex is particularly well-suited to influence motor, cognitive and emotional processing, as it sends projections to the motor cortex, receives input from other prefrontal regions, and is interconnected with ventral striatum, amygdala, and other limbic structures. Damage to areas within the associative loops results in a number of deficits on high-level and cognitive tasks, such as those requiring working memory and behavioral flexibility. Results from behavioral and electrophysiological studies investigating the role of associative loop in behavior are discussed in more detail in Section 1.6. In the rat, dorsal medial prefrontal cortical areas (including anterior cingulate, medial agranular regions, and prelimbic cortex) also project to the dorsomedial caudoputamen (associative striatum). The dorsomedial striatum then projects to medial portions of entopeduncular nucleus, which project to midline thalamic nuclei and back to medial frontal cortical regions, completing the associative loop.

1.2.3. Limbic

The ventral striatum, made up of the nucleus accumbens and olfactory tubercle, has many structural and histochemical similarities to the caudate nucleus and the putamen, and a loop through this structure has been similarly defined. The ventral striatum receives input from the "limbic" cortex including the hippocampus, entorhinal and perirhinal cortices, amygdala and anterior cingulate cortex, as well as portions of the medial orbitofrontal cortex. The ventral striatum then projects to the ventral pallidum and rostrodorsal substantia nigra, as well as to a rostrolateral region of the GPi. This region of GPi then projects to mediodorsal thalamic nuclei, which then completes the loop by sending projections back to the anterior cingulate. Lesions in this loop often influence motivation, making a subject unwilling to work for food reward and/or less responsive to pain stimuli.

1.2.4. Other cortico-basal ganglia loops

Alexander, Strick and colleagues have additionally defined a number of other cortico-basal ganglia loops (Alexander et al., 1986; Middleton and Strick, 2000; Middleton and Strick, 2002). For example, the occulomotor loop, involved in controlling eye movements, projects from the frontal eye field in the cortex through parts of the caudate, GPi/SNr, and thalamus. An orofacial loop and loops through inferotemporal and posterior parietal cortical areas have also been defined.

1.2.5. Interactions between loops

Once parallel loops have been identified, the question remains as to how they interact to control behavior and decision processes. Reviewed in this section are a number of ways in which the different loops may interact. First, the cortical regions in question are densely interconnected. Additionally, at each level of cortico-basal ganglia processing, large volumes of neurons project onto smaller volumes, suggesting that there is a high degree of convergence at each level. Thus, nearby but distinct regions of cortex may project to the same region of striatum, and this process is repeated in the striatopallidal and striatonigral connections.

It has also been shown that the parallel loops are partially overlapping. Joel and Wiener (1994) suggest that cortical regions project through basal ganglia to targets in both GPi and SNr, and these outputs remain segregated through different thalamic targets which then project not only to the cortical region they originated from but also to an adjacent region. Thus, the cortico-basal ganglia loops can be considered to have partially closed and partially open or overlapping architecture. These overlapping loops provide a means by which motor, associative and limbic information may be passed between loops (Joel and Weiner, 1994). Spiraling loops between striatum and substantia nigra have also been shown (Haber and Fudge, 1997). The dopamine neurons of the substantia nigra pars compacta project strongly to the ventral striatum, which then sends projections both to the region of SNc that enervates it and to an adjacent region of SNc. These partially-overlapping dopaminergic loops spiral toward dorsolateral (motor) striatum. This spiraling architecture has contributed to the view that regional processing within the striatum may be more of a graded continuum from ventromedial-based limbic circuits to dorsolateral-based motor circuits, rather than organized into strictly parallel loops (Voorn et al., 2004).

Finally, the thalamus may critically contribute to cross-loop interactions (Haber and Calzavara, 2009). The thalamus projects both focally to layer V of cortex as well as diffusely to layer I/II. Thalamic projections to layer V are likely more focal,(Flaherty and Graybiel, 1993) and cortical projections extend from layer V back to thalamus as well as to striatum. However, the projections to the superficial cortical layers are likely to influence a broad area of cortex, not only due to their diffuse nature, but also because dendrites from multiple layers and relatively distal regions of cortex are found in layers I/II. Finally, thalamic relay nuclei receive projections not only from the regions of cortex that they target, but also from nearby cortical regions, resulting in another spiraling pathway and suggesting that the thalamus itself may be a structure in which integration of activity in parallel loops occurs.

How multiple parallel loops contribute to behavioral control is still an open question. The most prominent hypothesis is that each striatal region performs a similar functional role for its respective cortical input. An older idea, consistent with the converging connections at each level of basal ganglia processing, suggests the basal ganglia essentially "funnel" activity from multiple cortical

areas to the motor system. Finally, a suggestion that has gained prominence recently is that hierarchical loops may exert successive control over behavior during different stages of learning, consistent with the spiraling striatonigral dopaminergic projections. With all of these models, the existence of parallel loops is generally not in question, rather the degree of overlap and the mechanisms of interaction are debated.

1.2.6. Summary

Multiple parallel loops exist connecting regions of cortex to specific regions of striatum, and these can be broadly classified into motor, associative and limbic networks. Topographical connections are preserved at the level of the pallidum, and basal ganglia output is directed toward thalamic nuclei that project back to the regions of cortex from which they originated, closing each loop. This parallel architecture is highly convergent and partially overlapping, providing several possible modes of interaction between loops. Communication between loops may also occur at the cortical level by means of direct projections between different cortical areas. How these parallel, convergent, and partially overlapping loops interact to control behavior is unknown, and a number of plausible (and not necessarily mutually-exclusive) hypotheses have been suggested.

1.3. Direct and indirect pathways through the basal ganglia

Based on the anatomical connectivity described in Section 1.2, two pathways can be defined through the striatum to the output structures of the basal ganglia. These pathways have been termed the "direct" and "indirect" pathways, and have opposing effects on the neural activity in thalamic target nuclei. More recently, this classic idea has been extended to include a third "hyperdirect" pathway through the basal ganglia, which bypasses the striatum and instead directs information through the STN. **Figure 1.2A** illustrates these three pathways. Their activities and their interaction in the control of movement and sequence behavior are discussed in the following section.

1.3.1. Classical definitions

As briefly outlined in Section 1.2, there are two main projection pathways from the striatum, the main input structure of the basal ganglia, to the globus pallidus internal segment (and analogous substantia nigra pars reticulata), the output structure of the basal ganglia. The first of these is the "direct pathway" from the striatum to the GPi. The second pathway is the "indirect pathway" from the striatum to the GPe, then STN, and finally the GPi. Each of these pathways is discussed in more detail below. It should be noted that direct and indirect pathway medium spiny neurons are mingled together within the striatum (Flaherty and Graybiel, 1993). Thus, neurons in the same region of putamen send projections to both external and internal segments of the globus pallidus.

The direct pathway arises from a distinct set of striatal medium spiny neurons expressing D1-class dopamine receptors. These inhibitory neurons then project directly to the output nucleus of the basal ganglia, the GPi/SNr. Neurons in these output nuclei fire at high rates, tonically inhibiting their thalamic targets. When the D1 neurons in the striatum are activated, they inhibit the firing of neurons in the GPi/SNr, which then releases thalamic targets from tonic inhibition, causing a net increase in firing among thalamic neurons.

The indirect pathway, as classically defined, arises from a second set of striatal medium spiny neurons (MSNs) expressing D2-class dopamine receptors. These inhibitory MSNs project to the external segment of the globus pallidus, which then sends inhibitory projections to the subthalamic nucleus. The subthalamic nucleus sends excitatory output to the GPi, which as stated above, provides tonic inhibition to the thalamus. When the D2-expressing MSNs of the striatum are activated, they inhibit the firing of GPe neurons, releasing STN neurons from inhibition so they can excite firing in GPi neurons. The activated GPi neurons then suppress firing of their thalamic targets.

By modulating activity in these two pathways, it is thought that the basal ganglia can select or deselect thalamic targets to produce or suppress movements and sequences of movements. **Figure 1.2B** illustrates the direct and indirect pathway, and their hypothesized interaction in the production of normal movements. Section 1.3.3 provides more detail on how actions may be selected and/or inhibited through activation of the direct and indirect pathways. In the following section, some of the issues arising from this simple conceptualization are discussed.

1.3.2. Issues with the classic direct/indirect pathway model

The direct/indirect pathway hypothesis arises largely from anatomical connectivity patterns and the clinical observation that in Parkinson's disease (PD) and hypokinetic disorders, patients have trouble initiating movements, whereas in Huntington's disease (HD) and other hyperkinetic disorders, patients cannot suppress movements. In Parkinson's disease, the dopamine neurons of the SNc are lost, causing a general decrease in firing in the striatum. The resulting increase in GPi firing is thought to result in difficulty initiating movements due to excessive inhibition of thalamic neurons. In Huntington's disease, the projection neurons of the striatum degenerate. Early in the disease, striatal neurons projecting to GPe are preferentially affected and chorea is an early motor symptom. It is thought that the decrease in striatal inhibition of the GPe causes a net decrease in inhibition on the thalamus, which then results in involuntary movements. This relatively simple conceptualization has provided substantial insight and suggested effective targets for therapies such as deep brain stimulation to treat movement disorders. A number of issues exist, however, and the picture of basal ganglia operation has become increasingly complex in recent years.

1.3.2.1. The "hyperdirect" pathway

More recently, it has been shown that there is a projection from GPe directly to GPi, and thus the indirect pathway may bypass the STN entirely. The question then remains as to the role of the STN in basal ganglia processing. The STN receives excitatory input from cortex and inhibitory input from GPe, and sends diffuse excitatory output to both GPe and GPi/SNr. Nambu et al. (2002) have studied the timing in this pathway from cortex to STN to GPi and developed a conceptualization of the basal ganglia which now incorporates the direct and indirect pathways as well as a "hyperdirect" pathway (**Figure 1.2A**). This hyperdirect pathway has a net diffuse inhibitory effect on thalamic targets of GPi neurons and shorter transmission delays than either the direct or the indirect pathways.

1.3.2.2. Direct and indirect pathways may not be segregated

D1 receptor expressing medium spiny neurons in the striatum have been shown to send collateral projections to GPe as well as GPi (Kawaguchi et al., 1990; Levesque and Parent, 2005; Wu et al., 2000), suggesting that the two pathways are not as segregated as the classic model would imply. It has been suggested that the functional role of the dual projection may be to ensure that GPi neurons are active only transiently, in response to a change in cortico-striatal activity, rather remaining active throughout movement production/suppression (Cohen and Frank, 2009). This hypothesis preserves

the overall direct-indirect scheme, but suggests that a short burst of activity should occur in GPi to initiate or inhibit an action.

1.3.2.3. Non-motor functions of the basal ganglia

In the years since the development of the direct/indirect pathway model of action selection, the role of the basal ganglia has been shown to extend beyond simple selection and suppression of movements. In particular, damage to the basal ganglia nuclei results not only in motor dysfunction, but also has been shown to impair certain types of learning and memory functions. It has been additionally observed that Huntington's disease patients often exhibit changes in personality and cognitive ability in conjunction with, or prior to, development of motor symptoms, and the striatum has been implicated in obsessive-compulsive disorder and a number of other non-motor diseases and disorders. Striatal lesions have also been shown in some experimental studies to impair cognitive and emotional abilities rather than motor function. Results from these studies are discussed in more detail in Section 1.6, but it should be obvious that the direct/indirect pathway model must now be extended to account for learning and non-motor functions of the basal ganglia.

1.3.3. Action selection through direct, indirect and hyperdirect pathways

In his extensive review, Jonathan Mink (1996) outlined a general mechanism by which the direct and indirect pathways may select desired motor programs and inhibit competing programs, respectively. Emphasizing the convergence of information onto output neurons in GPi/SNr as well as the specificity with which striatal neurons target individual neurons in GPi and GPe (Flaherty and Graybiel, 1994; Hazrati and Parent, 1992a; Hazrati and Parent, 1992b; Parent and Hazrati, 1993), Mink hypothesized that once "Motor Pattern Generators" (MPGs) are activated, activity increases in GPi for non-selected programs, increasing the "brake" on unwanted action. At the same time, GPi activity is decreased for the selected program, releasing the desired motor pattern from inhibition. Through this brake-release mechanism, "selected movements are enabled and competing postures and movements are prevented from interfering with the one selected."

Focusing of information occurs through the convergent projections across basal ganglia subregions, as well as through a number of mechanisms operating within each region. Notably, within the striatum, medium spiny projection neurons (MSNs) are thought to have bistable modes of operation. In the hyperpolarized state, MSNs exhibit membrane potentials around -80 mV and no action potentials are produced from this state (Jiang and North, 1991; Kawaguchi et al., 1989). Coincident excitatory input can drive MSNs into a less hyperpolarized state, with resting membrane potentials around -50 mV. Further depolarization from this "up state" can then result in the generation of action potentials. MSNs fire at low rates, and require coincident excitation of a large number of cortical neurons to drive them into an "up" state and moreover to induce action potentials. The specificity required in the large number of cortical neurons needed to excite a single MSN, focuses information from cortex to striatum. Further focusing of MSN activity is likely encouraged by inhibition from fast-firing striatal interneurons, which strongly inhibit the firing of MSNs. Information is further focused as a single striatal neuron often makes strong connections with its target GPi neurons. GPe receives similar targeted input from striatum, but serves to inhibit GPi – directly through inhibitory connections from GPe to GPi, or indirectly through the bisynaptic projection through STN. Similar information with opposite sign suggests that GPe projections may "oppose, limit, or focus" striatal input to GPi (Mink, 1996).

Cortico-STN activation in the "hyperdirect" pathway causes a fast, diffuse excitation of GPi output neurons, broadly inhibiting thalamic targets. Cortical activation of "direct" pathway neurons in the striatum then results in more focal inhibition of GPi targets, selecting the desired motor program by specifically releasing inhibition of particular thalamic neurons. Finally, cortical activation of the "indirect" pathway neurons in the striatum results in targeted inhibition of GPe followed by targeted release of inhibition at GPi, resulting in a net inhibition of thalamic targets. This indirect pathway activity may serve to further inhibit undesired motor programs, or may limit the activation of the desired program, by ensuring that neural activity in target thalamic cells is transient.

It is important to note that most movement-related activity in recorded striatal neurons has been shown to occur later than movement-related activity in motor cortical areas, as well as after the activation of EMG in related muscles, and often after the onset of movement itself. In addition, while striatal activity has been shown to correlate with the activation of muscles and the direction of movement, further correlations with other movement parameters including position, velocity, acceleration, force, and amplitude have not generally been observed. These data suggest that the basal ganglia do not directly control movements, but are consistent with the idea that they may serve to enable or inhibit controlling activity that resides elsewhere in the brain and/or spinal cord.

The inability to select desired movements could result in akinesia, whereas the simultaneous activation of multiple motor programs should result in inefficient and/or ineffective movements. Accordingly, the direct/indirect pathway model has been used to explain a number of motor deficits including those seen in diseases such as dystonia, chorea and Parkinson's Disease (Mink, 1996; Mink, 2003). Dystonia and chorea can be viewed as resulting from the inability to inhibit unwanted movements, and the incidence of these can be increased by lesions in the GPi and/or associated with a reduction in GPi firing rate. Parkinson's Disease patients exhibit a number of motor symptoms including difficulty initiating movements, co-contraction rigidity, and postural abnormalities. In accord with the direct/indirect pathway model for selection and inhibition of motor programs, a slight increase in GPi activity is generally observed with PD, suggesting that the inability to initiate movements could be caused by a difficulty in activating a desired motor program. Co-contraction rigidity and postural abnormalities cannot be explained by inability to initiate movements, and may be related to a decreased striatal dynamic range due to dopamine depletion in the SNc. This may result in an inability to further increase GPi activity sufficiently to suppress unwanted movements and postures. These hypotheses can account for some of the major features of movement disorders and accompanying firing rate changes within basal ganglia structures. However, further studies have revealed that the expected firing rate changes in most of these diseases are not observed or do not capture the complexity of the disease process, suggesting the need for additional elaboration of the classic direct/indirect pathway model.

Finally, the basal ganglia have been implicated in the automatization of actions and execution of well-learned sequences of movements. A number of studies have shown sequence-specific firing in the striatum of monkeys or rats performing well-learned or innate sequences (Aldridge and Berridge, 1998; Kermadi and Joseph, 1995). Matsumoto et al. (1999) found that unilateral lesions in the SNc impaired a monkey's ability to develop "smooth, efficient performance" of a sequence of instructed movements. When they manipulated the delivery of reward so that it came earlier than expected, a normal monkey nonetheless continued to complete the learned sequence. In a monkey with unilateral lesions of the dopamine neurons of the SNc, this perseverative responding was observed only in the arm contralateral to the intact side. When the monkey performed the task using the arm contralateral to the lesion, it did not complete the sequence when reward was delivered early, suggesting that dopamine was required for both smooth execution of single movements, and for the "chunking" of

those movements into a single motor program through repeated performance (for review, see Graybiel, 1998). The classic direct/indirect model fails to capture these learning functions of the basal ganglia.

Corticostriatal synaptic plasticity has been shown to be dopamine-dependent, providing a mechanism for learning to occur within basal ganglia networks. Additionally, dopamine may play a key role in action selection and suppression mechanisms by modulating the general excitability of striatal neurons as well as the ease with which they transition between Up and Down states (Grillner et al., 2005; Wilson, 1993). The role of dopamine on striatal activity is discussed further in Section 1.5.3.

It is important to note that GPi output targets not only thalamic nuclei, but also to pedunculopontine and related nuclei projecting into the reticulospinal motor pathway. Thus, the basal ganglia may affect movement and postural control not only by gating cortical output, but also by gating or otherwise influencing brainstem neuronal activity. Grillner et al. (2005) points out that the model outlined above for selection and suppression of motor programs through thalamic activation and inhibition may equally well describe selection and suppression at the level of the brainstem.

1.3.4. Summary

Classically, two pathways have been defined through the basal ganglia, with opposing effects on thalamic target nuclei. The "direct" pathway arises from D1 receptor expressing medium spiny neurons in the striatum and projects directly to the GPi, resulting in a net excitatory effect on the thalamus. The "indirect" pathway arises from D2 receptor expressing neurons in the striatum, and projects through the GPe to GPi, resulting in a net inhibitory effect on the thalamus. More recently, a "hyperdirect" pathway that projects from cortex through subthalamic nucleus to GPi has been defined. This pathway has a fast net-inhibitory effect on thalamic targets. Through the action of these three pathways, it is thought that thalamic targets can be activated and deactivated in a temporally-precise manner, suggesting that the basal ganglia may be involved in selection and inhibition of desired and unwanted actions to generate movement and motor sequences. The basal ganglia have further been implicated in motor learning and habit formation, requiring an update of the action selection model to incorporate learning-related plasticity mechanisms. Some of the mechanisms thought to be involved in the learning and memory functions of the basal ganglia are reviewed in Section 1.5. Additionally, a number of non-motor deficits are seen following lesions in basal ganglia nuclei and in a variety of basal ganglia disorders. Results of behavioral and electrophysiological studies and their implications for motor and non-motor functions of the basal ganglia are summarized in Section 1.6.

1.4. Striatal chemical architecture and interneurons

The striatum can be subdivided in several ways. Section 1.2 reviewed the anatomical connectivity patterns that broadly divide the striatum into limbic, associative and sensorimotor regions. Presented in Section 1.4.1 is the further division of the striatum into striosome and matrix compartments based on chemical expression patterns. These expression patterns differ between dorsal and ventral striatum, and the focus here is on the projection patterns and immunoreactivity of the dorsal striatal compartments only. These generally hold for both dorsomedial and dorsolateral subdivisions, though differences in degree as well as gradients in expression or innervation patterns exist. In Section 1.4.2 the different striatal neuron subtypes, their interconnections, and their effects on the firing patterns of medium spiny neurons are reviewed. Finally, discussed in Section 1.4.3 is the special role of

dopaminergic input from the substantia nigra pars compacta on the activity of different neuron subtypes and in synaptic plasticity at the corticostriatal synapse.

1.4.1. Striosome and matrix compartmentalization

In 1978, Graybiel and Ragsdale first described the patchy organization of subcompartments in the human striatum (Graybiel and Ragsdale, 1978). They stained for acetylcholinesterase (AChE) and observed that within the striatum, there appeared AChE-poor islands within otherwise well-stained striatum. They termed these islands "striosomes" (also often referred to as "patches" in the literature) and the well-stained regions the extrastriosomal "matrix." Striosomes make up approximately 10-15% of the striatal volume. Later studies revealed that a number of other neurochemicals, neuron subtypes and neurotransmitter receptors are differentially expressed in the two compartments. Holt et al. (1997) compare the striosome/matrix boundaries determined by stains for a number of these different chemicals. They note that the cholinergic stains, the first stains used to identify the two compartments, are least consistent with the compartmental organization of other neurochemicals, including enkephalin, substance P, tyrosine hydroxylase, calbindin, and parvalbumin. Notably, regions of intense mu-opioid receptor expression have been shown to coincide with striosomes (Herkenham and Pert, 1981) as have dopamine "islands" seen early during development (Graybiel, 1984), suggesting that these two neurotransmitters play a special role in striatal development and function. The preferential expression of these and other compartmentally distributed chemicals and receptors are summarized in Table 1.5.1. **Figure 1.3** depicts this compartmental organization of the striatum.

Table 1.5.1. Neurochemical expression in striosome and matrix compartments (adapted from Graybiel 1990)

Chemical Abbrev.	Full Name	Description	Medial / Lateral?	Striosome / Matrix?	Reference
AChE	Acetylcholinesterase	Cholinergic degradative enzyme		Matrix	Graybiel & Ragsdale, 1978
ChAT	Choline acetyltransferase	ACh synthetic enzyme		Matrix	
ACh	Acetylcholine	Neurotransmitter			
M1	Muscarinic ACh receptor type 1			Striosomes	Nastuk & Graybiel, 1988
M2	Muscarinic ACh receptor type 2			Both	Nastuk & Graybiel, 1988
D1	Dopamine receptor type 1			Striosomes	
D2	Dopamine receptor type 2		Lateral	Matrix	Joyce et al., 1986
TH	Tyrosine hydroxylase	DA synthetic enzyme, a marker for DA neurons		Matrix	Lavioe et al., 1989
CB1	Endocannabinoid receptor type 1		Lateral		Herkenham et al., 1991
MOR	Mu (μ) opioid receptors			Striosomes	Herkenham & Pert, 1981
NMDA		Ionotropic glutamate receptor		Matrix	Dure et al., 1992
AMPA		Ionotropic glutamate receptor		Matrix	Dure et al., 1992
kainate		Ionotropic glutamate receptor		Striosomes	Dure et al., 1992
CB	Calbindin	A calcium binding protein	Medial	Matrix	Holt et al., 1997; Gerfen et al., 1985
CR	Calretinin	A calcium binding protein			
PV	Parvalbumin	calcium binding protein		Matrix?	Holt et al., 1997
SP, subP	Substance P	Neuropeptide		Striosomes	
Dyn	Dynorphin	Neuropeptide		Striosomes?	Graybiel & Chesselet, 1984
Enk	Enkephalin	Neuropeptide		Striosomes?	Graybiel & Chesselet, 1984
neurotensin		Neuropeptide			
somatostatin		Neuropeptide		Matrix	
NADPHd	Dihydronicotinamide anenine dinucleotide phosphate diaphorase	enzyme		Matrix	Sandell et al., 1986

Graybiel and Ragsdale observed that striosomes were most prominent in the head of the caudate nucleus and fewer of these islands could be observed in the putamen (Graybiel and Ragsdale, 1978). In the rat, a similar distribution has been noted: striosomal compartments are more prominent in dorsomedial striatum than in dorsolateral striatum. A number of chemicals have differential expression patterns in dorsomedial and dorsolateral regions and such gradients in expression could be related in part to the striosome/matrix organization. However, a number of neurochemicals, including AChE and calbindin, have ventral-to-dorsal gradients in expression unrelated to the striosome/matrix distributions. Markers of cholinergic transmission (stronger expression dorsomedially) and D2-class receptors (stronger expression dorsolaterally) are among the chemicals with differential dorsomedial and dorsolateral distributions. Differential expression of the various neurochemicals and receptors in dorsomedial and dorsolateral striatum is also summarized in Table 1.5.1, when known.

In addition to a number of neurochemical differentiations that can be made, the projection patterns of striosome and matrix compartments differ. The matrix projections are organized as previously described: cortical and thalamic inputs project topographically to the matrix, and output from medium spiny neurons projects to the GPi and SNr by way of the direct and indirect pathways. Projections from the basal ganglia output nuclei complete the cortico-basal ganglia-thalamic loops. By contrast, striosomes receive input from the amygdala, midline thalamus and limbic cortical areas (Eblen and Graybiel, 1995; Levesque and Parent, 1998; Ragsdale and Graybiel, 1991; Russchen et al., 1985). Medium spiny neurons in striosomes are thought to project to the dopamine neurons in the SNc (Gerfen, 1985; but see Levesque and Parent, 2005). Striosomes are thus in a position to synthesize information from across the limbic system. They may influence the excitability of neurons in the matrix compartment as well as corticostriatal synaptic plasticity by controlling the levels of dopamine released by SNc neurons.

Within each compartment, there is further heterogeneity. Within striosomes, heterogeneous expression of neurochemicals has been observed in the striosomal border regions compared to their centers. Faull et al. (1989) observed differential expression of neurotensin in the striosome, matrix and striosomal border regions, and similar observations have been made for AChE, enkephalin and calbindin expression (Prensa et al., 1999). While neurochemical expression patterns are often graded within the matrix, further neurochemical differentiation of different functional domains has not been observed. However, afferents terminating in the matrix are distributed in a patchy manner, and these patchy regions have been termed "matrisomes." A single site in the cortex may send projections to multiple matrisomes within the striatum, information from related cortical sites (e.g. the "hand" representations in both primary motor and primary somatosensory cortex) can converge within single matrisomes, and neurons from multiple matrisomes may reconverge onto single targets within the globus pallidus (Flaherty and Graybiel, 1993; Flaherty and Graybiel, 1994; Gimenez-Amaya and Graybiel, 1991). The patchily distributed matrisomes thus appear to be discrete functional processing units within the striatal matrix, though their precise computational role in the restructuring and manipulation of cortical information remains unknown.

Medium spiny neurons in the striosomes and matrix are generally segregated, as axons and dendrites of these neurons seldom cross compartmental boundaries (Walker et al., 1993). This separation is not absolute, however, as approximately one quarter of MSNs may cross to some extent. Other types of striatal neurons more regularly cross compartmental borders. Notably, the cell bodies of cholinergic neurons are found in both compartments and these cells have dendritic fields that may span the striosome/matrix boundaries. Axons of ACh neurons are generally directed toward the matrix (Kawaguchi, 1992). The role of cholinergic neurons in striatal processing is discussed in more detail in the following section.

1.4.2. Striatal neuron subtypes

The preceding sections have focused primarily on the input/output characteristics of the striatum, and have therefore emphasized the medium spiny projection neurons, which are the most numerous cell type and the only neurons that transmit information out of the striatum. There are a number of other types of striatal neurons, however, which project only to other cells in the striatum. These include large aspiny cholinergic interneurons, fast-spiking parvalbumin-positive GABAergic interneurons, and two other non-parvalbumin positive GABAergic types. All of the interneuron types described have been found in both primates and rats, though their proportions are larger in primates than in rodents (Graveland and DiFiglia, 1985; Wu and Parent, 2000). In this section, the characteristics of the medium spiny projection neurons as well as these other striatal interneurons types are reviewed.

1.4.2.1. Medium spiny projection units

Medium spiny projection neurons (MSNs) make up over 95% of striatal neurons. These neurons receive input from outside the striatum in the form of excitatory cortical and thalamic input, as well as modulatory dopamine input from the SNc. As discussed previously, MSNs generally express either D1- or D2-type dopamine receptors, though a small percentage of MSNs has been shown to coexpress both receptor classes (Bertran-Gonzalez et al., 2008; Gerfen and Keefe, 1994; Levesque et al., 2003; Matamales et al., 2009; Shuen et al., 2008; Surmeier et al., 1993) The populations of D1 and D2 receptor expressing MSNs are approximately equal in size and project out of the striatum in the direct and indirect pathways of the basal ganglia, respectively. D1 and D2 neurons are similar morphologically, but project to different nuclei, express different neuropeptides, and are differentially responsive to dopamine and acetylcholine (Shen et al., 2007; Surmeier et al., 2007). D1 direct pathway neurons are immunoreactive for dynorphin and substance P, whereas D2 indirect pathway neurons coexpress adenosine A2A receptors and are immunoreactive for enkephalin (Gerfen, 1992 [review]; Schiffmann et al., 1991).

Cortical and thalamic inputs converge onto a striatal MSN with a ratio of approximately 1000:1. Coincident firing of a large number of excitatory neurons is thus thought to be needed to drive striatal MSNs into an "Up" state, from which it can produce action potentials. It is important to note, however, that up and down state transitions are more prominently observed *in vitro* and under anesthesia than in awake subjects. During waking, a single Gaussian distribution of membrane potentials has been observed, rather than the bimodal distribution commonly observed under anesthesia and during slow-wave sleep (Mahon et al., 2006). Standard theories of striatal function that rest on the Up/Down state transitioning of MSNs may thus require some revision to incorporate this recent result.

Cortical and thalamic input generally synapses on the spines of MSNs (Kemp and Powell, 1971), whereas dopaminergic terminals form synapses on the dendrites and spine necks of MSNs (Smith et al., 1994). This arrangement enables dopamine to modulate the excitatory input before it can have an effect at the soma. Additional input is received by MSNs from other striatal neurons. Especially well studied is the strong inhibitory input received at the soma from fast-firing GABAergic interneurons, discussed further in Section 1.4.2.3. The other GABAergic interneurons may strongly inhibit MSNs in a similar manner, but these details are less well-studied. Weaker inhibitory input from other MSNs makes contact at more distal sites on the dendrites and spines and may have a locally-restricted effect at single synapses (Tepper et al., 2004; Tunstall et al., 2002). Additional modulatory input comes from the large cholinergic neurons intrinsic to the striatum.

A number of other neurochemicals modulate synaptic transmission and neural excitability in the striatum, and thus have an impact on the activity of striatal projection neurons. These chemicals exert a number of additional effects on cell function through second messenger signaling cascades, and the range of their contributions to striatal computation is not yet fully understood. The neuropeptides enkephalin, dynorphin and substance P are co-released with GABA from the terminals of MSNs, and are differentially expressed in the D1/direct and D2/indirect pathway neuronal populations. Cannabinoids and opioids have likewise been shown to alter signaling at synapses targeting MSNs (Gerdeman et al., 2002; Miura et al., 2008).

The precise mechanisms of interaction of all these component inputs onto MSNs are still under investigation. However, current understanding of this system indicates that the excitatory drive that can depolarize and produce action potentials in MSNs comes from cortical and thalamic inputs. Due to the low resting membrane potentials in the "Down" state, combined with the necessity of highly convergent firing required to excite action potentials, MSNs fire sparsely. Inhibitory input from other MSNs may serve to modulate inputs and plasticity at individual synapses, but is likely too weak to have a large effect on the firing rates of target neurons. Instead, strong inhibitory input from GABAergic interneurons is thought to provide a mechanism for delaying or preventing spiking in target MSNs, allowing subsets of neurons to be focused, synchronized or even deactivated in specific contexts. Finally, the general excitability can be enhanced or reduced by the actions of the neuromodulators dopamine and acetylcholine, which are particularly well studied, and likely through the effects a number of other neurochemicals as well.

1.4.2.2. Cholinergic interneurons / Tonically-active neurons

Cholinergic interneurons are large aspiny cells within the striatum that stain strongly for markers of acetylcholine (ACh). These cells make up less than 1% of striatal neurons, but have "dense and extensive arborizations" which allow them to exert influence disproportionate to their numbers (Kreitzer, 2009). ACh release likely acts locally at synapses as well as more broadly through volume transmission (Contant et al., 1996). Volume transmission may be spatially and temporally restricted, however, as ACh is rapidly degraded by acetylcholinesterase - an extracellular enzyme richly expressed in the striatal matrix. Cholinergic cell bodies can be found both in the matrix and in striosomes, but are disproportionately located in the striosomal border regions, and their dendrites and axons often cross compartmental boundaries. The axons of ACh neurons arborize extensively in the matrix, creating the differential expression of AChE and ChAT in the two compartments.

Cholinergic interneurons receive sparse excitatory input, primarily from thalamus rather than cortex (Lapper and Bolam, 1992), and inhibitory input from MSNs. ACh cells synapse on MSNs and fast-firing (FF) neurons, and as mentioned above may act extrasynaptically as well. ACh acts at nicotinic and muscarinic receptors. Nicotinic receptors are relatively low affinity, requiring higher levels of ACh for activation. By contrast, muscarinic receptors are relatively high affinity, suggesting that low tonic levels may be sufficient for activation. In the striatum, nicotinic acetylcholine receptors are located presynaptically on the terminals of dopamine neurons, cortical and thalamic glutamatergic inputs, and fast-spiking GABAergic interneurons. These presynaptically-expressed nicotinic ACh receptors generally serve to enhance neurotransmitter release (Koos and Tepper, 2002; Schwartz et al., 1984; Zhou et al., 2002). Muscarinic ACh receptors come in several types. M1 muscarinic receptors are expressed in both direct and indirect pathway MSNs, and blocking M1 receptors in the striatum has been shown to reduce excitatory post-synaptic currents (EPSCs) at corticostriatal synapses (i.e., M1 activation has an excitatory effect on MSNs, (Wang et al., 2006). M4 receptors are

additionally expressed in direct pathway MSNs and may inhibit the effects of D1 activation in these neurons. M4 receptors additionally act as inhibitory autoreceptors on ACh neurons themselves.

The effects of ACh release on the firing of striatal output neurons are complex, and these effects may be mediated directly through receptors located on the MSNs themselves, or indirectly through the modulation of firing of striatal interneurons. For example, activation of postsynaptic nicotinic receptors on fast-spiking GABAergic interneurons directly depolarizes these cells, leading to an increase in inhibition on MSNs (Koos and Tepper, 2002), and an increased feedback inhibition onto the ACh neurons. Conversely, activation of presynaptic muscarinic receptors on the terminals of fast-spiking neurons has been shown to reduce the release of GABA from these terminals, reducing the inhibition on target MSNs (Koos and Tepper, 2002). Additionally, ACh neurons express both D5 and D2 dopamine receptors, enabling dopamine to increase or decrease the excitability of these neurons depending on the concentration expressed (Yan et al., 1997; Yan and Surmeier, 1997). A reduction in the firing of ACh neurons following activation of D2 receptors has been shown to contribute to LTD at corticostriatal synapses, suggesting that lowered ACh concentrations should result in reduced firing of MSNs (Wang et al., 2006). The above effects, acting through interneurons, should act equally on direct and indirect pathway MSNs in the striatum. Direct activation of M1 receptors on MSNs has a somewhat excitatory effect on these neurons, and this should be seen on both direct and indirect pathway neurons. By contrast, activation of M4 receptors may have an inhibitory effect differentially targeting direct pathway circuitry.

In vivo, ACh neurons throughout the striatum fire tonically at low frequencies, and firing rates are limited in ACh neurons by a long after-hyperpolarization (Kawaguchi, 1992; Wilson et al., 1990; Wilson and Goldberg, 2006). Due to their tonic firing, ACh neurons are often called "tonically-active neurons," or TANs, especially in electrophysiological studies in which it is difficult or impossible to identify the morphological characteristics of the neurons being studied. In recording experiments in monkeys, TANs have been shown to develop a pause in firing in response to behaviorally relevant stimuli (Aosaki et al., 1995; Aosaki et al., 1994; Blazquez et al., 2002; Joshua et al., 2008; Morris et al., 2004), which is followed by a rebound excitation. The pause requires intact thalamic and dopaminergic innervation to occur (Aosaki et al., 1994). ACh neurons also exhibit a phasic response at the time of reward delivery. These responses appear similar to those of the midbrain DA neurons, though studies have shown that unlike DA neurons, TANs generally do not encode reward prediction errors and have positive polarities regardless of reward delivery or omission (Morris et al., 2004) but see (Apicella et al., 2009). The temporal dynamics of their responses at reward delivery may be modulated by value of the outcome, however, rather than the magnitude of their neuronal response per se (Joshua et al., 2008). Interestingly, TANs have been shown to have stronger pause responses to stimuli directing a contralateral movement (Shimo and Hikosaka, 2001), perhaps indicating a more direct role for these interneurons in the planning, execution and/or evaluation of cue-evoked movements.

1.4.2.3. Fast-firing interneurons/Parvalbumin containing

The best studied of the GABAergic subtypes, the parvalbumin-positive (PV+) interneurons, make up 3-5% of striatal neurons, and are more prevalent laterally than medially (Kita et al., 1990). Like MSNs, PV+ neurons receive cortical and thalamic excitatory input, as well as dopaminergic modulation. Unlike MSNs, they also receive strong inhibitory feedback from GPe neurons projecting back to striatum (Bevan et al., 1998). They form chemical synapses with other PV+ neurons, but are also electrically coupled by gap junctions, providing a fast mechanism by which PV+ neurons may synchronize their firing (Kita et al., 1990).

PV+ neurons have a lower convergence ratio than MSNs, and thus do not require as high a degree of synchronous input to produce action potentials. PV+ interneurons fire at high rates when depolarized *in vitro* and *in vivo*, and are thus often referred to as fast-firing (or fast-spiking) neurons (FFs). Despite gap junction coupling, synchronized firing of even closely-located FFs has not been observed in awake animals (Berke, 2008), suggesting that firing in these neurons during behavior is dominated by other factors.

Fast-firing interneurons form synapses onto MSNs that are proximal and numerous, allowing FFs to delay or inhibit firing in target MSNs (Bennett and Bolam, 1994; Kita, 1993; Koos and Tepper, 1999; Mallet et al., 2005). *In vitro*, approximately one quarter of MSNs near an FF interneuron have been shown to make contact with the FF and are strongly inhibited in response to stimulated FF bursts. Combined with their high firing rates, these findings suggest that FFs are likely responsible for most or all of the inhibition seen in striatal projection neurons (Tepper et al., 2004). This likely occurs via a feed-forward mechanism by which FF neurons are excited by cortical and thalamic inputs; their firing then rapidly inhibits the firing of nearby MSNs.

Dopamine and acetylcholine both modulate the firing of FF neurons. FF neurons express primarily D5 receptors postsynaptically (Centonze et al., 2003b; Rivera et al., 2002), the activation of which increases neuronal excitability. D2 receptors are also expressed presynaptically in these neurons, and their activation can limit the release of GABA from FF terminals, likely without affecting the firing rate of the neuron. As discussed in the previous section, ACh can directly increase the firing rate of FF neurons through activation of postsynaptic nicotinic receptors, or can decrease the release of GABA through activation of presynaptic muscarinic receptors. ACh and DA may additionally act cooperatively to further increase firing in FF neurons: ACh may enhance the release of DA at terminals through activation of nicotinic receptors, which could then excite FF neurons through activation of D5 receptors, as described above.

1.4.2.4. Other interneuron subtypes

At least two other types of GABAergic interneurons have been identified (Bennett and Bolam, 1993; Chesselet and Graybiel, 1986; Cowan et al., 1990; Smith and Parent, 1986; Vincent and Johansson, 1983). The better-studied of the two expresses nitric oxide synthase (NOS), as well as a number of other chemicals including somatostatin, neuropeptide Y, and NADPH diaphorase (Chesselet and Graybiel, 1986; Kubota et al., 1993; Smith and Parent, 1986; Vincent and Johansson, 1983).

NOS-positive interneurons make up only a small percentage of striatal neurons (1-2%). They have extensive, but less dense, arborizations compared to FF and ACh neurons (Kawaguchi, 1993), and are more numerous in the ventral and medial striatum (Gerfen et al., 1985) than in dorsolateral striatum. NOS-positive cells also stain for acetylcholinesterase (AChE), and are one of the few striatal neuron types that express the NK-1 (substance P) receptor, suggesting they interact with both cholinergic neurons and direct pathway MSNs. Like ACh interneurons, their cell bodies can be found in the matrix or in striosomes, their dendrites often cross compartmental boundaries, and their axons primarily branch in the matrix. Like the PV+ interneurons, NOS-positive neurons receive glutamatergic inputs from cortex and thalamus and provide strong inhibitory synapses onto striatal MSNs (Koos and Tepper, 1999). They are similarly modulated by dopamine, and express D5 receptors (Rivera et al., 2002), the activation of which increases their firing rate (Centonze et al., 2002). Electrophysiologically, NOS-positive neurons exhibit large and persistent calcium-dependent plateau potentials as well as calcium-dependent low-threshold spikes, and are often referred to as

low-threshold spiking (LTS) neurons. The classic function for nitric oxide is as a vasodilator controlling blood flow, and this may be one of the roles NOS-positive cells in the striatum perform. NOS has also been shown to play a role in synaptic plasticity (Calabresi et al., 1999; Fino et al., 2009; Garthwaite, 2008; Kato and Zorumski, 1993).

Calretinin is expressed in another population of GABAergic interneurons. These neurons are medium-sized aspiny neurons most numerous in the rat rostro-medial striatum. Little else is known about these cells. These neurons may correspond to a subset of LTS neurons described by Koos and Tepper (1999) that lacked the persistent plateau potentials of NOS-positive neurons and exhibited a different spike morphology. These neurons were also observed to exert a powerful inhibitory influence on MSNs.

1.4.3. Dopaminergic modulation of striatal neurons

The dopamine-containing neurons of the substantia nigra project extensively to the dorsal striatum and have been shown to contribute to the action selection functions of the basal ganglia. Dopamine has also been shown to alter synaptic plasticity at the corticostriatal synapses, contributing an important mechanism by which learning can occur in this system. In this section, the anatomical connections between the dopamine neurons and the striatum are discussed in more detail, as are their firing properties. Additionally, the effects of dopamine on the firing of medium spiny and other neuron types in the striatum are reviewed.

1.4.3.1. Nigro-striatal connections

For an extensive review of the dopamine system, and especially its connections with dorsal and ventral striatum, see Joel & Weiner (2000). Briefly, the dopamine-containing neurons of mammalian brains are grouped together in midbrain regions labeled A8, A9 and A10. Region A10 corresponds to the ventral tegmental area (VTA), which projects to ventral striatum and frontal cortex. Area A9 corresponds to the substantia nigra pars compacta (SNc), which sends dopaminergic projections to the dorsal striatum. Area A8 corresponds to the retrorubral nucleus (RRN), which projects especially to ventral and lateral striatal regions as well as the amygdala. The projections of the dopamine system are highly divergent: in rats, there are ~7000 DA neurons in each hemisphere, projecting to 1000 times as many target neurons in cortex, striatum, amygdala, etc.

In rats, the projection from the dopamine neurons to different regions of striatum is roughly topographically organized. As mentioned above, VTA projects to ventral (limbic) striatal regions, while more lateral regions of the SNc project to lateral (motor) regions of dorsal striatum and more medial regions of the SNc project to medial (associative) regions of dorsal striatum. This spatial separation has not been observed in primates, however. Rather, clusters of neurons projecting to motor or associative regions of caudate and putamen tend to interdigitate. There are, however, some regions of the SNc that project solely to motor striatum, including the caudal SNc and the lateral columns that extend into the SNr.

Projections from striatum back to the dopamine neurons are also organized in a roughly topographical manner. The ventral striatum projects to the VTA and the dorsal SNc, while all three regions (ventral, dorsolateral and dorsomedial striatum) project to the ventral SNc and the retrorubral nucleus. In rats, lateral (motor) and medial (associative) striatal regions project to lateral and medial SNc, respectively. In primates, these projections are again interdigitating, though lateral and ventral

SNc is more predominantly targeted by the motor striatum, and the medial SNc is targeted by the limbic striatum.

Based on these observations, it can be stated that each region of striatum has reciprocal connections with the region of VTA or SNc which innervates it, but these reciprocal connections are not the whole story. The ventral striatum receives dopaminergic input from a restricted region of VTA, but sends projections to a much larger region of dopamine neurons in VTA and SNc. Associative striatum likewise sends return projections to dopaminergic neurons in the SNc which influence motor striatum. Motor striatum, on the other hand, influences only a relatively restricted region of SNc.

Gerfen et al. (1985) showed that in rats, the GABAergic neurons in the SNr are targeted primarily by matrix neurons of the striatum, whereas the DA neurons of the SNc are targeted by striosomal neurons. In rats, the dopaminergic projection to striatum may also be compartmentally segregated (Gerfen et al., 1987a; Gerfen et al., 1987b; Jimenez-Castellanos and Graybiel, 1987). Evidence suggests that the neurons of the dorsal SNc primarily target the matrix compartment, whereas the dopamine neurons in the ventral SNc and in the SNr primarily target striosomes. Based on these findings, the dopamine system has been divided into dorsal (VTA, RRN, and dorsal SNc) and ventral (ventral SNc and dopamine neurons of SNr) tiers targeting different striatal compartments. In primates, the striatal matrix projections preferentially target GABAergic neurons in the SNr, suggesting that striosomal neurons may preferentially target dopaminergic neurons, but this has not been shown (Levesque and Parent, 2005). Even less evidence exists to suggest that there are separate populations of dopaminergic neurons that project to striosome versus matrix compartments. Rather, if such populations exist, they are likely mingled together within the SNc, making the targeting of one or the other population difficult with traditional tracing techniques.

It is worth noting that DA neurons also receive input from and send projections to other nuclei of the basal ganglia. In particular, the ventral (limbic) pallidum sends projections to a wide region of dopaminergic neurons in a manner similar to ventral striatum. Electrophysiological evidence also suggests that there are direct connections from SNr to SNc, as well as a sparse connection from STN. Dopaminergic projections to the STN and pallidum have also been shown.

In summary, the limbic striatum projects to a large region of DA neurons in VTA and SNc, and thus has the potential to influence dopamine transmission to a number of cortical and subcortical areas, including the motor and associative regions of dorsal striatum. Dopamine innervates both patch and matrix compartments in the dorsal striatum, though this innervation likely arises from separate populations of DA neurons, especially in rats. In particular, DA projections to the dorsal striatal matrix compartment arise from regions influenced predominantly by limbic, rather than motor and associative, striatal regions. Striosomal DA, by contrast, is from neurons with reciprocal connections from the striosomes themselves, but which are also influenced by ventral striatal innervation.

1.4.3.2. Firing properties of dopamine neurons

Studies of firing properties of dopaminergic neurons seldom report variations between different regions (e.g., VTA versus SNc), and thus all DA neurons are considered to behave similarly. For reviews on the firing patterns of DA neurons, Joel and Weiner (2000) is again a good reference, as are either of two more recent reviews by Schultz (Schultz, 2007a; Schultz, 2007b).

Dopamine neurons exhibit two modes of firing, which they can rapidly switch between: single-spiking (or tonic) and burst modes. Increased excitation and/or decreased inhibition to DA neurons

results in burst firing, whereas the inverse tends to inhibit bursts. Burst firing in dopamine neurons depends on connections with the laterodorsal tegmental nucleus (Lodge and Grace, 2006), and can be elicited by stimulation of pedunculopontine nucleus (Lokwan et al., 1999) or prefrontal cortical areas (Tong et al., 1996). Excitatory input also comes from most regions which receive dopamine inputs, including frontal cortex, hippocampus and amygdala. The majority of input to dopamine neurons is inhibitory, and comes from the striatum, pallidum, and SNr (Tepper and Lee, 2007), as discussed in the previous section. Indirect release of inhibition additionally results from striatal and pallidal inhibition of SNr, which then disinhibits the firing of dopamine neurons in the SNc. Which effect dominates following striatal firing likely depends on the strength and targets of striatal/pallidal neurons. For a review of inhibitory (GABAergic) control of DA neurons, see Tepper and Lee (2007). The lateral habenula also sends inhibitory projections to the VTA/SNc, and neurons in this region have been shown to exhibit responses complementary to those of the DA neurons.

Joel and Wiener (2000) point out that neurons in the limbic striatum and in striosomes of the dorsal striatum can thus directly inhibit the firing of dopamine neurons, causing a decrease in DA release. Matrix neurons in the associative and motor regions, as well as those in limbic areas, may influence DA firing through multisynaptic connections. By inhibiting neurons in the SNr that then project to SNc, DA neurons can be released from inhibition, causing burst firing and an increase in DA release. Burst firing through this mechanism is likely to be both temporally and spatially restricted due to the low firing rates of striatal neurons, and the focal nature of striatonigral as well as SNr-to-SNc projections.

Schultz and colleagues have extensively studied the firing of dopamine neurons in primates during task performance (for review, see Schultz, 2007b). They find that 60-80% of dopamine neurons in SNc and VTA respond with a burst of firing ~60-100 msec after stimuli predicting reward, delivery of primary food and liquid rewards, or salient visual or auditory stimuli. This response is proportional to the discrepancy between the reward and its predicted value: food rewards that are completely predicted no longer produce a phasic DA response, and reward omission produces a dip in DA firing. Further study has shown that the firing of dopamine neurons tracks reward prediction errors (RPEs) particularly well for positive reward predictions, but not for negative outcomes, and DA signaling does not incorporate the costs of actions required to obtain reward (Bayer and Glimcher, 2005; Gan et al.). Phasic DA activity may also encode the predicted value of higher-level or cognitive rewards: DA neurons have been shown to fire in response to advance information about upcoming rewards in addition to the predicted value of rewards themselves (Bromberg-Martin and Hikosaka, 2009). Dopamine responses to reward-predicting stimuli are additionally modulated by the probability of reward delivery (Fiorillo et al., 2003) and its expected value (Tobler et al., 2005). DA responses correlated with reward prediction errors and uncertainty have led several investigators to develop reinforcement learning models of DA function, which may shed light on the numerous deficits in learning, memory and motor performance seen in DA system dysfunction. These models are discussed in more detail in Chapter 4.

1.4.3.3. Dopamine actions on striatal neurons

A number of deficits are observed following lesions in the dopamine system, including deficits in movement and procedural learning, in working memory, decision-making and strategy selection, as well as reduced motivational and emotional responses. These various motor, cognitive and emotional deficiencies are likely the result of dopamine's actions in a variety of target structures, including the dorsal striatum, prefrontal cortex and hippocampus, and the ventral striatum, respectively. In this section, more detail is given regarding the actions of dopamine in the dorsal striatum, but it is

important to realize that dopamine influences neural processing in a number of other structures, including those listed above. Thus, the effects of dopamine on a large network of interconnected brain regions may additionally influence dorsal striatal processing.

Conventional wisdom holds that D1-class (D1 and D5) receptors excite striatal MSNs, whereas D2-class (D2, D3, and D4) receptors inhibit them. Several lines of evidence suggest that this is a generally correct, though grossly oversimplified view. Activation of D1- or D2-class receptors influences neuron excitability in a state-dependent manner. It is also especially important to remember that the actions of dopamine not only directly affect striatal MSNs, but also act on striatal interneurons. Additionally, dopamine acts not only locally at synapses, but likely exerts an even greater influence extrasynaptically by diffusing away from its release sites (Cragg and Rice, 2004; Yung et al., 1995). Studies of the effects of dopamine on the firing of striatal projection neurons thus paint an extraordinarily complex picture.

In the Up state, activation of D1 receptors increases surface expression of AMPA and NMDA receptors – ionotropic glutamate receptors, the activation of which increases a neuron's response to excitatory inputs. D1 receptor activation also indirectly enhances currents evoked by NMDA receptor stimulation (Flores-Hernandez et al., 2002), and increases L-type Ca^{++} currents. In the Down state, however, D1 activation acts to reduce the response to current injection, by inhibiting Na+ currents and increasing K+ currents. Taken together, D1 receptor activation appears to increase the "signal-to-noise ratio" in the striatum by increasing the response to activation that is coherent or sustained enough to generate Up states, while inhibiting the response to transient uncoordinated excitation in the Down state.

Activation of D2 receptors results in a number of direct and indirect effects which decrease the overall excitability of striatal neurons when in the Up state. These include a decrease in AMPA receptor currents, as well as trafficking of AMPA receptors out of the cell membrane. Additionally, D2 receptor activation decreases L-type Ca^{++} currents and reduces the presynaptic release of glutamate, though the precise mechanisms by which the latter occurs are still debated. In the Down state, D2 receptors generally reduce K+ currents, thus encouraging transitions out of the Down state. However, once the transition to Up state has been achieved, activation of D2 receptors makes it more difficult for MSNs to fire action potentials.

Importantly, Up and Down state transitions are prominently observed *in vitro*, and have been observed *in vivo* during slow-wave sleep and under anesthesia (Goto and O'Donnell, 2001; Mahon et al., 2006). However, during awake states, the membrane potentials of MSNs show a unimodal distribution around -60 mV, rather than the bimodal distribution observed in other states (Mahon et al., 2006). The implications of this for D1 versus D2 effects on MSN firing during awake behavior are unknown.

D1 receptors are low-affinity, and thus thought to be activated by high levels of dopamine such as would be released by burst firing. D2 receptors are high-affinity and thought to be activated by lower tonic levels of dopamine release. As previously discussed, D1-class receptors are preferentially expressed in direct pathway neurons projecting to the GPi/SNr, whereas D2-class receptors are expressed in indirect pathway neurons projecting to the GPe. Thus, dopamine burst firing is thought to have a generally excitatory effect on the direct pathway, whereas tonic dopamine release is thought to have a generally inhibitory effect on the indirect pathway. Interestingly, glutamate facilitates dopamine release at corticostriatal synapses via presynaptic receptors on DA terminals (Chesselet, 1984), independent of dopamine impulse activity (Krebs et al., 1991; Nieoullon et al., 1978; Romo et

al., 1986). It is thus possible for excitatory cortical and thalamic input to modulate local dopamine concentrations without accompanying changes in dopamine neuron firing.

In addition to its effects on general excitability, dopamine has been shown to be a critical component of long-term plasticity at corticostriatal synapses. Such changes in synaptic strength are thought to enable the long-term storage of information acquired through learning. Early studies of corticostriatal synaptic plasticity showed that both D1 and D2 receptor activation were required for the induction of long-term depression (LTD) (Calabresi et al., 1992). Later studies confirmed that D2 receptor antagonism blocks LTD – interestingly, in both D1 and D2 receptor expressing MSNs, and mice lacking D2-receptors expressed long-term potentiation (LTP) instead of LTD (Calabresi et al., 1997). How D2 receptor activation enables LTD even in D1 receptor expressing MSNs is still debated; though one likely mechanism has been suggested by Wang et al. (2006). They found that activation of D2 receptors on ACh neurons reduces the firing of these neurons. The reduction in acetylcholine then allows the release of endocannabinoids by the postsynaptic cell, which act presynaptically to inhibit the release of glutamate, thus contributing to LTD in both types of MSNs. The contribution of D1 receptors to LTD, even in D2 receptor expressing MSNs, is even more confusing. Calabresi and colleagues have suggested that the activation of D1/D5 receptors on NOS-positive neurons may stimulate the release of nitric oxide, which has been shown to be critical for the induction of LTD (Calabresi et al., 1999).

LTP in corticostriatal synapses has been shown to depend on activation of D1 receptors, again in both D1 and D2 receptor expressing MSNs (Centonze et al., 2003a; Kerr and Wickens, 2001). These studies have shown that blocking D1-class (D1/D5) receptors blocks LTP, and that mice lacking D1 receptors do not express LTP. Conversely, inactivation of D2 receptors enhances LTP, and as mentioned above, mice lacking D2 receptors express LTP rather than LTD. The mechanisms by which D1/D5 activation affects LTP, even in D2-expressing MSNs, are unknown. The activation of D1 or D2 receptors may have different effects on the induction of LTP or LTD depending on the ongoing state of the cortico-striatal network. Interestingly, the expression of LTD or LTP is further dependent on striatal region and developmental age: the dorsolateral region of anterior striatum has been shown to switch from predominant induction of LTP to predominant LTD with development, whereas dorsomedial striatum tends to express NMDA-dependent LTP across all developmental ages (Partridge et al., 2000).

Spike-timing dependent plasticity (STDP) has also been demonstrated at corticostriatal synapses (Fino et al., 2005), and the mechanisms of this type of plasticity have been shown to critically depend on dopamine receptor activation in both types of MSNs. In STDP, LTP is observed when the presynaptic cell fires before the postsynaptic cell, but LTD is observed when the postsynaptic cell fires first. Shen et al. (2008) showed that in MSNs expressing D2 receptors, STDP protocols could result in LTP or LTD as expected, but blocking the D2 receptors disrupted LTD. Blocking A2a or NMDA receptors disrupted LTP in these neurons. Conversely, D2 agonism enhanced LTD expression and A2a agonism enhanced LTP, even when the opposite plasticity should have been observed based on spike timing. In D1 neurons, LTP was produced as expected in pre-post pairings, but LTD was not observed. Blocking D1 receptors enabled the expression of LTD in these neurons, and resulted in LTD even when LTP would normally be expected based on spike timing.

It is particularly important to realize that striatal interneurons also express dopamine receptors. Best-studied in this regard are the cholinergic interneurons. These interneurons express D2 receptors, and as discussed above, the activation of these receptors may contribute to LTD expression in both classes of MSNs. D5 receptors are likewise expressed in ACh interneurons, and their activation is

required for LTP in these cells (Suzuki et al., 2001). ACh in general acts to enhance the responsiveness of MSNs to excitatory input. Thus, by modulating the release of ACh in the striatum, dopamine exerts additional effects on the excitability of both classes of MSNs. D1-class receptors are also found postsynaptically on fast-firing GABAergic interneurons, which provide feedforward inhibition to MSNs, as well as on NOS-positive interneurons, which as noted may also play a critical role in LTD. Additionally, D2 receptors are commonly found presynaptically in the terminals of cortical cells, fast-firing neurons, and dopamine inputs (autoreceptors) and likely limit the release of neurotransmitter from these cells.

In summary, while it is true that the actions of D1- and D2-class receptor activation are generally excitatory and inhibitory, respectively, dopamine may act through a number of direct and indirect mechanisms to enhance or reduce the overall excitability of both classes of MSNs.

1.4.4. Summary

The striatum can be subdivided into limbic, motor and associative areas corresponding roughly to ventral, dorsolateral and dorsomedial regions, respectively. In addition to these broad regional subdivisions, compartmental structure exists by which patchy striosome compartments can be chemically distinguished from surrounding matrix tissue. The chemical makeup and projection patterns of striosomes suggest that they may integrate information from across the limbic system, and directly influence dopamine release. By contrast, projection neurons in the matrix send direct and indirect pathway projections through the basal ganglia that exert effects on target structures in the thalamus and brainstem. Modulatory control is exerted on the projection neurons in both striosomes and matrix by the combined action of a variety of interneurons as well as dopaminergic input from the SNc.

1.5. Behavioral and electrophysiological studies of striatal function

Anatomical and *in vitro* studies, such as those that have been the focus of this chapter so far, can provide hints regarding the striatal function and the cellular mechanisms underlying them. However, evidence from intact animals during natural and learned behaviors is critical for determining the functions of the basal ganglia. Lesion studies can provide evidence regarding which functions different striatal regions critically support, but electrophysiological studies are needed to determine how the firing of neurons in these regions may contribute to these functions. This section reviews behavioral and electrophysiological evidence from studies in awake behaving subjects, primarily rats and nonhuman primates, further illuminating the contribution of the striatum to ongoing behavior.

1.5.1. Lesion studies

Lesions in different brain regions are made by chemically or physically inactivating a target structure, so that the effects on subsequent behavior (which may be temporary or permanent) can be studied. Such studies often suffer from a lack of specificity, as fibers of passage are often ablated along with neurons situated in the target regions, though certain chemical methods may alleviate this problem. Additionally, regions outside the target zone are often somewhat affected and/or the target region may not be 100% affected. Nonetheless, the action of the specific region is severely limited, and a number of studies using different lesion methods and behavioral paradigms provide converging

evidence that dorsolateral and dorsomedial striatal regions are critically and differentially involved in certain functions. In this section, we review the major clinical observations and early monkey work indicating a role for the basal ganglia, and in particular the striatum, in motor control. We then review comparative lesion studies in rodents, which provide some of the best evidence for the dissociable roles for dorsolateral striatum in motor control, stimulus-response learning and habit formation, contrasted with a role for the dorsomedial striatum in flexible, goal-directed behavior, inhibition of habitual or prepotent behaviors, and memory functions.

1.5.1.1. Human and nonhuman primate evidence for basal ganglia involvement in motor control

Perhaps the earliest evidence that the basal ganglia are involved in movement and motor control comes from the study of Parkinson's disease (PD), Huntington's disease, and dystonia – disorders in which patients exhibit debilitating motor impairments. These diseases have different targets in the basal ganglia and produce markedly different symptoms. In Parkinson's disease, the neurons of the substantia nigra pars compacta degenerate, affecting the tonic and phasic expression of dopamine especially in the dorsal striatum. In Huntington's disease, the projection neurons of the dorsal striatum are differentially affected, altering the firing patterns of neurons projecting out of the basal ganglia from GPi and SNr. Dystonia can result from a number of mechanisms, including damage to basal ganglia circuits from trauma or environmental factors, or from genetic mutations resulting in altered basal ganglia function.

Following over 80% loss of dopaminergic terminals in PD, patients exhibit significant motor deficits, including impaired movement initiation, rigidity, and slowness of movement. The motor symptoms can be alleviated initially through dopamine-replacement therapy, but this does not stop continuing degeneration, and doses must therefore be continually adjusted and eventually become ineffective. Additionally, dyskinesias develop eventually in most patients given L-DOPA dopamine replacement therapy. Treatment using deep brain stimulation, a surgical intervention in which high frequency electrical stimulation is applied to the GPi or STN, in conjunction with significantly reduced medication doses, is then effective in further alleviating motor symptoms. A number of non-motor symptoms are also evident in PD – including anxiety, depression and cognitive impairments – highlighting the involvement of basal ganglia circuitry, and dopamine in particular, in cognitive and emotional function.

The motor deficits observed in Parkinson's disease are recreated in animal models of PD, in which the extent and timecourse of dopamine depletion can be better controlled. Further evidence that the basal ganglia are involved in habitual and efficient movement performance as well as the chunking of learned sequences of actions, comes from studies using these animal models. In particular, Matsumoto et al. (1999) trained two monkeys with unilateral MPTP lesions in the SNc – one given the lesion before training, one after – to perform sequences of arm movements. In the monkey trained before lesions were made, but not in the monkey with pre-training lesions, movements became efficient and stereotyped. When the monkeys were then "surprised" with an early reward delivery, they continued to complete the entire sequence with the arm contralateral to the intact side, but stopped at reward delivery when performing with the arm contralateral to the lesion. These results suggest that dopaminergic innervation of the striatum is necessary for the "chunking" of sequences of movements into a single efficient motor plan. Miyachi et al. (1997) further found a difference between the effects of lesions in different regions of the striatum on sequence learning. They found that lesions in the anterior caudate and putamen impaired sequence learning, but had little effect on the performance of learned sequences. By contrast, lesions in the medioposterior putamen impaired

the performance of well-learned sequences, but had no effect on initial learning. This study provides one of a number of results suggesting that different regions of the striatum are differentially engaged during different stages of learning, an idea that is revisited in the next section.

In Huntington's disease, MSN cell death, especially in the indirect pathway, is associated with erratic movements and chorea as well as prominent personality changes and cognitive decline. Tippett et al. (2007) suggest that mood dysfunction may be related to differential loss of striosomal neurons over matrix neurons, again emphasizing the multiple functional contributions of the basal ganglia. Dystonia is characterized by involuntary muscle contractions that result in abnormal postures or repetitive movements, and is sometimes responsive to anticholinergic drugs or in severe cases, relieved by deep brain stimulation of basal ganglia targets. In monkeys, Kato & Kimura (1992) reproduced some of the motor deficits observed in Huntington's disease or dystonia by making reversible lesions in the striatum and other basal ganglia nuclei. Tremblay and colleagues have similarly demonstrated motor deficits following bicuculine stimulation to basal ganglia nuclei (Francois et al., 2004; Grabli et al., 2004).

1.5.1.2. Striatal lesions in rodents

The above clinical observations and monkey lesion work highlight the role of the basal ganglia in motor control, movement sequence generation and habit formation and hint at their involvement in a number of non-motor functions. More recently, these functions have been investigated in lesion studies in rodents, in which different regions of basal ganglia can be selectively targeted and their effects compared. In an extensive review published recently, White (2009) summarized the results of a number of striatal lesion studies in rodents. For a more thorough treatment, this review is recommended. Below, the discussion is limited to a selection of lesion studies that focus on the differential involvement of dorsomedial (associative) versus dorsolateral (motor) striatal regions in behavior.

In a series of instrumental experiments, Yin, Knowlton and Balleine showed that the dorsolateral striatum was critical for the expression of habitual outcome-insensitive behavior, whereas the dorsomedial striatum was critical for the expression of outcome-sensitive goal-directed behavior (Yin et al., 2004; Yin et al., 2005a; Yin et al., 2006; Yin et al., 2005b). In this task, rats are required to press a lever in order to receive a food reward. Behavior is initially goal-directed: manipulations that reduce the value of the reward (LiCl treatment, feeding to satiety) result in reduced lever pressing. Following several days of training on the task, behavior is no longer sensitive to such manipulations: the rats will continue to press the lever even for undesired rewards. Yin and colleagues showed that in rats given dorsolateral striatal lesions after training, behavior remains goal-directed rather than habitual (unlike control rats, these animals stop pressing the lever), and that the expression of habitual behavior depends on the activity of NMDA receptors. Further, in animals given dorsomedial lesions prior to training, lever-pressing behavior is insensitive to the reward value even early in training, when normal rats would exhibit goal-directed actions. This result for the medial striatum is region-specific: lesions in the posterior dorsomedial striatum produce these results, but lesions in the anterior dorsomedial striatum do not. For reviews of these topics, see Yin and Knowlton (2006), and Balleine et al. (2007).

In another series of lesion experiments, Featherstone and McDonald provide evidence that the dorsolateral striatum is particularly critical for learning or performance of conditional stimulus-response discriminations, whereas dorsomedial lesions may impair the ability to discriminate contexts or inhibit prepotent responses (Featherstone and McDonald, 2004a; Featherstone and

McDonald, 2004b; Featherstone and McDonald, 2005a; Featherstone and McDonald, 2005b). Corbit and Janak (2007) found similar results using a Pavlovian Instrumental Transfer task (PIT). During the initial stages of PIT, animals are presented with repeated pairings of stimuli and reward (Pavlovian training), and separately with training in which they learn that pressing one lever leads to a delivery of a particular reward and pressing the other lever leads to delivery of a different reward (instrumental training). During the test phase, rats are presented with one of the two levers, and additionally the stimuli used in the Pavlovian training are randomly presented. Normal rats press the presented lever more when stimuli previously associated with the same reward as the available lever are presented. Lesions in the dorsolateral striatum reduce this tendency, whereas lesions in the dorsomedial striatum result in increased responding in the presence of the stimulus that was previously paired with a different reward, in addition to that paired with the same reward. These results support the idea that dorsomedial striatum may be critical for disambiguating similar contexts or for inhibiting responses.

A number of studies show performance deficits during reversal learning following lesions in the dorsomedial striatum (Pisa and Cyr, 1990; Ragozzino and Choi, 2004), supporting the idea that the dorsomedial striatum is involved in suppression of inappropriate habits, but not as critical for their initial acquisition. Water maze studies have provided some additional support for these ideas. Rats exhibited increased thigmotaxis (swimming in circles near the edge of the pool) in the water maze after lesions were made in the dorsomedial striatum, which may result from an inability to suppress an inappropriate behavioral strategy (Devan et al., 1999). Whishaw et al. (2007) showed that rats with lesions in the dorsolateral striatum were impaired at a food-reaching task, whereas rats with dorsomedial striatal lesions performed better than controls. Thus, while a number of studies have shown dissociable effects of dorsomedial versus dorsolateral striatal lesions, these results demonstrate further that the dorsomedial striatum may competitively interfere with dorsolateral control of motor behaviors.

Supporting the idea that dorsomedial striatum may be critical for disambiguating closely-associated contexts, Adams et al. (2001) found that pretraining lesions in either dorsolateral or dorsomedial striatum impaired rats ability to perform a conditional discrimination task. However, in a study which did not differentiate between dorsolateral and dorsomedial functions, Atallah et al. (2007) found that inactivation of the dorsocentral striatum impaired performance, but not learning of an odor-approach discrimination task, suggesting that learning deficits must be interpreted with some caution. Nonetheless, the results of Adams et al. show a critical role for dorsomedial and dorsolateral striatum in the acquisition or performance of a conditional discrimination task.

Further supporting a role for the dorsomedial striatum that is distinct from that of the dorsolateral striatum, and related to its presumed cognitive functions, several studies have seen deficits in working memory, as might be expected from an area closely connected to prefrontal cortex (Cook and Kesner, 1988; DeCoteau and Kesner, 2000). For example, (Divac et al., 1978) found impairments on a delayed alternation task, suggesting that dorsomedial striatum was required for remembering the previous response across the delay between trials. Supporting these early findings, Kesner & Gilbert (2006) found that medial caudate lesions impair rats' ability to remember their previous response across a delay in a match-to-sample task.

1.5.1.3. The functional roles of frontal cortical areas projecting to striatum

Another critical clue to the function of various striatal regions comes from the functions of the cortical regions providing glutamatergic input to them, as it has generally been shown that lesions in connected cortical and striatal sites produce similar deficits. While the neuronal control of movement by motor and premotor cortical areas is relatively well understood, the various functions of prefrontal associative cortical areas are less obvious. In this section, we focus on the different executive functions that have been attributed to different regions of prefrontal cortex. Three broad prefrontal regions have been defined: anterior cingulate/medial prefrontal cortex, dorsolateral prefrontal cortex, and orbitofrontal cortex. Each of which can be subdivided further and has been the topic of extensive focused research. As the goal of this section is to summarize the functions in which the associative striatum may be implicated, we limit the discussion here to a very brief and very general review. This necessarily leaves out all details regarding the operation of the various regions, and the reader is directed to a number of excellent review articles and the references therein for further information.

The anterior cingulate cortex (ACC) is located on the medial wall of the prefrontal cortex, and can be further divided into dorsal (paralimbic) and ventral (limbic) tiers (Paus, 2001). Motor and premotor areas project to the dorsal tier, whereas limbic input from thalamus, amygdala and ventral striatum projects to the ventral tier, and the ventral tier then projects to the dorsal tier. ACC sends prominent projections to another frontal cortical region, the dorsolateral prefrontal cortex. Lesions of the ACC, or of the dopaminergic input to the ACC, result in deficits in the initiation of voluntary movements, and/or the inability to suppress triggered movements. Studies in humans have shown EEG activity attributed to neural activation in the ACC at the time of an incorrect response, around the time of response after rules change, and around the time of response in high-conflict situations. Electrophysiological recording in awake behaving monkeys has shown that ACC neurons respond around the time of reward delivery to unexpected rewards as well as unexpected errors (Matsumoto et al., 2007). Lesions in the ACC in rats can result in movement impairments similar to those seen in humans, and disrupt action selection based on goal-directed or outcome-sensitive responding. ACC lesions have been shown to bias rats toward choosing a small low-effort reward rather than a larger high-cost reward. Different regions of the ACC have also been implicated in weighing the value of new information, in integrating costs and reward values, and in integrating reward history to guide action selection. Due to the diversity of deficits observed following ACC lesions, and the variety of responses exhibited by ACC neurons, nailing down a precise role for ACC has been difficult. Recent theories are beginning to favor a role for this region in the processing of uncertainty, which may fit with the observations of increased ACC activity after rule switching, after errors, and in high-conflict situations. From a computational perspective, the processing of uncertainty has been shown to be useful for arbitrating between different response strategies (Daw et al., 2005). For reviews related to the function of the anterior cingulate cortex, see Paus (2001), Rushworth et al. (2004), Walton et al. (2007), Rushworth and Behrens (2008) and Carter and van Veen (2007).

The orbitofrontal cortex (OFC) is the far frontal region of cortex "overlying the orbits" and has been extensively studied in the context of drug addiction and more recently in behavioral studies investigating its role in goal-directed behavior. Lesions in the OFC cause social deficits and increase perseveration (rats with lesions in the OFC fail to alter their responding following devaluation) and/or impulsiveness (rats are no longer willing to tolerate a delay to receive a larger reward). The OFC has been implicated in "supporting behavior mediated by representations of outcome" and neural activity in the OFC has been correlated with the conjunctive encoding of cue plus outcome information. It is hypothesized that the OFC may assign a "common currency" or value to different

stimuli, which can then be used to guide behavior. In this view, perseveration and impulsivity could be considered an inability to update cue-outcome associations, and/or an inability to integrate this information over time. More recently, it has been proposed that the OFC may signal "outcome expectancy," or the value of a given state of the world. Notably, neural activity in the OFC has been shown to change with satiation, and OFC lesions impair normal responding when only the value of the outcome has changed (but response contingencies have not changed). The OFC may play a role in the calculation of reward prediction errors by providing DA neurons with information regarding the expected outcome value. This hypothesized role in expectancy calculation may explain the perseveration seen following OFC damage. In addiction, OFC activation is seen in response to drug stimuli that elicit craving, perhaps consistent with its hypothesized role in encoding expected reward. For reviews on the role of the OFC in addiction, see Everitt et al. (2007). For reviews regarding the OFC in normal behavior, see Schoenbaum et al. (2009), Murray et al. (2007), and Walton et al. (2007).

Lesions in the dorsolateral prefrontal cortex (DLPFC) impair working memory and deficits in task switching, and EEG activity attributable to this region correlates with increasing cognitive demand. Neural correlates of motor planning, task rules, and reward anticipation have been observed in recordings from the DLPFC in rodents and nonhuman primates. It is thought that the DLPFC is responsible for maintaining, manipulating and/or amplifying task relevant features "in the service of planning, problem solving, and predicting forthcoming events." For reviews related to the DLPFC, see Seamans et al. (2008) and Mansouri et al. (2009). Seamans et al. (2008) and Uylings et al (2003) deal especially with the issue of defining prefrontal cortical areas in the rat, and contain a number of useful references. Both suggest that prelimbic cortex may serve as a rudimentary DLPFC in rodents, underlying functions such as working memory and direction of attention toward relevant task features.

Based on structural and functional arguments, there is some debate about whether rats can truly be said to have a prefrontal cortex. Prefrontal cortex is defined in humans based on the presence of a granular layer IV (distinguishing prefrontal from motor and premotor areas), which is absent from rodent frontal cortical regions. Further, it is unclear that rodents are capable of performing many of the executive functions exhibited by humans, even in a limited sense. These issues are discussed by a number of authors (Seamans et al., 2008; Uylings et al., 2003), but at this time it seems generally accepted that rodents do possess cortical structures which are at least broadly analogous to the major prefrontal cortical regions studied in humans, including medial prefrontal, orbitofrontal and dorsolateral prefrontal areas.

1.5.2. Electrophysiological recordings from striatum of awake behaving animals

The previous section reviewed evidence from lesion studies implicating the dorsolateral and dorsomedial striatum in different aspects of procedural learning and motor performance. Existing evidence points to a role for the dorsolateral striatum in motor control and motor sequencing, in procedural learning and habit formation, and in stimulus-response learning. By contrast, the dorsomedial striatum has been implicated in behavioral flexibility and the performance of goal-directed, as opposed to habitual, behavior. Lesion studies can provide only limited information regarding the neural mechanisms by which these behaviors arise. Electrophysiological studies can provide further insight into how the neural activity in these striatal regions may contribute to the expression of different behaviors. It is important to realize that such studies are correlative in nature

and therefore cannot establish causality between neural activity and behavior. Nonetheless, when combined with the results of prior lesion studies, these recording experiments can provide useful information regarding the relationship between neural activity in different brain regions and behavior. In this section, we summarize the results of experiments in which neural activity was recorded from the dorsal striatum during task performance.

1.5.2.1. Neural recording

Two types of neural signals are generally recorded in electrophysiology experiments: unit activity and local field potentials. Unit activity captures the spiking of individual neurons, or a few individual neurons, such that the timing of single spikes in relation to different task events can be recreated. By contrast, local field potentials capture the activity of a large number of neurons near the recording electrode. These signals consist of not only spiking activity, but also dendritic currents arising from input activity. LFPs are thus thought to represent a highly averaged and low-pass filtered summation of the dendritic and spiking activity, providing a more global representation of neuronal activity within a region. In the following sections, we review studies in which striatal single unit activity in behaving animals was recorded, then we summarize the findings related to LFPs recorded from the basal ganglia of humans, nonhuman primates and rats.

1.5.2.1.1. Striatal single unit activity in rodents

Chronic recording studies in awake behaving rats have generally focused on recording from the motor regions of striatum, during tasks requiring sequential movements, skilled motor performance, and/or stimulus-response learning. In one such study, Jog et al. (1999) found that during learning on a conditional T-maze task, activity in the dorsolateral (sensorimotor) striatum comes to emphasize task-boundaries, developing what could be considered a neural correlate of procedural chunking. Related work later revealed that ensemble activity becomes increasingly stable and the signal-to-noise ratio increases for both single units and neuronal ensembles as the T-maze task becomes well-learned (Barnes et al., 2005). Similar results have also been found by other groups (Carelli et al., 1997; Schmitzer-Torbert and Redish, 2004; Tang et al., 2007; Tang et al., 2009; West et al., 1990). These ensemble firing patterns remain stable following the introduction of new conditional stimuli to be learned (Kubota et al., 2009). Different populations of dorsolateral striatal neurons have been shown to respond during maze running versus during reward delivery (Schmitzer-Torbert and Redish, 2004). Some controversy exists regarding whether striatal cells encode spatial parameters (Eschenko and Mizumori, 2007; Mizumori et al., 2004; Yeshenko et al., 2004), but Berke et al. (2009) convincingly argues against this, and Schmitzer-Torbert et al. (2008) suggest that such correlations may be due to task design rather than striatal encoding of spatial position per se.

Most studies have focused on the activity of striatal projection neurons, as these are the most numerous recorded, but a few studies have reported activities of other neuron types. In particular, Berke (2008) showed a diversity of firing patterns among fast-firing neurons during task performance – notable because these neurons were expected to fire synchronously due to gap junction coupling. Kubota et al. (2009) showed that fast-firing neurons are among the few neurons in the striatum to change their firing following the introduction of novel stimuli, though whether these changes were specifically related to the novelty of the presented stimulus or to the particular stimulus modality is unknown. A few other studies also report the activities of TANs, LTS, or other types of interneurons during performance of various tasks (Berke, 2008; Schmitzer-Torbert and Redish, 2008).

Fewer recordings have been made from dorsomedial striatum specifically. Supporting a role in computing stimulus or action values, neurons in medial striatum have been shown to change their

firing following changes in stimulus-response or stimulus-outcome contingencies (Kimchi and Laubach, 2009b). Neurons have additionally been shown to correlate with "response bias" following reversal, indicating a possible role for uncertainty in medial striatal activity (Kimchi and Laubach, 2009a).

Two studies have directly compared medial and lateral activity. Yin et al. (2009) found that dorsomedial striatal activity is enhanced early in training on a skill learning task and then returns to baseline after extended training. By contrast, they found that dorsolateral striatal activity is elevated late in training. Kimchi et al. (2009) found that neurons in both regions were modulated early during training on an instrumental task and that the number of units modulated and the degree of modulation increased in both regions with training. These studies provide limited and conflicting information regarding the simultaneous and potentially competitive interactions of the two regions during behavior, and suggest that task parameters are critical determinants of neural activation.

1.5.2.1.2. Striatal recordings in primates

The basal ganglia are known to be critical for motor control and sequence generation and early electrophysiological studies often focused on finding neural correlates of such movement and sequence-related activity. The striatum is also a major recipient of the reward prediction error signals from the SNc, receives information from almost all parts of cortex, and is known to be important for stimulus-response learning and habit development. As reinforcement learning theory has become more prominently applied to brain function, recent research has additionally focused on the role that the striatum may play in such functions. Thus, recent research has focused on finding neural correlates of reinforcement learning in the striatum, especially signals related to reward expectation, including the computation of state values and/or action values. Select studies are briefly reviewed in this section.

Hikosaka and colleagues have studied extensively the role of the striatum in the control of eye movements (for review, see Hikosaka (2007), and have shown that caudate projection neurons exhibit firing related to target position, which is modulated by the reward expected (Kawagoe et al., 1998). Interestingly, the development of value-related responses in the striatum is coincident with the development of reward-prediction signals in the dopamine neurons of the SNc. Dopamine neurons show a phasic response to the stimuli predicting large reward, while they show a dip in response to stimuli predicting no reward. Hikosaka and colleagues have shown that the application of D1 receptor antagonists increases the response time of monkeys during large-reward trials, whereas D2 antagonists slows the responses during small-reward trials, suggesting a dissociation between the role of the direct and indirect pathways in the initiation of movements (Nakamura and Hikosaka, 2006). The direct pathway appears to play an important role in initiating movements in response to positive reward expectancy, while the indirect pathway may be critical for initiating responses to small or negatively valued rewards. Lau and Glimcher (2008) further found two classes of value-encoding neurons in the striatum of monkeys performing a matching task. "Action-value" neurons fired prior to movement in relation to the values of the available actions, whereas "chosen-value" neurons fired after movement in relation to the value of the performed action. Work from these and other authors (Pasquereau et al., 2007; Samejima et al., 2005), suggests that the basal ganglia contribute to the control of motor behaviors based on reward information, especially the expected values of actions.

The timing of striatal neuronal responses, especially in relation to the performance of motor responses, is of critical importance in determining what role the striatum may play in the control of movement and the selection of actions. As mentioned previously, many studies have confirmed that

the majority of neurons in the striatum fire after the onset of movement, suggesting that the striatum may not be directly controlling muscle activity. As noted above, Lau and Glimcher (2008) did find a subset of neurons that fired prior to execution of movement, which therefore could have been involved in the action-selection process. They also found a large proportion of neurons that encoded the values of the chosen responses, or encoded the direction of the previous response, again supporting the idea that the majority of neurons in the striatum are active following the onset of movement. Lau and Glimcher argue that those neurons that fire after the movement are more likely to play "an evaluative role in learning itself" – perhaps by evaluating the success of a chosen action, or by evaluating the accuracy of currently held beliefs about the value of a chosen action. In another study, these authors found that neurons that fired following movement and following reward delivery/omission, encoded either the direction of the preceding action, or the outcome of the trial, but not generally both (Histed et al., 2009; Lau and Glimcher, 2007), providing evidence that both pieces of information required for action evaluation are available in separate populations of neurons in the striatum.

Pasupathy, Miller and colleagues have investigated the timing of striatal activity in relation to cortical activity during a switching task and found that the striatum develops neural activity related to the currently relevant stimulus-response or stimulus-outcome contingencies faster than the cortex following a reversal in these contingencies. The development of activity related to the new contingencies in the cortex matches the behavioral performance, and the authors suggest that the basal ganglia may train the cortex following such a reversal (Pasupathy and Miller, 2005). Elaborating on these findings, Histed et al. (2009) showed that the direction selectivity at the time of the saccade (response) occurred earlier in the trial and earlier after a switch in the caudate compared to the PFC, again suggesting a leading role for the striatum following a switch in stimulus-response or stimulus-outcome contingencies. Histed et al. also showed delay-period activity in both caudate and PFC related to the outcome on the previous trial, suggesting that both areas could use this maintained outcome information to modulate neural activity and behavioral performance on the subsequent trial.

The timing of striatal activation across different stages of training is also of interest, especially in light of the results of rodent lesion studies that have shown that lesions in dorsomedial/associative striatum result in habitual performance whereas lesions in dorsolateral/sensorimotor striatum result in goal-directed behavior. As it is thought that the normal progression of habit development is that behavior is initially goal-directed and then becomes habitual after extended training, the engagement of associative versus sensorimotor regions during different stages of learning should be related to behavioral strategies exhibited during habit development. Consistent with the idea that the associative networks drive goal-directed behavior early in training, and the sensorimotor networks drive behavior later in training, a number of studies, including human fMRI and rodent and nonhuman primate electrophysiology, have shown that during the course of normal procedural learning, associative regions of cortex and striatum are more active initially whereas sensorimotor regions of cortex and striatum are more active in the late stages of training (Miyachi et al., 2002; Tricomi et al., 2009; Yin et al., 2009). However, as noted in the previous section, other researchers have not observed this pattern (e.g., Kimchi et al., 2009), and the specific task demands are likely to be critical in determining whether either, both or neither loop is engaged.

Combined, these studies suggest that the striatum may play a critical role in evaluating chosen actions, based on their expected outcomes, and/or in evaluating and updating the expectations themselves. Different striatal populations are engaged in these processes during different stages of

training, and downstream targets of the basal ganglia, including different cortical or brainstem regions, may use the stored value information to adjust behavioral performance.

1.5.2.2. Local field activity

Different characteristic oscillations have been observed as markers of different brain states, especially different states of sleep and waking. For example, slow-wave sleep is characterized by coherent low-frequency oscillations (1-4 Hz delta waves and/or 7-12 Hz spindles) in multiple brain regions. The hippocampus exhibits strong 5-12 Hz theta oscillations during awake behavior and REM sleep. Thalamocortical oscillations in the 5-12 Hz range are also observed during waking, either during quiet resting or active sensory processing, depending on the experimental design and the brain region under investigation. Higher-frequency activity (> 35 Hz) is also often observed, generally during active cortical or hippocampal processing. The various frequency bands have been given names, though the frequency ranges indicated by these labels often overlap as the observed rhythms are modulated by task and brain region. Nonetheless, on the low end are the delta rhythms (~1-5 Hz), then theta (~ 4-8 Hz), alpha (8-12 Hz), beta (14-35 Hz), and gamma (> 35 Hz) frequencies. A prominent 5-12 Hz oscillation called "mu" is also studied in the motor cortex. The oscillations literature is quite confusing, and a thorough review is beyond the scope of the summary here. Gyorgy Buzsaki's *Rhythms of the Brain* (Buzsaki, 2006) provides a good introduction to the subject, especially with regard to cortical and hippocampal oscillations, and also contains a number of relevant references.

Of particular interest to basal ganglia researchers are lower-frequency rhythms, as delta, theta and beta-band oscillations are abnormally prominent in recordings from STN and GP in Parkinson's Disease patients. Dopamine depletion is additionally known to produce an increase in the power of low-frequency oscillations in nonhuman primates and rats. How these oscillations arise and how they relate to normal functioning of the basal ganglia is unclear, however. More recently, attention has turned to the role that low-frequency oscillations may play during waking in normal subjects, as characteristic profiles have been observed during behavior and periods of awake rest. Below, we briefly summarize some of these recent findings, as background for the experiments presented in Chapter 3.

Implantation of deep brain stimulators for treatment of Parkinson's Disease (PD) has provided the opportunity to record electrophysiological signals from the basal ganglia of human patients. These recordings focus on the STN and GPi, as these are the target structures for stimulation, and have shown that in PD patients, these basal ganglia nuclei exhibit prominent rhythmicity at low frequencies (< 20 Hz). These findings have been further studied in nonhuman primate and rodent models of PD, in which it is possible to observe both normal and pathological states. These studies have found that dopamine depleted animals exhibit much higher amplitude rhythmicity at low frequencies than do normal animals. Further, individual neurons recorded from STN, GPi and striatum after DA depletion are more rhythmic at these low frequencies, exhibit synchronized firing with each other, and fire or burst in synchrony with the ongoing local field oscillations (Dejean et al., 2008; Goldberg et al., 2004; Raz et al., 2001). In this state, the movement-related firing characteristic of each of these regions deteriorates. STN and GPe neurons have been shown to exhibit pacemaker activity *in vitro*, leading some to suggest that an internal basal ganglia pacemaker may be the source of the abnormal oscillations seen in the DA depleted state, though the natural rhythm of this pacemaker is much lower (~0.4-1.2 Hz) than the oscillation frequencies typically observed in PD. Alternatively, cortical rhythms may be more easily transmitted through the basal ganglia network in DA depleted conditions. Magill and colleagues (Magill et al., 2000; Magill et al., 2001) showed that

this was true in anesthetized preparations and more recently DeJean et al. (2008) have shown similar results in awake rats. It has thus been suggested that the low-frequency rhythms seen in the basal ganglia in PD may be similar to the "idling" rhythms seen in cortex and associated with quiet wakeful states, and that the inability to break out of this state may contribute to the motor symptoms, in particular the paucity of movement and difficulty with movement initiation, seen in PD. Stimulation of the STN may be effective in alleviating the motor symptoms of PD by driving the network at high frequencies and thus suppressing this abnormal synchrony.

Costa et al. (2006) show that DA depletion or hyperdopaminergia can produce both firing rate and synchrony changes in conjunction with motor impairments. DA depletion renders rats akinetic, and concurrent increases are observed in striatal firing rate as well as increased power in the delta and beta bands and an increase in the percentage of units entrained to these low frequencies. Dopamine replacement restores normal oscillatory activity patterns. Hyperdopaminergia results in hyperkinesia, and increased power in the theta and gamma frequency bands are observed in the cortex and striatum in this state, combined with a reduction in entrainment of single units to the LFPs. Burkhardt et al. (2009) suggest that the synchrony and firing rate effects of dopamine depletion may be somewhat dissociable, with increased synchrony occurring after administration of D1, but not D2 antagonists. Conversely, D2 antagonism (but not D1) reduced firing rates and LFP power in the striatum.

For reviews of abnormal low-frequency rhythmicity in Parkinson's Disease and dopamine depletion, see Bevan et al. (2002) and Bergman et al. (1998).

Sharott et al. (2009) further investigated the entrainment of different subtypes of striatal units to the ongoing LFP in halothane anesthetized rats. They found that a large percentage of neurons of each type were entrained to a 2-4 Hz delta rhythm, whereas only FFs were significantly entrained to higher-frequency gamma oscillations. Further, FFs tended to show significant pairwise correlated firing with MSNs and other FFs, whereas other interneuron types exhibited very limited correlated firing.

In the basal ganglia of normal awake behaving animals, our knowledge regarding the characteristic activity of LFPs and the relationship of single units to the local field potentials is much more limited. However, a number of authors have provided some interesting clues. Courtemanche et al. (2003) found that beta-band oscillations (~16 Hz) were prominent in the striatum of normal behaving monkeys and modulated in a task-related manner. Further, about half of medium spiny and tonically active neurons were entrained to these rhythms during task performance. In rats, Berke et al. (2004) also found low-frequency rhythms, this time in the theta band (~8 Hz), which were coherent with hippocampal theta rhythms during awake behavior. Entrainment of medium spiny units to the ongoing low-frequency oscillations was graded within the striatum such that more units were entrained ventromedially than dorsolaterally. Other authors have recently suggested that even lower frequency rhythms may be relevant in the striatum of awake behaving rats: Schmitzer-Torbert and Redish (2008) and Kimchi and Laubach (2009) both reported significant power for frequencies under 5 Hz, and medium spiny neurons were entrained to these rhythms in-task. These studies indicate that low-frequency rhythms are prominent in the striatum of normal subjects during behavior, are modulated in a task-dependent manner, and are relevant to firing of local single units. The relevant frequencies during any particular task, however, likely depend critically on the species and region under investigation.

There has been substantial emphasis on low-frequency rhythms within the striatum, but higher-frequency gamma rhythms (> 35 Hz) are also prominently observed, especially during task

performance. Two high-frequency oscillations are most commonly observed, the lower centered around 50 Hz, and the higher around 80 Hz. Masimore et al. (2005) showed that the lower of these two frequencies exhibited a peak in power especially around movement initiation and following reward delivery. Berke (2009) and van der Meer et al. (2009) have both recently shown that the 50- and 80-Hz gamma frequencies tend not to co-occur and that the prominent frequency switches between gamma-50 and gamma-80 around reward delivery. Both authors further demonstrated that fast-firing interneurons, especially in the ventral striatum, are often entrained to either or both of these high frequency rhythms, suggesting that these neurons may play a critical role in the timing of striatal processing at short timescales.

Finally, in work that extends from the experiments described in Chapter 3, Tort et al. (2008) showed that low-frequency theta and high-frequency gamma rhythms interact with each other during task performance. This study showed that during task performance, when theta power was highest, gamma oscillations tended to "nest" within the theta cycle – gamma power was highest at the trough of the theta cycle. Interestingly, the relevant theta and gamma frequencies differed between the striatum and hippocampus in this study. In the striatum, the theta frequency was lower – around 5-8 Hz – and modulated a higher-frequency rhythm between 80 and 120 Hz. In the hippocampus, a higher-frequency 7-12 Hz theta modulated two different high-frequency rhythms – one centered around 80 Hz and one centered around 160 Hz. These results suggest that while theta-gamma nesting may be a common mechanism contributing to the regulation and timing of neural computation in numerous brain regions, the striatal and hippocampal processes which give rise to this cross-frequency interaction are specific to local networks.

In summary, while it is unknown precisely what cellular mechanisms are responsible for generating the oscillations observed in local field potentials in the striatum, it is clear that these rhythms reflect neural activity in normal subjects and through the entrainment of single striatal units, are likely to be relevant to local striatal computation during task performance. In particular, the suggestion that gamma-frequency oscillations arise from local processes, especially the activity of fast-firing interneurons in the striatum, continues to gain strength. It remains unclear, however, whether the low-frequency rhythms observed in striatum are truly locally generated, or whether these are the result of striatal interactions with cortex and thalamus, regions which are known to produce oscillations in these low-frequency bands. Dopamine depletion, such as in Parkinson's Disease, may alter the state of the cortico-basal ganglia-thalamic network such that the entire circuit exhibits abnormal and strong low-frequency oscillations, which may further contribute to the motor impairments observed in PD.

1.5.2.3. Summary

In humans, the debilitating motor deficits observed in Parkinson's Disease and Huntington's Disease have drawn attention to the importance of the basal ganglia in motor control. In Huntington's Disease, cognitive symptoms are also prominent, and more recently, the basal ganglia have been implicated in a number of non-motor diseases, such as Tourette Syndrome and obsessive-compulsive disorder. Lesion studies have demonstrated differential contributions of various striatal regions and their connected cortical regions to motor versus cognitive functioning. In particular, the sensorimotor striatum has been shown to be critical for stimulus-response learning and the development of motor habits. The associative striatum has been shown to be critical for the expression of goal-directed and flexible behavior, perhaps by providing uncertainty-based arbitration between different behavioral strategies. In rats, electrophysiological studies of dorsal striatum have primarily focused on the sensorimotor regions, where activity timed to the action boundaries of tasks has been observed, and

such activity develops in conjunction with improved behavioral performance. In the monkey, electrophysiological studies, influenced by reinforcement learning ideas, have shown that striatal neurons exhibit firing correlated with action values and outcome values following the onset of movement, and suggest that the striatum may play a particularly important role in the evaluation of actions. Finally, in Parkinson's Disease, the strength of abnormal oscillations has been correlated with the severity of some motor symptoms, drawing attention to the potential function of low-frequency rhythms in normal and disease states. It has recently been demonstrated in both monkeys and rats that low-frequency rhythms are task-relevant in normal subjects, and may modulate the local computation occurring at higher frequencies. This higher-frequency gamma processing is likely to depend on rhythms generated by fast-firing interneurons, which may critically determine the timing of spiking among the projection unit populations in the striatum. The abnormal entrainment of the entire cortico-basal ganglia network to low frequency rhythms may disrupt this kind of normal processing and contribute to the motor symptoms seen in PD.

1.6. Conclusions

The previous chapter summarized the basic anatomical, behavioral and electrophysiological findings related to basal ganglia structure and function in health and disease. This evidence is beginning to converge on a role for the sensorimotor basal ganglia loop in the selection and/or evaluation of actions through the reinforcement of stimulus-response associations and motor programs by modulation of corticostriatal synaptic plasticity. The connectivity of different basal ganglia nuclei into direct, indirect and hyperdirect pathways, and the existence of multiple cortico-basal ganglia-thalamocortical loops suggests that associative and limbic striatal regions perform similar computations on different incoming information. The compartmental organization of the striatum into striosomes and matrix, combined with the different neuronal subtypes and neuromodulatory input constrains how these computations may be performed.

This chapter has focused on the basic organizational principles and experimental findings most relevant to the electrophysiology experiments described in Chapters 2 and 3. This necessarily leaves out a number of interesting topics. In particular, an extensive literature exists regarding the role of the basal ganglia in addiction, in which drugs of abuse may take advantage of the same processes that contribute to normal procedural learning and habit formation and drive them into abnormal modes of operation. Likewise omitted are a number of fMRI and other imaging studies in humans, many of which support the results presented in this chapter by extending the findings from animal studies to human subjects. The imaging literature is especially diverse as these non-invasive techniques have been used to investigate all aspects of basal ganglia engagement in human subjects. These include activation in different disease states, in addiction and drug seeking, and during normal procedural or reinforcement-based learning.

In short, an extensive amount of work has been done in the last 50+ years investigating the specific functions of the basal ganglia in health and disease. A number of theories regarding how the basal ganglia impact movement generation have been developed and modified during this time and a coherent picture is beginning to emerge. This latest understanding of basal ganglia function draws on previous anatomical, behavioral, and electrophysiological findings, and incorporates computational reinforcement learning theories. The picture, however, is still far from conclusive and far from complete. In the following chapters, the results of two electrophysiological recording experiments are presented which expand on current knowledge of striatal function. The first, presented in Chapter 2, compares activity simultaneously recorded from dorsolateral (sensorimotor) and dorsomedial

(associative) striatum in rats as they acquire and are overtrained on a T-maze task. A complex and dynamic pattern of neuronal activation was found, which differed dramatically between the two regions, both across task performance and across training. These results further highlight the different functional roles of the two striatal regions, demonstrate that both can be active simultaneously during learning, and suggest a novel way by which their activation may lead to behavioral expression. Chapter 3 presents the results of experiments in which local field potentials were simultaneously recorded from both the dorsal striatum and the hippocampus during T-maze acquisition. These results demonstrate that both learning and memory structures are active during task performance, that both structures exhibit low-frequency oscillations during behavior in normal subjects, and that the cross-structure coordination of these rhythms is critical for successful learning on the task. Chapter 4 returns to the issue of how different regions of the striatum may contribute to reinforcement learning by summarizing different concepts from the computational field, presenting a model that extends these ideas to the findings presented in Chapter 2, and suggesting novel experiments which may further clarify the participation of sensorimotor and associative striatal regions in different aspects of behavioral control.

Figures

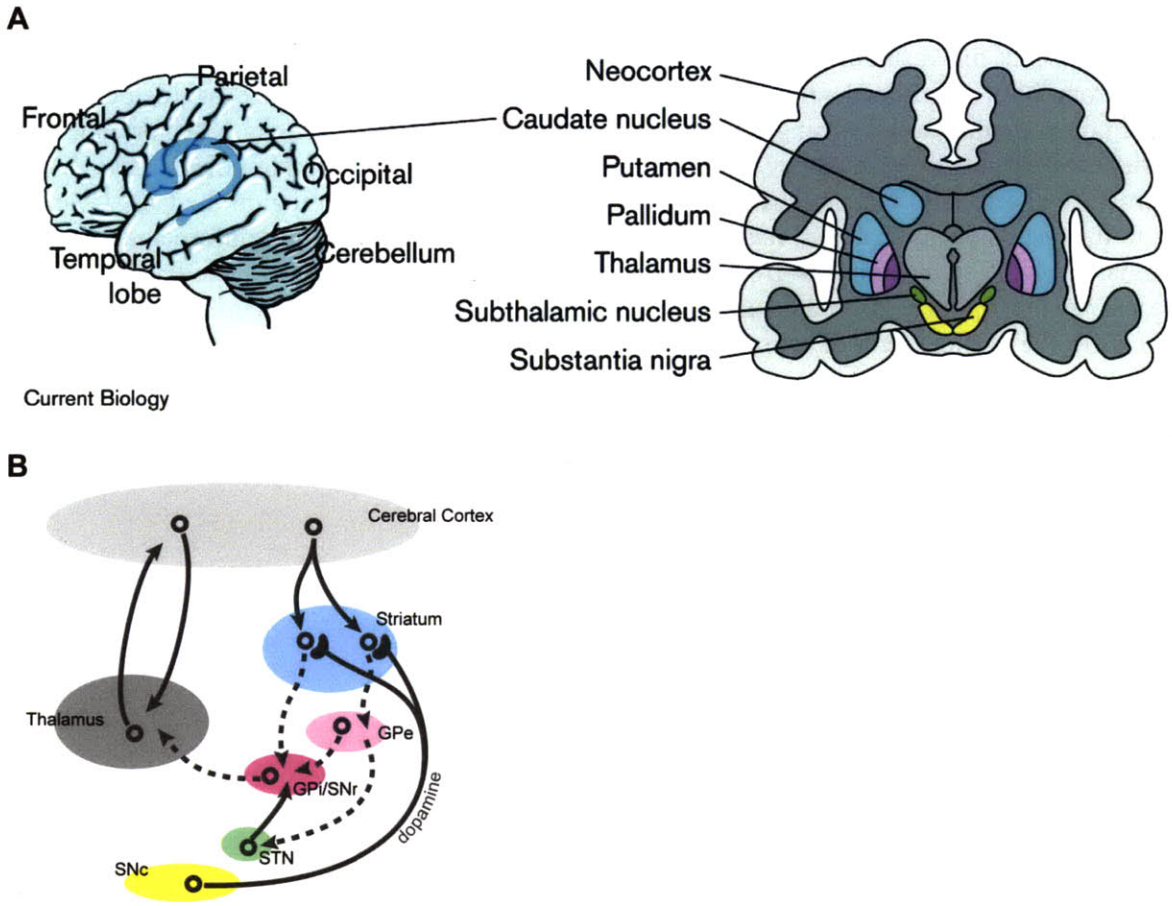

Figure 1.1. Anatomy and connectivity of the basal ganglia.
(A) The basic anatomy of the brain showing the major regions within the basal ganglia: the striatum (blue), which is made up of the caudate nucleus and the putamen; the pallidum (pink), which is made up of outer and inner segments; the subthalamic nucleus (green); and the substantia nigra (yellow). From Graybiel, A.M. (2000) "The basal ganglia." *Current Biology*. (B) Schematic representation of the loop architecture of cortico-basal ganglia-thalamocortical circuits, with structures color-coded as in A. Except for labelled dopamine projections from SNc to striatum, excitatory glutamatergic projections are indicated by solid lines and inhibitory GABAergic projections are denoted by dashed lines.

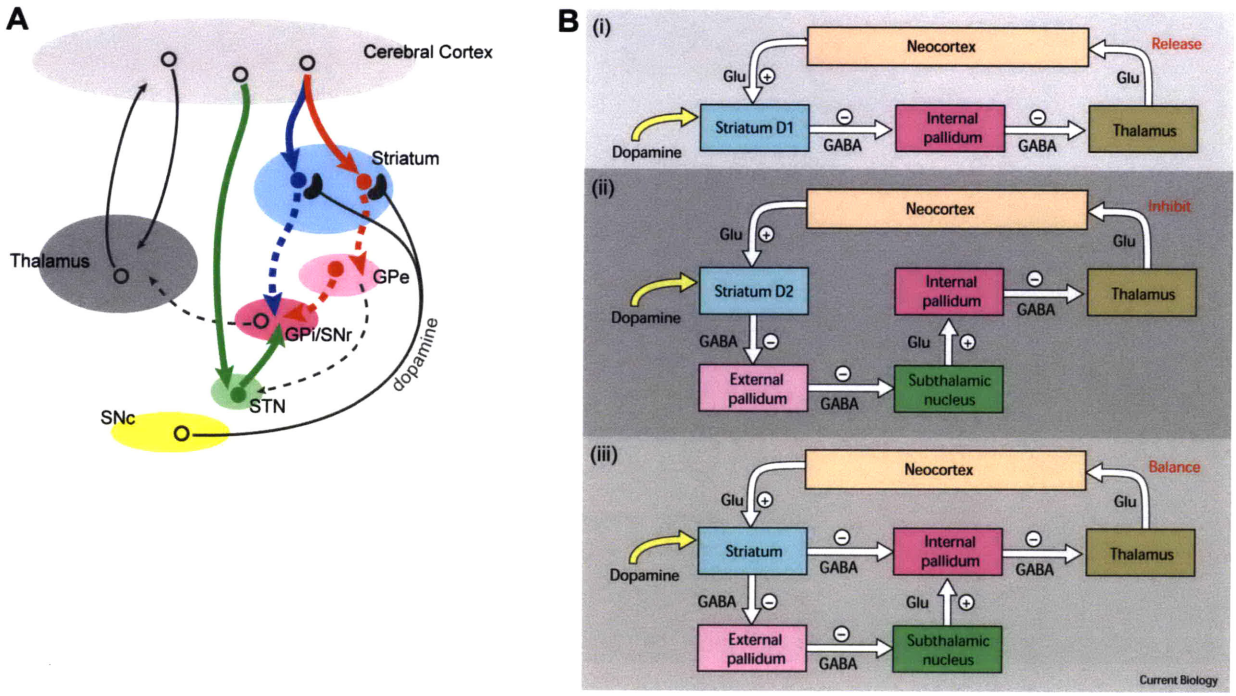

Figure 1.2. The direct, indirect and hyperdirect pathways of the basal ganglia
(A) Schematic illustration of the direct (blue), indirect (red) and hyperdirect (green) pathways from cortex to GPi/SNr. (B) The brake-accelerator model for basal ganglia motor disorders. (i) The direct pathway (leading to release of movement) consists of two successive GABAergic connections, from the striatum to the internal pallidum and from the internal pallidum to the thalamus. This flow diagram sugests that excitatory (glutamate; Glu) inputs from the neocortex to the striatum would disinhibit thalamic neurons. Dopamine modulates the system mainly in the striatum, where it activates D1-class and D2-class dopamine receptors. (ii) In the indirect pathway (leading to inhibition of movement), there is an extra step after the external pallidum, so that the subthalamic nucleus excites the internal pallidum. (iii) Balance is achieved when these antagonistic systems are combined under normal circumstances. From Graybiel, A. M. (2000) "The basal ganglia." *Current Biology.*

Figure 1.3. Compartmental organization of the striatum.
A thin slice through the striatum of the human brain stained for a marker of acetylcholine. Patchy gray zones are acetylcholine-poor striosomes. From Graybiel, A.M. (1984) Neurochemically specified subsystems in the basal ganglia. In: *Functions of the Basal Ganglia*, D. Evered and M. O'Connor, Eds. London: Pitman, pp. 114-149.

References

Adams, S., Kesner, R. P., and Ragozzino, M. E. (2001). Role of the medial and lateral caudate-putamen in mediating an auditory conditional response association. Neurobiol Learn Mem *76*, 106-116.

Aldridge, J. W., and Berridge, K. C. (1998). Coding of serial order by neostriatal neurons: a "natural action" approach to movement sequence. J Neurosci *18*, 2777-2787.

Alexander, G. E., DeLong, M. R., and Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu Rev Neurosci *9*, 357-381.

Aosaki, T., Kimura, M., and Graybiel, A. M. (1995). Temporal and spatial characteristics of tonically active neurons of the primate's striatum. J Neurophysiol *73*, 1234-1252.

Aosaki, T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A. M., and Kimura, M. (1994). Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. J Neurosci *14*, 3969-3984.

Apicella, P., Deffains, M., Ravel, S., and Legallet, E. (2009). Tonically active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context. Eur J Neurosci *30*, 515-526.

Atallah, H. E., Lopez-Paniagua, D., Rudy, J. W., and O'Reilly, R. C. (2007). Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. Nat Neurosci *10*, 126-131.

Balleine, B. W., Delgado, M. R., and Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. J Neurosci *27*, 8161-8165.

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., and Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature *437*, 1158-1161.

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

Bennett, B. D., and Bolam, J. P. (1993). Characterization of calretinin-immunoreactive structures in the striatum of the rat. Brain Res *609*, 137-148.

Bennett, B. D., and Bolam, J. P. (1994). Synaptic input and output of parvalbumin-immunoreactive neurons in the neostriatum of the rat. Neuroscience *62*, 707-719.

Bergman, H., Feingold, A., Nini, A., Raz, A., Slovin, H., Abeles, M., and Vaadia, E. (1998). Physiological aspects of information processing in the basal ganglia of normal and parkinsonian primates. Trends Neurosci *21*, 32-38.

Berke, J. D. (2008). Uncoordinated firing rate changes of striatal fast-spiking interneurons during behavioral task performance. J Neurosci *28*, 10075-10080.

Berke, J. D. (2009). Fast oscillations in cortical-striatal networks switch frequency following rewarding events and stimulant drugs. Eur J Neurosci *30*, 848-859.

Berke, J. D., Breck, J. T., and Eichenbaum, H. (2009). Striatal versus hippocampal representations during win-stay maze performance. J Neurophysiol *101*, 1575-1587.

Berke, J. D., Okatan, M., Skurski, J., and Eichenbaum, H. B. (2004). Oscillatory entrainment of striatal neurons in freely moving rats. Neuron *43*, 883-896.

Bertran-Gonzalez, J., Bosch, C., Maroteaux, M., Matamales, M., Herve, D., Valjent, E., and Girault, J. A. (2008). Opposing patterns of signaling activation in dopamine D1 and D2 receptor-expressing striatal neurons in response to cocaine and haloperidol. J Neurosci *28*, 5671-5685.

Bevan, M. D., Booth, P. A., Eaton, S. A., and Bolam, J. P. (1998). Selective innervation of neostriatal interneurons by a subclass of neuron in the globus pallidus of the rat. J Neurosci *18*, 9438-9452.

Bevan, M. D., Magill, P. J., Terman, D., Bolam, J. P., and Wilson, C. J. (2002). Move to the rhythm: oscillations in the subthalamic nucleus-external globus pallidus network. Trends Neurosci *25*, 525-531.

Blazquez, P. M., Fujii, N., Kojima, J., and Graybiel, A. M. (2002). A network representation of response probability in the striatum. Neuron *33*, 973-982.

Bromberg-Martin, E. S., and Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. Neuron *63*, 119-126.

Burkhardt, J. M., Jin, X., and Costa, R. M. (2009). Dissociable effects of dopamine on neuronal firing rate and synchrony in the dorsal striatum. Front Integr Neurosci *3*, 28.

Buzsaki, G. (2006). Rhythms of the Brain (New York, USA, Oxford University Press).

Calabresi, P., Gubellini, P., Centonze, D., Sancesario, G., Morello, M., Giorgi, M., Pisani, A., and Bernardi, G. (1999). A critical role of the nitric oxide/cGMP pathway in corticostriatal long-term depression. J Neurosci *19*, 2489-2499.

Calabresi, P., Maj, R., Pisani, A., Mercuri, N. B., and Bernardi, G. (1992). Long-term synaptic depression in the striatum: physiological and pharmacological characterization. J Neurosci *12*, 4224-4233.

Calabresi, P., Saiardi, A., Pisani, A., Baik, J. H., Centonze, D., Mercuri, N. B., Bernardi, G., and Borrelli, E. (1997). Abnormal synaptic plasticity in the striatum of mice lacking dopamine D2 receptors. J Neurosci *17*, 4536-4544.

Carelli, R. M., Wolske, M., and West, M. O. (1997). Loss of lever press-related firing of rat striatal forelimb neurons after repeated sessions in a lever pressing task. J Neurosci *17*, 1804-1814.

Carter, C. S., and van Veen, V. (2007). Anterior cingulate cortex and conflict detection: an update of theory and data. Cogn Affect Behav Neurosci *7*, 367-379.

Centonze, D., Bracci, E., Pisani, A., Gubellini, P., Bernardi, G., and Calabresi, P. (2002). Activation of dopamine D1-like receptors excites LTS interneurons of the striatum. Eur J Neurosci *15*, 2049-2052.

Centonze, D., Grande, C., Saulle, E., Martin, A. B., Gubellini, P., Pavon, N., Pisani, A., Bernardi, G., Moratalla, R., and Calabresi, P. (2003a). Distinct roles of D1 and D5 dopamine receptors in motor activity and striatal synaptic plasticity. J Neurosci *23*, 8506-8512.

Centonze, D., Grande, C., Usiello, A., Gubellini, P., Erbs, E., Martin, A. B., Pisani, A., Tognazzi, N., Bernardi, G., Moratalla, R., *et al.* (2003b). Receptor subtypes involved in the presynaptic and postsynaptic actions of dopamine on striatal interneurons. J Neurosci *23*, 6245-6254.

Chesselet, M. F. (1984). Presynaptic regulation of neurotransmitter release in the brain: facts and hypothesis. Neuroscience *12*, 347-375.

Chesselet, M. F., and Graybiel, A. M. (1986). Striatal neurons expressing somatostatin-like immunoreactivity: evidence for a peptidergic interneuronal system in the cat. Neuroscience *17*, 547-571.

Cohen, M. X., and Frank, M. J. (2009). Neurocomputational models of basal ganglia function in learning, memory and choice. Behav Brain Res *199*, 141-156.

Contant, C., Umbriaco, D., Garcia, S., Watkins, K. C., and Descarries, L. (1996). Ultrastructural characterization of the acetylcholine innervation in adult rat neostriatum. Neuroscience *71*, 937-947.

Cook, D., and Kesner, R. P. (1988). Caudate nucleus and memory for egocentric localization. Behav Neural Biol *49*, 332-343.

Corbit, L. H., and Janak, P. H. (2007). Inactivation of the lateral but not medial dorsal striatum eliminates the excitatory impact of Pavlovian stimuli on instrumental responding. J Neurosci *27*, 13977-13981.

Costa, R. M., Lin, S. C., Sotnikova, T. D., Cyr, M., Gainetdinov, R. R., Caron, M. G., and Nicolelis, M. A. (2006). Rapid alterations in corticostriatal ensemble coordination during acute dopamine-dependent motor dysfunction. Neuron *52*, 359-369.

Courtemanche, R., Fujii, N., and Graybiel, A. M. (2003). Synchronous, focally modulated beta-band oscillations characterize local field potential activity in the striatum of awake behaving monkeys. J Neurosci *23*, 11741-11752.

Cowan, R. L., Wilson, C. J., Emson, P. C., and Heizmann, C. W. (1990). Parvalbumin-containing GABAergic interneurons in the rat neostriatum. J Comp Neurol *302*, 197-205.

Cragg, S. J., and Rice, M. E. (2004). DAncing past the DAT at a DA synapse. Trends Neurosci *27*, 270-277.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci *8*, 1704-1711.

DeCoteau, W. E., and Kesner, R. P. (2000). A double dissociation between the rat hippocampus and medial caudoputamen in processing two forms of knowledge. Behav Neurosci *114*, 1096-1108.

Dejean, C., Gross, C. E., Bioulac, B., and Boraud, T. (2008). Dynamic changes in the cortex-basal ganglia network after dopamine depletion in the rat. J Neurophysiol *100*, 385-396.

Devan, B. D., McDonald, R. J., and White, N. M. (1999). Effects of medial and lateral caudate-putamen lesions on place- and cue-guided behaviors in the water maze: relation to thigmotaxis. Behav Brain Res *100*, 5-14.

Divac, I., Markowitsch, H. J., and Pritzel, M. (1978). Behavioral and anatomical consequences of small intrastriatal injections of kainic acid in the rat. Brain Res *151*, 523-532.

Eblen, F., and Graybiel, A. M. (1995). Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. J Neurosci *15*, 5999-6013.

Eschenko, O., and Mizumori, S. J. (2007). Memory influences on hippocampal and striatal neural codes: effects of a shift between task rules. Neurobiol Learn Mem *87*, 495-509.

Everitt, B. J., Hutcheson, D. M., Ersche, K. D., Pelloux, Y., Dalley, J. W., and Robbins, T. W. (2007). The orbital prefrontal cortex and drug addiction in laboratory animals and humans. Ann N Y Acad Sci *1121*, 576-597.

Faull, R. L., Dragunow, M., and Villiger, J. W. (1989). The distribution of neurotensin receptors and acetylcholinesterase in the human caudate nucleus: evidence for the existence of a third neurochemical compartment. Brain Res *488*, 381-386.

Featherstone, R. E., and McDonald, R. J. (2004a). Dorsal striatum and stimulus-response learning: lesions of the dorsolateral, but not dorsomedial, striatum impair acquisition of a simple discrimination task. Behav Brain Res *150*, 15-23.

Featherstone, R. E., and McDonald, R. J. (2004b). Dorsal striatum and stimulus-response learning: lesions of the dorsolateral, but not dorsomedial, striatum impair acquisition of a stimulus-response-based instrumental discrimination task, while sparing conditioned place preference learning. Neuroscience *124*, 23-31.

Featherstone, R. E., and McDonald, R. J. (2005a). Lesions of the dorsolateral or dorsomedial striatum impair performance of a previously acquired simple discrimination task. Neurobiol Learn Mem *84*, 159-167.

Featherstone, R. E., and McDonald, R. J. (2005b). Lesions of the dorsolateral striatum impair the acquisition of a simplified stimulus-response dependent conditional discrimination task. Neuroscience *136*, 387-395.

Fino, E., Glowinski, J., and Venance, L. (2005). Bidirectional activity-dependent plasticity at corticostriatal synapses. J Neurosci *25*, 11279-11287.

Fino, E., Paille, V., Deniau, J. M., and Venance, L. (2009). Asymmetric spike-timing dependent plasticity of striatal nitric oxide-synthase interneurons. Neuroscience *160*, 744-754.

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

Flaherty, A. W., and Graybiel, A. M. (1993). Output architecture of the primate putamen. J Neurosci *13*, 3222-3237.

Flaherty, A. W., and Graybiel, A. M. (1994). Input-output organization of the sensorimotor striatum in the squirrel monkey. J Neurosci *14*, 599-610.

Flores-Hernandez, J., Cepeda, C., Hernandez-Echeagaray, E., Calvert, C. R., Jokel, E. S., Fienberg, A. A., Greengard, P., and Levine, M. S. (2002). Dopamine enhancement of NMDA currents in dissociated medium-sized striatal neurons: role of D1 receptors and DARPP-32. J Neurophysiol *88*, 3010-3020.

Francois, C., Grabli, D., McCairn, K., Jan, C., Karachi, C., Hirsch, E. C., Feger, J., and Tremblay, L. (2004). Behavioural disorders induced by external globus pallidus dysfunction in primates II. Anatomical study. Brain *127*, 2055-2070.

Gan, J. O., Walton, M. E., and Phillips, P. E. (2010). Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. Nat Neurosci *13*, 25-27.

Garthwaite, J. (2008). Concepts of neural nitric oxide-mediated transmission. Eur J Neurosci *27*, 2783-2802.

Gerdeman, G. L., Ronesi, J., and Lovinger, D. M. (2002). Postsynaptic endocannabinoid release is critical to long-term depression in the striatum. Nat Neurosci *5*, 446-451.

Gerfen, C. R. (1985). The neostriatal mosaic. I. Compartmental organization of projections from the striatum to the substantia nigra in the rat. J Comp Neurol *236*, 454-476.

Gerfen, C. R. (1992). The neostriatal mosaic: multiple levels of compartmental organization. Trends Neurosci *15*, 133-139.

Gerfen, C. R., Baimbridge, K. G., and Miller, J. J. (1985). The neostriatal mosaic: compartmental distribution of calcium-binding protein and parvalbumin in the basal ganglia of the rat and monkey. Proc Natl Acad Sci U S A *82*, 8780-8784.

Gerfen, C. R., Baimbridge, K. G., and Thibault, J. (1987a). The neostriatal mosaic: III. Biochemical and developmental dissociation of patch-matrix mesostriatal systems. J Neurosci *7*, 3935-3944.

Gerfen, C. R., Herkenham, M., and Thibault, J. (1987b). The neostriatal mosaic: II. Patch- and matrix-directed mesostriatal dopaminergic and non-dopaminergic systems. J Neurosci *7*, 3915-3934.

Gerfen, C. R., and Keefe, K. A. (1994). Neostriatal dopamine receptors. Trends Neurosci *17*, 2-3; author reply 4-5.

Gimenez-Amaya, J. M., and Graybiel, A. M. (1991). Modular organization of projection neurons in the matrix compartment of the primate striatum. J Neurosci *11*, 779-791.

Goldberg, J. A., Rokni, U., Boraud, T., Vaadia, E., and Bergman, H. (2004). Spike synchronization in the cortex/basal-ganglia networks of Parkinsonian primates reflects global dynamics of the local field potentials. J Neurosci *24*, 6003-6010.

Goto, Y., and O'Donnell, P. (2001). Network synchrony in the nucleus accumbens in vivo. J Neurosci *21*, 4498-4504.

Grabli, D., McCairn, K., Hirsch, E. C., Agid, Y., Feger, J., Francois, C., and Tremblay, L. (2004). Behavioural disorders induced by external globus pallidus dysfunction in primates: I. Behavioural study. Brain *127*, 2039-2054.

Graveland, G. A., and DiFiglia, M. (1985). The frequency and distribution of medium-sized neurons with indented nuclei in the primate and rodent neostriatum. Brain Res *327*, 307-311.

Graybiel, A. M. (1984). Correspondence between the dopamine islands and striosomes of the mammalian striatum. Neuroscience *13*, 1157-1187.

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. Neurobiol Learn Mem *70*, 119-136.
Graybiel, A. M., and Ragsdale, C. W., Jr. (1978). Histochemically distinct compartments in the striatum of human, monkeys, and cat demonstrated by acetylthiocholinesterase staining. Proc Natl Acad Sci U S A *75*, 5723-5726.
Grillner, S., Hellgren, J., Menard, A., Saitoh, K., and Wikstrom, M. A. (2005). Mechanisms for selection of basic motor programs--roles for the striatum and pallidum. Trends Neurosci *28*, 364-370.
Haber, S. N., and Calzavara, R. (2009). The cortico-basal ganglia integrative network: the role of the thalamus. Brain Res Bull *78*, 69-74.
Haber, S. N., and Fudge, J. L. (1997). The primate substantia nigra and VTA: integrative circuitry and function. Crit Rev Neurobiol *11*, 323-342.
Hazrati, L. N., and Parent, A. (1992a). Differential patterns of arborization of striatal and subthalamic fibers in the two pallidal segments in primates. Brain Res *598*, 311-315.
Hazrati, L. N., and Parent, A. (1992b). The striatopallidal projection displays a high degree of anatomical specificity in the primate. Brain Res *592*, 213-227.
Herkenham, M., and Pert, C. B. (1981). Mosaic distribution of opiate receptors, parafascicular projections and acetylcholinesterase in rat striatum. Nature *291*, 415-418.
Hikosaka, O. (2007). Basal ganglia mechanisms of reward-oriented eye movement. Ann N Y Acad Sci *1104*, 229-249.
Histed, M. H., Pasupathy, A., and Miller, E. K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. Neuron *63*, 244-253.
Holt, D. J., Graybiel, A. M., and Saper, C. B. (1997). Neurochemical architecture of the human striatum. J Comp Neurol *384*, 1-25.
Jiang, Z. G., and North, R. A. (1991). Membrane properties and synaptic responses of rat striatal neurones in vitro. J Physiol *443*, 533-553.
Jimenez-Castellanos, J., and Graybiel, A. M. (1987). Subdivisions of the dopamine-containing A8-A9-A10 complex identified by their differential mesostriatal innervation of striosomes and extrastriosomal matrix. Neuroscience *23*, 223-242.
Joel, D., and Weiner, I. (1994). The organization of the basal ganglia-thalamocortical circuits: open interconnected rather than closed segregated. Neuroscience *63*, 363-379.
Joel, D., and 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*, 451-474.
Jog, M. S., Kubota, Y., Connolly, C. I., Hillegaart, V., and Graybiel, A. M. (1999). Building neural representations of habits. Science *286*, 1745-1749.
Joshua, M., Adler, A., Mitelman, R., Vaadia, E., and Bergman, H. (2008). Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. J Neurosci *28*, 11673-11684.
Kato, K., and Zorumski, C. F. (1993). Nitric oxide inhibitors facilitate the induction of hippocampal long-term potentiation by modulating NMDA responses. J Neurophysiol *70*, 1260-1263.
Kato, M., and Kimura, M. (1992). Effects of reversible blockade of basal ganglia on a voluntary arm movement. J Neurophysiol *68*, 1516-1534.
Kawagoe, R., Takikawa, Y., and Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. Nat Neurosci *1*, 411-416.
Kawaguchi, Y. (1992). Large aspiny cells in the matrix of the rat neostriatum in vitro: physiological identification, relation to the compartments and excitatory postsynaptic currents. J Neurophysiol *67*, 1669-1682.
Kawaguchi, Y. (1993). Physiological, morphological, and histochemical characterization of three classes of interneurons in rat neostriatum. J Neurosci *13*, 4908-4923.
Kawaguchi, Y., Wilson, C. J., and Emson, P. C. (1989). Intracellular recording of identified neostriatal patch and matrix spiny cells in a slice preparation preserving cortical inputs. J Neurophysiol *62*, 1052-1068.
Kawaguchi, Y., Wilson, C. J., and Emson, P. C. (1990). Projection subtypes of rat neostriatal matrix cells revealed by intracellular injection of biocytin. J Neurosci *10*, 3421-3438.
Kemp, J. M., and Powell, T. P. (1971). The termination of fibres from the cerebral cortex and thalamus upon dendritic spines in the caudate nucleus: a study with the Golgi method. Philos Trans R Soc Lond B Biol Sci *262*, 429-439.
Kermadi, I., and Joseph, J. P. (1995). Activity in the caudate nucleus of monkey during spatial sequencing. J Neurophysiol *74*, 911-933.

Kerr, J. N., and Wickens, J. R. (2001). Dopamine D-1/D-5 receptor activation is required for long-term potentiation in the rat neostriatum in vitro. J Neurophysiol *85*, 117-124.
Kesner, R. P., and Gilbert, P. E. (2006). The role of the medial caudate nucleus, but not the hippocampus, in a matching-to sample task for a motor response. Eur J Neurosci *23*, 1888-1894.
Kimchi, E. Y., and Laubach, M. (2009a). The dorsomedial striatum reflects response bias during learning. J Neurosci *29*, 14891-14902.
Kimchi, E. Y., and Laubach, M. (2009b). Dynamic encoding of action selection by the medial striatum. J Neurosci *29*, 3148-3159.
Kimchi, E. Y., Torregrossa, M. M., Taylor, J. R., and Laubach, M. (2009). Neuronal correlates of instrumental learning in the dorsal striatum. J Neurophysiol *102*, 475-489.
Kita, H. (1993). GABAergic circuits of the striatum. Prog Brain Res *99*, 51-72.
Kita, H., Kosaka, T., and Heizmann, C. W. (1990). Parvalbumin-immunoreactive neurons in the rat neostriatum: a light and electron microscopic study. Brain Res *536*, 1-15.
Koos, T., and Tepper, J. M. (1999). Inhibitory control of neostriatal projection neurons by GABAergic interneurons. Nat Neurosci *2*, 467-472.
Koos, T., and Tepper, J. M. (2002). Dual cholinergic control of fast-spiking interneurons in the neostriatum. J Neurosci *22*, 529-535.
Krebs, M. O., Desce, J. M., Kemel, M. L., Gauchy, C., Godeheu, G., Cheramy, A., and Glowinski, J. (1991). Glutamatergic control of dopamine release in the rat striatum: evidence for presynaptic N-methyl-D-aspartate receptors on dopaminergic nerve terminals. J Neurochem *56*, 81-85.
Kreitzer, A. C. (2009). Physiology and pharmacology of striatal neurons. Annu Rev Neurosci *32*, 127-147.
Kubota, Y., Liu, J., Hu, D., DeCoteau, W. E., Eden, U. T., Smith, A. C., and Graybiel, A. M. (2009). Stable encoding of task structure coexists with flexible coding of task events in sensorimotor striatum. J Neurophysiol *102*, 2142-2160.
Kubota, Y., Mikawa, S., and Kawaguchi, Y. (1993). Neostriatal GABAergic interneurones contain NOS, calretinin or parvalbumin. Neuroreport *5*, 205-208.
Lapper, S. R., and Bolam, J. P. (1992). Input from the frontal cortex and the parafascicular nucleus to cholinergic interneurons in the dorsal striatum of the rat. Neuroscience *51*, 533-545.
Lau, B., and Glimcher, P. W. (2007). Action and outcome encoding in the primate caudate nucleus. J Neurosci *27*, 14502-14514.
Lau, B., and Glimcher, P. W. (2008). Value representations in the primate striatum during matching behavior. Neuron *58*, 451-463.
Levesque, M., Bedard, A., Cossette, M., and Parent, A. (2003). Novel aspects of the chemical anatomy of the striatum and its efferents projections. J Chem Neuroanat *26*, 271-281.
Levesque, M., and Parent, A. (1998). Axonal arborization of corticostriatal and corticothalamic fibers arising from prelimbic cortex in the rat. Cereb Cortex *8*, 602-613.
Levesque, M., and Parent, A. (2005). The striatofugal fiber system in primates: a reevaluation of its organization based on single-axon tracing studies. Proc Natl Acad Sci U S A *102*, 11888-11893.
Lodge, D. J., and Grace, A. A. (2006). The laterodorsal tegmentum is essential for burst firing of ventral tegmental area dopamine neurons. Proc Natl Acad Sci U S A *103*, 5167-5172.
Lokwan, S. J., Overton, P. G., Berry, M. S., and Clark, D. (1999). Stimulation of the pedunculopontine tegmental nucleus in the rat produces burst firing in A9 dopaminergic neurons. Neuroscience *92*, 245-254.
Magill, P. J., Bolam, J. P., and Bevan, M. D. (2000). Relationship of activity in the subthalamic nucleus-globus pallidus network to cortical electroencephalogram. J Neurosci *20*, 820-833.
Magill, P. J., Bolam, J. P., and Bevan, M. D. (2001). Dopamine regulates the impact of the cerebral cortex on the subthalamic nucleus-globus pallidus network. Neuroscience *106*, 313-330.
Mahon, S., Vautrelle, N., Pezard, L., Slaght, S. J., Deniau, J. M., Chouvet, G., and Charpier, S. (2006). Distinct patterns of striatal medium spiny neuron activity during the natural sleep-wake cycle. J Neurosci *26*, 12587-12595.
Mallet, N., Le Moine, C., Charpier, S., and Gonon, F. (2005). Feedforward inhibition of projection neurons by fast-spiking GABA interneurons in the rat striatum in vivo. J Neurosci *25*, 3857-3869.
Mansouri, F. A., Tanaka, K., and Buckley, M. J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. Nat Rev Neurosci *10*, 141-152.
Masimore, B., Schmitzer-Torbert, N. C., Kakalios, J., and Redish, A. D. (2005). Transient striatal gamma local field potentials signal movement initiation in rats. Neuroreport *16*, 2021-2024.

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

Matsumoto, M., Matsumoto, K., Abe, H., and Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci *10*, 647-656.

Matsumoto, N., Hanakawa, T., Maki, S., Graybiel, A. M., and Kimura, M. (1999). Role of [corrected] nigrostriatal dopamine system in learning to perform sequential motor tasks in a predictive manner. J Neurophysiol *82*, 978-998.

Middleton, F. A., and Strick, P. L. (2000). Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Res Brain Res Rev *31*, 236-250.

Middleton, F. A., and Strick, P. L. (2002). Basal-ganglia 'projections' to the prefrontal cortex of the primate. Cereb Cortex *12*, 926-935.

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

Mink, J. W. (2003). The Basal Ganglia and involuntary movements: impaired inhibition of competing motor patterns. Arch Neurol *60*, 1365-1368.

Miura, M., Masuda, M., and Aosaki, T. (2008). Roles of micro-opioid receptors in GABAergic synaptic transmission in the striosome and matrix compartments of the striatum. Mol Neurobiol *37*, 104-115.

Miyachi, S., Hikosaka, O., and Lu, X. (2002). Differential activation of monkey striatal neurons in the early and late stages of procedural learning. Exp Brain Res *146*, 122-126.

Miyachi, S., Hikosaka, O., Miyashita, K., Karadi, Z., and Rand, M. K. (1997). Differential roles of monkey striatum in learning of sequential hand movement. Exp Brain Res *115*, 1-5.

Mizumori, S. J., Yeshenko, O., Gill, K. M., and Davis, D. M. (2004). Parallel processing across neural systems: implications for a multiple memory system hypothesis. Neurobiol Learn Mem *82*, 278-298.

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

Murray, E. A., O'Doherty, J. P., and Schoenbaum, G. (2007). What we know and do not know about the functions of the orbitofrontal cortex after 20 years of cross-species studies. J Neurosci *27*, 8166-8169.

Nakamura, K., and Hikosaka, O. (2006). Role of dopamine in the primate caudate nucleus in reward modulation of saccades. J Neurosci *26*, 5360-5369.

Nambu, A., Tokuno, H., and Takada, M. (2002). Functional significance of the cortico-subthalamo-pallidal 'hyperdirect' pathway. Neurosci Res *43*, 111-117.

Nieoullon, A., Cheramy, A., and Glowinski, J. (1978). Release of dopamine evoked by electrical stimulation of the motor and visual areas of the cerebral cortex in both caudate nuclei and in the substantia nigra in the cat. Brain Res *145*, 69-83.

Parent, A., and Hazrati, L. N. (1993). Anatomical aspects of information processing in primate basal ganglia. Trends Neurosci *16*, 111-116.

Partridge, J. G., Tang, K. C., and Lovinger, D. M. (2000). Regional and postnatal heterogeneity of activity-dependent long-term changes in synaptic efficacy in the dorsal striatum. J Neurophysiol *84*, 1422-1429.

Pasquereau, B., Nadjar, A., Arkadir, D., Bezard, E., Goillandeau, M., Bioulac, B., Gross, C. E., and Boraud, T. (2007). Shaping of motor responses by incentive values through the basal ganglia. J Neurosci *27*, 1176-1183.

Pasupathy, A., and Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature *433*, 873-876.

Paus, T. (2001). Primate anterior cingulate cortex: where motor control, drive and cognition interface. Nat Rev Neurosci *2*, 417-424.

Pisa, M., and Cyr, J. (1990). Regionally selective roles of the rat's striatum in modality-specific discrimination learning and forelimb reaching. Behav Brain Res *37*, 281-292.

Prensa, L., Parent, A., and Gimenez-Amaya, J. M. (1999). [Compartmentalized organization of human corpus striatum]. Rev Neurol *28*, 512-519.

Ragozzino, M. E., and Choi, D. (2004). Dynamic changes in acetylcholine output in the medial striatum during place reversal learning. Learn Mem *11*, 70-77.

Ragsdale, C. W., Jr., and Graybiel, A. M. (1991). Compartmental organization of the thalamostriatal connection in the cat. J Comp Neurol *311*, 134-167.

Raz, A., Frechter-Mazar, V., Feingold, A., Abeles, M., Vaadia, E., and Bergman, H. (2001). Activity of pallidal and striatal tonically active neurons is correlated in mptp-treated monkeys but not in normal monkeys. J Neurosci *21*, RC128.

Rivera, A., Alberti, I., Martin, A. B., Narvaez, J. A., de la Calle, A., and Moratalla, R. (2002). Molecular phenotype of rat striatal neurons expressing the dopamine D5 receptor subtype. Eur J Neurosci *16*, 2049-2058.

Romo, R., Cheramy, A., Godeheu, G., and Glowinski, J. (1986). In vivo presynaptic control of dopamine release in the cat caudate nucleus--III. Further evidence for the implication of corticostriatal glutamatergic neurons. Neuroscience *19*, 1091-1099.

Rushworth, M. F., and Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci *11*, 389-397.

Rushworth, M. F., Walton, M. E., Kennerley, S. W., and Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. Trends Cogn Sci *8*, 410-417.

Russchen, F. T., Bakst, I., Amaral, D. G., and Price, J. L. (1985). The amygdalostriatal projections in the monkey. An anterograde tracing study. Brain Res *329*, 241-257.

Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. Science *310*, 1337-1340.

Schiffmann, S. N., Jacobs, O., and Vanderhaeghen, J. J. (1991). Striatal restricted adenosine A2 receptor (RDC8) is expressed by enkephalin but not by substance P neurons: an in situ hybridization histochemistry study. J Neurochem *57*, 1062-1067.

Schmitzer-Torbert, N., and Redish, A. D. (2004). Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple T task. J Neurophysiol *91*, 2259-2272.

Schmitzer-Torbert, N. C., and Redish, A. D. (2008). Task-dependent encoding of space and events by striatal neurons is dependent on neural subtype. Neuroscience *153*, 349-360.

Schoenbaum, G., Roesch, M. R., Stalnaker, T. A., and Takahashi, Y. K. (2009). A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nat Rev Neurosci *10*, 885-892.

Schultz, W. (2007a). Behavioral dopamine signals. Trends Neurosci *30*, 203-210.

Schultz, W. (2007b). Multiple dopamine functions at different time courses. Annu Rev Neurosci *30*, 259-288.

Schwartz, R. D., Lehmann, J., and Kellar, K. J. (1984). Presynaptic nicotinic cholinergic receptors labeled by [3H]acetylcholine on catecholamine and serotonin axons in brain. J Neurochem *42*, 1495-1498.

Seamans, J. K., Lapish, C. C., and Durstewitz, D. (2008). Comparing the prefrontal cortex of rats and primates: insights from electrophysiology. Neurotox Res *14*, 249-262.

Selemon, L. D., and Goldman-Rakic, P. S. (1985). Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. J Neurosci *5*, 776-794.

Sharott, A., Moll, C. K., Engler, G., Denker, M., Grun, S., and Engel, A. K. (2009). Different subtypes of striatal neurons are selectively modulated by cortical oscillations. J Neurosci *29*, 4571-4585.

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

Shen, W., Tian, X., Day, M., Ulrich, S., Tkatch, T., Nathanson, N. M., and Surmeier, D. J. (2007). Cholinergic modulation of Kir2 channels selectively elevates dendritic excitability in striatopallidal neurons. Nat Neurosci *10*, 1458-1466.

Shimo, Y., and Hikosaka, O. (2001). Role of tonically active neurons in primate caudate in reward-oriented saccadic eye movement. J Neurosci *21*, 7804-7814.

Shuen, J. A., Chen, M., Gloss, B., and Calakos, N. (2008). Drd1a-tdTomato BAC transgenic mice for simultaneous visualization of medium spiny neurons in the direct and indirect pathways of the basal ganglia. J Neurosci *28*, 2681-2685.

Smith, Y., Bennett, B. D., Bolam, J. P., Parent, A., and Sadikot, A. F. (1994). Synaptic relationships between dopaminergic afferents and cortical or thalamic input in the sensorimotor territory of the striatum in monkey. J Comp Neurol *344*, 1-19.

Smith, Y., and Parent, A. (1986). Neuropeptide Y-immunoreactive neurons in the striatum of cat and monkey: morphological characteristics, intrinsic organization and co-localization with somatostatin. Brain Res *372*, 241-252.

Surmeier, D. J., Ding, J., Day, M., Wang, Z., and Shen, W. (2007). D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. Trends Neurosci *30*, 228-235.

Surmeier, D. J., Reiner, A., Levine, M. S., and Ariano, M. A. (1993). Are neostriatal dopamine receptors co-localized? Trends Neurosci *16*, 299-305.

Suzuki, T., Miura, M., Nishimura, K., and Aosaki, T. (2001). Dopamine-dependent synaptic plasticity in the striatal cholinergic interneurons. J Neurosci *21*, 6492-6501.

Tang, C., Pawlak, A. P., Prokopenko, V., and West, M. O. (2007). Changes in activity of the striatum during formation of a motor habit. Eur J Neurosci *25*, 1212-1227.

Tang, C. C., Root, D. H., Duke, D. C., Zhu, Y., Teixeria, K., Ma, S., Barker, D. J., and West, M. O. (2009). Decreased firing of striatal neurons related to licking during acquisition and overtraining of a licking task. J Neurosci *29*, 13952-13961.

Tepper, J. M., Koos, T., and Wilson, C. J. (2004). GABAergic microcircuits in the neostriatum. Trends Neurosci *27*, 662-669.

Tepper, J. M., and Lee, C. R. (2007). GABAergic control of substantia nigra dopaminergic neurons. Prog Brain Res *160*, 189-208.

Tippett, L. J., Waldvogel, H. J., Thomas, S. J., Hogg, V. M., van Roon-Mom, W., Synek, B. J., Graybiel, A. M., and Faull, R. L. (2007). Striosomes and mood dysfunction in Huntington's disease. Brain *130*, 206-221.

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

Tong, Z. Y., Overton, P. G., and Clark, D. (1996). Stimulation of the prefrontal cortex in the rat induces patterns of activity in midbrain dopaminergic neurons which resemble natural burst events. Synapse *22*, 195-208.

Tort, A. B., Kramer, M. A., Thorn, C., Gibson, D. J., Kubota, Y., Graybiel, A. M., and Kopell, N. J. (2008). Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a T-maze task. Proc Natl Acad Sci U S A *105*, 20517-20522.

Tricomi, E., Balleine, B. W., and O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. Eur J Neurosci *29*, 2225-2232.

Tunstall, M. J., Oorschot, D. E., Kean, A., and Wickens, J. R. (2002). Inhibitory interactions between spiny projection neurons in the rat striatum. J Neurophysiol *88*, 1263-1269.

Uylings, H. B., Groenewegen, H. J., and Kolb, B. (2003). Do rats have a prefrontal cortex? Behav Brain Res *146*, 3-17.

van der Meer, M. A., and Redish, A. D. (2009). Low and High Gamma Oscillations in Rat Ventral Striatum have Distinct Relationships to Behavior, Reward, and Spiking Activity on a Learned Spatial Decision Task. Front Integr Neurosci *3*, 9.

Vincent, S. R., and Johansson, O. (1983). Striatal neurons containing both somatostatin- and avian pancreatic polypeptide (APP)-like immunoreactivities and NADPH-diaphorase activity: a light and electron microscopic study. J Comp Neurol *217*, 264-270.

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

Walker, R. H., Arbuthnott, G. W., Baughman, R. W., and Graybiel, A. M. (1993). Dendritic domains of medium spiny neurons in the primate striatum: relationships to striosomal borders. J Comp Neurol *337*, 614-628.

Walton, M. E., Rudebeck, P. H., Bannerman, D. M., and Rushworth, M. F. (2007). Calculating the cost of acting in frontal cortex. Ann N Y Acad Sci *1104*, 340-356.

Wang, Z., Kai, L., Day, M., Ronesi, J., Yin, H. H., Ding, J., Tkatch, T., Lovinger, D. M., and Surmeier, D. J. (2006). Dopaminergic control of corticostriatal long-term synaptic depression in medium spiny neurons is mediated by cholinergic interneurons. Neuron *50*, 443-452.

West, M. O., Carelli, R. M., Pomerantz, M., Cohen, S. M., Gardner, J. P., Chapin, J. K., and Woodward, D. J. (1990). A region in the dorsolateral striatum of the rat exhibiting single-unit correlations with specific locomotor limb movements. J Neurophysiol *64*, 1233-1246.

Whishaw, I. Q., Zeeb, F., Erickson, C., and McDonald, R. J. (2007). Neurotoxic lesions of the caudate-putamen on a reaching for food task in the rat: acute sensorimotor neglect and chronic qualitative motor impairment follow lateral lesions and improved success follows medial lesions. Neuroscience *146*, 86-97.

White, N. M. (2009). Some highlights of research on the effects of caudate nucleus lesions over the past 200 years. Behav Brain Res *199*, 3-23.

Wilson, C. J. (1993). The generation of natural firing patterns in neostriatal neurons. Prog Brain Res *99*, 277-297.

Wilson, C. J., Chang, H. T., and Kitai, S. T. (1990). Firing patterns and synaptic potentials of identified giant aspiny interneurons in the rat neostriatum. J Neurosci *10*, 508-519.

Wilson, C. J., and Goldberg, J. A. (2006). Origin of the slow afterhyperpolarization and slow rhythmic bursting in striatal cholinergic interneurons. J Neurophysiol *95*, 196-204.

Wu, Y., and Parent, A. (2000). Striatal interneurons expressing calretinin, parvalbumin or NADPH-diaphorase: a comparative study in the rat, monkey and human. Brain Res *863*, 182-191.

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

Yan, Z., Song, W. J., and Surmeier, J. (1997). D2 dopamine receptors reduce N-type Ca2+ currents in rat neostriatal cholinergic interneurons through a membrane-delimited, protein-kinase-C-insensitive pathway. J Neurophysiol *77*, 1003-1015.

Yan, Z., and Surmeier, D. J. (1997). D5 dopamine receptors enhance Zn2+-sensitive GABA(A) currents in striatal cholinergic interneurons through a PKA/PP1 cascade. Neuron *19*, 1115-1126.

Yeshenko, O., Guazzelli, A., and Mizumori, S. J. (2004). Context-dependent reorganization of spatial and movement representations by simultaneously recorded hippocampal and striatal neurons during performance of allocentric and egocentric tasks. Behav Neurosci *118*, 751-769.

Yin, H. H., and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. Nat Rev Neurosci *7*, 464-476.

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

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2005a). Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. Eur J Neurosci *22*, 505-512.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. Behav Brain Res *166*, 189-196.

Yin, H. H., Mulcare, S. P., Hilario, M. R., Clouse, E., Holloway, T., Davis, M. I., Hansson, A. C., Lovinger, D. M., and Costa, R. M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. Nat Neurosci *12*, 333-341.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., and Balleine, B. W. (2005b). The role of the dorsomedial striatum in instrumental conditioning. Eur J Neurosci *22*, 513-523.

Yung, K. K., Bolam, J. P., Smith, A. D., Hersch, S. M., Ciliax, B. J., and Levey, A. I. (1995). Immunocytochemical localization of D1 and D2 dopamine receptors in the basal ganglia of the rat: light and electron microscopy. Neuroscience *65*, 709-730.

Zhou, F. M., Wilson, C. J., and Dani, J. A. (2002). Cholinergic interneuron characteristics and nicotinic properties in the striatum. J Neurobiol *53*, 590-605.

Abbreviations

A2A	an adenosine receptor expressed in indirect pathway striatal neurons
ACC	anterior cingulate cortex
ACh	acetylcholine
AChE	acetylcholinesterase
AMPA	α-amino-3-hydroxyl-5-methyl-4-isoxazole-propionate (activates ionotropic AMPA receptors)
Ca++	calcium
ChAT	choline acetyltransferase
DA	dopamine
DLPFC	dorsolateral prefrontal cortex
EMG	electromyography (measures electrical activity of muscles)
FF	fast-firing (striatal interneuron)
fMRI	functional magnetic resonance imaging
GABA	gamma-aminobutyric acid (an inhibitory neurotransmitter)
GP	globus pallidus
GPe	globus pallidus pars externa, external segment of the globus pallidus
GPi	globus pallidus pars interna, internal segment of the globus pallidus
HD	Huntington's disease
K+	potassium
LFP	local field potential
LTD	long-term depression
LTP	long-term potentiation
LTS	low-threshold spiking (striatal interneuron)
MSN	medium spiny neuron
Na+	sodium
NMDA	N-methyl-D-aspartic acid (activates ionotropic NMDA receptors)
NOS	nitric oxide synthase
OFC	orbitofrontal cortex
PD	Parkinson's disease
PFC	prefrontal cortex
PIT	Pavlovian instrumental transfer
PV	parvalbumin
RPE	reward prediction error
RRN	retrorubral nucleus
SNc	substantia nigra pars compacta
SNr	substantia nigra pars reticulata
STDP	spike-timing dependent plasticity
STN	subthalamic nucleus
TAN	tonically-active neuron (striatal interneuron)
VTA	ventral tegmental area

2. Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning

Catherine A. Thorn[1,2], Hisham Atallah[1], Mark Howe[1,3] and Ann M. Graybiel[1,3]

[1]McGovern Institute for Brain Research, [2]Department of Electrical Engineering and Computer Science, and [3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

This chapter consists of a manuscript and figures that were accepted for publication in *Neuron*, to be published June 10, 2010.

Acknowledgments
This work was funded by NIH MH60379, ONR N000140410208, the Stanley H. and Sheila G. Sydney Fund, European Union grant 201716 and a fellowship (CT) from the McGovern Institute for Brain Research. We thank Patricia Harlan, Christine Keller-McGandy, Henry Hall, Gila Fakterman and Kyle Smith for their help.

Summary

The basal ganglia are implicated in a remarkable range of functions influencing emotion and cognition as well as motor behavior. Current models of basal ganglia function hypothesize that parallel limbic, associative and motor cortico-basal ganglia loops contribute to this diverse set of functions, but little is yet known about how these loops operate and how their activities evolve during learning. To address these issues, we recorded simultaneously in sensorimotor and associative regions of the striatum as rats learned different versions of a conditional T-maze task. We found highly contrasting patterns of activity in these regions during task performance and found that these different patterns of structured activity developed concurrently, but with sharply different dynamics. Based on the region-specific dynamics of these patterns across learning, we suggest a working model whereby dorsomedial associative loops can modulate the access of dorsolateral sensorimotor loops to the control of action.

2.1. Introduction

The basal ganglia, long known to be critical for normal motor control, are now also recognized as influencing cognitive and motivational aspects of behavior (Balleine et al., 2009; Dagher and Robbins, 2009; Graybiel, 2008). Moreover, the striatum, the largest structure in the basal ganglia, is thought to be critical for learning functions across these domains, especially reinforcement-based learning (Daw et al., 2005; Samejima and Doya, 2007). Reflecting this wide functional scope, basal ganglia dysfunction has been identified in disorders ranging from Parkinson's disease and Huntington's disease to neuropsychiatric disorders including obsessive-compulsive disorder, Tourette syndrome, and major psychosis (DeLong and Wichmann, 2007; Graybiel and Mink, 2009).

Candidates for functionally distinct motor and cognitive circuits have been identified in behavioral experiments in humans and animals (Graybiel, 2008; Middleton and Strick, 2000; Worbe et al., 2009). In rodents, sensorimotor loops connect somatosensory and motor cortical areas with the dorsolateral striatum, and lesions of these loops, including lesions centered in the dorsolateral striatum, impair the acquisition and performance of motor sequences and stimulus-response (S-R) tasks, as well as the habitual responding in instrumental tasks that follows earlier goal-directed performance (Balleine et al., 2009; White, 2009). Correspondingly, in some sensorimotor tasks, neurons in this dorsolateral region have been shown to fire in relation to motor behaviors, and this activity continues to be modulated late in training (Barnes et al., 2005; Kimchi et al., 2009; Kubota et al., 2009; Schmitzer-Torbert and Redish, 2004; Tang et al., 2007; Yin et al., 2009). It has been suggested that the dorsolateral striatum is important for the chunking of motor patterns as habits are formed and stamped in (Barnes et al., 2005; Graybiel, 2008).

By contrast, associative loops interconnect the medial prefrontal cortex with regions of the dorsomedial striatum. Lesions made within these loops, including lesions of the dorsomedial striatum, impair goal-directed responding in instrumental tasks (Yin and Knowlton, 2006) and impair reversal learning (Ragozzino, 2007). These lesions do not generally affect behavioral performance during learning of simple S-R tasks (Ragozzino, 2007; White, 2009), but may impair the learning and performance of more complicated paradigms (Adams et al., 2001; Corbit and Janak, 2007; Featherstone and McDonald, 2005; Kantak et al., 2001). Neurons in the dorsomedial striatum undergo changes in activity early during motor learning and their firing has been shown to change according to flexible stimulus-value assignments, as well as with response bias (Kimchi and Laubach, 2009a; Kimchi and Laubach, 2009b; Yin et al., 2009). Based on this evidence, it is thought that the associative cortical-basal ganglia loop, including the dorsomedial striatum, is involved in flexible goal-directed behavioral control.

How the parallel dorsolateral and dorsomedial striatum-based loops interact to produce habitual versus goal-directed behaviors is still unclear. Available evidence suggests that behavior often evolves during trial-and-error learning from being flexible and goal-directed to being habitual. As this transition occurs, neural control by dorsal striatal circuits is thought to shift from associative circuits that take account of the outcome contingencies of actions to those that are less flexible and that underpin habit formation and repetitive behaviors and thoughts (Graybiel, 2008; Yin et al., 2008). However, lesions of the dorsomedial striatum can result in the expression of habitual behavior even early in training, and lesions of the dorsolateral striatum can result in goal-directed responding even after extended training (Yin and Knowlton, 2006). These and related results suggest that the two control systems operate independently, and perhaps simultaneously or even competitively (Balleine et al., 2009; Wassum et al., 2009).

To determine the patterns of neural activity that occur in these dorsolateral and dorsomedial striatal districts during procedural learning in freely moving animals, we made simultaneous tetrode

recordings of single-unit activity in both the dorsolateral and dorsomedial parts of the striatum as rats acquired a T-maze task. The task was designed to require not only skilled motor performance, but also flexible responding based on sensory cues signaling the baited end-arm, thus taxing both sensorimotor and cognitive circuitry. Moreover, we trained the rats on two different task versions concurrently, with instruction cues of either auditory or tactile modalities, and we varied the difficulty of the tactile version in order further to differentiate changes in neural activity along sensory, motor, and cognitive domains. Finally, given evidence that a classical lithium chloride devaluation procedure shows that training on a similar T-maze task behavior is initially goal-directed and becomes habitual with over-training (Smith and Graybiel, *unpublished data*), we tracked neural activity chronically from the naïve state to the extensively over-trained state. In this way, we sought to identify activity that was associated with the early flexible action-outcome phase of behavioral control and activity that was related to repetitive late-stage habitual performance.

We focused on the activity patterns of neurons characterized as striatal projection neurons to ensure that the activities recorded would reflect those of the corresponding cortico-basal ganglia loops. Our findings demonstrate that the sensorimotor and associative cortico-basal ganglia loops are active simultaneously during learning, but that they develop strikingly different task-related patterns that are characterized by different dynamics across training sessions.

2.2. Results

We recorded from 6750 well-isolated striatal neurons in eight Long-Evans rats over 196 training sessions. All recordings were made concurrently in the dorsolateral and dorsomedial striatum (**Figure 2.1A**). We studied two groups of rats. The 5 rats in Group 1 acquired the auditory version of the task (> 72.5% correct performance for 10 consecutive training sessions) in 10-26 sessions (median = 13; **Figures 2.1B** and **2.S1**) but failed to acquire the tactile discrimination. Group 2 rats (n = 3) were trained using tactile cues with more readily discriminated textures, so that these animals could reach the performance criterion on both the auditory and tactile task versions. The Group 2 rats acquired the auditory discrimination in 9-22 sessions (median = 16) and the tactile discrimination in 18-28 sessions (median = 23; **Figures 2.1C** and **2.S1**). The combined values for both groups of rats are shown in **Figure 2.1D**. Running times decreased across training ($p < 0.001$, 2-way ANOVA), and mean running times during the tactile-cued trial blocks were slightly longer than those during the auditory-cued trial blocks (**Figure 2.1E**, $p < 0.001$, 2-way ANOVA).

Ninety percent (n = 6082) of recorded neurons were classified as putative medium spiny projection neurons (**Figures 2.1F** and **2.S2A-C**), and were accepted for further analysis if they fired more than 150 spikes in a session. Medium spiny neurons were further classified as "task-responsive" neurons (TRNs) if their firing rates during any peri-event window were greater than 2 standard deviations above their pretrial baseline firing rates for at least 3 consecutive 20-ms bins. The TRNs made up approximately two-thirds of the recorded projection neurons, and this proportion did not change with training (**Figure 2.1G**, lateral and medial: $p > 0.1$, Chi-square test). Tetrodes were not moved except as necessary at the beginning of each session to maintain high-quality single unit recordings. Thus, some neurons may have been recorded over multiple days. Employing the method of Emondi et al. (2004), we estimated that up to one third of our sample could be potential repeated units. Repeating the main analyses after removing these neurons did not qualitatively alter the results (**Figure 2.S3A** and **2.S3B**), and we therefore included all units for the analyses reported.

2.2.1. Simultaneously recorded dorsolateral and dorsomedial striatal ensemble activities differ during training on the T-maze tasks

We found that markedly different patterns of task-related ensemble activity in the dorsolateral and dorsomedial striatum emerged after the first stages of training. To gain a global picture of this population activity, we normalized firing rates for each neuron by calculating a z-score for each 20-ms bin of a ±300-ms peri-event time histogram constructed around each of 9 task events. For each stage, z-scores were averaged across all included units to calculate ensemble activity for the entire population (**Figure 2.2**).

During training, TRNs in the dorsolateral striatum (**Figure 2.2A**, top) developed strong ensemble responses at action boundaries of the task (locomotion onset, turn, and goal). Activity during mid-run was reduced after the first stages of training. In sharp contrast, ensemble TRN activity recorded in the dorsomedial striatum (**Figure 2.2A**, bottom) was strongest mid-run, especially around the time of instruction cue onset and turn start, and was weakest at task start and task end, almost opposite to the dorsolateral pattern. The dorsolateral and dorsomedial activities began to diverge early in training, and were strongly different especially during the middle training stages (**Figures 2.2B** and **2.2C** and **Table 2.1**). We further examined the ensemble activity of subsets of the dorsolateral and dorsomedial TRNs that responded to particular task events (**Figure 2.S2D**). These results highlight the

preferential firing of dorsolateral ensembles around the beginning and end of the trial, in contrast to the strong dorsomedial activity mid-task.

Despite the fact that only the Group 2 animals successfully learned the tactile as well as the auditory version of the T-maze task, the ensemble activity patterns for the two groups of animals were similar (**Figures 2.3A-B** and **Table 2.2**) as was their motor performance on the maze (**Figure 2.3C**). For both groups, TRN ensemble activity in the two regions did not differ substantially during the first training block (stages A1-A5), when neither group had reached the learning criterion for either task, but medial-lateral differences developed during the second training block (stages B1-B5) as the Group 2 animals, but not the Group 1 animals, acquired the tactile task (**Figure 2.3B** and **Table 2.2**). Laterally, the Group 2 rats had stronger goal responses, even in early sessions, than did the Group 1 rats, and the start activity of Group 2 rats accentuated the warning click rather than locomotion onset in the second training block. Medially, the Group 1 rats, which did not learn the tactile version, exhibited stronger pattern expression during the second training block than did the learners in Group 2. The ensemble activity patterns were otherwise comparable for the two groups. Ensemble patterns were also generally consistent across individual animals (**Figures 2.3D** and **2.S1**), despite differences in response selection on the tactile task (**Figure 2.S1**). Further, these patterns remained even after removing the animals in each group that exhibited the strongest patterned activity (**Figure 2.S3C-F**). Thus, the data from all rats were combined for subsequent analyses.

To quantify the strength of the dorsolateral and dorsomedial ensemble patterns over training, we calculated a spike probability distribution from the ensemble z-scores and estimated the entropy of this distribution as a measure of randomness in the population firing across trial-time for each training stage. In the dorsolateral striatum, ensemble activity became progressively more structured across training, as indicated by the reduced entropy in later training stages compared to that in stage A1 (**Figure 2.4A**). By contrast, the entropy of the dorsomedial activity was lowest during the middle training stages (block 2) and then returned to initial levels as training continued (**Figure 2.4D**). **Figure 2.4B** and **2.4E** show similarly contrasting trends in ensemble pattern development across training, expressed as changes in z-scores relative to the first training stage around each task event. Similar results for the two striatal regions were also obtained for calculations based on spike count distributions as opposed to z-score normalized firing patterns (**Figure 2.S4**). We found that a single linear regression provided the best fit to the dorsolateral entropy estimates, and that a segmented regression with a breakpoint at stage B1 best fit the dorsomedial entropy estimates. Using these optimal regressions, we next tested each 20-ms bin in each peri-event window for changes in the neural activity across training. **Figure 2.4C** shows that dorsolateral TRN activity prior to warning click and at goal reaching increased significantly across training stages, while activity around locomotion onset and out-of-start events declined with training. Dorsomedial TRN activity around cue onset and turn start increased during the first part of training, while activity around goal reaching declined, and during the later stages of training, these trends reversed (**Figure 2.4F-G**).

These findings suggested that task-related projection neurons in the dorsolateral and dorsomedial regions of the striatum, parts of different cortico-basal ganglia loops, develop different structured activities concurrently during the course of learning, and that the dynamics of the activity changes are different throughout learning.

2.2.2. Dorsolateral and dorsomedial ensembles preferentially respond to different stimulus modalities only around the time of cue onset

Surprisingly, despite the differences in percent correct performance on the auditory and tactile task-versions, ensemble neural activity during the auditory and tactile trials was similar in both dorsolateral and dorsomedial regions (**Figures 2.5A**, **2.5B** and **2.S5A**). We observed differences in

ensemble activity only around the time of instruction cue onset: dorsolateral ensembles showed higher activity in response to the presentation of the tactile cues, whereas dorsomedial ensembles preferentially responded to the onset of the auditory cues (**Figure 2.5C**). At the single unit level, modest numbers of TRNs differentiated between the two modalities: up to ca. 15% around the cue onset and turn start events (**Figure 2.5D**). In the dorsomedial striatum, these units tended to exhibit higher firing rates during auditory trials ($p < 0.001$, Chi-square). These percentages did not change with training in either region (**Figure 2.5E**, $p > 0.1$, Chi-square).

Fewer than 5% of the recorded neurons in either region changed their firing rates significantly in response to the instruction cue presentations (**Figure 2.5F**), and units discriminative for each stimulus in any peri-event window were also rare (**Figure 2.S5F-O**). Within this small stimulus-selective population, dorsomedial units favored the more salient 8 kHz tone and dorsolateral units favored the tactile stimuli (lateral and medial: $p < 0.001$, Chi-square test). Finally, we found only a few neurons with stimulus value-correlated firing that could not be accounted for by other parameters such as stimulus selectivity, modality selectivity, or turn-specific activity (**Figure 2.S5B-E**).

2.2.3. Dorsolateral and dorsomedial striatal neurons similarly encode turn response and trial outcome parameters

Given evidence that the dorsolateral striatum is critical for forming S-R associations and the dorsomedial striatum for forming associations related to reinforcement outcome, we tested for corresponding biases in neural activity in these two striatal districts. We compared the proportion of units in each region firing differentially in relation to either the different responses that the rats could select (right and left turns) or to the different reinforcement outcomes that could occur (reward or lack of reward). Unexpectedly, we found no large-scale differences between the dorsolateral and dorsomedial striatal districts in encoding either motor responses or trial outcomes.

Similar percentages of units in the two striatal regions (ca. 15-35%) differentiated between right and left turns during task events following turn onset (**Figure 2.6A**), and the mean number of spikes with which these units differentiated right from left turns were also similar across regions (**Figure 2.6B**). Importantly, the activities of these neurons were not predictive of turn direction prior to turn onset in either region. Dorsolaterally, but not dorsomedially, neuronal responses favored turns to the side contralateral to the implant (lateral: $p < 0.001$, medial: $p > 0.1$, Chi-square test). The percentage of turn-discriminative neurons did not change with training (**Figure 2.6C**, lateral and medial: $p > 0.1$, Chi-square).

We identified a few reward-sensitive neurons with differential firing restricted to the time around goal reaching, when the rat presumably could detect the presence or absence of reward in the food well (**Figure 2.6D**). The proportions of such units, though small in both regions, was larger dorsolaterally ($p = 0.003$, Chi-square), and did not change with training (**Figure 2.6E**, lateral and medial: $p > 0.1$, Chi-square). Nor did population activity differ between correct and incorrect trials (**Figure 2.S6A**).

Our peri-event analyses suggest that independent populations of neurons encode stimulus, response and reinforcement outcome parameters (**Figure 2.S6B**). Based on previous work (Histed et al., 2009; Kim et al., 2007; Kimchi and Laubach, 2009a), we searched for, but did not find (**Figure 2.S6C-E**), a significant number of units with differential activity dependent on the response executed in the previous trial (right or left turn) or the outcome of the previous trial (correct or incorrect). Additional analyses (**Figure 2.S6F-K**) also suggested that changes in response values or reward values (Kimchi and Laubach, 2009b; Samejima et al., 2005) were not a dominant factor in neuronal responding in our task (**Figure 2.S6**).

2.2.4. Reduced in-task activity characterizes subpopulations of projection neurons in both dorsolateral and dorsomedial striatum

Approximately one third of the medium spiny neurons recorded did not meet our criteria for classification as TRNs (**Figure 2.1F**). We called this population of units "non-task-responsive neurons" (NTRNs). The population of NTRNs exhibited markedly lower activity during the task than during the pre-trial baseline period. The reduced in-task firing was similar for the dorsolateral and dorsomedial NTRN ensembles (**Figures 2.7A-C**). The entropy of the NTRN ensemble activity declined slightly during the first stage of training and then fell sharply at the start of the last block of training, when, both medially and laterally, the pre-task activity was differentially enhanced compared to in-task activity (**Figure 2.7C**). Thus the NTRNs, though lacking phasic in-task activity similar to that of the TRNs, nevertheless had activity that was modulated by task context. We did not detect differences in the percentages of NTRNs medially and laterally, nor changes in these percentages across training (data not shown).

2.2.5. Dorsolateral and dorsomedial activity patterns are correlated with different behavioral parameters

To identify potential relationships between the activity patterns of the TRNs and the behavioral parameters measured as the animals were trained, we used the entropy of the ensemble activity patterns in the dorsolateral and dorsomedial regions as a measure of the strength of pattern expression during each training stage and then computed the correlation coefficients between this neural measure and the measures of behavioral performance. We found significant correlations between the strength of the dorsolateral striatal ensemble pattern and percent correct performance (calculated separately for auditory, tactile, and all trials) as well as significant correlations with running time (**Figure 2.8A**): the task-bracketing pattern of ensemble activity that appeared in the dorsolateral striatum became stronger as percent correct performance and running speeds improved over the course of training.

Strikingly, for the dorsomedial striatum, we found no significant correlations between pattern strength and any of these behavioral measures on either the auditory or tactile versions of the task (**Figure 2.8A**). These negative findings suggested that the strength of the dorsomedial activity pattern was not linearly related to any measured behavioral parameter. The findings did not, however, exclude either a non-linear association between them or a relationship of the neural activity to combinations of behavioral parameters. We tested for two of these.

First, prior studies have shown that spike activity in the associative striatum is highest during the period in training when behavioral performance is improving most rapidly, principally during the task-times in which feedback about performance is available (Williams and Eskandar, 2006). To test whether this effect could contribute to the modulation of spike activity that we found in the dorsomedial striatal data set, we fit a third order polynomial to the total percent correct performance per learning stage for all rats and calculated the derivative of this polynomial to find the slope of the learning curve for each stage. For the population as a whole, we found a significant correlation between the entropy of the dorsomedial activity and the slope of the total percent correct learning curve (**Figure 2.S7A**). However, when Group 1 and Group 2 rats were analyzed separately, we found that only the Group 2 rats showed a strong correlation between the slope of the behavioral performance curve and entropy of the dorsomedial striatal activity. Group 1 rats failed to exhibit this correlation: the dorsomedial activity patterns in this group were most strongly expressed toward the end of training, when their behavioral performance had reached asymptote and was no longer changing (**Figure 2.S7**). These results suggest that neither a close correlation with percent correct or

motor performance, nor a close correlation with the rates of change in these parameters, accounted for the patterns of activity that we recorded during training in the dorsomedial striatum in the two groups of animals.

A second possibility was that the development of the patterned ensemble activity in the dorsomedial striatum might be more closely related to the difference in performance levels on the auditory and tactile task versions than to the overall performance improvement. We found that this was so: there was a strong correlation between the disparity in performance levels on the two task-versions and the entropy for the dorsomedial activity pattern, but no such correlation for the dorsolateral striatal activity pattern (**Figure 2.8A**). Remarkably, this finding held for both Group 1 and Group 2, considered separately (**Figure 2.S7**), suggesting that the performance disparity could be key to understanding the dynamics of the TRN ensemble patterns that emerged in the dorsomedial striatum through training. Repeating these correlational analyses for individual rats gave similar results (**Tables 2.S1** and **2.S2**, and **Figure 2.S7**).

The results for the NTRNs differed from those seen for the TRN ensembles. The changes in entropy of the NTRN ensemble activities were significantly correlated with improvements in both percent correct performance and running time across training (**Figure 2.8B**). This was true both dorsolaterally and dorsomedially indicating that, unlike the TRNs, the activities of NTRNs in dorsomedial and dorsolateral regions of the striatum were similarly correlated with behavioral performance.

2.3. Discussion

Our findings demonstrate that highly contrasting patterns of task-related ensemble activity emerge in the sensorimotor and the associative parts of the striatum as rats learn T-maze tasks instructed by auditory and tactile cues. The sensorimotor striatum developed ensemble spike activity that was heightened at the action boundaries of the task. The associative striatum developed heightened ensemble spike activity mainly during the middle of the task, when the animals chose between alternate actions based on instruction cues. These striatal activity patterns developed simultaneously across training. Remarkably, however, the dynamics of the learning-related changes in these two striatal regions were sharply different, and they were differently related to the behavior of the rats. In the sensorimotor striatum, the emerging ensemble activity pattern steadily increased as training progressed, and was clearly correlated with improving performance. In the associative striatum, the activity pattern first waxed and then waned as training progressed, and was not correlated with individual behavioral parameters but instead, with the difference in performance on the two versions of the T-maze task. Based on this conjoint reorganization of activity patterns in the sensorimotor and associative striatum during learning, and the differing dynamics of these activities across learning, we suggest that the simultaneous activity of these two striatal regions may be critical in determining the development and expression of habitual behavior.

2.3.1. Dorsolateral and dorsomedial striatal regions have different task-related patterns of activity

Our findings strongly support previous evidence for functional differences between the sensorimotor and associative striatum. As observed in previous studies with a single-modality version of the T-maze task used here (Barnes et al., 2005), we found that the phasic ensemble activity of dorsolateral striatal neurons was, after training, high at action boundaries, including around trial start and goal reaching; and we also found heightened activity at turn. The developing intensity of the dorsolateral pattern was strongly correlated with behavioral improvements in percent correct and decreases in running times across training. These results are consistent with the idea that the phasic ensemble activity in the dorsolateral striatum strengthens as performance on the task improves and behavior becomes highly stereotyped and, as related evidence suggests (Smith and Graybiel, *unpublished data*), highly habitual.

It was during the critical decision period of the task that phasic task-related activity increased in the dorsomedial striatum and ramped up until the decision was executed. The expression of this mid-task dorsomedial activity was most strongly correlated with the disparity in the performance accuracy of the rats on the auditory and tactile task-versions. This remarkable difference between the behavioral correlates of the neural activities in the dorsolateral and dorsomedial striatum suggests that the two regions, and their corresponding cortico-basal ganglia circuits, have distinct functions during the course of behavioral learning of the conditional T-maze task.

We examined several alternative possibilities to account for the striking experience-dependent modulation of the dorsomedial striatal activity across the different stages of training. A first possibility, favored here, is that the different plasticity demands that the animals faced in the successive training phases accounted for the heightened modulation of activity in the dorsomedial striatum during training. The dorsomedial mid-run activity gradually strengthened during the first training block, in which the rats were attempting to learn both task-versions, but it became intense during the second block when the auditory task-version had been acquired but the tactile version had

not. Then the dorsomedial pattern weakened in the third block as both task-versions were mastered. Thus, the dorsomedial ensemble activity pattern was strongest during the time when the acquisition demands on the animals were in conflict for the two task-versions. Moreover, the heightened activity during this conflict period was greater for the Group 1 animals, which never learned the more difficult tactile version. This changing pattern of activity in the dorsomedial striatum stood in contrast to the relative stability of the structured activity in the dorsolateral striatum: there, the patterned activity was relatively constant after the initial phase of the training.

These findings suggest that, at a population level, the strength of the activity patterns in the dorsomedial striatum rose and fell during the successive training blocks in relation to the training demands imposed by the task. During the second phase of training, when the auditory task had been acquired but the tactile task had not, differing plasticity demands were required for the two task versions. For the auditory version, further neuronal plasticity should only have consolidated the already-mastered S-R associations. By contrast, the animals still needed to acquire the S-R associations necessary to gain reward on the tactile version of the task. Thus, new learning in the tactile task was required for improving performance, but new learning on the auditory task (as opposed to continued consolidation) would have been detrimental to the already acquired auditory version. The heightened dorsomedial ensemble activity during this phase of acquisition suggests that the dorsomedial region may have been sensitive to these conflicting plasticity demands during the successive training blocks.

A second possibility, consistent with reinforcement learning models, is that response uncertainty due to a lack of adequate experience with a task could be related to an animal's willingness to make exploratory actions, and therefore to the rate at which learning occurs (Rushworth and Behrens, 2008). Our finding that the expression of structured activity in the dorsomedial striatum was correlated with the slope of the behavioral performance curve in some animals warrants further consideration of this idea. Assuming, in accord with the behavioral findings, that the S-R associations to the conditional cues were built up slowly through experience for each task-version, there must have been a period during acquisition when the direction of turn that would lead to reward was uncertain in each of the task-versions, and this time-period would have been different for the two tasks. At first glance, response uncertainty should have been highest early in training, when none of the four conditional cues had been mastered. However, some initial exposure to the task might have been required for mastering the task mechanics and determining that there were rules to be learned, and thus uncertainty-related activity might have developed slightly later in training. Even in this view, however, it is not clear why such activity should be highest during the second training block, when two of the four stimulus-response associations had been mastered. Nor should these activities be identical during auditory and tactile versions, as we found them to be, because again, one version was well learned while only the other version remained uncertain. Thus, we think it unlikely that this type of uncertainty can fully account for the patterns of activity we observed.

Notably, the enhanced dorsomedial striatal mid-run activity was present not only in the animals that failed to learn the difficult version of the tactile task, but was also present, though less strong, in the animals that acquired the easier tactile task. This result is important: it was not a failure to learn the tactile version that accounted for the heightened dorsomedial striatal activity.

We also considered the possibility that the heightened dorsomedial activity reflected differential engagement of this striatal region in switching behavior, needed every 20 trials as the auditory and tactile trial sets were interchanged. This view is in accord with evidence that the dorsal striatum is differentially active in relation to switches in stimulus modality or stimulus value (Kimchi and Laubach, 2009b; Kubota et al., 2009). However, population firing as well as the firing rates of the majority of single units were unaffected by cue modality, and the dorsomedial activity clearly rose during training and then fell as training progressed, despite the fact that the switching demands of the

task were similar across all sessions. The heightened dorsomedial activity that we observed mid-task and mid-training thus appears unlikely to reflect the within-session switches in the stimulus modality.

We did not have explicit ways of testing definitively for a relationship between the firing of the striatal units and the decision process itself, nor outcome expectancy from the action taken, as opposed to the right or left turn responses emitted. We did test whether the ensemble activity or the individual unit activities during the presumptive decision period predicted the direction or success of the upcoming turn. They did not. It thus seems likely that this activity, though occurring during the decision period, was not directly responsible for the action that the rats subsequently executed in a given trial, even if it was, as we suspect, related to the decision process. The dorsomedial activity is thus likely to be a global or state-level property not related to moment-to-moment conditions.

The proposal that conflicting behavioral and plasticity demands could have evoked the activity modulation in the dorsomedial striatum raises the possibility that the population activity reflected a global monitoring signal tracking the disparity between auditory and tactile task performance during training. This possibility accords well with what is known about the functions of the medial frontal and cingulate cortical areas that project to this striatal region. These neocortical regions have long been implicated in various types of performance monitoring, especially during tasks with ambiguous stimuli or conflicting response choices (Carter et al., 1998; Rushworth, 2008; Schall et al., 2002), or tasks with delayed and/or uncertain rewards (Cardinal, 2006; Rushworth, 2008). Firing rates of neurons in the dorsomedial striatum have been found to be related to response bias during performance of a go/no-go discrimination, suggesting that these responses might be heightened in conjunction with increased uncertainty (Kimchi and Laubach, 2009a). Combined with our findings, a pattern emerges of similar functional engagement throughout entire cortical-basal ganglia loop circuits interconnecting associative cortical regions and associative districts in the striatum.

2.3.2. Individual units in the dorsolateral and dorsomedial striatum similarly encode stimulus, response and outcome parameters

Behavioral evidence strongly favors the view that the dorsomedial striatum mediates outcome-sensitive behavior and the dorsolateral striatum mediates outcome-insensitive (habitual, S-R) behavior (Balleine et al., 2009; Graybiel, 2008). The simultaneous recordings that we made allowed us to look for unit activity that might be correlated with aspects of these two postulated control functions for learning, including neural activity discriminating the stimuli (tactile or auditory), the responses (left or right turns) and the reinforcement outcome (reward or no-reward). Surprisingly, despite the striking differences between the ensemble activity patterns in the two regions, we found only modest differences in the proportions of single dorsolateral and dorsomedial neurons that differentiated between cue modalities, turn directions and trial outcomes. In both regions a majority of neurons discriminated between right and left turn responses; a large minority of neurons responded differently to the two modalities; and only a very small proportion of neurons were sensitive to trial outcome.

We did observe preferential responding by dorsolateral ensembles to the onset of the tactile conditional cues, whereas single dorsomedial units and ensembles preferentially responded to the onset of the auditory cues. These results, and the preference for contralateral turns in the dorsolateral but not dorsomedial striatum, are consistent with the differential projections of somatosensory and motor cortex to more lateral regions of dorsal striatum and auditory cortex to more medial regions (McGeorge and Faull, 1989). For the few neurons responding to the presentation or lack of reward at goal-reaching, the outcome-sensitive sample was larger in the dorsolateral striatum than in the dorsomedial striatum.

Together, these results suggest that comparable subsets of neurons in dorsolateral and dorsomedial regions of the striatum encode stimulus, response, reinforcement outcome, context, and/or performance parameters. Consistent with other studies (Barnes et al., 2005; Berke et al., 2009; Kimchi and Laubach, 2009a; Kimchi and Laubach, 2009b), we found that neurons responsive to the instruction cues and trial outcomes were sparse for both task-versions as well as across learning. Moreover, the neurons that did discriminate between instruction cue modalities (stimulus), turn directions (response), and reward at trial end (outcome) were largely independent populations (Lau and Glimcher, 2007; Schmitzer-Torbert and Redish, 2004). The unexpected similarity in single unit selectivities in the dorsolateral and dorsomedial striatal regions, combined with the (at most) sparse encoding of combinations of these parameters, suggests that the currently accepted stimulus-response control functions of the dorsolateral striatum and response-outcome control functions of the dorsomedial striatum are not distinguished by the conjunctive representations of stimulus, response and reinforcement outcome by spike activity in the two striatal regions.

In a series of analyses, we found no clear evidence for the activity of more than a few neurons in either striatal region as being related to stimulus or outcome value. Interestingly, in two rats, we observed stronger discrimination among turn-discriminative populations of neurons as training progressed, providing some evidence that action-value encoding may be an important function of striatal neurons. In these rats, right-turn-related firing increased in the dorsolateral striatum, whereas left-turn-related firing increased in the dorsomedial striatum, hinting that the encoding of action-value contingencies might differ between the two regions. The lack of conclusive evidence for value encoding in our experiment is somewhat surprising given previous studies (Kimchi and Laubach, 2009b; Lau and Glimcher, 2008; Samejima et al., 2005). However, our experiments were not designed to study value, and our estimates of value rely heavily on the assumption that stimulus values and response values are correlated with the percent correct performance of the rats throughout training. They must therefore must be interpreted with some caution. We also failed to find single unit activity related to previous trial outcome or to the response executed in the previous trial (Histed et al., 2009; Kim et al., 2007; Kimchi and Laubach, 2009a). From a reinforcement learning perspective, the function of reward-contingent neuronal firing would be to update the value estimates associated with a chosen action or stimulus-action combination. The resulting synaptic plasticity changes may not necessarily result in immediate changes in firing on the subsequent trial.

2.3.3. Modes of neural firing in associative and sensorimotor striatum

Prior studies have compared dorsolateral and dorsomedial striatal activity during motor skill learning (Yin et al., 2009) and during performance of instrumental behavior (Kimchi et al., 2009). The specific patterns that we have found to emerge in the associative and sensorimotor zones suggest two main modes of activity in the corresponding cortico-basal ganglia loops. First, we found that the dorsolateral task-bracketing pattern of ensemble activity can emerge early during training, before either motor performance or percent correct performance reach asymptote. Such early plasticity accords with the findings of Kimchi et al., but contrasts with those of Yin et al., during learning of markedly different tasks. For the dorsomedial striatum, we found that early increases and then later decreases of mid-run activity emerged with training. Kimchi et al. observed early changes in dorsomedial striatal activity that were sustained or enhanced with training, whereas Yin et al. observed heightened activity only during the initial stages of learning. In agreement with the former study, our findings demonstrate that dorsomedial striatal activity can develop in conjunction with dorsolateral activity and remain active long after the initial stages of learning. In agreement with the latter study, we observed a decline in dorsomedial striatal activation once our task was well learned. However, the relationship of our findings to these previous reports is complex. In contrast to these

other studies, we used a task with a navigational component, our dorsomedial recording sites were anterior to those previously reported, and we trained the animals on two task-versions in single training sessions. Nevertheless, combined, these studies suggest that the acquisition of habitual behavior is characterized by the simultaneous operation of cortico-basal ganglia loops based in the dorsomedial and dorsolateral striatum, and that the modes of activation strongly depend on the demands of the task to be learned.

Interestingly, despite the view that dorsomedial striatal regions can mediate goal-directed or flexible responding early in training, few studies have yielded evidence for deficits in initial learning in rats with dorsomedial striatal lesions (Ragozzino, 2007; White, 2009). These previous results are consistent with the idea that multiple learning and memory systems interact in the expression of behavior, and suggest that performance deficits might not appear unless the task were to tax associative circuitry. Supporting this idea, one of the rare studies that did find learning deficits with dorsomedial striatal lesions suggested that the dorsomedial caudoputamen is essential for learning two responses to two similar arbitrary cues, in a paradigm with substantial similarities to the T-maze task used here (Adams et al., 2001). This result favors our suggestion that dorsomedial striatum – and its corresponding cortico-basal ganglia loops – could be important for performance monitoring, perhaps especially in disambiguating closely related contexts such that the correct action is chosen. This view is compatible both with the dorsomedial activity being related to the conflicting plasticity demands faced by the rats as they learned and the proposal that the conflict in task-version demands in itself produced the markedly heightened activity during the second phase of training in our experiment. In a number of other studies, it may be possible to interpret changes in behavioral performance following lesions of the dorsomedial striatum as being related, at least in part, to the inability to disambiguate closely related contexts (Corbit and Janak, 2007; Featherstone and McDonald, 2005; Kantak et al., 2001).

2.3.4. Both task-responsive and non-task-responsive neuronal subpopulations are modulated during learning

Both dorsolaterally and dorsomedially, a large population of putative projection neurons fired mainly during the baseline period rather than during the maze-runs themselves. We called these "non-task-responsive" neurons (NTRNs), recognizing nonetheless that the context specificity of these neurons and their modulation over the course of training suggests that they were in fact task-sensitive. We did not record after goal-reaching, during the time of reward consumption, due to noise artifact produced by chewing. It is possible that NTRNs (or the TRNs) responded at this time. Thus, we identified the NTRNs as those neurons lacking detectible phasic, in-task responses during the recording periods. These results confirm previous findings from our laboratory for the non-task-responsive neurons recorded in the dorsolateral striatum of rats and mice (Barnes et al., 2005; Kubota et al., 2009), as well as related findings by West and colleagues (Tang et al., 2007). Our findings further suggest that the distinction between neuronal populations with and without significant phasic activity during the task holds across at least two regions of the striatum.

Approximately half of the recorded neurons were classified as TRNs, and about a quarter of the neurons were medium spiny units classified as NTRNs. These estimates are approximate: neurons silent during the task would not have been counted unless we detected their activity during the baseline period. Using a less strict criterion for classifying task-responsiveness, the same as that used by Barnes et al. (Barnes et al., 2005), we found, as they did, that the phasic and quiet neuronal populations were nearly equal in size. With this classification, we also began to detect weak phasic activity in the population of neurons presumably without responses, and thus chose to report our results using the more conservative classifier.

Nevertheless, these results raise the possibility that the two classes of neurons might correspond, at least in part, to the direct and indirect pathway neurons of the striatum. Yin and colleagues reported evidence suggesting that in rats performing a rotarod motor learning task, the striatal neurons that undergo major changes during learning correspond to D2-class dopamine receptor-bearing indirect pathway neurons (Yin et al., 2009). We found large-scale changes in both the TRNs and the NTRNs, but we did find a greater quieting of the NTRNs in the dorsolateral striatum, which is enriched in D2-class dopamine receptors, than in the dorsomedial striatum, which expresses lower levels of D2-class receptors. Moreover, we found that for both dorsolateral and dorsomedial NTRN ensembles, the in-task decrease in activity was correlated with behavioral performance improvements, including increasing percent correct and decreasing running speeds. Selective targeting of neuronal subtypes during recording, now becoming feasible, will help to settle the identity of these two populations of striatal neurons.

2.3.5. Simultaneous activation of dorsolateral and dorsomedial striatum has implications for understanding cortico-basal ganglia loop functions

The central issue that we attempted to address in this study is how, during the course of habit learning, the neural activities in two key striatal regions change. Our results suggest that there are fundamental differences in the patterns of activity in associative and sensorimotor cortico-basal ganglia loops in the task-times of maximal ensemble activity during learning, in the dynamics of the activity changes across learning, and in the relation of the activity each region to the behavioral parameters that we were able to measure. We conclude that cortico-basal ganglia loops can operate simultaneously and with contrasting behavior-related dynamics during procedural learning.

The strikingly different dynamics of the acquired activity patterns in the two striatal regions are of special interest and raise a key question. Why, if the task-bracketing pattern appeared in the sensorimotor striatum early during training, and is a correlate of habitual performance (Barnes et al., 2005), did it not drive habitual behavior from its earliest time of appearance? As a working hypothesis, we propose that the differing dynamics of the activity patterns we observed in the dorsomedial and dorsolateral striatum hold a clue to the answer (**Figure 2.8C**). We suggest that even if the dorsolateral activity could have directed behavior from early in training, this dorsolateral activity was able to gain access to such executive capacity only after activity subsided in associative cortico-basal ganglia loops engaging the dorsomedial striatum (**Figure 2.8C**).

According to this model, exploration driven by frontostriatal associative circuits would be the default mode for behavior in a new learning environment. During the middle training blocks in the T-maze task, strong dorsolateral task-bracketing activity would have indicated that the neural bases for a habit existed, but equally strong or stronger dorsomedial activation would have prevented its expression. Finally, following mastery of all aspects of the task, the subsiding of dorsomedial activation would have enabled dorsolaterally-based habitual behavior to be expressed (**Figure 2.8C**). Though perhaps overly explicit, the core idea of this model is that there is a permissive role of the associative striatum in the evolution of behavior toward habitual performance. Such a permissive function would not require a direct transfer of information from the dorsomedial to the dorsolateral striatum. Rather, through their output connections, they could set up a competition at downstream targets (including regions of the neocortex or brainstem), enabling the disruption of habitual responses that would otherwise be driven by dorsolateral striatum-based loops.

This conceptualization, which considers the dynamics of simultaneously active sensorimotor and associative striatal circuits during training, has implications for many popular models of cortico-basal ganglia loop function (Daw et al., 2005; Graybiel, 2008; Horvitz, 2009; Samejima and Doya, 2007;

Yin et al., 2008). By extension, our suggestion that the dorsomedial striatum has a permissive function relative to the dorsolateral striatal circuits that release or inhibit action also has potential clinical implications. Our findings suggest the possibility that in dysfunctions such as seen in addiction, it is the lack of normal associative striatal cortico-basal ganglia circuit activation that contributes more to the pathology than the development of sensorimotor S-R associations per se, though this S-R activity may be most obvious in the addicted state (Graybiel, 2008; Kalivas, 2008; Robbins et al., 2008; Volkow et al., 2009). The classical idea that the prefrontal cortex can act as an inhibitory gate on motor cortex could thus be extended to the entire associative cortico-striatal loop circuitry. The flexibility of activity in the dorsomedial striatum, seen here in the waxing and waning of activity during the course of training, thus could be critical to the emergence of less flexible, habitual patterns of behavior.

2.4. Experimental procedures

Subjects and housing. Eight adult (300-350 g) male Long-Evans rats were housed in individual cages in a reverse light-cycle cubicle (lights on: 9 pm-9 am), and were handled and trained during their active cycle. All experimental procedures were approved by the Committee on Animal Care at the Massachusetts Institute of Technology. To accustom the rats to handling, the experimenter held the animals near their home cages for 15-20 min daily for one week prior to T-maze acclimation. During this week, a food restriction protocol was begun so that the rats maintained a weight greater than 90% of their free-feeding weight.

T-maze and acclimation. The T-maze consisted of 2 raised polycarbonate runways (height = 9 in.), joined in the shape of a T. Dimensions of the long arm were 3 x 48 inches, dimensions of the short arm were 3 x 29 inches. A drawbridge gate, which was raised and lowered manually, separated the starting block at the end of the long arm from the rest of the maze. Black polycarbonate walls (height = 16 in.), or wooden walls painted black, surrounded the maze at a distance of 4.5-8 inches. Photobeams were embedded ca. every 7 inches into the walls of the maze to control behavioral software and monitor the position of the rat in the maze. Position of the rat was additionally monitored during training by an overhead CCD camera tracking the location of an LED attached to the implanted headstage.

Rats were placed on the T-maze for 5-10 sessions prior to implant surgery to acclimate them to the maze and experimental room. During these initial sessions, chocolate-flavored sprinkles were placed throughout the maze, and rats were allowed to explore and eat freely during 20-30 min sessions. Once each rat was adequately moving around the maze, chocolate sprinkles were placed only in food wells at the ends of the goal arms. The rat received 1-2 sessions during which it could explore and retrieve chocolate from these baited goals. Finally, 1-3 sessions were given during which the rat was required to wait behind the start gate. The gate was opened, after which the rat was free to retrieve chocolate from either baited goal, but was required to return to the starting block again and wait behind the raised gate while the goal arms were rebaited. The rat received up to 10 such trials during these late acclimation sessions.

Implant surgery. Following T-maze acclimation, each rat was anesthetized with a ketamine/xylazine mixture (100 mg/kg ketamine + 10 mg/kg xylazine), and a headstage loaded with 11-12 tetrodes, 5-7 targeting the medial striatum (AP = 1.7 mm, ML = -1.8 mm) and 5-6 targeting the lateral striatum (AP = 0.5 mm, ML = -3.5 mm), was implanted and secured with dental cement and jeweler's screws. During the week following surgery, tetrodes were lowered to their target depths (3.5-4.5 mm, both medial and lateral sites).

Behavioral training. Following recovery from surgery and the lowering of tetrodes to their target depths, behavioral training began. During training, tetrodes were not moved except in small increments (<100 μm) as necessary to maintain high-quality recordings. Rats were trained on two versions (auditory and tactile) of the T-maze task in single daily sessions. The experiments were performed on two groups of rats.

The rats in the first group (n = 5) were required to sit in a starting block until a warning click was presented and the gate opened. They then were free to run on the maze toward the goal arms. When they broke a photobeam approximately halfway down the long arm, either an auditory cue or a tactile cue was presented that signaled which direction to turn to receive reward. The auditory cues were 1 and 8 kHz tones that remained on until the rat reached the end of a goal arm. The tactile cues

were black vinyl runner mat strips (McMaster-Carr, NJ), with rough texture on one side and a smooth texture on the other, that were placed on the maze so that they covered the long arm of the T-maze from the Cue On photobeam to its end. Cue modality was switched every 20 trials. Starting cue type was alternated daily so the rat received either a session of 20A-20T-20A-20T or 20T-20A-20T-20A. Sessions were ended after 80 trials or 3 hours. Acquisition training continued until the rat performed above 72.5% correct on the auditory version of the task. Overtraining for these animals continued until the rat performed above 72.5% correct on the auditory cues for 10 consecutive sessions, regardless of its performance in the trials with tactile cues.

These 5 rats were trained on a difficult tactile discrimination, so that they failed to learn the tactile cues but were able to learn, in the same daily sessions, the tone discrimination in the auditory version of the task. A second group of rats (n = 3) received a different set of tactile cues that were more easily discriminable. These cues were brittle plastic lighting covers painted black, with a rough texture on one side and a smooth texture on the other (Home Depot). So that the durations of the cue presentation could be nearly equal for the two modalities, the auditory cues were turned off by photobeam control when the animal reached the Turn End photobeam. To reduce the possibility that the rats could use odor cues to solve the tactile task, identical inserts were interchanged every 1-5 tactile trials. Training continued until the rats performed at or above 72.5% correct for 10 consecutive sessions on both tasks. This experimental design meant that we had two groups of rats: Group 1 rats that learned the auditory but not the tactile task, and Group 2 rats that learned both tasks.

Neural recordings. A Cheetah recording system (Neuralynx, MT) was used to record single unit and local field potentials (LFPs) from each tetrode throughout training. For single units, spikes were recognized as occurring after the voltage crossed a pre-set threshold on any one of the 4 tetrode channels. Single unit signals were amplified (1000-10000), filtered (600-6000 Hz), and sampled at 30 kHz or 32 kHz for approximately 1 ms (1.056 or 0.998 ms) around the time of threshold crossing. For LFP recording, the signal on a selected channel of each tetrode was split and was fed to an amplifier (gain: 1000, filtered: 1-475 Hz, sampling rate: 1 kHz, or 1.89 kHz).

Lesions and histology. Following overtraining, rats were anesthetized with 0.3 mL of 50 mg/mL sodium pentobarbital solution (ca. 40-50 mg/kg), and current was passed through each tetrode to make lesions marking the ends of the tetrode tracks (25 μA, 10 sec). Two to three days later, rats were deeply anaesthetized with a lethal dose (0.8-1.0 mL, or ca. 100-145 mg/kg) of sodium pentobarbital, and brains were fixed by transcardial perfusion with 4% paraformaldehyde in 0.1M $KNaPO_4$ buffer. Brains were post-fixed and cut transversely at 30 μm on a sliding microtome. Sections were processed for Nissl substance and examined microscopically to identify the lesions and tetrode tracks.

Data analysis

Spike Sorting & Unit Classification. Spikes were sorted into individual units with Plexon Offline Sorter (Plexon Inc., Dallas, TX) by manually defining contours around clusters. For three rats, cross-channel whitening (Emondi et al., 2004) was performed on the original recorded spike signals before sorting, and spikes were sorted manually on both the whitened data and the unprocessed data. Whitening procedures did not significantly affect the number or quality of units sorted and were thus not used for the remaining rats.

Units were classified as putative medium spiny neurons, fast-firing interneurons or tonically active interneurons according to procedures described in detail elsewhere (Barnes et al., 2005). Briefly, once spikes were sorted, neuron type was determined based on firing rate, autocorrelograms, interspike interval histograms and peri-event raster plots. Units were manually graded for quality and accepted for analysis based on examination of autocorrelograms and overlaid spike waveform traces.

A unit was further classified as a task-responsive neuron (TRN) if the firing rate in any ±300-ms peri-event window was more than 2 standard deviations above its baseline firing rate for 3 consecutive 20-ms bins. Units not classified as task-responsive were deemed "non-task-responsive" neurons (NTRNs).

Learning Stages. To compare data from multiple animals, data for both groups was staged according to percent correct performance. Stage A1 = first 1 or 2 consecutive sessions with >40 trials; Stage A2 = second 1 or 2 consecutive sessions with >40 trials, Stage A3-5 = subsequent 1 or 2 consecutive sessions prior to reaching criterion on the auditory cues, Stages B1-B5 = 1 or 2 consecutive sessions of >72.5% correct performance on auditory cues and <72.5% correct performance on tactile cues, Stages C1-C5 = 2 consecutive sessions of >72.5% correct on both auditory and tactile versions of the task.

Z-scores. For each unit, a normalized firing rate was calculated in the following manner. The total number of spikes across all trials in a session was calculated for each 20-ms bin in a ±300 ms peri-event window around each of 9 task events (baseline, warning click, gate, locomotion onset, out-of-start, cue onset, turn start, turn end, and goal reaching) and divided by the number of trials included. The mean, S_{mean}, and standard deviation, S_{std}, were calculated across all 261 bins (29 bins x 9 events). For each unit, the spike count in each bin was then normalized by the mean and standard deviation across all bins to obtain a Z-score for each bin: $Z_{bin} = (S_{bin} - S_{mean}) / S_{std}$. A similar calculation was performed for each unit using only auditory trials and only tactile trials to compare normalized firing rates between the two modalities. For each stage, the mean z-score and standard error of the mean (SEM) for each bin was calculated across all units included in each stage, and smoothed with a 3-point averaging filter, to obtain the population-averaged activity.

Comparing ensemble activations. To quantify the difference between dorsolateral and dorsomedial ensemble patterns, as well as the difference between the ensemble activity of Group 1 and Group 2 rats, three complimentary measures were used. First, for each 20-ms bin a t-test was performed to compare the mean z-scores of dorsolateral neurons to those of dorsomedial neurons (or Group 1 versus Group 2 neurons). Activation was considered significantly different for $p < 0.01$. The difference between the two patterns was expressed as the percentage of significantly differing bins: $100 * N_{sig} / N_{total}$, where $N_{total} = 261$. Next, we calculated the difference between the mean z-score in each bin, squared this difference and summed over all bins to obtain a residual sum of squares measure comparing dorsolateral and dorsomedial (or Group 1 and Group 2) ensemble activity: $RSS = \sum (Z_{bin,l} - Z_{bin,m})^2$. Finally, we computed the symmetrized Kullback-Leibler divergence between dorsolateral and dorsomedial activity by first calculating a spiking distribution for each region from the ensemble z-scores: $P_{bin} = (Z_{bin} + \alpha) / \sum(Z_{bin} + \alpha)$, where the constant $\alpha = 1$ was added to the ensemble z-scores such that values for all bins were greater than 0. Taking the dorsolateral-dorsomedial comparison as an example, K-L divergence was then computed as: $KL = \sum P_{bin,l} * ln(P_{bin,l} / P_{bin,m}) + \sum P_{bin,m} * ln(P_{bin,m} / P_{bin,l})$.

Entropy. To quantify the strength of the pattern expressed in each stage, the spiking distribution for each stage, s, was obtained from the ensemble z-scores as described above: $P_{bin}(s) = (Z_{bin}(s) + \alpha) / \sum(Z_{bin}(s) + \alpha)$, where $\alpha = 1$. Entropy for each stage was then determined for the resulting firing distribution: $H(s) = -\sum P_{bin}(s) * ln(P_{bin}(s))$. The mean entropy and 95% confidence intervals for each stage were determined using 1000 bootstrap samples from the neuronal populations in each stage.

Characterizing pattern development across training. We next determined whether a single regression across all training stages or a piecewise-continuous segmented regression with a single breakpoint

provided a better fit to the observed entropy trends across learning stages. First, a single linear regression was performed on the entropy across training stages. A segmented linear regression was deemed a better fit than the single linear regression if the following criteria were met: 1) the slopes of both segments were significantly different from 0 at $p < 0.05$, 2) the coefficient of determination for the segmented regression was much greater than the R^2 value for the single regression ($CD > 4 * R^2$). For dorsolateral entropy, no breakpoint was found that provided a piecewise fit that was better than the single regression. For the dorsomedial striatum, only one potential breakpoint met these criteria (stage B1).

After determining the type of regression to use, and the location of the breakpoint for the dorsomedial data, we analyzed the trends across training for each 20-ms bin across the peri-event task-time. For dorsolateral striatum, we performed a linear regression on the z-scores in each bin across the 15 stages of training to obtain the slope of the regression and the 95% confidence limits. For the dorsomedial striatum, we performed a segmented linear regression with a breakpoint at stage B1, obtaining the slope of the regressions and 95% confidence limits for each of the two segments (stages A1-B1 and stages B1-C5).

T-test Comparison to Identify Units with Differing Firing Rates on Trial Subsets. To determine whether a unit exhibited different response rates during auditory and tactile trials, the following method was used. For each unit, the number of spikes in a ±300-ms window around each of 9 task events was calculated for each auditory trial. Similarly, for each tactile trial, the number of spikes in a ±300-ms window around each event was computed. The mean spike counts for the two conditions (auditory trials vs. tactile trials) were then compared for each of the 9 events using a standard t-test assuming unequal variances (Matlab's ttest2 function with the 'unequal' option) and accepted as significantly different if $p < 0.01$. Similar calculations were performed to compare responses of each unit during right turn trials vs. left turn trials, and correct trials vs. incorrect trials, as well as to compare responses during trials following right vs. left turns and following correct vs. incorrect trial outcomes. A minimum of 10 trials was required in each condition before running the t-test, thus, late in training several units were excluded from the correct/incorrect discrimination testing as the number of incorrect trials became too small for meaningful analysis. We additionally determined the percentage of neurons expected to make each discrimination by chance. To determine this percentage, we ran the t-tests again for each unit and each discrimination, but prior to testing, trials in each session were randomly assigned to each comparison group (e.g., auditory or tactile) such that the sizes of the original groups were maintained.

The above analyses were performed to determine whether single neurons discriminated between auditory/tactile trials, right/left turn trials, and correct/incorrect trials. We additionally tested for sensitivity to the outcome (correct/incorrect) of the previous trial and the behavioral response (right/left turn) executed on the previous trial. For the latter analysis, we first removed neurons recorded during sessions in which an animal's response choices on the current trial were significantly ($p < 0.01$) correlated with those of the previous trial. Forty-eight of 174 sessions were removed (or 2335 units of 5977) before testing for single unit discriminations based on previous trial response.

Statistical Evaluation of Dorsolateral and Dorsomedial Discriminatory Populations: To determine whether the proportion of TRNs discriminating auditory trials from tactile trials or right from left turns differed between lateral and medial striatal regions, we calculated the total number of TRNs in each region that were discriminative during any task event during any stage of training (n_L and n_M for lateral and medial regions, respectively). The number of neurons that would be expected was then estimated as $(n_L + n_M) / (N_L + N_M)$, where N_L and N_M were the total number of TRNs recorded in all stages in lateral and medial regions, respectively. Pearson's Chi-square test was then used to compare the observed counts and the expected counts, and accepted as significant for $p < 0.01$. Because the

number of incorrect trials decreases dramatically during later stages of training, we used a number of alternate statistical methods to verify our discrimination results. First, we used bootstrapping to draw 50 trials in each of the correct and incorrect categories, and compared the population means using both t-test and ANOVA. Next, we determined the number of trials in the smaller of the two categories, and drew this number of trials from both correct and incorrect conditions, using a t-test to compare population means. Finally, we increased the minimum number of trials required for analysis to 20 trials in each category instead of 10 (thus excluding more units). For all of these tests, we found that the number of neurons discriminating during the goal-reaching task epoch was significantly larger than during other periods of the task, reaching 5-10% of our recorded TRN population in each region around goal reaching and 0-2% around events from baseline to turn end. We found that the trends across learning varied among statistical tests used for the pre-goal time periods, leading us to conclude that the statistical measures we used to determine outcome-sensitivity were potentially inaccurate during these time periods due to the small number of incorrect trials late in training and/or the small number of cells meeting the criteria for discrimination during these task epochs. This was confirmed by the trial-shuffling procedure described in the previous section, which determined that approximately 2% of neurons are found to preferentially respond during correct trials by chance, and this percentage increases with increasing percent correct performance. Thus for the correct vs. incorrect condition, the number of units differentiating around goal-reaching was used when further analysis was required (e.g., for trends across training, comparison of medial vs. lateral proportions), as this number was consistently greater than that expected by chance and did not vary significantly across test conditions.

Chi-square tests were used to determine whether the populations of discriminative TRNs in each striatal region exhibited a preference for one condition over the other. Using the auditory/tactile condition as an example, for each region, the number of TRNs with higher firing rates in auditory trials than in tactile trials was calculated, along with the number of TRNs with higher firing rates in tactile trials than in auditory trials ($n_{A,m}$ and $n_{T,m}$ for medial, $n_{A,l}$ and $n_{T,l}$ for lateral). The expected number of units was then computed as $(n_{A,m} + n_{T,m}) / 2$ for medial and $(n_{A,l} + n_{T,l}) / 2$ for lateral. Pearson's Chi-square was then used to compare the observed number of units in each region to the expected. The same procedure was used to compare the proportion of units discriminating right vs. left response conditions. For the correct vs. incorrect comparison, only the units discriminating around goal reaching were counted in the comparison.

To determine whether the populations of neurons responsive in relation to different discrimination conditions were independent, neurons that were identified as potentially repeated over multiple days were first removed from the sample (Emondi et al., 2004). For each of the 2-combination comparisons (modality + response, modality + outcome, and response + outcome), a 2 x 2 table was constructed where each cell contained the number of TRNs observed to make each of the discriminations during at least one task event (not necessarily both discriminations during a single event). The probabilities of making the individual discriminations were then calculated (e.g., $p_{Auditory,m} = (n_{Auditory,m}) / N_M$, $p_{Right,m} = n_{Right,m} / N_M$) and used to compute the expected number of observations for each cell (e.g., $n_{Aud+Right} = N_M * p_{Auditory,m} * p_{Right,m}$). When the observed and expected values for each cell were greater than 10 for all cells, a Pearson's Chi-square test of independence was used to compare observed and expected values. When the expected values were less than 10 for multiple cells, a Yates' Chi-square was used instead. A Bonferroni correction was used to adjust α for multiple comparisons, and the p values were accepted as significant for $p < 0.0033$.

Correlation of Ensemble Pattern Strength and Behavioral Measures. We computed the Pearson's linear correlation coefficients (Matlab's corr function) between entropy and a number of behavioral

measures to determine whether the strength of pattern expression in the dorsolateral and dorsomedial striatal regions was related to the overt behavior of the animals. For calculations based on population-averaged activity, we used entropy derived from the population z-scores, as described above. Behavioral measures used included percent correct on auditory trials only, percent correct on tactile trials only, and total percent correct. We also performed the correlational analysis on the difference in performance between auditory and tactile trials, as well as running times from cue onset to goal reaching. For the behavioral measures, each unit was assigned the percent correct, performance disparity (auditory percent correct – tactile percent correct), and mean running time values of the rat and session from which it was recorded. For each stage, the mean percent correct, performance disparity, and running time were taken over the population of recorded dorsolateral units and dorsomedial units separately.

For correlation analyses involving individual rats, lateral ensemble unit activity was calculated as mean spikes/unit in a ±300-ms window around goal reaching using all dorsolateral medium spiny units for each animal. All sessions with more than 35 total trials and at least 2 good medium spiny units recorded were included in the analysis. For dorsomedial ensembles, unit activity was calculated as mean spikes/unit in ±300-ms windows around cue onset and turn onset using all dorsomedial medium spiny units for each animal. Behavioral measures used in calculating correlations were the same as those listed above. For the correlational analyses involving individual animals, both ensemble activity and each of the behavioral measures were smoothed with a 3-point smoothing filter prior to computing the correlation coefficients.

To investigate the relationship of the slope of the behavioral performance curve to the entropy of the medial data set, a third-order polynomial was fit to the average staged percent correct performance for all rats, and the derivative of this polynomial was taken as an estimate of the slope of the learning curve for each stage. The entropy of the spike distribution of medial TRN ensembles was calculated as described above for each stage, and likewise fit with a third order polynomial. A linear correlation was performed between the slope of the learning curve and the medial entropy data. Similar correlations were additionally performed for Group 1 and Group 2 rats separately, and for each individual animal. For the individual animals, if the polynomial fit did not capture the main features of the entropy data, a higher-order polynomial that provided a better estimate was used instead. In one case (D16), a 3-point running average provided the best fit to the entropy data and was used in place of a polynomial fit.

Stimulus-value, response-value, and outcome-value analysis. To investigate whether neural activity might be related to changes in stimulus or response value across training, we looked for evidence that the populations of neurons sensitive to these parameters differentiated them more robustly as training progressed. Taking the auditory/tactile comparison as an example, for each unit, we calculated the number of spikes in a ±300-ms window around each task event for each trial (s_A and s_T, respectively). When the mean number of spikes was found to be significantly different between auditory and tactile conditions, we calculated the mean number of spikes differentiating the two conditions as $S_{ATdiff} = s_A - s_T$. For each stage, we then took the average of S_{ATdiff} over all modality-sensitive neurons in that stage to gain a global picture of whether these populations changed the magnitude of their sensitivity across training in either dorsolateral striatum or dorsomedial striatum. We also examined auditory-preferring units ($s_A - s_T > 0$) and all tactile-preferring units ($s_A - s_T < 0$) separately. Similar analysis was performed to determine whether neuronal populations differentiated right from left turns more robustly as training progressed.

To further address the question of stimulus-value or stimulus-response encoding (which cannot be completely dissociated in our task), we performed two additional analyses. First, we asked whether the percentage of stimulus-selective neurons increased with training. For each unit, a t-test was used to compare the spike counts in +/-300-ms windows around each task event during trials in

which a particular stimulus type was presented (e.g. 8 kHz tone) to all other trials. Stimulus selectivity was determined for each stimulus type for $p < 0.01$.

Next, we looked for medium spiny neurons that exhibited firing correlated with the current value of the stimulus. For each unit, the spike count in each 300-ms peri-event window was calculated for each trial, along with the value of the stimulus presented on that trial (estimated by the percent correct responses associated with that stimulus over the entire session). A linear correlation was then performed on the firing rates during each trial compared to the stimulus values for each trial. Using the percent correct performance associated with each stimulus over the entire session to estimate its value, it was impossible to dissociate changes in stimulus-response contingencies from stimulus-value contingencies, and it is likely that the increase in the percentage of units with firing correlated with the value of the presented stimulus is related to the increased propensity of the rats to respond with a particular turn direction in response to a particular instruction cue as training progressed. We thus additionally tested each unit for single stimulus selectivity, stimulus modality discrimination, or turn direction discrimination as previously described.

We additionally looked for evidence that changes in outcome value due to increasing satiety within a session might modulate the firing of medium spiny ensembles. The mean spike counts and SEM during each successive block of 20 trials was computed for TRN and NTRN neurons in each region. Comparisons across blocks were performed for all rats combined, and Group 1 and Group 2 rats considered separately.

Tables

Table 2.1. Difference between dorsolateral and dorsomedial patterns

	% Bins	RSS	K-L Div.
A1-2	2.68	2.78	0.04
A3-5	28.35	6.14	0.10
B1-2	37.93	18.12	0.28
B3-5	68.97	25.64	0.43
C1-2	20.69	10.93	0.18
C3-5	34.48	12.98	0.21

For each group of training stages, the difference in dorsolateral and dorsomedial patterns is expressed as the percentage of 20-ms bins with significantly differing z-score activations (t-test, $p < 0.01$), the sum of squared dorsolateral-dorsomedial residuals across all bins, and the Kullback-Leibler divergence of the firing distributions across task time computed for each region. See also **Figure 2.2**.

Table 2.2. Comparison of Group 1 and Group 2 activations

	Block 1			Block 2		
	% Bins	**RSS**	**K-L Div.**	**% Bins**	**RSS**	**K-L Div.**
Group 1 vs. Group 2						
Dorsolateral	15.33	6.96	0.122	18.4	8.62	0.13
Dorsomedial	12.64	4.95	0.09	15.3	5.64	0.11
Dorsolateral vs. Dorsomedial						
Group 1	15.71	6.62	0.102	55.2	21.2	0.33
Group 2	14.56	3.81	0.058	56.3	20.8	0.34

Difference measures for Group 1 versus Group 2 activities in each region (top), and the dorsolateral-dorsomedial difference measures for each group (bottom), as in **Table 2.1**. See also **Figure 2.3**.

Table S1

	Auditory % Correct		Tactile % Correct		Total % Correct		Run Time		Auditory-Tactile % Correct	
	R	p	R	p	R	p	R	p	R	p
D15	Not enough units									
D16	**-0.801**	**0.009**	-0.455	0.219	**-0.757**	**0.018**	**0.724**	**0.027**	-0.592	0.093
D19	0.110	0.627	-0.375	0.086	-0.005	0.983	0.203	0.365	0.220	0.325
D22	**0.599**	**0.031**	0.090	0.771	**0.645**	**0.017**	-0.277	0.360	0.531	0.062
D23	0.253	0.363	-0.229	0.412	0.190	0.497	-0.475	0.073	0.330	0.230
D25	**0.685**	**0.000**	**0.916**	**0.000**	**0.830**	**0.000**	**-0.383**	**0.040**	-0.256	0.180
D27	0.365	0.079	**0.469**	**0.021**	**0.440**	**0.031**	0.083	0.699	-0.357	0.087
D28	**0.596**	**0.000**	**0.458**	**0.006**	**0.541**	**0.001**	**-0.386**	**0.022**	0.316	0.064

R and p values for correlations between dorsolateral ensemble unit activity around goal reaching and various measures of behavioral performance, calculated for individual rats, related to Figure 8. Red type indicates significant correlation in the expected direction, pink type indicates significant correlation in the opposite direction.

Table S2

	Auditory % Correct		Tactile % Correct		Total % Correct		Run Time		Auditory-Tactile % Correct	
	R	p	R	p	R	p	R	p	R	p
D15	0.091	0.666	**-0.524**	**0.007**	-0.073	0.730	0.251	0.226	0.281	0.174
D16	**0.901**	**0.000**	0.479	0.115	**0.915**	**0.000**	**-0.827**	**0.001**	**0.598**	**0.040**
D19	**0.665**	**0.001**	-0.172	0.456	**0.562**	**0.008**	-0.273	0.231	**0.711**	**0.000**
D22	**-0.852**	**0.000**	0.374	0.207	**-0.772**	**0.002**	**0.869**	**0.000**	**-0.874**	**0.000**
D23	0.262	0.346	-0.340	0.215	0.158	0.573	-0.451	0.092	0.373	0.171
D25	0.153	0.427	0.266	0.164	0.210	0.275	0.022	0.910	-0.152	0.430
D27	0.046	0.832	-0.215	0.313	-0.104	0.628	-0.226	0.289	**0.478**	**0.018**
D28	-0.001	0.997	-0.197	0.257	-0.104	0.550	**-0.438**	**0.008**	**0.421**	**0.012**

Correlations between ensemble activity of dorsomedial medium spiny units around cue and turn onset and various behavioral parameters, related to Figure 8. Color conventions as in Table S1.

Figures

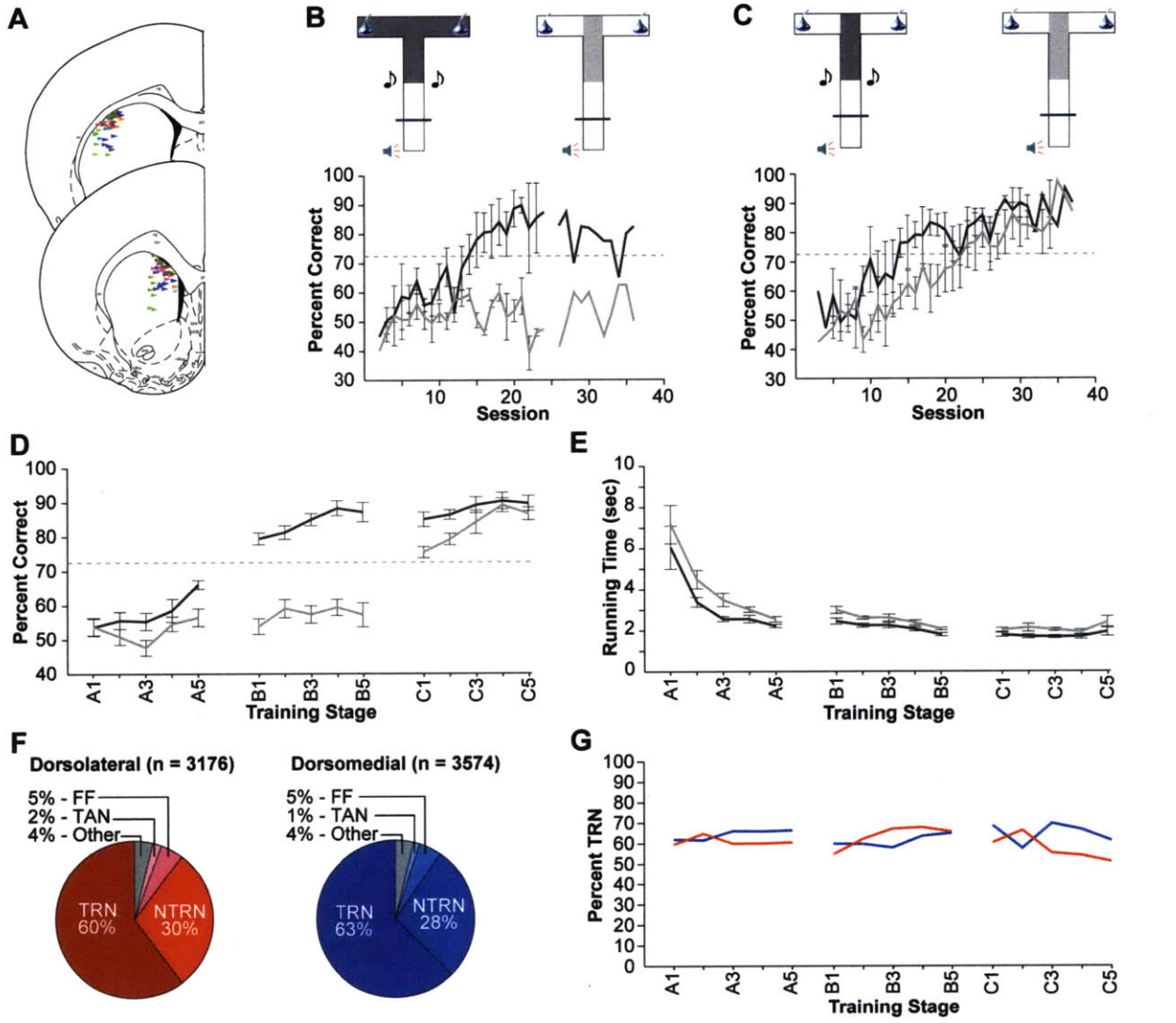

Figure 2.1. Behavioral training and neuronal recording

(A) Final tetrode locations for dorsolateral (top) and dorsomedial (bottom) recording sites. Different colors indicate sites from different animals. (B and C) Diagrams of T-maze task-versions (top) and percent correct performance across training sessions (bottom) for Group 1 (B, n = 5) and Group 2 (C, n = 3) animals. Dark gray denotes auditory instruction cue presentation, light gray, tactile instruction cue presentation. Only one animal in Group 1 continued training beyond 23 sessions, and session 25 for this animal was excluded from analysis as too few trials were performed. (D and E) Percent correct performance (D) and cue-to-goal running times (E) averaged across all rats, for auditory (dark gray) and tactile (light gray) task-versions. Stages denoted as: stage A1 = first 1-2 days of training; stage A2 = second 1-2 sessions of training; stages A3-A5 = evenly sampled 1-2 sessions of training prior to criterial performance (72.5%) on either task version; stages B1-B5: evenly sampled 1-2 sessions of training following criterial performance on the auditory version but prior to criterion on the tactile version; stages C1-C5: 2 consecutive sessions following criterial performance on both auditory and tactile task versions. Error bars indicate SEM. (F) Percent recorded units from dorsolateral (left, red) and dorsomedial (right, blue) striatum, classified as different putative neuronal subtypes. TRN = task-responsive medium spiny neurons; NTRN = non-task-responsive medium spiny neurons; FF = fast-firing interneurons; TAN = tonically active neurons. (G) Percent of TRNs across training stages. See also Figure S1.

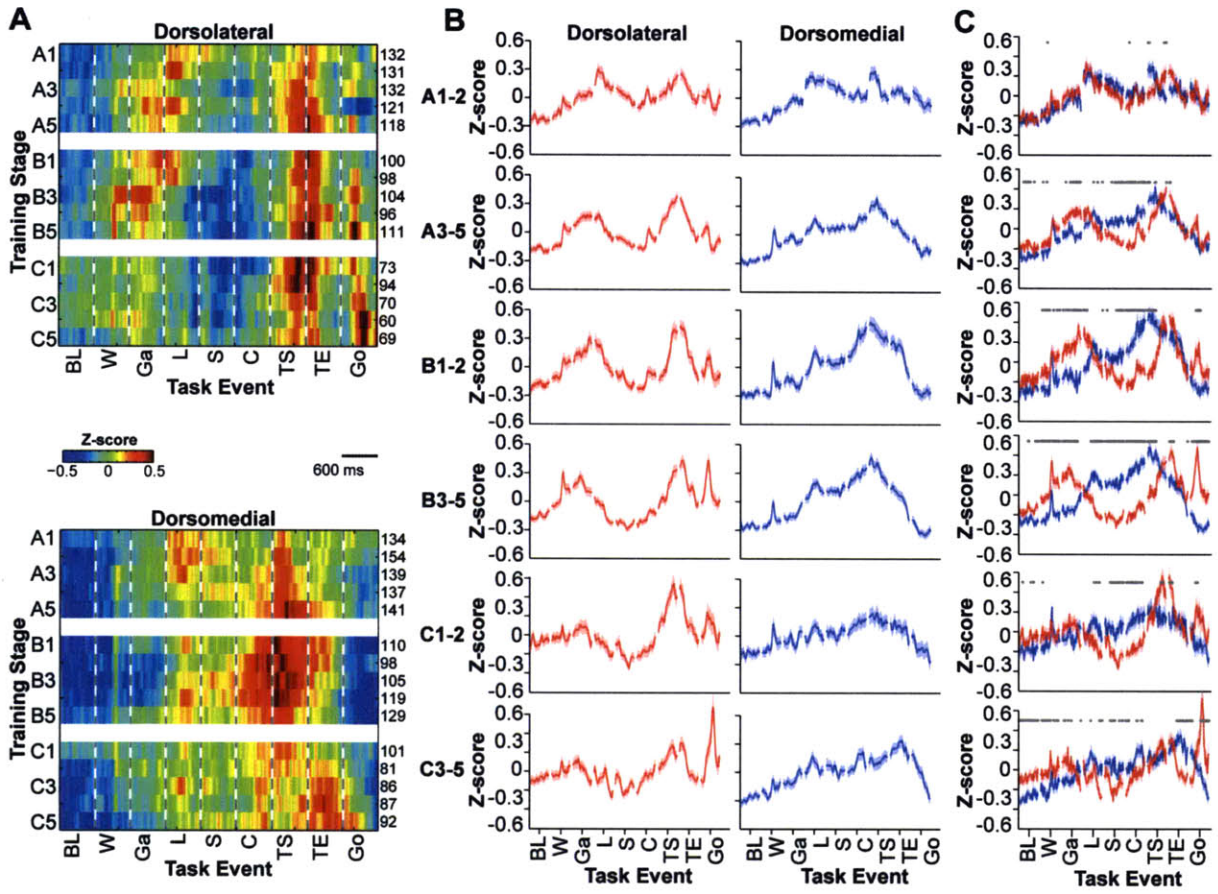

Figure 2.2. Ensemble neural activity differs between dorsolateral and dorsomedial striatal recording sites during T-maze training

(A) Ensemble z-score plots illustrating population activity across trial time and training stages for dorsolateral (top) and dorsomedial (bottom) TRNs. Scale for both plots shown in center. Numbers to the right of each row indicate the number of units included in that stage. (B and C) Mean z-scores (solid lines) and SEMs (shaded) plotted across task time for dorsolateral (red) and dorsomedial (blue) TRNs separately (B) and overlaid (C) for successive phases of training. Task events abbreviated as: BL = baseline (1 sec prior to warning click); W = warning click; Ga = gate opening; L = locomotion onset; S = out of start; C = cue onset; TS = turn start; TE = turn end; Go = goal reaching. Gray dots in C indicate significant difference between dorsolateral and dorsomedial activity during the corresponding 20-ms bin ($p < 0.01$, t-test). See also Table 1 and Figure S2.

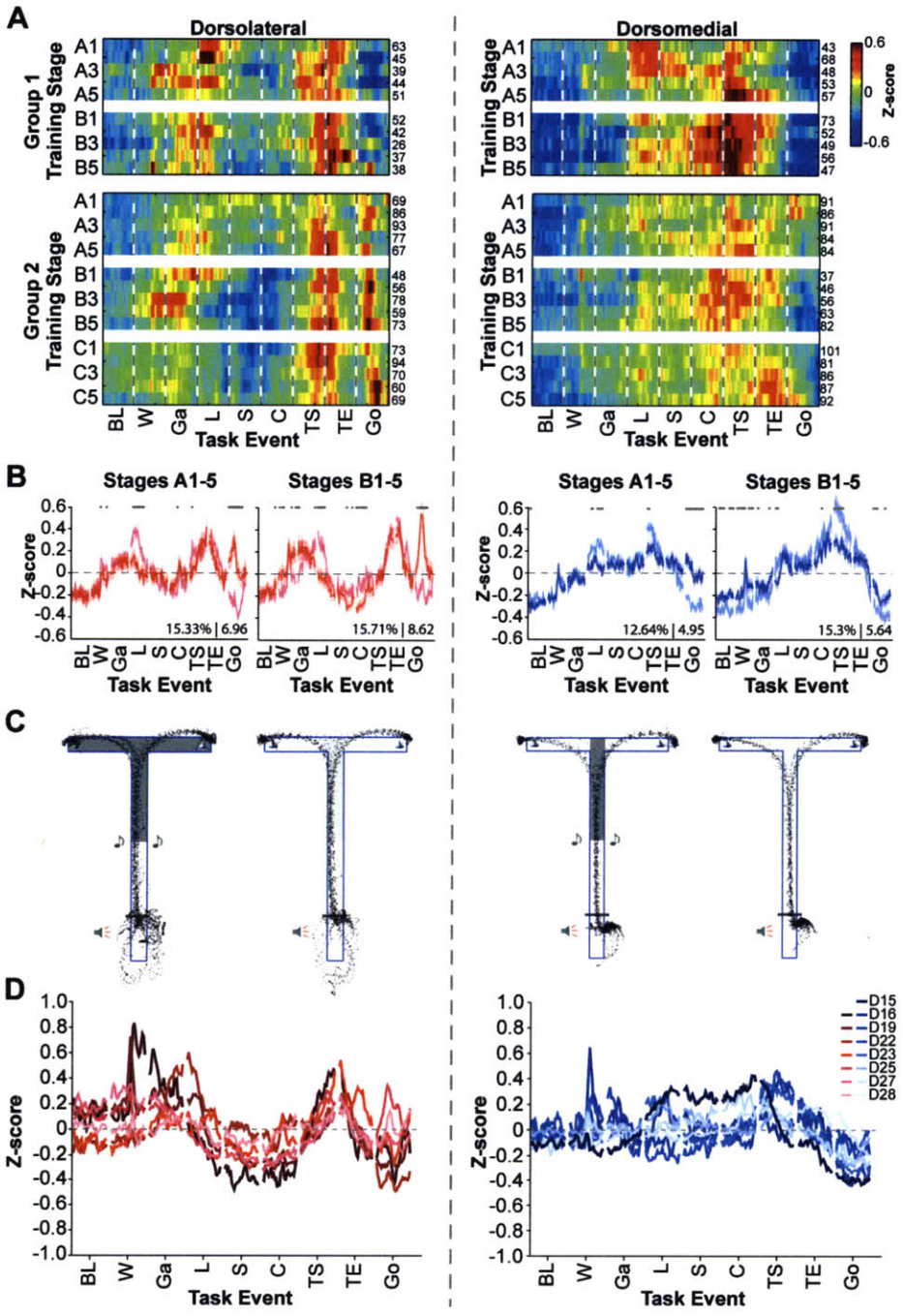

Figure 2.3. Group 1 and Group 2 rats have similar ensemble TRN firing patterns
(A) Ensemble z-score plots for TRN populations in dorsolateral (left) and dorsomedial (right) striatum recorded from rats in Group 1 (top) and Group 2 (bottom). Conventions as in Figure 2A. (B) Mean z-scores and SEMs across task-time for Group 1 (light color) and Group 2 (dark color) neuronal populations in dorsolateral (left, red) and dorsomedial (right, blue) striatum during stages A1-A5 and stages B1-B5. Gray dots as in Figure 2C for Group 1 versus Group 2 activity. (C) Representative run trajectories during the performance of the two task versions recorded during the final training session for a Group 1 animal (left, D22 session 19) and a Group 2 animal (right, D25 session 33). (D) Mean z-scores for TRN ensembles recorded from each rat, left/red: dorsolateral, right/blue: dorsomedial. See also Table 2 and Figure S3.

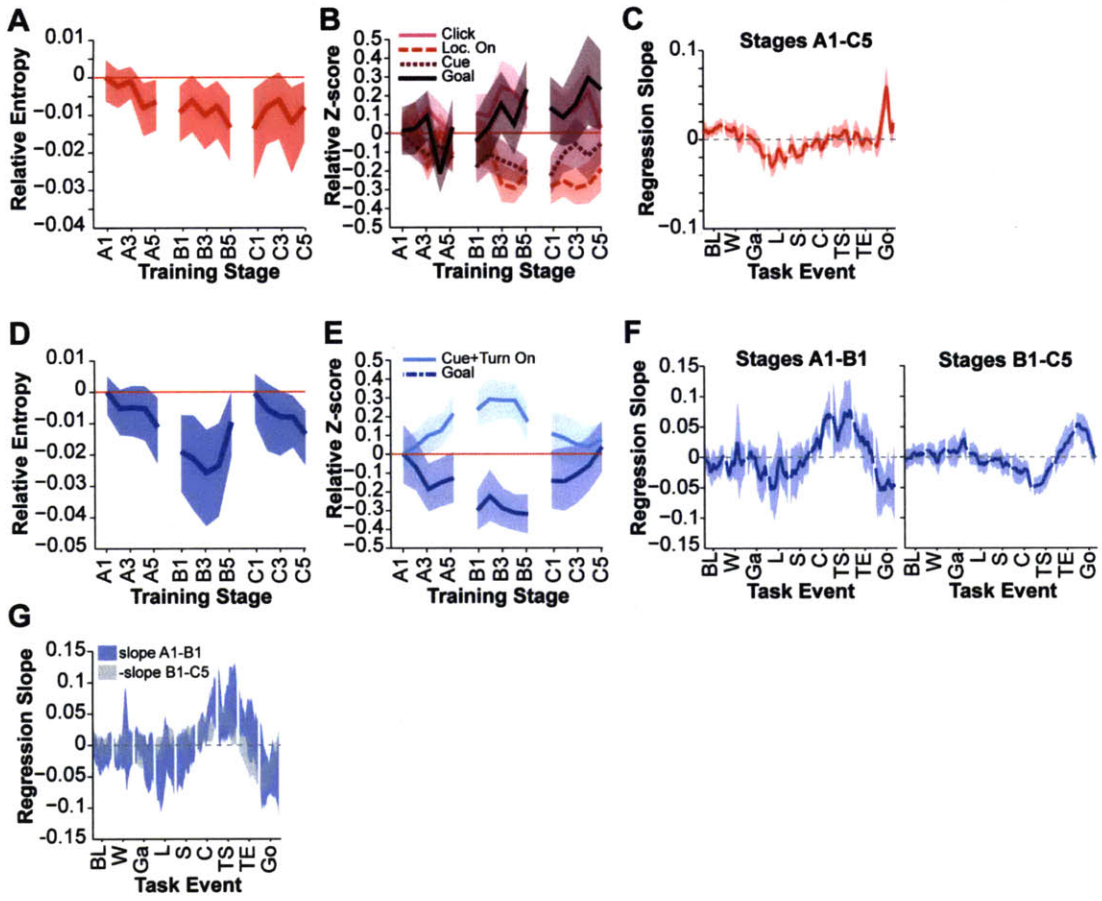

Figure 2.4. Ensemble TRN activity displays different training-related dynamics in dorsolateral and dorsomedial striatum

(A and D) Mean entropy and 95% confidence interval of the ensemble firing distribution for each stage of training relative to stage A1 for dorsolateral (A) and dorsomedial (D) striatum. (B and E) Mean z-scores and 95% confidence interval around specific task events for dorsolateral (B) and dorsomedial (E) ensembles across training stages, relative to stage A1. Means and confidence intervals were computed using 1000 bootstrap samples over the neuronal population for each stage. (C and F) Z-score regression slopes for each 20 ms bin and 95% confidence intervals for dorsolateral striatum (C), using a single linear regression across all stages, and for dorsomedial striatum (F), using a segmented linear regression with a single breakpoint at training stage B1. (G) Overlaid 95% confidence intervals of dorsomedial regression slopes for stages A1-B1 and the negative of the regression slopes for stages B1-C5. See also Figure S4.

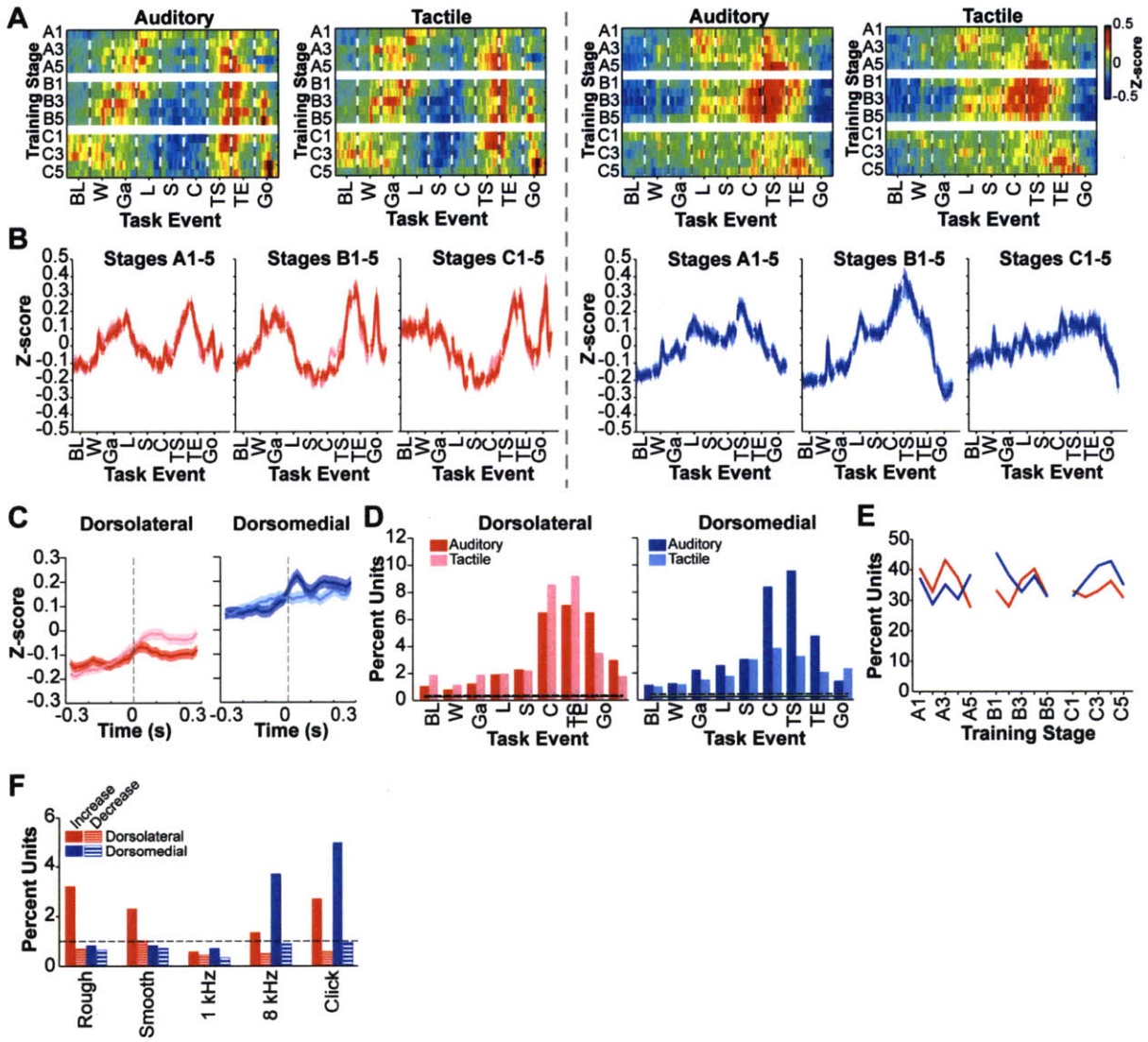

Figure 2.5. Ensemble TRN activity differs only around cue onset during auditory and tactile trials
(A) Pseudocolor z-score plots comparing dorsolateral (left) and dorsomedial (right) striatal TRN ensemble activity during auditory and tactile trials (as labeled). (B) Mean z-scores and SEMs across task-time for auditory (dark color) and tactile (light color) trials, plotted for each training block for dorsolateral (left, red) and dorsomedial (right, blue) ensembles. (C) Mean z-scores and SEM across all stages for dorsolateral (left) and dorsomedial (right) TRNs during ±300-ms around cue onset, (D) Percentage of units differentiating between auditory and tactile task-versions for dorsolateral (left, red) and dorsomedial (right, blue) TRNs. Dark and light bars indicate percentage of units with higher firing during auditory or tactile conditions, respectively. Solid and dashed black lines indicate percentage of auditory- and tactile-preferring neurons obtained after shuffling trials. (E) Percentages of modality-discriminative TRNs in dorsolateral (red) and dorsomedial (blue) striatal regions, plotted across training stage. (F) Percentage of TRNs responding with significant increases or decreases in firing to the onset of each of the four discriminative stimuli and the warning click in the dorsolateral (red) and the dorsomedial (blue) striatum. Dashed line indicates percentage expected by chance. See also Figure S5.

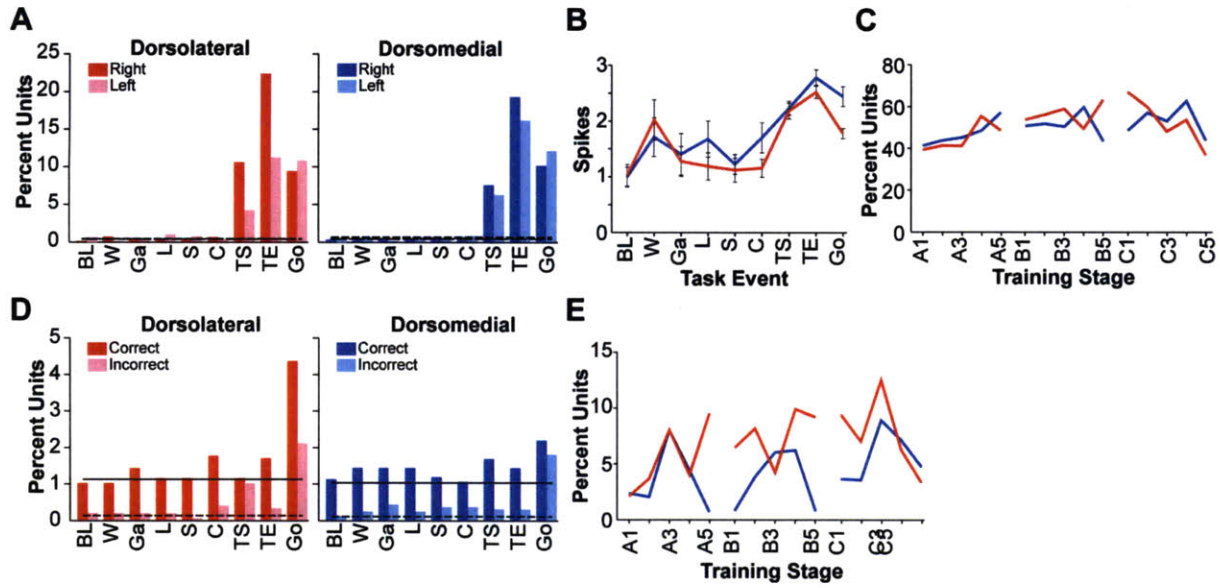

Figure 2.6. Dorsolateral and dorsomedial striatal TRNs similarly discriminate turn responses and trial outcomes

(A) Percentage of TRNs with higher firing rates during right or left turn responses around each task event for dorsolateral (left, red) and dorsomedial (right, blue) striatum. Solid and dashed black lines indicate proportion of right- and left-preferring neurons obtained after shuffling trials. (B) Mean numbers of spikes and SEM with which turn-discriminative TRNs in dorsolateral (red) and dorsomedial (blue) striatum differentiate turn direction during each event-epoch. (C) Percentage of turn-discriminative TRNs across training. (D) Percentage of dorsolateral and dorsomedial TRNs differentiating correct and incorrect trials during each task epoch. Solid and dashed black lines as in A, for correct- and incorrect-preferring populations, respectively. (E) Percentage of outcome-discriminative TRNs across training. See also Figure S6.

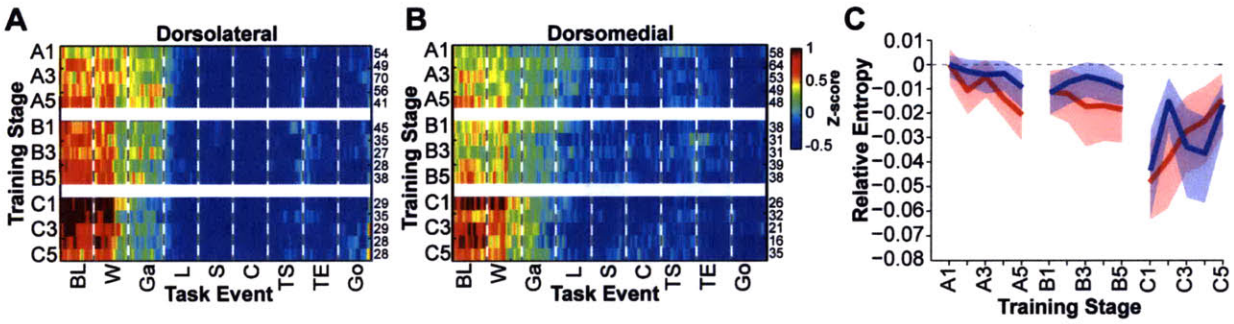

Figure 2.7. The activity of non-task-responsive striatal ensembles is also modulated during training
(A and B) Pseudocolor z-score plots showing ensemble neural activity for dorsolateral (A) and dorsomedial (B) NTRNs. Conventions as in Figure 2A. (C) Entropy estimates across training for dorsolateral (red) and dorsomedial (blue) NTRN ensembles.

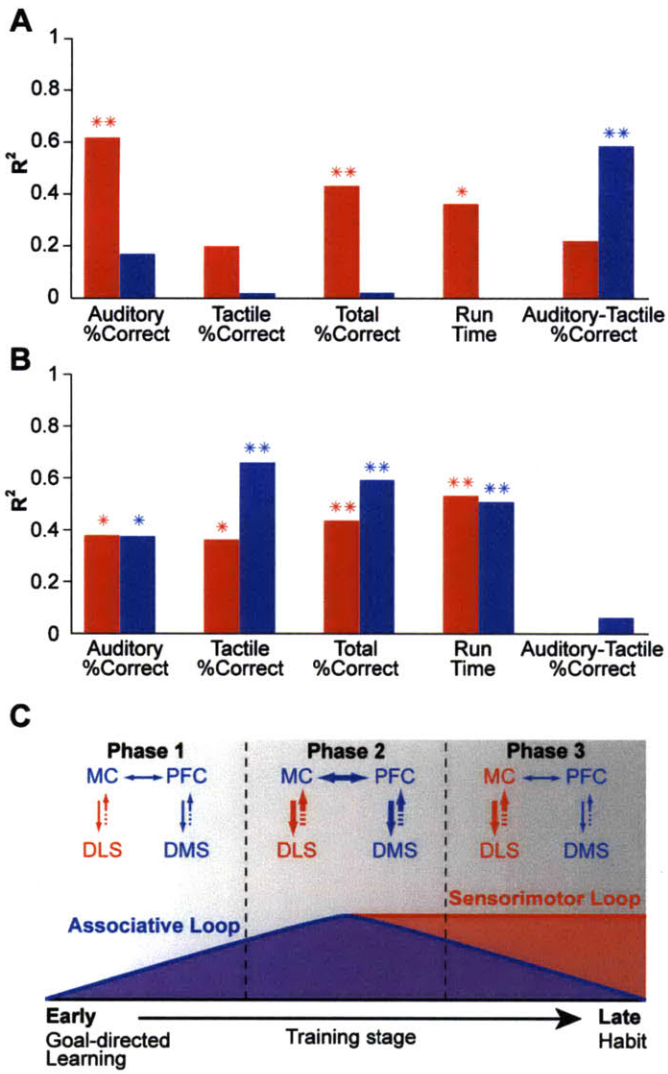

Figure 2.8. Ensemble activity patterns of dorsolateral and dorsomedial striatal TRNs are correlated with different performance measures

(A) R2 values for correlations between entropy of ensemble activity and behavioral parameters (as labeled), shown in red for dorsolateral TRN ensembles and in blue for dorsomedial TRN ensembles. *: $p < 0.05$, **: $p < 0.01$. (B) R2 values for correlations between NTRN entropy and behavioral performance measures (conventions as in A). (C) Schematic model illustrating hypothesized dorsomedial and dorsolateral cortico-basal ganglia loop interactions across different phases of learning. Activity in both striatal regions and their corresponding loops becomes structured simultaneously during Phase 1. In Phase 3, the reduction in structured dorsomedial striatal activity permits sensorimotor circuits to drive execution of habitual behavior. Broken arrows indicate multisynaptic connections from striatum to neocortex through pallidum and thalamus. MC: motor cortex; PFC: prefrontal cortex; DLS: dorsolateral striatum; DMS: dorsomedial striatum. See also Figure S7 and Tables S1 and S2.

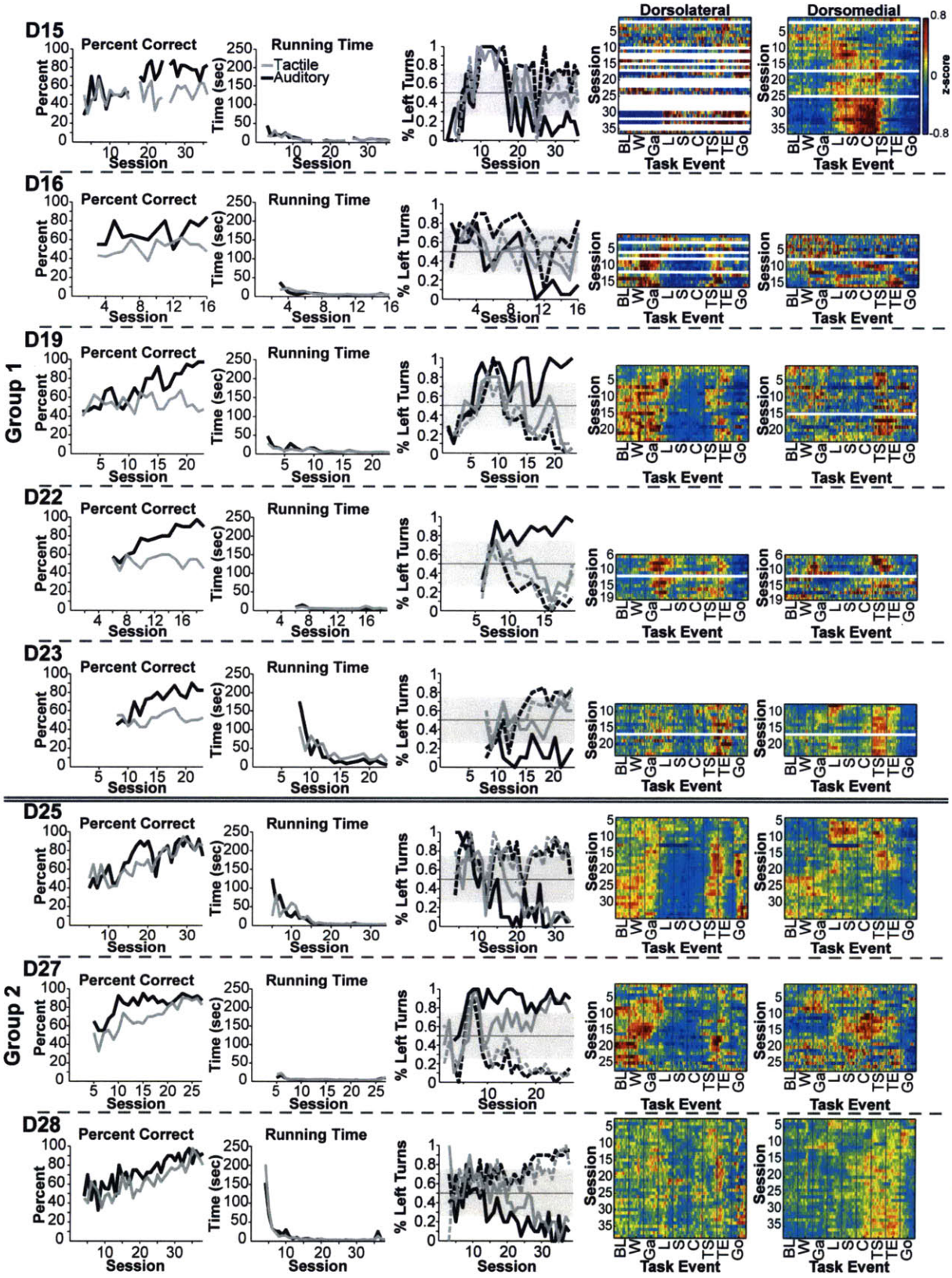
Group 1
Group 2
D15
D16
D19
D22
D23
D25
D27
D28
Percent Correct
Running Time
Tactile
Auditory
Percent
Time (sec)
% Left Turns
Session
Dorsolateral
Dorsomedial
Task Event
BL
W
Ga
L
S
C
TS
TE
Go
z-score
0.8
0
-0.8

Figure S1.

Figure 2.S1. Behavioral and neural data for individual rats, related to Figure 1.
Percent correct (far left) and cue-to-goal running time (center left) across training sessions for auditory (dark gray) and tactile (light gray) task-versions. Percent left turns (center) performed in response to each stimulus type (solid black line: 8 kHz tone; dashed black line: 1 kHz tone; solid gray line: Rough texture; dashed gray line: Smooth texture). Ensemble activity of putative projection neurons across task time for all training sessions in the dorsolateral (center right) and dorsomedial (far right) striatum. Five rats (Group 1: D15, D16, D19, D22 and D23) did not acquire the tactile discrimination, whereas 3 rats (Group 2: D25, D27, and D28) acquired both the auditory and tactile discriminations.

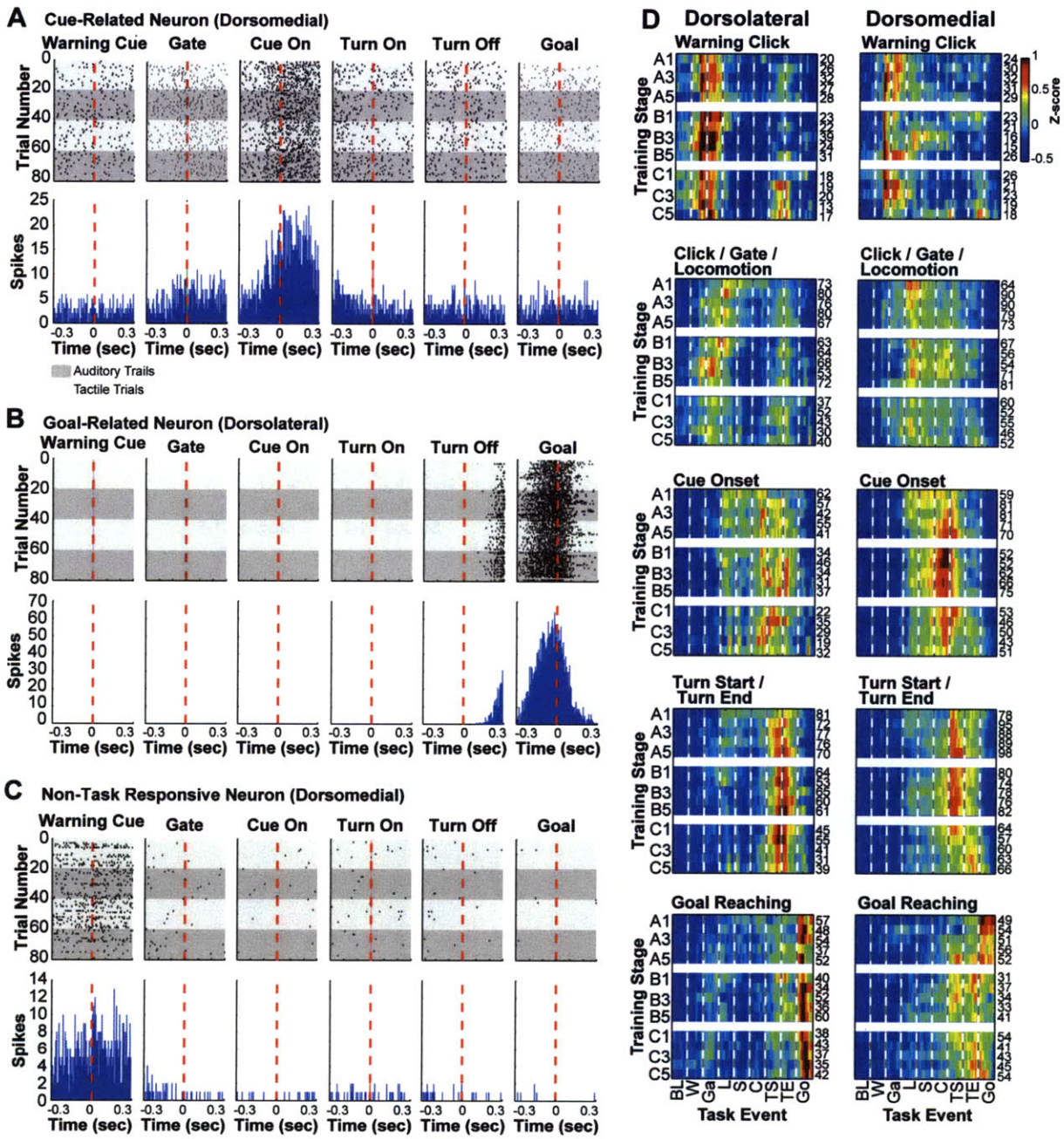

Figure 2.S2. Activity of individual medium spiny neurons and task-related neuronal ensembles, related to Figure 2. (A) A dorsomedial neuron responsive around the time of cue onset. (B) A goal-responsive neuron recorded in the dorsolateral striatum. (C) A dorsomedial "non-task-responsive" neuron. Histograms show the number of spikes per 10-ms bin occurring over all trials in the session. Dark and light shading in raster plots indicates auditory and tactile trials, respectively. (D) Ensemble activity for dorsolateral (left) and dorsomedial (right) TRNs responsive to various task events, as labeled [top to bottom: Warning Click, Beginning task events (Click, Gate Opening, and/or Locomotion Onset), Cue Onset, Turning (Turn On and/or Turn Off), and Goal Reaching].

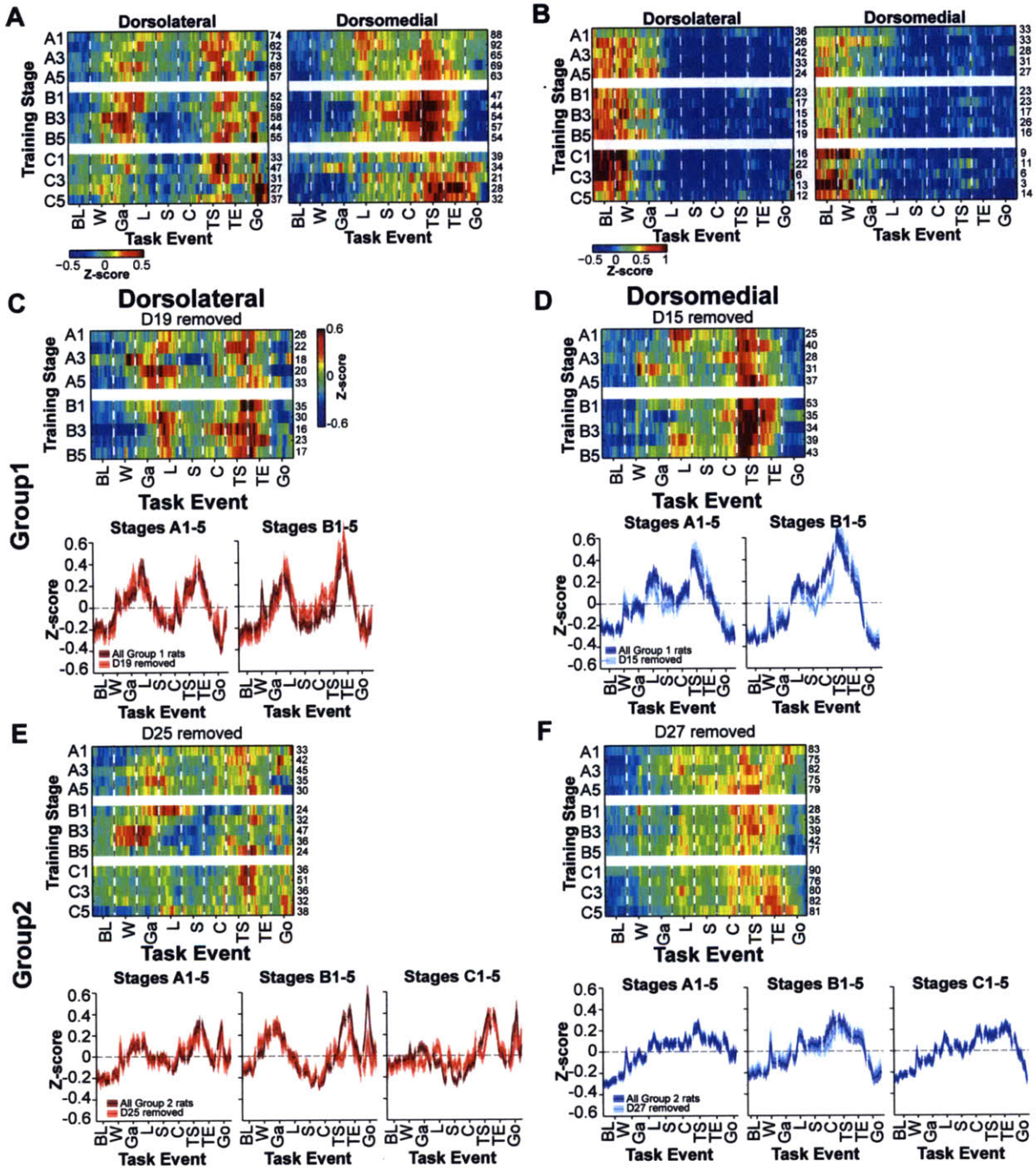

Figure 2.S3. Ensemble patterns are robust, related to Figure 3. (A-B) Ensemble activity of TRN (A) and NTRN (B) populations in dorsolateral (left) and dorsomedial (right) striatum after removal of units that were putatively identified as recorded during consecutive training sessions. (C-F) Ensemble patterns remain when rats expressing the strongest patterned activity are excluded. (C) Top: ensemble activity across task time and training for dorsolateral TRNs recorded from Group 1 animals D15, D16, D22, and D23 (D19 excluded). Bottom: mean z-score and SEM for dorsolateral ensembles recorded from Group 1 rats during training blocks A and B (light color: D19 excluded; dark color: all Group 1 rats). (D) Similar plots as in C for dorsomedial ensembles recorded from Group 1 rats, D15 excluded. (E-F) Similar plots as in C-D for Group 2 dorsolateral ensembles, excluding D25 (E) and dorsomedial ensembles, excluding D27 (F).

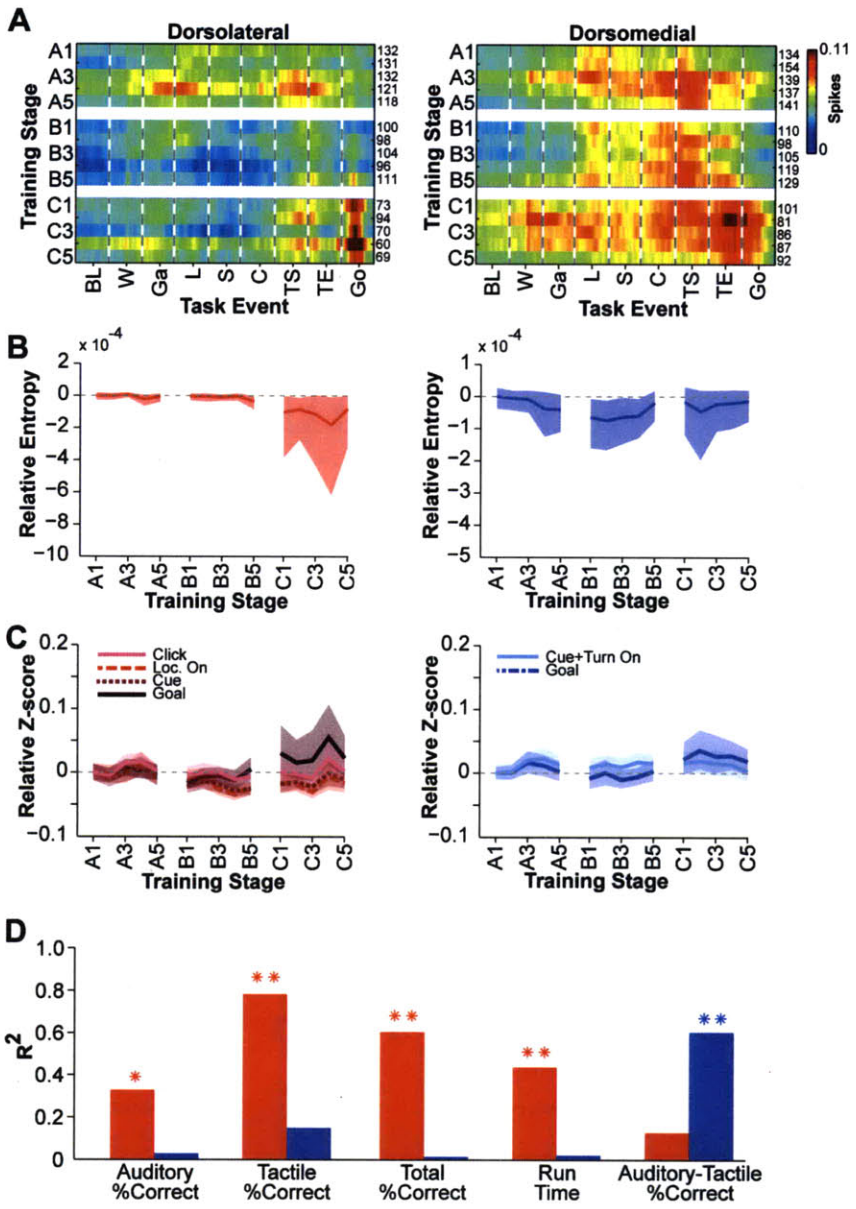

Figure 2.S4. Ensemble pattern dynamics using raw spike counts without normalization, related to Figure 4. (A) Mean single-unit spike counts per trial across task-time and training stages for dorsolateral (left) and dorsomedial (right) TRN ensembles. (B) Entropy and standard deviation calculated from spike distributions of TRN ensembles across task-time for each stage in the dorsolateral (left, red) and dorsomedial (right, blue) striatum, shown as deviation from entropy of a uniform distribution. (C) Mean spike counts relative to stage 1 and standard deviation for individual task events for dorsolateral (left) and dorsomedial (right) TRN ensembles across training. (D) R2 values for correlations between entropy shown in B and various behavioral parameters (as labeled) for lateral (red) and medial (blue) TRN ensembles. *: $p < 0.05$, **: $p < 0.01$.

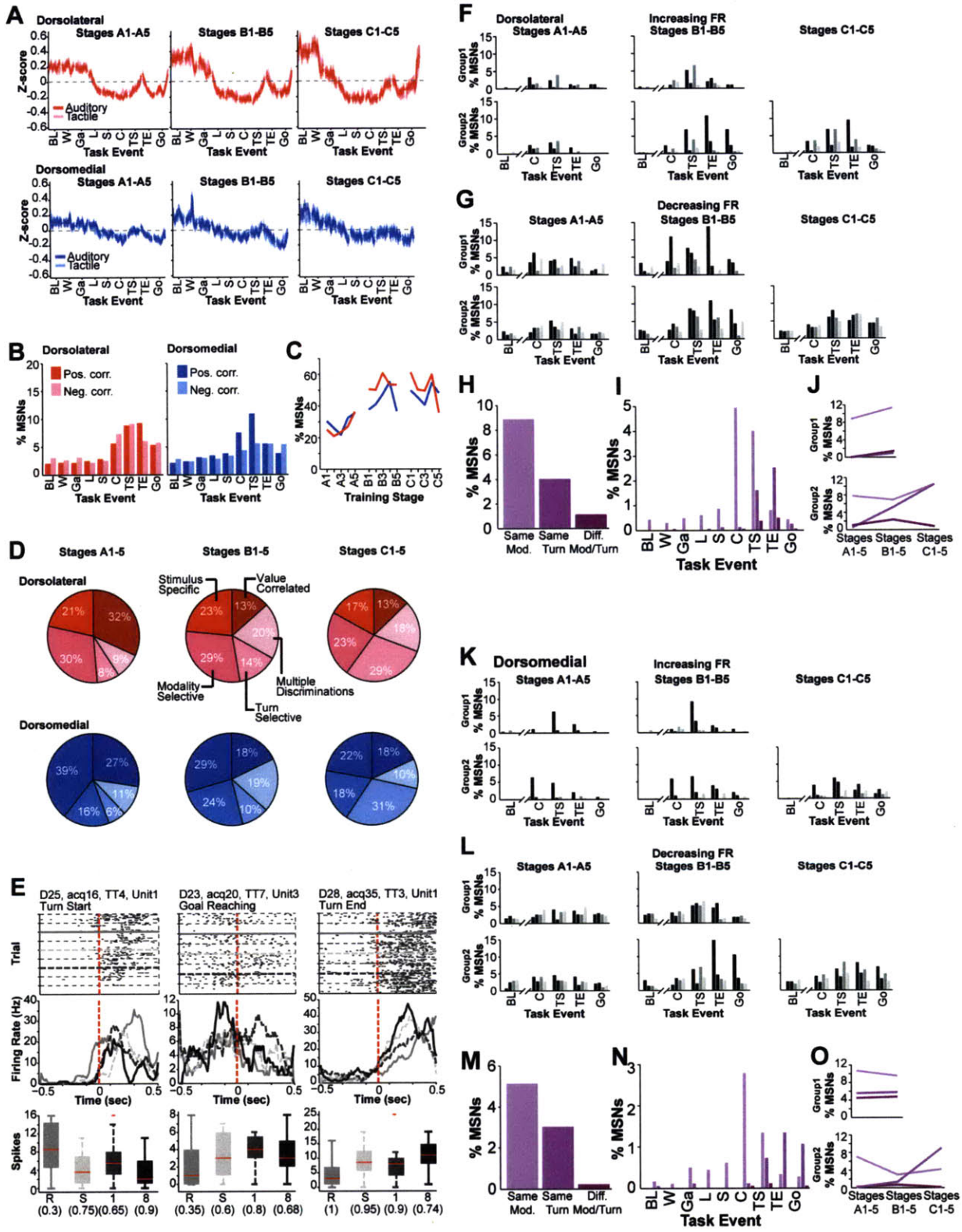
A
Dorsolateral
Stages A1-A5
Stages B1-B5
Stages C1-C5
Z-score
Auditory
Tactile
Task Event
Dorsomedial
B
Dorsolateral
Dorsomedial
Pos. corr.
Neg. corr.
% MSNs
Task Event
C
% MSNs
Training Stage
D
Stages A1-5
Stages B1-5
Stages C1-5
Dorsolateral
Stimulus Specific
Value Correlated
Modality Selective
Multiple Discriminations
Turn Selective
Dorsomedial
E
D25, acq16, TT4, Unit1
Turn Start
D23, acq20, TT7, Unit3
Goal Reaching
D28, acq35, TT3, Unit1
Turn End
Trial
Firing Rate (Hz)
Time (sec)
Spikes
R
S
1
8
(0.3)
(0.75)
(0.65)
(0.9)
(0.35)
(0.6)
(0.8)
(0.68)
(1)
(0.95)
(0.9)
(0.74)
F
Dorsolateral
Increasing FR
Stages A1-A5
Stages B1-B5
Stages C1-C5
Group1 % MSNs
Group2 % MSNs
Task Event
G
Decreasing FR
H
% MSNs
Same Mod.
Same Turn
Diff. Mod/Turn
I
Task Event
J
Stages A1-5
Stages B1-5
Stages C1-5
K
Dorsomedial
L
M
N
O

Figure 2.S5.

Figure 2.S5. Neural activity related to stimulus modality, stimulus value, or specific stimulus identity is rare, related to Figure 5.

(A) Mean z-score and SEM of dorsolateral (top) and dorsomedial (bottom) NTRN ensemble activity across task-time in training blocks A, B, and C (as labeled) during auditory (dark color) versus tactile (light color) trials.

(B-E) Activity of a small percentage of neurons is correlated with stimulus value in both striatal areas. (B) Percent medium spiny neurons (MSNs) in dorsolateral (left, red) and dorsomedial (right, blue) striatum with significant positive correlations (dark bars) or negative correlations (light bars) with stimulus value, during each event epoch. Value-correlated firing is observed primarily around cue onset and turn events. (C) The percentage of MSNs in dorsolateral (red) and dorsomedial (blue) striatum with firing significantly correlated with stimulus value increases across training stages. (D) For the majority of units, stimulus value-correlated firing can be explained by sensitivity to other task parameters during each block of training (as labeled), and the proportion of units with otherwise-explained activity increases with training. Using the percent correct performance associated with each stimulus over the entire session to estimate its value, it was impossible to disambiguate changes in stimulus-response contingencies from changes in stimulus-value contingencies, and it is likely that the increase units with value-correlated firing shown in (C) is due to the increased propensity of the rats to respond with a particular turn direction in response to a particular instruction cue as training progressed, suggested by (D). (E) Examples of neurons with stimulus value-correlated firing that was not explained by sensitivity to other task parameters. Trials are ordered in raster plots by stimulus type presented (trials 1-20: rough; trials 21-40: smooth; trials 41-60: 1 kHz; trials 61-80: 8 kHz). Line plots below rasters show mean firing rate of neuron across all trials of each stimulus type (solid light gray line: rough; dashed light gray line: smooth; dashed dark gray line: 1 kHz; solid dark gray line: 8 kHz). Box and whisker plots indicate firing rate distributions for trials of each stimulus type (R: rough; S: smooth; 1: 1 kHz; 8: 8 kHz). Numbers in parentheses indicate percentage of correct responses performed for that stimulus type in the session during which the neuron was recorded.

(F-O) Single unit activity related to individual stimuli is sparse. (F-G) Percentage of dorsolateral medium spiny neurons (MSNs) with significantly increased (F) or decreased (G) responses to each of the 4 conditional stimuli around each task event from cue onset to goal reaching, plotted separately for Group 1 and Group 2 animals during each training block. Dark gray indicates auditory cues (8 and 1 kHz tones), light gray indicates tactile cues (rough and smooth textures). Note that percentages are generally low, and the percentage of responsive units around cue onset remain stable with training, whereas percentages of units selective following turn onset increase. (H) Percentage of dorsolateral MSNs with significant firing rate modulation to 2 of the 4 conditional cues that responded similarly to the 2 cues of similar modality (light bar), to the 2 cues indicating the same turn direction (center bar), or to the two cues of opposite modality indicating different turn directions (dark bar). (I) Percentage of dorsolateral MSNs with significant responses to 2 conditional cues expressing similar responses to cues of the same modality (light bar), same turn direction (medium bar), or opposing modalities and turns (dark bar), for each task epoch. (J) Percent units responsive to 2 of 4 cues for each training block for Group 1 (top) and Group 2 (bottom) rats. Colors as in M. Note that the proportion of neurons sensitive to cues of the same modality does not increase with training, whereas the proportion of neurons sensitive to cues indicating the same turn direction does increase, but only in Group 2 rats (J). This latter set of neurons is active after the onset of turning (I), suggesting that their stimulus sensitivity reflects the increased propensity of the animal to make a specific turn in response to that stimulus, rather than the value of the stimulus itself. Increased propensity to turn in a specific direction may likewise explain training-related increases in stimulus-selective neurons around turn and goal events observed in F-G. (K-O) Same as top F-J for dorsomedial striatum.

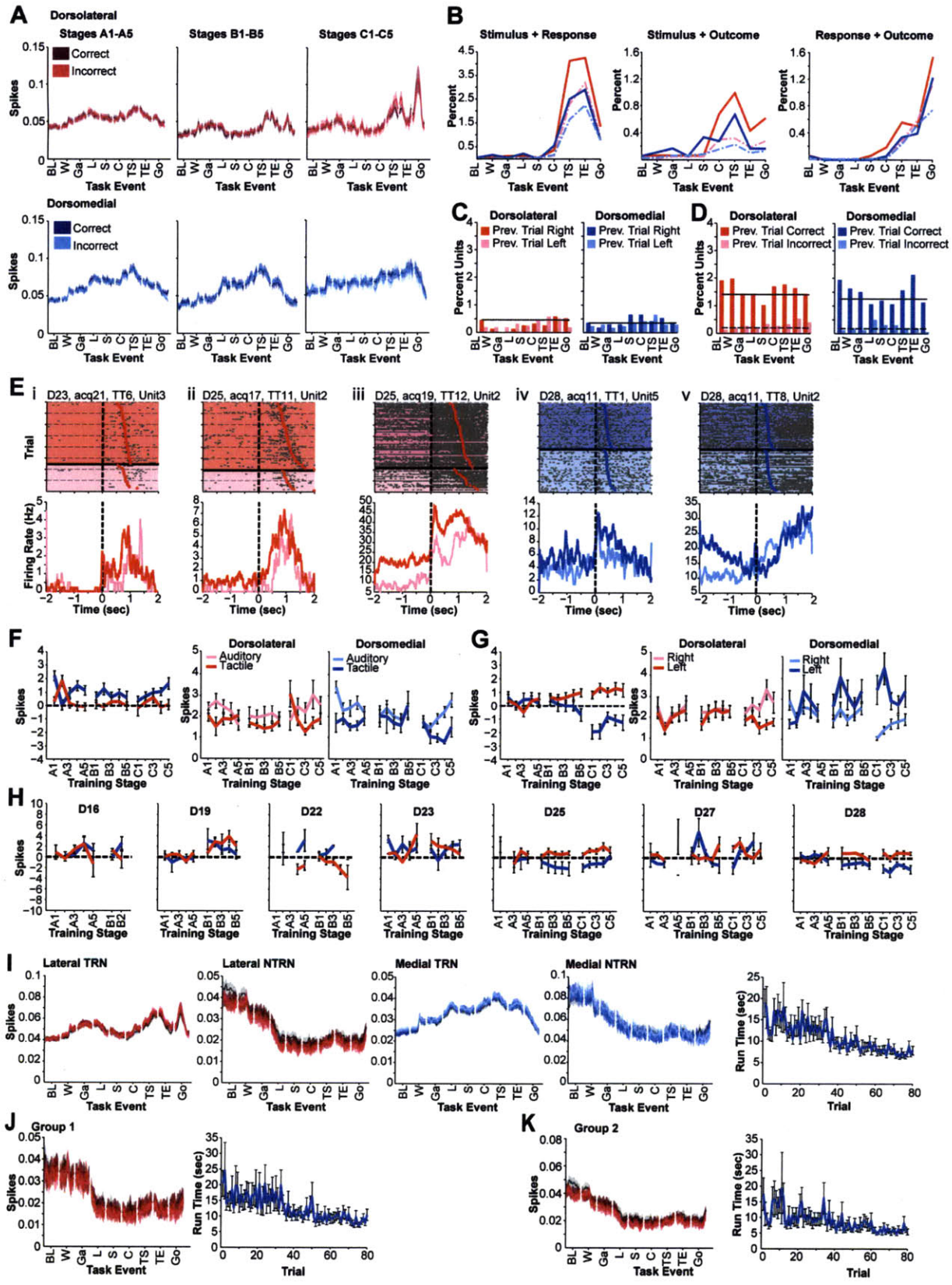
A
Dorsolateral
Stages A1-A5
Stages B1-B5
Stages C1-C5
Correct
Incorrect
Spikes
Task Event
Dorsomedial
B
Stimulus + Response
Stimulus + Outcome
Response + Outcome
Percent
C
Dorsolateral
Dorsomedial
Prev. Trial Right
Prev. Trial Left
Percent Units
D
Prev. Trial Correct
Prev. Trial Incorrect
E
i D23, acq21, TT6, Unit3
ii D25, acq17, TT11, Unit2
iii D25, acq19, TT12, Unit2
iv D28, acq11, TT1, Unit5
v D28, acq11, TT8, Unit2
Trial
Firing Rate (Hz)
Time (sec)
F
Auditory
Tactile
Training Stage
G
Right
Left
H
D16
D19
D22
D23
D25
D27
D28
I
Lateral TRN
Lateral NTRN
Medial TRN
Medial NTRN
Run Time (sec)
Trial
J
Group 1
K
Group 2

Figure 2.S6.

Figure 2.S6. Further analysis of neural activity related to response value or outcome value, related to Figure 6.

(A) Mean spike counts per trial and SEM across task time during each training block (Stages A1-5, B1-5, and C1-5 as labeled) did not differ between correct (dark color) auditory trials compared to incorrect (light color) auditory trials for dorsolateral (top, red) and dorsomedial (bottom, blue) ensembles.

(B) Solid lines indicate proportion of TRNs in the dorsolateral (red) and dorsomedial (blue) striatum that differentiated the combinations of modality and turn (left), of modality and outcome (center), and of turn and outcome (right), which did not differ from the proportions expected assuming statistical independence (dashed lines).

(C-D) Single neurons did not discriminate behavioral responses (C) or reinforcement outcomes (D) that occurred in the previous trial. Solid and dashed lines indicate proportion of units obtained after shuffling trials.

(E) Sample neurons discriminating based on previous trial outcome. Top: rasters aligned on warning click presentation. Trials are ordered first by correct (dark shading) or incorrect (light shading) outcome on the previous trial and then by timing of the gate opening event (filled dots). Bottom: average firing rate of neuron across all previous-trial-correct (dark line) and previous-trial-incorrect (light line) trials. Red indicates dorsolateral unit, blue indicates dorsomedial unit. Units i, ii and iv are putative medium spiny neurons; units iii and v are putative fast-firing interneurons.

(F-H) Analysis of stimulus-value and response-value tracking by TRN ensembles. (F) The number of spikes with which TRNs discriminated between auditory and tactile cue modalities was stable across training for all discriminative neurons in dorsolateral (red) and dorsomedial (blue) striatum (far left), as well as for auditory-preferring and tactile-preferring subpopulations in dorsolateral (center) and dorsomedial striatum (right). (G) The number of spikes with which TRNs discriminate right from left turns increases across training (far left) both laterally (red) and medially (blue). Dorsolateral ensembles develop stronger firing during right turns (center), whereas dorsomedial ensembles develop stronger firing to left turns (right). (H) Plots similar to those shown in G for each individual animal (D15 had too few turn-sensitive neurons for analysis). Note that only D25 and D28 show changes across training consistent with those seen for the entire population.

(I-K) Dorsolateral striatal NTRN ensemble firing may be modulated by trial number within a session. (I) Average population activity for dorsolateral TRNs (far left), dorsolateral NTRNs (center left), dorsomedial TRNs (center) and dorsomedial NTRNs (center right) during successive blocks of 20 trials within a session. Gray line and shading denotes mean spike count per unit and SEM during first 20 trials of each training session. Average activity during subsequent blocks of 20 trials is denoted by progressively lighter colored lines and shading, such that the darkest color denotes trials 21-40 and the lightest shade denotes trials 61-80. Only dorsolateral NTRNs show activity modulated by trial block. Far right: Mean trial running time (averaged across all sessions) decreases for later trials, suggesting animals may be more motivated or behave more stereotypically as trials accumulate. Error bars denote SEM. (J-K) NTRN ensemble activity (left) and trial running time for Group 1 (J) and Group 2 (K) animals considered separately. Conventions as in I.

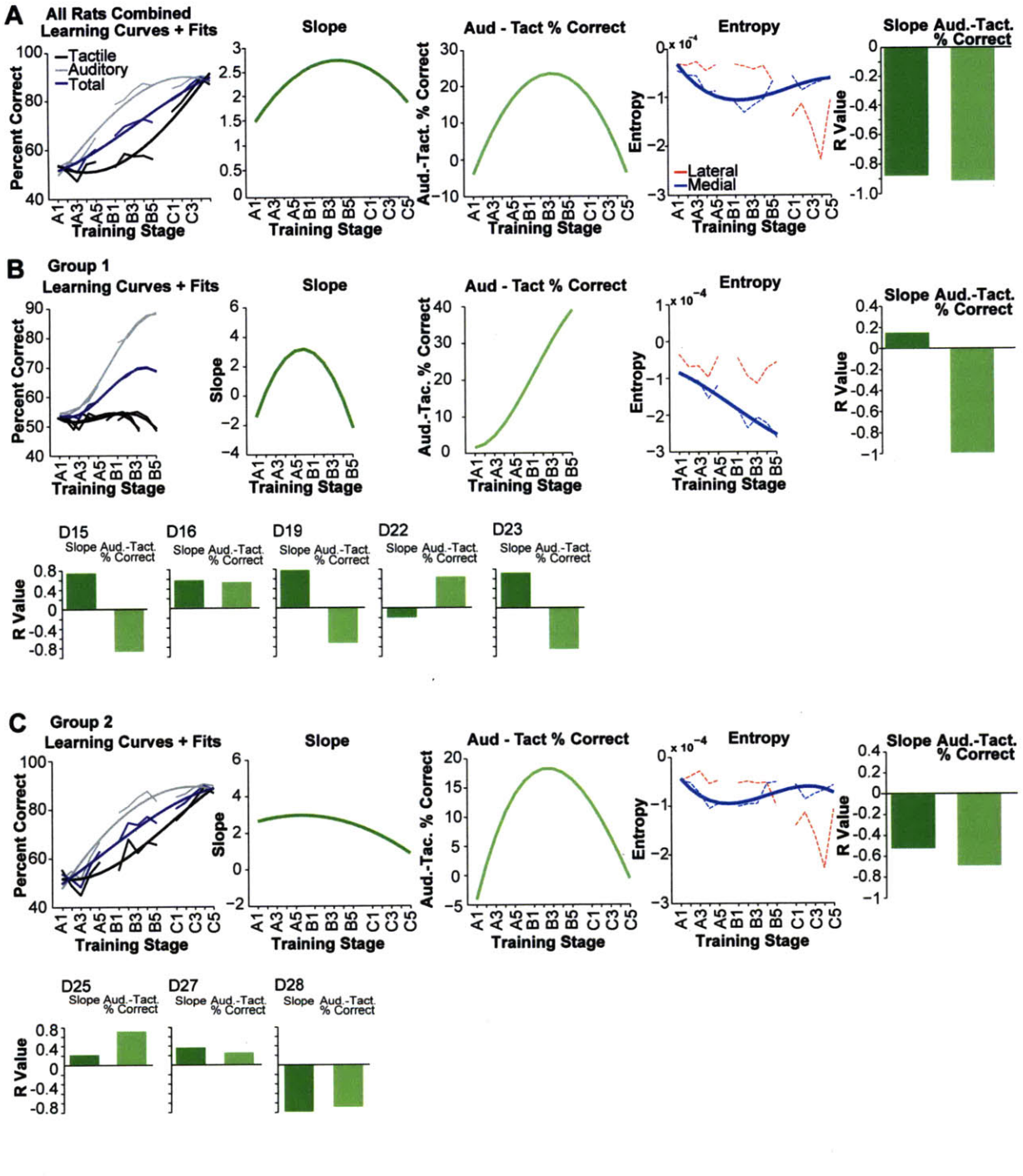
A
All Rats Combined
Learning Curves + Fits
Tactile
Auditory
Total
Percent Correct
Training Stage
Slope
Aud - Tact % Correct
Aud.-Tact. % Correct
Entropy
Lateral
Medial
R Value
Slope Aud.-Tact. % Correct
B
Group 1
Learning Curves + Fits
D15
D16
D19
D22
D23
C
Group 2
Learning Curves + Fits
D25
D27
D28

Figure 2.S7.

Figure 2.S7. Medial ensemble activity patterns are more strongly correlated with the difference in performance between auditory and tactile trials than with the slope of the behavioral performance curves, related to Figure 8. (A) Staged percent correct performance for all rats and third-degree polynomial fits (far left), slope of the total percent correct curve estimated by taking the derivative of the fit (center left), the difference between auditory and tactile percent correct performance across training stages (center), and entropy (center right) of dorsolateral (dashed red line) and dorsomedial (dashed blue line) ensemble activity across training. Solid blue line shows the third-degree polynomial fit to the medial entropy curve. Far right: R-values for correlations between medial entropy fit and slope (dark green) or between medial entropy fit and difference in percent correct performance during auditory and tactile trials (light green). (B) Top: similar plots as in A, for Group 1 rats. Bottom: plots as in far right A for individual rats in Group 1. (C) Top: plots as in A for Group 2 rats. Bottom: plots as in far right A for each rat in Group 2. Note that as a group, only Group 2 rats show a significant negative correlation between medial entropy and the slope of the behavioral performance curve, whereas both groups show significant correlations between entropy and auditory-tactile performance discrepancy.

References

Adams, S., Kesner, R. P., and Ragozzino, M. E. (2001). Role of the medial and lateral caudate-putamen in mediating an auditory conditional response association. Neurobiol Learn Mem *76*, 106-116.

Balleine, B. W., Liljeholm, M., and Ostlund, S. B. (2009). The integrative function of the basal ganglia in instrumental conditioning. Behav Brain Res *199*, 43-52.

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., and Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature *437*, 1158-1161.

Berke, J. D., Breck, J. T., and Eichenbaum, H. (2009). Striatal versus hippocampal representations during win-stay maze performance. J Neurophysiol *101*, 1575-1587.

Cardinal, R. N. (2006). Neural systems implicated in delayed and probabilistic reinforcement. Neural Netw *19*, 1277-1301.

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

Corbit, L. H., and Janak, P. H. (2007). Inactivation of the lateral but not medial dorsal striatum eliminates the excitatory impact of Pavlovian stimuli on instrumental responding. J Neurosci *27*, 13977-13981.

Dagher, A., and Robbins, T. W. (2009). Personality, addiction, dopamine: insights from Parkinson's disease. Neuron *61*, 502-510.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci *8*, 1704-1711.

DeLong, M. R., and Wichmann, T. (2007). Circuits and circuit disorders of the basal ganglia. Arch Neurol *64*, 20-24.

Emondi, A. A., Rebrik, S. P., Kurgansky, A. V., and Miller, K. D. (2004). Tracking neurons recorded from tetrodes across time. J Neurosci Methods *135*, 95-105.

Featherstone, R. E., and McDonald, R. J. (2005). Lesions of the dorsolateral or dorsomedial striatum impair performance of a previously acquired simple discrimination task. Neurobiol Learn Mem *84*, 159-167.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. Annu Rev Neurosci *31*, 359-387.

Graybiel, A. M., and Mink, J. W. (2009). The basal ganglia and cognition. In The Cognitive Neurosciences IV, M. Gazzaniga, ed. (Cambridge, MA, MIT Press).

Histed, M. H., Pasupathy, A., and Miller, E. K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. Neuron *63*, 244-253.

Horvitz, J. C. (2009). Stimulus-response and response-outcome learning mechanisms in the striatum. Behav Brain Res *199*, 129-140.

Kalivas, P. W. (2008). Addiction as a pathology in prefrontal cortical regulation of corticostriatal habit circuitry. Neurotox Res *14*, 185-189.

Kantak, K. M., Green-Jordan, K., Valencia, E., Kremin, T., and Eichenbaum, H. B. (2001). Cognitive task performance after lidocaine-induced inactivation of different sites within the basolateral amygdala and dorsal striatum. Behav Neurosci *115*, 589-601.

Kim, Y. B., Huh, N., Lee, H., Baeg, E. H., Lee, D., and Jung, M. W. (2007). Encoding of action history in the rat ventral striatum. J Neurophysiol *98*, 3548-3556.

Kimchi, E. Y., and Laubach, M. (2009a). The dorsomedial striatum reflects response bias during learning. J Neurosci *29*, 14891-14902.

Kimchi, E. Y., and Laubach, M. (2009b). Dynamic encoding of action selection by the medial striatum. J Neurosci *29*, 3148-3159.

Kimchi, E. Y., Torregrossa, M. M., Taylor, J. R., and Laubach, M. (2009). Neuronal correlates of instrumental learning in the dorsal striatum. J Neurophysiol *102*, 475-489.

Kubota, Y., Liu, J., Hu, D., DeCoteau, W. E., Eden, U. T., Smith, A. C., and Graybiel, A. M. (2009). Stable encoding of task structure coexists with flexible coding of task events in sensorimotor striatum. J Neurophysiol *102*, 2142-2160.

Lau, B., and Glimcher, P. W. (2007). Action and outcome encoding in the primate caudate nucleus. J Neurosci *27*, 14502-14514.

Lau, B., and Glimcher, P. W. (2008). Value representations in the primate striatum during matching behavior. Neuron *58*, 451-463.

McGeorge, A. J., and Faull, R. L. (1989). The organization of the projection from the cerebral cortex to the striatum in the rat. Neuroscience *29*, 503-537.

Middleton, F. A., and Strick, P. L. (2000). Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Res Brain Res Rev *31*, 236-250.

Ragozzino, M. E. (2007). The contribution of the medial prefrontal cortex, orbitofrontal cortex, and dorsomedial striatum to behavioral flexibility. Ann N Y Acad Sci *1121*, 355-375.

Robbins, T. W., Ersche, K. D., and Everitt, B. J. (2008). Drug addiction and the memory systems of the brain. Ann N Y Acad Sci *1141*, 1-21.

Rushworth, M. F. (2008). Intention, choice, and the medial frontal cortex. Ann N Y Acad Sci *1124*, 181-207.

Rushworth, M. F., and Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci *11*, 389-397.

Samejima, K., and Doya, K. (2007). Multiple representations of belief states and action values in corticobasal ganglia loops. Ann N Y Acad Sci *1104*, 213-228.

Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. Science *310*, 1337-1340.

Schall, J. D., Stuphorn, V., and Brown, J. W. (2002). Monitoring and control of action by the frontal lobes. Neuron *36*, 309-322.

Schmitzer-Torbert, N., and Redish, A. D. (2004). Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple T task. J Neurophysiol *91*, 2259-2272.

Tang, C., Pawlak, A. P., Prokopenko, V., and West, M. O. (2007). Changes in activity of the striatum during formation of a motor habit. Eur J Neurosci *25*, 1212-1227.

Volkow, N. D., Fowler, J. S., Wang, G. J., Baler, R., and Telang, F. (2009). Imaging dopamine's role in drug abuse and addiction. Neuropharmacology *56 Suppl 1*, 3-8.

Wassum, K. M., Cely, I. C., Maidment, N. T., and Balleine, B. W. (2009). Disruption of endogenous opioid activity during instrumental learning enhances habit acquisition. Neuroscience *163*, 770-780.

White, N. M. (2009). Some highlights of research on the effects of caudate nucleus lesions over the past 200 years. Behav Brain Res *199*, 3-23.

Williams, Z. M., and Eskandar, E. N. (2006). Selective enhancement of associative learning by microstimulation of the anterior caudate. Nat Neurosci *9*, 562-568.

Worbe, Y., Baup, N., Grabli, D., Chaigneau, M., Mounayar, S., McCairn, K., Feger, J., and Tremblay, L. (2009). Behavioral and movement disorders induced by local inhibitory dysfunction in primate striatum. Cereb Cortex *19*, 1844-1856.

Yin, H. H., and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. Nat Rev Neurosci *7*, 464-476.

Yin, H. H., Mulcare, S. P., Hilario, M. R., Clouse, E., Holloway, T., Davis, M. I., Hansson, A. C., Lovinger, D. M., and Costa, R. M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. Nat Neurosci *12*, 333-341.

Yin, H. H., Ostlund, S. B., and Balleine, B. W. (2008). Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. Eur J Neurosci *28*, 1437-1448.

3. Striatal theta-band oscillations are coherent with hippocampal theta during T-maze learning

This chapter consists of two published manuscripts and associated figures:

1. *DeCoteau, W. E., *Thorn, C., Gibson, D. J., Courtemanche, R., Mitra, P., Kubota, Y., and Graybiel, A.M. (2007). Oscillations of local field potentials in the rat dorsal striatum during spontaneous and instructed behaviors. *J Neurophysiol* 97(5):3800-5
* These authors contributed equally to the work.

2. *DeCoteau, W. E., *Thorn, C., Gibson, D. J., Courtemanche, R., Mitra, P., Kubota, Y., and Graybiel, A.M. (2007). Learning-related coordination of striatal and hippocampal theta rhythms during acquisition of a procedural maze task. *Proc Natl Acad Sci USA* 104(13):5644-9.
* These authors contributed equally to the work.

3.1. Oscillations of local field potentials in the rat dorsal striatum during spontaneous and instructed behaviors

William E. DeCoteau,[1] Catherine Thorn,[2,3] Daniel J. Gibson,[2,4] Richard Courtemanche,[5] Partha Mitra,[6] Yasuo Kubota,[2,4] and Ann M. Graybiel[2,4,]

[1]Department of Psychology, St. Lawrence University, Canton, New York 13617; [2]McGovern Institute for Brain Research, [3]Department of Electrical Engineering and Computer Science, and [4]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; [5]Department of Exercise Science and Center for Studies in Behavioral Neurobiology, Concordia University, Montreal, Quebec H4B 1R6 Canada; and [6]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724

Acknowledgments

We thank Patricia Harlan for her help with the histology and Henry F. Hall, who is responsible for the photography.

Grants

This work was supported by NIH/NIMH grants MH60379 and MH071744.

ABSTRACT

Oscillatory activity is a candidate mechanism for providing frequency coding for the generation, storage and replay of sequential representations of events and episodes. We recorded local field potentials (LFPs) and spike activity in the striatum, a basal ganglia structure implicated in behavioral action-sequence learning and performance, as rats engaged in spontaneous and instructed behaviors in a T-maze task. We found that during voluntary behaviors, striatal LFPs exhibit prominent theta-band oscillations together with rhythms at higher and lower frequencies. Analysis of the theta-band activity demonstrated that these oscillations are strongly modulated during task performance and increase as the animals choose and execute their turning responses in the cue-instructed T-maze task. These theta rhythms are locally generated and are coherent across large parts of the striatum. We suggest that modulation of oscillatory activity in the striatum may be a key feature of neural processing related to the control of voluntary behavior.

3.1.1. Introduction

Theta rhythms are prominent features of hippocampal spike and local field potential (LFP) activity recorded as rats engage in active behaviors (Buzsaki, 2005; Hasselmo, 2005; Vertes, 2005) and have increasingly been observed in other cortical and subcortical regions (Hasselmo, 2005). Such rhythmic activity is thought to have a major function in organizing the encoding and retrieval of sequential information in cortico-hippocampal circuits.

In sharp contrast, oscillatory spike activity is normally weak in the striatum and becomes strong only in dopamine-depleted states (Boraud et al., 2005; Courtemanche et al., 2003; Goldberg et al., 2004; Raz et al., 2001). Despite the lack of oscillatory spiking in most striatal neurons, prominent oscillations do occur in the up- and down-state transitions of striatal projection neurons, and these membrane transitions are correlated with those of cortical neurons in anesthetized preparations and can exhibit oscillatory behavior that synchronizes with LFPs (Goto and O'Donnell, 2001; Stern et al., 1997). Consistent with these findings, oscillatory LFP activity has been observed in the caudoputamen and related basal ganglia structures in the rat (Berke et al., 2004; Boraud et al., 2005; Magill et al., 2005; Masimore et al., 2005). Moreover, in normal, non-parkinsonian monkeys, it was shown that prominent rhythmic LFP activity occurs in the striatum and is strongly modulated as the monkeys perform sensorimotor tasks to receive reward (Courtemanche et al., 2003).

Here, we asked whether such behavioral modulation of oscillatory LFP activity occurs in the striatum of non-parkinsonian rats and, if so, what the characteristics of the task-dependent modulations were. To do this, we recorded LFP and spike activity in the dorsal caudoputamen as the rats rested, explored their environment, or performed a goal-directed instructed behavior in a T-maze. We found that behaviorally modulated oscillations are prominent features of LFP activity in the striatum, including striatal theta rhythms. We suggest that such rhythmic activity is likely to influence information processing in basal ganglia–based neural circuits.

3.1.2. Results

3.1.2.1. Oscillations in striatal local field potentials occur in the awake, behaving rat and are modulated by behavioral activity

Robust theta-band activity was evident in the LFPs recorded in the caudoputamen during periods of spontaneous movement through the T-maze (**Figure 3.1.1A-B**) as well as during locomotion during task performance in the maze (**Figure 3.1.1A** and **3.1.1C**). During active running, the power in the oscillatory signal was greatest at the 7- to 14-Hz band, conventionally defined in the rat as theta activity (Jones and Wilson, 2005; McNaughton et al., 2006; O'Keefe and Recce, 1993). Oscillatory activity was also present in the delta range (<5 Hz), beta range (about 14–22 Hz), and gamma range (about 30–50 Hz) as well (**Figure 3.1.1B** and **3.1.1D**). Theta activity was less prominent during grooming and during wakeful rest, but less-rhythmic, higher-frequency oscillations were still observable (**Figure 3.1.1B**). Our analyses focused on the theta band and on activity during performance of the T-maze task (**Figure 3.1.1**).

The power of these striatal LFP rhythms was strongly modulated as the rats performed the T-maze runs (**Figure 3.1.1C** and **Table 3.1.1**). We examined recordings made during the session in which the rats reached asymptotic running times in the T-maze task (5.6 ± 1.9 to 4.0 ± 0.7 s). At this point, the rats had achieved 37.5 to 90% correct performance. Theta-band activity at 9 Hz increased during the maze runs, peaked after the rats heard the instruction tone and around the start of turning, and then fell after turning (**Figure 3.1.1C**). Just before goal reaching, there was activity in many trials

at a slightly higher band (about 11–14 Hz), considered theta in the rodent literature but in human studies identified as alpha (**Figure 3.1.1C**). Beta-band activity appeared especially during the tone and turn periods. Low-frequency delta rhythms were recorded throughout, but they peaked around gate opening. High-frequency gamma activity occurred, often in brief bursts (data not shown; Masimore et al., 2005). These basic features of the LFP rhythms recorded during the maze runs were consistent in general form across animals (**Table 3.1.1**), although we did observe some variations in the LFP patterns on a trial-by-trial basis in each rat (data not shown).

Striatal theta-band power was only weakly correlated with running speed (R = 0.03–0.48, P = 0.000–0.300; **Figure 3.1.1E**) and we found no consistent correlation between spectral power and velocity in the 11- to 14- or 14- to 22-Hz bands. We did, however, note a moderate inverse correlation for the low gamma range (30–50 Hz) activity (**Figure 3.1.1E**). Spectral power was not correlated with acceleration in any of the frequency bands studied (**Figure 3.1.1E**); nor was there a consistent relationship between the magnitude of striatal theta-band activity and either the turning direction of the rats or the accuracy of their turns in reaching the baited goal (data not shown).

TABLE 3.1.1. *Spectral power of theta-band oscillations in the striatum during T-maze task performance*

Frequency Band (Hz)		Task Event						
		Baseline	Click	Gate	Tone	Turn Start	Turn End	Goal
7-11	Mean	51.81	54.55*	56.60*	55.93*	55.84*	55.70*	52.84
	Upper Limit	52.54	55.27	57.44	56.73	56.49	56.26	53.44
	Lower Limit	51.07	53.84	55.75	55.13	55.19	55.14	52.24
11-14	Mean	45.68	47.13	48.52*	45.96	45.59	45.65	47.56*
	Upper Limit	46.50	47.90	49.34	46.78	46.31	46.47	48.41
	Lower Limit	44.86	46.35	47.69	45.15	44.87	44.83	46.72

* Indicates increases in power from the pre-trial baseline period judged by 95% confidence limits calculated with the jackknife procedure.

3.1.2.2. Oscillations in striatal local field potentials are generated locally and are coherent with spike activity in a subset of striatal neurons

As a control for the possibility that electrotonic spread of voltage signal from remote oscillators could account for the rhythmic activity recorded in the striatum, we recorded LFP activity in the caudoputamen using as a local reference an adjacent tetrode about 300–600 μm away. The spectral content of oscillatory activity under these recording conditions was similar to that recorded with the amplifier ground as reference (**Figure 3.1.2A** and **3.1.2B**). Moreover, in a parallel study, we found that striatal theta is not consistently correlated with hippocampal theta during such maze behavior, showing that volume conduction from the hippocampus is not responsible for the striatal rhythms (DeCoteau et al., 2007).

To test whether there was spike rhythmicity related to the LFP rhythms in the striatum, we computed spike-LFP coherence in five different frequency bands (1–5, 7–11, 11–14, 14–22, and 25–50 Hz) and we compared the results for each band to data for the same sessions in which trials were shuffled. The percentages of putative projection neurons with significant ($P < 0.05$) spike-LFP coherence were low (6–17%). In the example shown in **Figure 3.1.2C**, they were highest for the theta

(7–11 Hz) band, and the highest proportions were found for the tone-turn period of the task (roughly 17%). Thus the spiking of some striatal projection neurons was coordinated with the striatal theta rhythms we observed in the LFP recordings, but, in agreement with other studies in a range of species, they were a minority (Berke et al., 2004; Boraud et al., 2005; Courtemanche et al., 2003; Goldberg et al., 2004).

3.1.2.3. Functionally distinct zones of the striatum exhibit coherent LFP oscillations during performance of the T-maze task

In two of the rats, we recorded oscillatory LFP activity simultaneously in the dorsomedial (associative) caudoputamen, which receives prefrontal and limbic corticostriatal inputs, and in the dorsolateral (sensorimotor) caudoputamen, which receives cortical inputs from sensorimotor cortex (**Figure 3.1.3A**). During free-run sessions, theta-band activity was highly synchronous within and across these striatal regions, with coherence values of 0.9 during the instructed maze runs (**Figure 3.1.3B**) and cross-covariance close to 1 at zero lag (**Figure 3.1.S1A**). During instructed run sessions, there was appreciably reduced variance in the theta-band coherence during the tone-turn period, as shown in the plots for tone on and turn start (**Figure 3.1.3C**, red arrows). Coherence values were low at frequencies <7 Hz (**Figure 3.1.3D**) and fell to <0.5 at frequencies >100 Hz (**Figure 3.1.S1B**). In some sessions (**Figure 3.1.3D**, top), peaks of coherence in the beta-band occurred near the beginning and the end of the runs. Prominent coherence between the medial and lateral theta rhythms was still visible after subtraction of local reference signals from the medial and lateral recordings and was heightened around the tone-turn period. Overall, the coherence levels were lower under these conditions (**Figure 3.1.3D**, bottom).

The phase relationships between the LFP oscillations in the two striatal regions were remarkably stable across task time for any one frequency band (**Figure 3.1.3E**). The phase angles measured at 9 Hz varied from -3 ± 4 to $9 \pm 7°$ (95% confidence limits). Functionally distinct zones of the striatum thus exhibit coherent LFP oscillations across a broad range of frequencies during instructed goal-directed behaviors, with the most stable coherence being at theta-band frequencies and during the tone-turn period.

3.1.3. Discussion

The timing of neuronal activity in the striatum is critical for motor and cognitive control: striatal output neurons affect the levels of phasic release and inhibition in cortico-basal ganglia pathways. Our findings demonstrate that oscillatory activity is a prominent feature of locally generated field potential activity in the rat's striatum and show that, across a range of frequency bands, the power of these LFP oscillations varies with spontaneous behavior and during instructed navigation in T-maze tasks. Temporal codes based on oscillatory modulation of neuronal activity have been invoked in functions ranging from sensory representation and neuronal network coordination to expectancy coding, timing, and sequence learning and memory (Baker et al., 1999; Buhusi and Meck, 2005; Buzsaki, 2005; Engel et al., 2001; Gray, 1994; Laurent et al., 2001; Lisman, 1999; Mehta et al., 2002). Our findings suggest that task-dependent modulation of oscillatory activity in the striatum could be an important factor influencing cortico-basal ganglia loop function during active behavior.

Theta rhythmicity was most conspicuous during spontaneous and instructed running and was weak during grooming and during wakeful rest. These characteristics held whether the LFP recordings were in the medial (associative) caudoputamen or in the lateral (sensorimotor) striatum. We found that roughly 15% of the striatal neurons classified as projection neurons exhibited oscillatory spike activity that was coherent with the LFP oscillations at theta-range frequencies at statistically significant levels. This oscillatory spiking and the results of our bipolar recording

experiments strongly support the view that the striatal theta was locally generated rather than being the result of electrotonic spread.

The theta-band oscillations recorded in these different regions were largely coherent. Theta rhythmicity is thus a general characteristic of LFP activity in the striatum of rats actively exploring and moving in their environment. These results suggest that the theta-band rhythmicity in the striatum does not depend exclusively on region-specific functions of particular striatum-based circuits. Rather, the LFP rhythms appear to be a shared feature of the temporal structuring of field activity in the striatum (Courtemanche et al., 2003; Magill et al., 2006). The striatal LFP oscillations we observed, and their marked coherence across medial and lateral striatal recording sites, could reflect different states imposed on striatal output neuron membrane potentials by other sites such as neocortex, thalamus, and pallidum, or by local circuits in the striatum operating in conjunction with these (Aldridge and Gilman, 1991; Boraud et al., 2005; Courtemanche et al., 2003; Lebedev and Nelson, 1999). For example, the fast-firing parvalbumin (PV)-containing interneurons of the striatum are inhibited by the external pallidum, itself part of an oscillatory pallido-subthalamic network under partial control by the neocortex (Bevan et al., 2002), and these neurons powerfully inhibit striatal output neurons. Moreover, these PV neurons have been proposed to be part of an intrastriatal electronically coupled inhibitory network appropriate for organizing the temporal activity of striatal neurons and for selecting input combinations leading to their activation (Berretta et al., 1997; Tepper et al., 2004). Coherent LFP oscillations could serve as a dynamic filter in the striatum (Courtemanche et al., 2003), setting a threshold for spike discharge in striatal projection neurons receiving cortical, thalamic, and other inputs. This view also accords with the proposal that oscillatory activity in corticostriatal circuits is part of a neural timing mechanism for encoding short intervals (Buhusi and Meck, 2005).

The lack of a consistent relation between either velocity or acceleration and the power of the theta-band activity suggests that the striatal LFP rhythms may not be strictly linked to sensorimotor parameters, but rather to other behavioral-state characteristics engaged during exploration and instructed running. The modulation of both power and cross-striatal coherence during the tone-turn period of the T-maze task is consistent with this conclusion. During this period, the animals were required to use the tone cues to choose which way to turn to reach the baited goal. The heightened power and coherence could, in this view, be related to behavioral decision and execution.

There is strong precedent for the presence of oscillatory activity in other nuclei of the basal ganglia, particularly in the pallidum and recurrent subthalamo-pallidal circuits (Bevan et al., 2002; Plenz and Kital, 1999; Ruskin et al., 1999; Terman et al., 2002; Wichmann et al., 2002). In these basal ganglia circuits, oscillatory activity is greatly augmented by dopamine-depleting lesions mimicking parkinsonian states and in Parkinson's disease itself (Boraud et al., 2005; Brown et al., 2001; Goldberg et al., 2004; Levy et al., 2002; Ni et al., 2000; Raz et al., 2001). The functions of such oscillatory activity in normal basal ganglia circuits are unknown. However, our findings, together with those in behaving monkeys (Courtemanche et al., 2003), provide strong evidence that they are systematically modulated by behavioral context in the striatum and are coordinated across functionally different striatal regions. This result accords with the possibility that they reflect a dynamic process integral to a range of cortico-basal ganglia circuits.

3.1.4. Methods

Eight adult male Sprague–Dawley rats served as subjects. All procedures were approved by the Massachusetts Institute of Technology Committee on Animal Care and were in accordance with the National Research Council's Guide for the Care and Use of Laboratory Animals. Headstages carrying 12 independently movable tetrodes targeting either the dorsomedial striatum (AP = +1.7 mm, ML = 1.8 mm relative to bregma, n = 6) or the dorsomedial and dorsolateral (AP = +0.5 mm, ML = 3.5 mm) striatum (n = 2) were secured on the skull with dental acrylic and anchor screws, one of which served as animal ground. As described more fully in Jog et al. (2002) and Barnes et al. (2005), headstages were designed so that a bundle of four to six tetrodes penetrated the brain tissue in a circular configuration (OD 600 μm) with inter-tetrode spacing of about 300–600 μm. Tetrodes were then lowered until unit and LFP signals were identified within the estimated depth (3.6–4.6 mm).

During recording, rats engaged in spontaneous behaviors and performed a procedural task in a T-maze under dim red light. In the T-maze task (Barnes et al., 2005), the start gate opened 200–400 ms after a click warning cue signaled the beginning of the trial. When the rat had traveled halfway to the choice point, a 1- or 8-kHz tone instructing the correct turn direction sounded and was left on until the end of the trial. The rats received chocolate sprinkles at the correct goal. Before each training session of about 40 trials, neural activity was recorded as rats freely behaved (e.g., locomotion, grooming, and quiet rest) in the same T-maze.

Neuronal and behavioral data were acquired with a Cheetah system (Neuralynx, Bozeman, MT). For unit recording, amplified (gain: 2,000–10,000) and band-pass filtered (600–6,000 Hz) signals above a preset voltage threshold were sampled at 32 kHz. Either a dedicated reference electrode or a tetrode channel without spike activity served as reference. For LFP recording, amplified (gain: 1,000) and filtered (1–475 Hz) signals were continuously digitized at 1 kHz. During training, the animal ground (one of the skull screws) or the external ground (ground of the amplifier used for neuronal recording) was used as reference. In control sessions given to test whether locally recorded LFPs were generated by a distant source, a tetrode channel about 300–600 μm away served as reference, instead of the animal or external ground. Movement-related behavioral events were marked with the aid of video tracker data (sampled at 60 Hz). During training on the T-maze task, photobeams (Med Associates, St. Albans, VT) detected the times of gate opening and goal reaching and triggered the instruction tone.

The LFP data were analyzed with open-source Chronux algorithms (http://chronux.org), in-house software, the Matlab Signal Processing Toolkit (The MathWorks, Natick, MA), and other libraries (Courtemanche et al., 2003; Pesaran et al., 2002). The multitaper method was used to estimate frequency spectra (Pesaran et al., 2002). Spectrograms were constructed by plotting spectral power during a series of overlapping constant-width time windows.

Coherence between two simultaneously recorded signals was computed as $C = S_{12}/\mathrm{sqrt}(S_1 \times S_2)$, where S_{12} denotes the averaged cross-spectrum computed from the FFTs of the tapered waveforms for each taper and trial, and S_1 and S_2 denote the averaged power spectra of the two signals. Confidence limits (95%) were estimated for coherence magnitude by a jackknife procedure (which does not assume coherence to be normally distributed).

For the bipolar recording data shown in **Figure 3.1.2**, the differences between LFP voltage and reference voltage were computed by a differential amplifier with 100-dB common-mode rejection. For the recording data illustrated in **Fig. 3.1.3D** (bottom), the value of the signal on a local reference electrode was subtracted from the value of the LFP off-line in Matlab. Because the recording amplifier gains were specified to ±1% precision, the common-mode rejection ratio in this configuration was 34 dB.

For band-limited spectral power, a single-taper unpadded spectrum was calculated for each trial and electrode in a 0.75-s window centered on each event marker. The power components were then summed for each frequency band. These time series were linearly interpolated at 1 kHz (sampling rate for LFP recording).

Pearson's linear correlation coefficients between the band-limited power and the speed and acceleration of locomotion were computed with Matlab's *corr* function for a 1-s window moving in 0.1-s steps. To calculate speed and acceleration, video tracker data were linearly interpolated and smoothed with a Hanning window (2,001 samples wide).

Single units sorted with Offline Sorter (Plexon, Dallas, TX) and accepted on the bases of spike waveform overlays and autocorrelograms (Barnes et al. 2005) were included in the calculations of spike-LFP coherence. For these coherence calculations, spike trains were represented as impulse trains at the same sampling rate as the LFPs by placing a 1 at the sample closest to each spike time and a 0 at all other samples. Coarse-grained coherograms were computed around each of six task events between the spike train (n = 53–385) and each of five filtered LFP bands (1–5, 7–11, 11–14, 14–22, and 25–50 Hz). The distribution of maximum coherence magnitudes over all time–frequency points and all LFP channels was compared with the distribution computed after shuffling the order of the trials for the spike data to detect significant non-shuffled coherence at $P < 0.05$.

Tetrode tracks and microlesions marking the final tetrode position were identified in sections of formalin-fixed tissue cut at 24 μm and stained for Nissl substance.

Figures

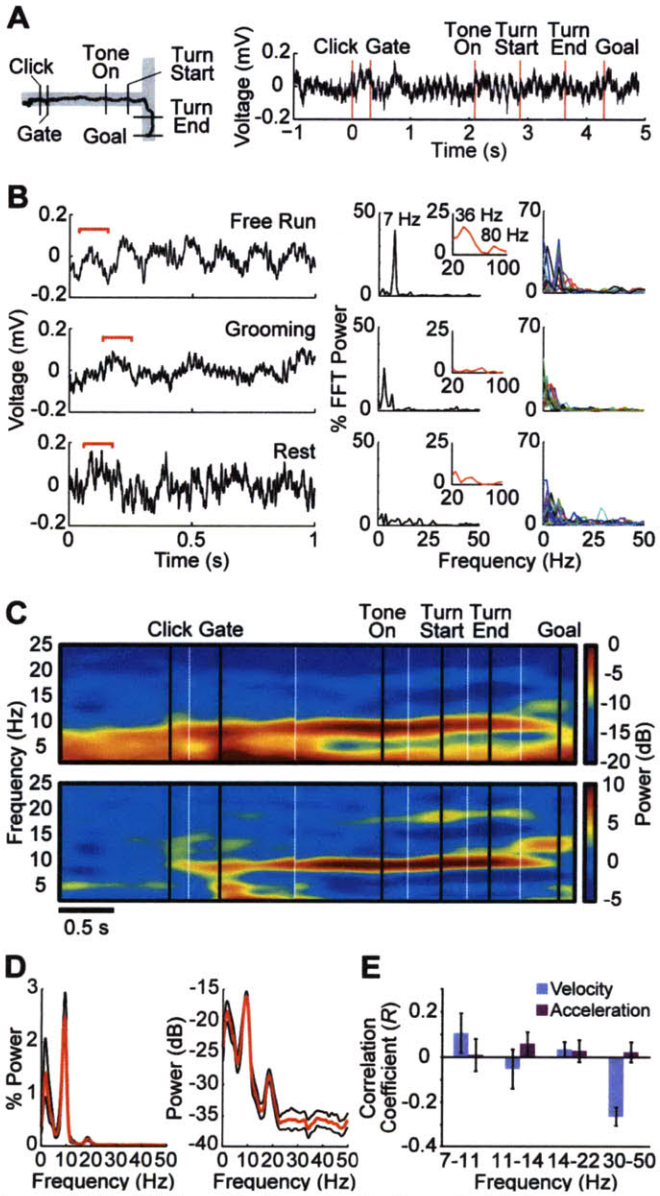

Figure 3.1.1. LFP oscillations in the dorsal striatum are modulated by behavioral conditions.
A: T-maze with overhead tracker data (left) and raw striatal LFP trace (right) recorded during a single representative trial (rat S36, acquisition day 10). B: Data from 1 sec periods of recording in the caudoputamen (CP) during spontaneous running (top, rat S19, medial CP, acquisition day 6), spontaneous grooming (middle, rat S31, medial CP, acquisition day 5), and quiet rest during which no movement was detected by video tracker (bottom, rat S19, medial CP, acquisition day 6). Raw voltage traces band-pass filtered at 1-475 Hz (left), Fast Fourier Transform (FFT) plots for this period (middle) and overlay plots of spectral traces for 15 one-second samples recorded within the same recording session (right). C: Spectrograms of session-averaged data for the entire task-time showing strong delta- and theta-band oscillations during turn approach, as well as beta-band activity, and peaks in high theta (11-14 Hz) activity near start and prior to goal-reaching. Task-time was reconstructed by abutting individual peri-event windows (bracketed by white vertical lines) with widths reflecting median inter-event intervals. Data are plotted as raw power (top) and as normalized power relative to pre-trial baseline activity (bottom) on pseudocolor log scales (right). Labeled task event-times are indicated by black vertical lines. D: Spectral estimates of oscillatory power during 0.75 s window after tone onset, plotted on normalized linear (left) and log (right) scales. Mean power (red) smoothed with a single taper (width = 1.8) is shown together with upper and lower 95% confidence limits (black).

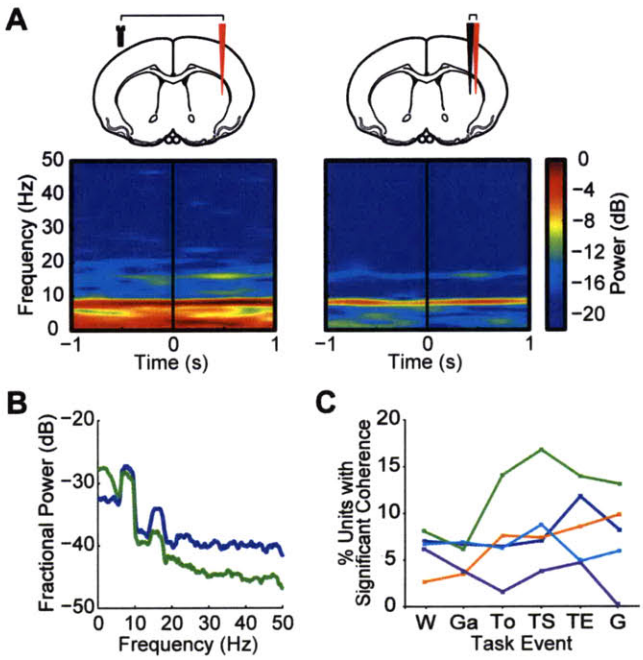

Figure 3.1.2. Striatal theta-band LFP activity is present under bipolar recording conditions with local referencing, and some striatal units exhibit theta-band spike rhythmicity suggesting that the oscillatory signal is recorded from local striatal current sources.

A: Schematic drawing of recording scheme (top) and spectrograms (bottom) of striatal unipolar recording with a ground screw reference (left) compared to bipolar recording with a local reference (right). Location of recording electrode is indicated in red, and the ground channel in black (screw, left and wedge, right). For spectrograms, data are aligned on turn onset (± 1 s) and are averaged across 10 trials of the same training session for the same striatal recording channel. B: Spectral estimates showing fractional power (percent of total power) for the unipolar (green) and bipolar (blue) recording conditions shown in A. C: Percentage of striatal units phase-locked to striatal LFP signals in 1-5 Hz (dark blue), 7-11 Hz (green), 11-14 Hz (orange), 14-22 Hz (light blue) and 25-50 Hz (purple) bands recorded in rat S36.

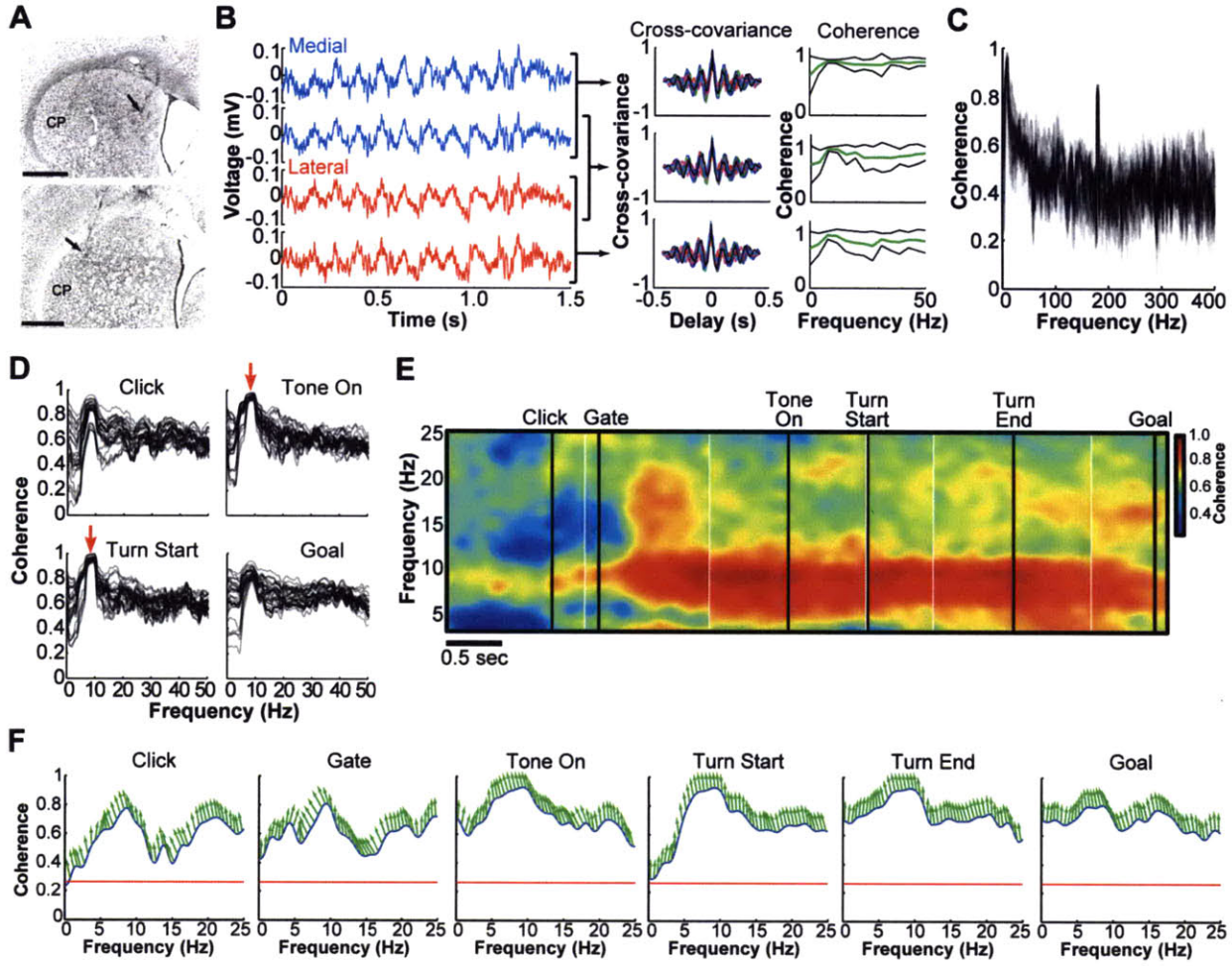

Figure 3.1.3. LFP oscillations recorded at medial and lateral sites in the caudoputamen are highly synchronous.

A: Photographs illustrating recording sites in the medial (top) and dorsolateral (bottom) caudoputamen. B: LFPs recorded simultaneously from tetrodes at two sites in the medial caudoputamen (blue) and at two sites in the lateral caudoputamen (red) during an episode of spontaneous locomotion in the T-maze (rat S19, acquisition day 6). Left, raw voltage traces (filtered at 1-475 Hz) recorded at each site. Right, cross-covariance and coherence plots for pairs of medial, medial-lateral, and lateral striatal sites, as indicated by brackets. The coherence plots show the mean (green) ± 1 standard deviation (black) for the session data. C: Average coherence plots for the same session shown in C, illustrating decreased variability in theta-band (7-11 Hz) coherence for the middle of the task (tone onset and turn start, indicated by red arrows) compared to coherence values for the beginning and end of task performance. Each plot illustrates data for the ± 0.5 s interval around each labeled task event, smoothed with 3 tapers (smoothing width = 2). Traces showing average coherence for 25 medial-lateral electrode pairs are overlaid. D: Coherogram reconstructed from 6 peri-event medial-lateral striatal coherograms, smoothed with 2 tapers (smoothing width = 3). Coherence was calculated for LFP signals recorded on two electrodes with remote references (top) and for the same signals converted to bipolar data by subtracting activity recorded on a nearby electrode (reference) from the activity on each electrode (bottom). Pseudocolor scales at right show the average coherence values. E: Average coherence (black) and coherence phase between medial and lateral striatal LFP signals (green arrows; up: 0 degrees, down: 180 degrees, left: 90 lead or 270 lag of medial striatum) measured during 0.75 s peri-event intervals around each task event. Red horizontal lines represent the threshold levels for significant coherence.

References

Aldridge, J. W., and Gilman, S. (1991). The temporal structure of spike trains in the primate basal ganglia: afferent regulation of bursting demonstrated with precentral cerebral cortical ablation. Brain Res *543*, 123-138.

Baker, S. N., Kilner, J. M., Pinches, E. M., and Lemon, R. N. (1999). The role of synchrony and oscillations in the motor output. Exp Brain Res *128*, 109-117.

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., and Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature *437*, 1158-1161.

Berke, J. D., Okatan, M., Skurski, J., and Eichenbaum, H. B. (2004). Oscillatory entrainment of striatal neurons in freely moving rats. Neuron *43*, 883-896.

Berretta, S., Parthasarathy, H. B., and Graybiel, A. M. (1997). Local release of GABAergic inhibition in the motor cortex induces immediate-early gene expression in indirect pathway neurons of the striatum. J Neurosci *17*, 4752-4763.

Bevan, M. D., Magill, P. J., Terman, D., Bolam, J. P., and Wilson, C. J. (2002). Move to the rhythm: oscillations in the subthalamic nucleus-external globus pallidus network. Trends Neurosci *25*, 525-531.

Boraud, T., Brown, P., Goldberg, J. A., Graybiel, A. M., and Magill, P. J. (2005). Oscillations in the basal ganglia: The good, the bad, and the unexpected. In The Basal Ganglia VIII, J. P. Bolam, C. A. Ingham, and P. J. Magill, eds. (New York, Springer Science and Business Media), pp. 3-24.

Brown, P., Oliviero, A., Mazzone, P., Insola, A., Tonali, P., and Di Lazzaro, V. (2001). Dopamine dependency of oscillations between subthalamic nucleus and pallidum in Parkinson's disease. J Neurosci *21*, 1033-1038.

Buhusi, C. V., and Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. Nat Rev Neurosci *6*, 755-765.

Buzsaki, G. (2005). Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. Hippocampus *15*, 827-840.

Courtemanche, R., Fujii, N., and Graybiel, A. M. (2003). Synchronous, focally modulated beta-band oscillations characterize local field potential activity in the striatum of awake behaving monkeys. J Neurosci *23*, 11741-11752.

DeCoteau, W. E., Thorn, C., Gibson, D. J., Courtemanche, R., Mitra, P., Kubota, Y., and Graybiel, A. M. (2007). Learning-related coordination of striatal and hippocampal theta rhythms during acquisition of a procedural maze task. Proc Natl Acad Sci U S A *104*, 5644-5649.

Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci *2*, 704-716.

Goldberg, J. A., Rokni, U., Boraud, T., Vaadia, E., and Bergman, H. (2004). Spike synchronization in the cortex/basal-ganglia networks of Parkinsonian primates reflects global dynamics of the local field potentials. J Neurosci *24*, 6003-6010.

Goto, Y., and O'Donnell, P. (2001). Synchronous activity in the hippocampus and nucleus accumbens in vivo. J Neurosci *21*, RC131.

Gray, C. M. (1994). Synchronous oscillations in neuronal systems: mechanisms and functions. J Comput Neurosci *1*, 11-38.

Hasselmo, M. E. (2005). What is the function of hippocampal theta rhythm?--Linking behavioral data to phasic properties of field potential and unit recording data. Hippocampus *15*, 936-949.

Jones, M. W., and Wilson, M. A. (2005). Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. PLoS Biol *3*, e402.

Laurent, G., Stopfer, M., Friedrich, R. W., Rabinovich, M. I., Volkovskii, A., and Abarbanel, H. D. (2001). Odor encoding as an active, dynamical process: experiments, computation, and theory. Annu Rev Neurosci *24*, 263-297.

Lebedev, M. A., and Nelson, R. J. (1999). Rhythmically firing neostriatal neurons in monkey: activity patterns during reaction-time hand movements. J Neurophysiol *82*, 1832-1842.

Levy, R., Hutchison, W. D., Lozano, A. M., and Dostrovsky, J. O. (2002). Synchronized neuronal discharge in the basal ganglia of parkinsonian patients is limited to oscillatory activity. J Neurosci *22*, 2855-2861.

Lisman, J. E. (1999). Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions. Neuron *22*, 233-242.

Magill, P. J., Pogosyan, A., Sharott, A., Csicsvari, J., Bolam, J. P., and Brown, P. (2006). Changes in functional connectivity within the rat striatopallidal axis during global brain activation in vivo. J Neurosci *26*, 6318-6329.

Magill, P. J., Sharott, A., Harnack, D., Kupsch, A., Meissner, W., and Brown, P. (2005). Coherent spike-wave oscillations in the cortex and subthalamic nucleus of the freely moving rat. Neuroscience *132*, 659-664.

Masimore, B., Schmitzer-Torbert, N. C., Kakalios, J., and Redish, A. D. (2005). Transient striatal gamma local field potentials signal movement initiation in rats. Neuroreport *16*, 2021-2024.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., and Moser, M. B. (2006). Path integration and the neural basis of the 'cognitive map'. Nat Rev Neurosci *7*, 663-678.

Mehta, M. R., Lee, A. K., and Wilson, M. A. (2002). Role of experience and oscillations in transforming a rate code into a temporal code. Nature *417*, 741-746.

Ni, Z., Bouali-Benazzouz, R., Gao, D., Benabid, A. L., and Benazzouz, A. (2000). Changes in the firing pattern of globus pallidus neurons after the degeneration of nigrostriatal pathway are mediated by the subthalamic nucleus in the rat. Eur J Neurosci *12*, 4338-4344.

O'Keefe, J., and Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. Hippocampus *3*, 317-330.

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., and Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. Nat Neurosci *5*, 805-811.

Plenz, D., and Kital, S. T. (1999). A basal ganglia pacemaker formed by the subthalamic nucleus and external globus pallidus. Nature *400*, 677-682.

Raz, A., Frechter-Mazar, V., Feingold, A., Abeles, M., Vaadia, E., and Bergman, H. (2001). Activity of pallidal and striatal tonically active neurons is correlated in mptp-treated monkeys but not in normal monkeys. J Neurosci *21*, RC128.

Ruskin, D. N., Bergstrom, D. A., Kaneoke, Y., Patel, B. N., Twery, M. J., and Walters, J. R. (1999). Multisecond oscillations in firing rate in the basal ganglia: robust modulation by dopamine receptor activation and anesthesia. J Neurophysiol *81*, 2046-2055.

Stern, E. A., Kincaid, A. E., and Wilson, C. J. (1997). Spontaneous subthreshold membrane potential fluctuations and action potential variability of rat corticostriatal and striatal neurons in vivo. J Neurophysiol *77*, 1697-1715.

Tepper, J. M., Koos, T., and Wilson, C. J. (2004). GABAergic microcircuits in the neostriatum. Trends Neurosci *27*, 662-669.

Terman, D., Rubin, J. E., Yew, A. C., and Wilson, C. J. (2002). Activity patterns in a model for the subthalamopallidal network of the basal ganglia. J Neurosci *22*, 2963-2976.

Vertes, R. P. (2005). Hippocampal theta rhythm: a tag for short-term memory. Hippocampus *15*, 923-935.

Wichmann, T., Kliem, M. A., and Soares, J. (2002). Slow oscillatory discharge in the primate basal ganglia. J Neurophysiol *87*, 1145-1148.

3.2. Learning-related coordination of striatal and hippocampal theta rhythms during acquisition of a procedural maze task

William E. DeCoteau,[1] Catherine Thorn,[2,3] Daniel J. Gibson,[2,4] Richard Courtemanche,[5] Partha Mitra,[6] Yasuo Kubota,[2,4] and Ann M. Graybiel[2,4,]

[1]Department of Psychology, St. Lawrence University, Canton, New York 13617; [2]McGovern Institute for Brain Research, [3]Department of Electrical Engineering and Computer Science, and [4]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; [5]Department of Exercise Science and Center for Studies in Behavioral Neurobiology, Concordia University, Montreal, Quebec H4B 1R6 Canada; and [6]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724

Acknowledgments
This work was funded by NIH/NIMH grants MH60379 and MH071744. We thank Uri Eden for his help with the statistical analyses, Patricia Harlan for her help with the histology, and Henry F. Hall, who is responsible for the photography.

Summary

The striatum and hippocampus are conventionally viewed as complementary learning and memory systems, with the hippocampus specialized for fact-based episodic memory and the striatum for procedural learning and memory. Here we directly tested whether these two systems exhibit independent or coordinated activity patterns during procedural learning. We trained rats on a conditional T-maze task requiring navigational and cue-based associative learning. We recorded local field potential (LFP) activity with tetrodes chronically implanted in the caudoputamen and CA1 field of the dorsal hippocampus during 9 to 13 days of training. We show that simultaneously recorded striatal and hippocampal theta rhythms are modulated differently as the rats learned to perform the T-maze task, but become highly coherent during the choice period of the maze runs in rats that successfully learned the task. Moreover, in the rats that acquired the task, the phase of the striatal-hippocampal theta coherence was modified towards a consistent antiphase relationship, and these changes occurred in proportion to the levels of learning achieved. We suggest that rhythmic oscillations, including theta-band activity, could influence not only neural processing in cortico-basal ganglia circuits, but also dynamic interactions between basal ganglia-based and hippocampus-based forebrain circuits during the acquisition and performance of learned behaviors. Experience-dependent changes in coordination of oscillatory activity across brain structures thus may parallel the well-known plasticity of spike activity that occurs as a function of experience.

3.2.1. Introduction

The striatum and the hippocampus are both forebrain structures implicated in the learning and memory of behavioral sequences, but behavioral sequences of different sorts. The striatum, as part of basal ganglia circuitry, is associated with learning sequences of actions that make up goal-directed procedures and habits (Graybiel, 1998; Graybiel, 2005; Hikosaka et al., 1999; Packard and Knowlton, 2002). The hippocampus and adjoining cortical structures are recognized as critical for encoding and storing sequences based on episodic, context-cued events (Dragoi and Buzsaki, 2006; Ergorul and Eichenbaum, 2006; Hasselmo, 2005; McNaughton et al., 2006). Lesion studies have dissociated striatum-dependent and hippocampus-dependent forms of learning and memory (DeCoteau and Kesner, 2000; Packard and McGaugh, 1996; White and McDonald, 2002), supporting the view that these systems work independently or even competitively. In humans, there is evidence that one system can substitute for another (Rauch et al., 2007). Other evidence, however, suggests that "hippocampal" deficits can follow damage in regions of the dorsal striatum interconnected with hippocampal/limbic circuits (Devan and White, 1999; Yin and Knowlton, 2004). Furthermore, part of the ventral striatum receives direct projections from the hippocampus.

Rhythmic activity in the theta range (ca. 7–14 Hz in the rodent) has been proposed to be crucial for the mnemonic coding in the hippocampus and related limbic structures. Pathways interconnecting the hippocampus and neocortex are thought to use these rhythms for transferring and coordinating neural representations in cortico-hippocampal circuits in relation to sequential spatial behavior (Buzsaki, 2005; Eichenbaum, 2000; Gervasoni et al., 2004; Hasselmo, 2005; Hyman et al., 2005; Jones and Wilson, 2005; Lisman, 1999; Mehta et al., 2002; O'Keefe and Recce, 1993; Siapas et al., 2005; Skaggs et al., 1996). Temporal spike precession relative to the hippocampal theta rhythms has further been suggested as a way to gain temporal resolution in sequence encoding in the hippocampus and in directly interconnected zones of the prefrontal cortex (Dragoi and Buzsaki, 2006; Hasselmo, 2005; Jones and Wilson, 2005; Siapas et al., 2005).

These findings are mainly based on experiments in rats navigating tracks and mazes. We tested for, and found, robust theta-band oscillations in the striatum of rats engaged in similar-navigation tasks (DeCoteau et al., 2007). Striatal theta rhythms were strongly modulated during performance of a procedural T-maze task, suggesting that such rhythmic activity could thus be important components of basal ganglia activity influencing the organization of the sequential behavioral performance.

These findings raised the intriguing possibility that the LFP oscillations in the striatum and the hippocampus might themselves be interrelated as animals learn and perform sequences of actions. To test this possibility, we recorded LFP and spike activity chronically both in the caudoputamen and in the dorsal hippocampus as rats were trained to perform this conditional T-maze task, and we measured the coherence between striatal and hippocampal theta-band LFP activity across successive weeks of training. Our findings suggest that changing patterns of striatal-hippocampal theta coherence are a cardinal feature of the neural activity that accompanies procedural learning.

3.2.2. Results

3.2.1.1. Striatal and hippocampal LFP oscillations are differentially modulated during T-maze performance.

We recorded simultaneously in the medial caudoputamen and in the CA1 field of the dorsal hippocampus (**Figure 3.2.1A**) in 6 rats as they performed the maze task illustrated in **Figure 3.2.1B**. We first analyzed the LFP activities recorded during the early phase of maze training, when the

animals first reached asymptotic running times (**Figure 3.2.1**, **3.2.2** and **3.2.4B**). There were marked contrasts between the striatal and hippocampal LFP rhythms recorded during the maze runs (**Figure 3.2.2A**). Hippocampal theta (7-11 Hz) rose gradually as the rats left the start zone and began to run, then peaked toward the tone-turn interval (the decision point in the task), and then gradually diminished as the goal was approached. By contrast, striatal theta-band oscillations reached an early peak near trial start, continued at nearly constant levels and then declined quite abruptly. In the high-theta 11–14 Hz band (considered as alpha in humans), the contrast between hippocampal and striatal oscillations was even clearer. Hippocampal power did not vary significantly, judged by 95% confidence limits, whereas the power of striatal 11-14 Hz activity had significant peaks at the start and end points of the runs. The power of the hippocampal rhythms in the 14–22 Hz (beta) band rose significantly toward the tone-turn decision point and then fell, but striatal 14–22 Hz power remained nearly constant. Finally, in the 30–50 Hz band, a sharp peak occurred in the hippocampal LFPs around the sounding of the warning click that indicated the beginning of each trial, but only a very small peak appeared then in the striatal LFPs.

We also tested whether the power of the striatal and hippocampal theta-band oscillations were differentially related to two measures of motor behavior. Hippocampal theta power was highly correlated with running speed (R = 0.52-0.81, $P < 0.001$; **Figure 3.2.2B**), but striatal theta-band power was much more weakly correlated with running speed (R = 0.03-0.48, P = 0.000-0.300; **Figure 3.2.2B**, (see also DeCoteau et al., 2007). Neither striatal nor hippocampal theta activity was strongly related to acceleration (R = 0.04 – 0.43, P = 0.000-0.200; **Figure 3.2.2C**). These results demonstrate that contrasting patterns of oscillatory LFP activity in the striatum and hippocampal CA1 field accompanied different segments of behavior in the maze task, with the power profiles of the oscillations different for the two structures in each of the frequency bands that we analyzed. The theta rhythms in the two regions also exhibited different relations to the rats' velocity profiles.

3.2.1.2. Striatal and hippocampal theta-band rhythms exhibit highly task-dependent patterns of coherence.

Throughout the training period, there were striking modulations of coherence between the striatal and hippocampal theta rhythms as the rats ran the maze (**Figure 3.2.3** and **3.2.4**). First, the magnitude of coherence was modulated during the task in the 4 rats that learned the task (9 to 13 sessions per rat, 43 total training sessions, **Figure 3.2.4A** and **3.2.4B**). The striatal and hippocampal theta rhythms in the rats exhibited individually varying levels of coherence (0.13 to 0.79) during the baseline period before the runs in these rats, but in each rat, the coherence values rose at the tone-turn period, when the rats were required to make a decision about the expected goal arm and then to execute this decision by its running direction (mean = 0.70, range = 0.27 – 0.96, **Figure 3.2.3**, **3.2.4C** and **3.2.6**. These higher levels of coherence were largely maintained up to the period before goal reaching (mean = 0.64, range = 0.09 – 0.91; **Figure 3.2.3**, **3.2.4*C*** and **3.2.6**. The increases in coherence magnitude from the baseline period to the tone period was significant in all 4 rats (P = 0.0000 – 0.0294, t test, **Figure 3.2.4E**), as were those from the baseline period to the goal period in 3 of 4 rats (P = 0.0000 – 0.0003). We observed such strong modulations of theta coherence for LFPs recorded in both medial and lateral striatal sites in which recordings were made simultaneously with recordings of hippocampal LFPs (**Figure 3.2.7**). These elevated coherence magnitude values were not accounted for by correlations with either speed or acceleration during the tone-post tone period (**Figure 3.2.4G-I** and **3.2.8**).

These patterns of coherence between the striatal and hippocampal theta rhythms were not modulated during the course of learning. We did not find a systematic change in the magnitude of coherence during the baseline, tone and goal periods in relation to stages of learning (R = -0.0087 – 0.2326, P = 0.2335 – 0.9650, **Figure 3.2.9A**, performance accuracy (R = 0.0801 – 0.1976, P =

0.2155 – 0.6185, **Figure 3.2.9B**, or running speed (R = -0.1673 – -0.0031, P = 0.2959 – 0.9849). Nor did we find a significant correlation between the pattern of coherence, with a rise during the decision part of the task, and any of these behavioral measures (R = -0.0889 – 0.0883, P = 0.5853 – 0.6552).

Two of the rats studied did not learn the maze task. In contrast to the results for the 4 rats that learned the maze task, we did not find comparable increases in coherence at the tone-turn period in the rats that failed to reach the criterion for behavioral acquisition (72.5% correct for at least 2 consecutive days). The magnitude of striatal-hippocampal theta coherence was similar for the 4 learners and these 2 non-learners during the baseline period (**Figure 3.2.4D**, P = 0.3695, ANOVA), but the levels of coherence were significantly lower in the non-learners during the tone and goal periods (P = 0.0000 and 0.0352, respectively, ANOVA, **Figure 3.2.4D**). Thus, there was significant difference in increase of coherence from the baseline to tone periods (learners: $n_{session}$ = 41, mean ± SEM = 0.264 ± 0.025, non-learners: $n_{session}$ = 19, mean = -0.053 ± 0.032, P < 0.0001, ANOVA, **Figure 3.2.4E**). The coherence profiles did not parallel either the velocity or the acceleration profiles of the learners and non-learners (see **Figure 3.2.8** and *supporting text*).

This difference held even during early training sessions, in which the 4 learners and the 2 non-learners did not differ in performance accuracy (P = 0.43, ANOVA) or run times (P = 0.08, ANOVA, **Figure 3.2.10**). There again was a significant increase in coherence values from baseline to tone for the learners but not for the non-learners (learners: $n_{session}$ = 18, mean = 0.299 ± 0.041; non-learners: $n_{session}$ = 9, mean = -0.134 ± 0.040, P < 0.0001, ANOVA; **Figure 3.2.4F** and **3.2.10**). Thus, even before the percent correct values for the learners and non-learners diverged, the 4 learners showed increases of coherence between the striatal and hippocampal theta rhythms at the decision point of the maze, whereas the 2 non-learners did not. This result raised the possibility that the coherence peak at the instruction tone period did not reflect the current accuracy of performance or running speed of the rats, but whether they would learn the task.

3.2.1.3. The phase relations of coherent striatal and hippocampal theta rhythms are modified as a function of learning.

Each of the 4 rats that learned the maze task had a characteristic mean coherence phase profile for the striatal-hippocampal theta band oscillations recorded during the trial runs (**Figure 3.2.5A** and **3.2.5B**). Overall, they had coherence phase angles near 180° (mean ± SEM = 171.1° ± 3.5), i.e., anti-phase. We analyzed group delays between striatal and hippocampal theta rhythms. There were small but statistically significant group delays in individual sessions, but these delays did not show a consistent pattern across days or rats, failing to provide evidence that one structure consistently led the other.

Significant shifts in the phase relationship between the striatal and hippocampal theta oscillations occurred during the maze runs. To examine these, we first analyzed the data recorded in the 4 learner rats during training sessions before and up to running time asymptote. We calculated the phase difference at ca. 9 Hz between the striatal and hippocampal rhythms at baseline, tone, and goal for those sessions in which coherence values were significant (P < 0.01, 1-tailed t test). We then compared the shift in coherence phase from the baseline period to the instruction tone, and from the instruction tone period to goal-reaching. For example, in the record shown in **Figure 3.2.3B** (rat S17, acquisition day 8), the coherence in the theta band became significant around the time of the instruction tone. From this time to the time of goal-reaching, there was a phase advance (precession) of striatal theta relative to hippocampal theta of ca. 45°.

We found such phase precession during the choice-to-goal period in all rats during this early period (**Figure 3.2.5C**). The precession values varied from 12° to 72°, corresponding to 3.7 to 22.2 ms. By contrast, during the first half of the task (the baseline-to-tone period), the phase differences between the striatal and hippocampal theta rhythms changed in the opposite direction: they recessed

by 1° to 81° degrees, corresponding to 0.3 to 25.0 ms (**Figure 3.2.5C**). Thus, phase differences between the theta rhythms in the two regions were modulated during the course of the maze runs in such a way that they tended to increase as the rats approached the choice period and then decreased as they ran to the goal.

The successive phase recession and precession of striatal theta relative to hippocampal theta that was so prominent early during training decreased as the rats learned the maze task. As a group, the learners showed larger changes in phase difference both for the baseline-to-tone and tone-to-goal periods early in training than late in training (R = -0.4570, P = 0.0492, **Figure 3.2.5D** for the baseline-to-tone recession; and R = -0.5308, P = 0.0063, **Figure 3.2.5G** for the tone-to-goal precession). The amount of precession of the striatal theta-band oscillations was inversely related to the percent correct performance of these rats (R = -0.5451, $P < 0.001$, **Figure 3.2.5H** for the tone-to-goal precession). The decreases in phase shift during training were not simply due to a shortening of the time available for phase angles to shift. The recession and precession decreased as the rats' running times decreased (**Figure 3.2.5F** and **3.2.5I**), but correlations between the amount of phase shift and inter-event duration were not significant (**Figure 3.2.9C** and **3.2.9D**; baseline-tone: R = -0.3233, P = 0.1072; tone-goal: R = 0.3016, P = 0.0738).

The 2 non-learners differed from the 4 learners in coherence phase angles between the striatal and hippocampal theta rhythms recorded during training. First, the non-learners had significantly smaller phase angles (mean ± SEM = 78.1° ± 5.7, $P < 0.0001$, ANOVA). Second, the amount of recession and precession in the non-learners was not correlated with the percent correct performance (baseline-to-tone recession: R = -0.4208, P = 0.1522; tone-to-goal precession: R = 0.3232, $P <$ 0.3055).

We observed, but did not analyze in detail, further complexity in the phase relations both within the theta band and at other frequencies. At any one time point in the maze run, the coherence phase angles between the striatal and hippocampal LFPs were clearly different at different frequencies, and there were multiple, frequency-dependent shifts in the coherence patterns between striatal and hippocampal theta as the rats ran the maze (**Figure 3.2.6**).

3.2.2. Discussion

Oscillatory modulation of neuronal activity has been implicated in a wide range of functions including sensory processing, network coordination, expectancy coding, sequence learning, episodic memory, and interval timing (Baker et al., 2006; Baker et al., 1999; Buhusi and Meck, 2005; Buzsaki, 2005; Engel et al., 2001; Fell et al., 2003; Gray, 1994; Hasselmo, 2005; Huxter et al., 2003; Laurent et al., 2001; Lisman, 1999; Mauk and Buonomano, 2004; Mehta et al., 2002; Nerad and Bilkey, 2005; Rizzuto et al., 2003; Senkowski et al., 2007; Siapas et al., 2005). We demonstrate here that during goal-directed behavior, striatal theta-band oscillations have structured, task-dependent and learning-dependent coherence relationships with the theta rhythms concurrently recorded in the CA1 field of the dorsal hippocampus. We suggest that oscillatory modulation of neuronal activity in the striatum could contribute to the interplay between basal ganglia-based circuits and concurrently active hippocampal circuits. The marked patterning of striatal-hippocampal theta coherence phase in rats that learned the task further suggests that adjustment of conjoint activity between the basal ganglia and hippocampus may be a critical part of the learning process as such goal-directed behaviors are acquired.

3.2.2.1. Striatal and hippocampal LFP oscillations have different task-dependent patterns of modulation but can become coherent during the maze runs.

By simultaneously recording LFP activity in the striatum and hippocampus, we directly compared the theta-band LFP rhythms in these two structures under identical behavioral conditions. Both striatal and hippocampal theta rhythms were maximal as the rats ran the maze, were reduced at rest, and fell at the end of the maze runs, but their magnitudes were modulated differently during the course of the maze runs. Remarkably, the task-modulation of the striatal and hippocampal LFP oscillations was different not only for theta rhythms, but also for each frequency subrange from delta to gamma.

Despite this different task-dependent modulation of the striatal and hippocampal LFP rhythms, they exhibited periods of high coherence as the rats performed the T-maze task. For any one frequency band, the levels of coherence varied across task-time, and the levels of coherence differed for different frequency bands.

3.2.2.2. The coherence phase between striatal and hippocampal theta-band LFP activity is modulated as a function of learning.

Two patterns in the coherence between striatal and hippocampal theta oscillations suggest that the relationship between these rhythms is modulated during learning. First, during the maze runs, striatal theta in the learners tended to recess and to precess relative to hippocampal theta. The coherence phase changes emphasized the decision point of the task. Striatal theta-band activity recessed (slowed) relative to hippocampal theta as the rats approached the instruction tone period, but then precessed (quickened) relative to hippocampal theta as the rats ran to the goal. The amounts of phase recession and phase precession in the learners were inversely related to success of their performance: the higher the percent correct and the shorter the run time, the smaller the adjustments of the phase angle between the theta rhythms in the striatum and hippocampus. Accordingly, the recession and the precession of striatal theta relative to hippocampal theta were largest early in training and decreased later as the animals learned.

A plausible interpretation of these findings is that early in training, when improvement in running speed and percent correct performance had not yet been achieved, the phase relationships between the striatal and hippocampal theta-band rhythms were adjusted relative to each other during the maze runs, reflecting exploration during the maze runs to achieve an optimal relationship at the most salient event (making a cue-based decision). But as performance accuracy and running speed increased as a result of learning (the exploitation phase), these adjustments became unnecessary because the phase relationship between the striatal and hippocampal theta rhythms was set near the start of the maze and was then maintained during the rest of the trial. In the two non-learners, the coherence phase relation between the striatal and hippocampal theta rhythms was highly variable, possibly reflecting continuous phase adjustment. Conceivably, these rats would have gone on to learn the task; our data show only that, up to the time recording failed, neither had reached the relatively steady anti-phase pattern for striatal-hippocampal coherence.

3.2.2.3. Modulation of striatal theta rhythms and their coherence with hippocampal theta rhythms peak during the choice period of the task.

For the rats that learned the task, the magnitude of coherence between the striatal and CA1 theta rhythms rose to a peak as they reached the instruction tone part of the task, and the coherence

remained high, or fell only slightly, as the rats made a decision about a turning direction and turned. The non-learner sample was small (n = 2), but neither showed this pattern. These findings raise the possibility that the increased coordination between the striatal and hippocampal rhythms during the decision period of the task was modulated by, or was required for, learning the instructional significance of the tone. The fact that this pattern was present for the learners even during early acquisition, favors the second of these alternatives. Neither during the decision period nor during other time-windows was the coherence well correlated with running speed or acceleration. These findings are consistent with the possibility that the coherence was modulated by cognitive processing, and that the coherence of striatal theta and hippocampal theta at the decision point of the maze may have been necessary for learning the task.

The magnitude and phase of coherence between the striatal and hippocampal theta oscillations varied even among the learners, and the coherence phase between two LFP signals could also fluctuate differently at different frequencies within the theta band. Other patterns of coherence held between the striatal and hippocampal LFP oscillations at different frequencies. This variability and the multiple coupling of the striatal and hippocampal theta rhythms suggest that the striatum and hippocampus are not locked in a single temporal relation; rather, their relationship is dynamic and highly task-dependent. Our findings raise the possibility that this dynamic relationship is shaped by, and may influence, the learning of goal-directed behaviors.

3.2.2.4. Network dynamics of striatal and hippocampal theta rhythms suggest experience-dependent plasticity of oscillatory activity during learning.

It is remarkable that the coherence between striatal and hippocampal theta rhythms reached levels as high as above 0.9, given that the caudoputamen and dorsal hippocampus are thought not to be directly connected. The high levels of coherence that we found thus suggest that a broader network of interconnected regions shares these dynamic patterns of coherence. Because we did not find a clear relation between the levels of coherence between striatal and hippocampal theta rhythms and velocity or acceleration at the choice periods of the task, but we did find a relation between both coherence magnitude and within-trial phase to learning and performance measures, we suggest that cognitive demands of the task influenced the relationship between the striatal and hippocampal rhythms. The fact that the choice period was the time of peak coherence in the learners, and was also the apparent reference point for the phase adjustments during learning, suggests that the coherence relationships of the striatal and hippocampal theta rhythms could be an integral part of mastering the maze task. If so, dynamic patterns of coherence across these brain structures may be a critical component of the decision and learning process of goal-directed behaviors. Task-selective, cross-structure relationships have been reported for the hippocampus and amygdala and for pairs of cortical areas (McNaughton et al., 2006; Moore et al., 2006; Pesaran et al., 2002; Seidenbecher et al., 2003). Our findings suggest that cross-structure coherence patterns are built through experience and may be required for learning, and that these changing coherence patterns may influence the degree of coordination with which the striatum and the hippocampus operate during goal-directed behaviors.

Phase precession of spike activity in the prefrontal cortex relative to hippocampal theta rhythms has been observed in rats running linear tracks and choice mazes as well as during foraging (Hyman et al., 2005; Jones and Wilson, 2005), and in the choice paradigm the prefrontal-hippocampal theta coherence is maximal in the decision period of the task (Jones and Wilson, 2005), as we show here for striatal-hippocampal coherence. The prefrontal cortex does directly project to the striatum and could influence striatal LFP rhythms, but the prefrontal inputs do not, according to available anatomical evidence, reach the full breadth of medial and lateral sites in the caudoputamen in which we found high striatal-hippocampal theta coherence. These findings again emphasize the possibility

that a distributed system of forebrain structures, ranging from striatum to the prefrontal cortex to the hippocampus, become coordinated in their rhythmic activities as goal-directed behaviors are learned and performed. We suggest that experience-dependent plasticity includes not only adjustment of firing rates, but also regulation of cross-structure oscillatory activity.

3.2.3. Methods

Seven adult male Sprague Dawley rats implanted with headstages carrying 12 tetrodes targeting either the dorsomedial striatum and the dorsal hippocampus (n = 6) or the dorsolateral striatum, the dorsomedial striatum and the dorsal hippocampus (n = 1) were trained for 9 to 13 days on a T-maze task that required a right or left turn at the choice point as instructed by 1 and 8 KHz tone cues indicating rewarded end goal baited with chocolate sprinkles. About 40 trials were given daily until rats made correct responses in >72.5% of trials during two consecutive sessions. In each training session, single unit and LPF activities were recorded with gain and filter settings appropriate for each recording (Barnes et al., 2005). Neuronal activity and movement of the rats in the maze (detected by video tracker and photobeam crossing) were monitored throughout the training session, and data were stored for later off-line analysis. Multitaper spectral analysis of LPF coherence and power was performed using Matlab, with window durations and taper parameters adjusted as needed to characterize the features studied (**Table 3.2.1**). Standard histology was conducted following the completion of study to verify recording sites. Detailed descriptions of methods are available in *Supporting Text.*

3.2.4. Supporting text

3.2.1.1. Supporting results

Changes in coherence magnitude are not related to changes in running velocity or acceleration. Coherence rose significantly from gate opening to just before the instruction tone for each of the 4 rats that learned the task, but not for the 2 rats that did not learn the task (**Figure 3.2.4D-F**). One possibility to account for this result is that the learners' running velocities were higher than those of the non-learners' during this time period. The average coherence magnitudes of the striatal and hippocampal theta rhythms did show a significant correlation with the corresponding averaged velocity values for the learners, but not for the non-learners, when values for all event-windows were included in the analysis (**Figure 3.2.8A** and **3.2.8B**; learners: $R = 0.5283$, $P < 0.0001$; non-learners: $R = 0.0632$, $P = 0.3326$). However, for the learners, during the event-windows in which the coherence was high (the periods around tone onset and turn onset), striatal-hippocampal theta coherence and velocity were not significantly correlated (**Figure 3.2.8C** and **3.2.8E**; tone: $R = -0.0195$, $P = 0.8626$; turn: $R = 0.0923$, $P = 0.4163$). There also was not a significant correlation between the theta-band coherence and acceleration for the learners during the tone period (**Figure 3.2.8C**, $R = 0.04$, $P = 0.2495$). During the turn period, a significant correlation appeared (**Figure 3.2.8F**, $R = 0.253$, $P < 0.05$). Finally, the profiles for mean coherence magnitude across task-time were different from the profiles for mean instantaneous velocity and mean instantaneous acceleration (**Figure 3.2.4G-I**).

3.2.1.1. Detailed methods

Chronic tetrode implantation. Experiments were conducted on 7 adult male Sprague Dawley rats maintained on a 12:12 hr light cycle (lights on 7 AM). All procedures met the approval of the Massachusetts Institute of Technology Committee on Animal Care and were in accordance with the National Research Council's Guide for the Care and Use of Laboratory Animals. Headstages carrying 12 independently movable tetrodes were implanted over small openings in the calvarium and underlying dura mater. They were secured with dental acrylic and anchor screws, one of which served as animal ground. Six rats were implanted with headstages having 6 tetrodes targeting the dorsomedial striatum and 6 tetrodes targeting the dorsal hippocampus. One additional rat received a three-site implant targeting the dorsolateral striatum, the dorsomedial striatum and the dorsal hippocampus (4 tetrodes each) and was used to compare striatal-hippocampal theta for medial and lateral recording sites in the striatum. Coordinates for the dorsolateral striatum were AP = +0.5 mm, ML = -3.5 mm relative to bregma; those for the medial striatum were AP = +1.7 mm, ML = -1.8 mm relative to bregma. Hippocampal recordings were made at AP = -3.3 mm and ML = -2.2 relative to bregma. Following surgery, tetrodes were lowered in small steps (< 80 μm per day) over a period of 5–7 days until they reached the estimated striatal (DV = 3.6–4.6 mm) and hippocampal (DV = 2.4–2.8 mm) targets, and until single unit and LFP signals were identified on the recording channels. Thereafter, the tetrodes were moved only when unit activity could not otherwise be recorded, and then the movements were in ca. 10 μm steps.

Behavioral procedures. Recordings were made as the rats learned and performed a procedural task (Barnes et al., 2005) in an elevated T-maze (height: 22 cm) consisting of a long track (127 x 7.5 cm) and two short arms (33 x 7.5 cm) made of black Plexiglas. The entire maze was surrounded by black walls (height: 41 cm). A gate 20 cm from the start point prevented the rat from leaving the start zone

during intertrial intervals. A removable Plexiglas plate (2.8 x 6 cm) with a circular well (diameter: 2.5 cm) was fixed with magnets at the goal end of each choice arm for delivery of the reward (chocolate sprinkles). An audio speaker for presentation of auditory stimuli was located behind the choice point of the maze. All recording sessions were carried out under dim red illumination.

In the T-maze task, the rats were presented with auditory cues (1 or 8 KHz pure tones, 80 dB) instructing them to turn right or left at the choice point in order to receive reward at the goal (Barnes et al., 2005). Each trial began with a click sound that signaled the beginning of the trial. The start gate opened 200–400 ms after the warning cue, and the rat was free to move toward the goal. When the rat had traveled half-way to the turning point in the maze (ca. 50 cm from the start gate), one of the tones instructing correct turn directions was turned on and was left on until the rat reached the goal or the trial was aborted. The rats received chocolate flavored sprinkles (General Mills) upon reaching the instructed goal. Trials were terminated 0.5–1 s after goal reaching or after the rat, in non-completion error trials, failed to run to a goal. After each completed trial, the rat was guided back to the start location for the next trial, which began after an intertrial interval of 1–2 min. Up to 40 trials were given during daily training sessions that lasted 1–1.5 hrs (usually 5 days/week). Learning criterion was reached when a rat reached 72.5% correct choices for 2 successive training sessions. Training sessions given before reaching the criterion were considered as acquisition sessions, and those given after this point were considered as overtraining sessions. Rats were trained for a total of 6 to 25 days. Chocolate sprinkles were scattered through the chamber 2–4 times per session to prevent odor cues at goals from dominating.

Data collection. All neuronal and behavioral data were acquired with a Cheetah data acquisition system (Neuralynx, Tucson, AZ). During each recording session, one or two 24-channel preamplifiers (utility gain: 1) were attached to the headstage, and neural signals were sent to 8 sets of 8-channel programmable amplifiers. To record single unit activity, signals were amplified (gain: 2000–10000) and band-pass filtered (600–6000 Hz). Spikes (signals above a preset voltage threshold) were sampled at 32 KHz per channel. Either a dedicated reference electrode or a tetrode channel without spike activity served as reference. To record LFP activity, amplified (gain: 1000) and filtered (1–475 Hz) signals were continuously sampled at 1 KHz. The animal ground or the external ground of the recording system was used as reference for LFP recording except in control experiments, in which a local (adjacent tetrode channel) ground was used to conduct bipolar recordings. Activity recorded on selected channels was monitored on-line with an oscilloscope and a speaker throughout the recording sessions, and all unit and LFP data were stored for off-line analysis.

During each recording period, the position of the rat was monitored by a video tracker (Cheetah, sampled at 60 Hz) that supplied video images from an overhead CCD camera. The tracker detected an LED light-source on the tetrode headstage, and the data were used to determine the onset and offset of locomotion and the beginning and end of turns. The video images were stored on tape for off-line inspection of behaviors. In addition, during performance of the T-maze task, a separate computer detected signals marking the breakage of photobeams (Med Associates, St. Albans, VT) placed along the maze to determine times of gate opening and goal reaching and to trigger the tone stimulus presented before the choice point. Transistor-transistor logic (TTL) pulses marking these events were sent to the Cheetah computer to be time-stamped, and time-stamps for these task events, the video tracker and neuronal data were synchronized to allow analysis of unit and LFP data relative to rats' behaviors.

LFP analysis. The frequency content and synchronization of LFPs were analyzed with open-source Chronux algorithms (http://chronux.org), in-house software, the Matlab Signal Processing Toolkit (MathWorks, Natick, MA), and other libraries (Courtemanche et al., 2003; Pesaran et al., 2002). Frequency spectra were estimated via the multitaper method (Pesaran et al., 2002), with the time

window, the number of tapers and time-bandwidth product (referred to as "width" in the text) of the tapers chosen to suit the size in time and frequency of the features being studied (**Table 3.2.1**). Power spectra were computed for each taper and each trial separately before averaging over tapers and trials. Spectrograms were constructed by dividing a raw waveform into a series of overlapping constant-width time windows with equally spaced centers, and displaying spectral power in each window as color in a series of vertical strips corresponding to the time window centers. For each time window analyzed, the DC component was removed, and the raw waveforms were padded at the end with zeros to produce a finer frequency grid.

Coherence between two simultaneously recorded signals was estimated as follows. First the FFTs of the tapered waveforms were computed individually for each taper and each trial. Cross-spectra were computed from the FFTs for each taper and trial and then were averaged over tapers and trials. Coherence was computed as $C = S_{12}/\mathrm{sqrt}(S_1 * S_2)$, where S_{12} denotes the averaged cross-spectrum and S_1 and S_2 denote the averaged power spectra of the two signals. Phase information was analyzed in time-and-frequency windows for which the coherence was statistically significant at the $P < 0.01$ level (t-test against the null hypothesis of independent random noise on both channels). Estimates of 95% confidence limits for coherence magnitude were computed by a jackknife procedure. For coherence phase, confidence limits were estimated by the formula:

$$\mathrm{mean} \pm 2 * \mathrm{sqrt}((1/K*Ntr)*(1/(C\hat{}2) - 1))$$

where "mean" is the mean coherence phase at a given frequency, K is the number of tapers, Ntr is the number of trials, and C is the magnitude of the coherence at that frequency. When the data in a fixed-width time window near the beginning or end of trial were aggregated over trials, and the window had to be truncated to accommodate the shortest trial, excessively brief trials were eliminated using an algorithm that maximized the total duration of all data going into the average.

Band-limited spectral power around trial events was computed as follows. First, a single-taper (Hamming window) unpadded spectrogram of each trial was computed for each electrode by moving a 0.75 s window in 0.05 s steps across trial-time for the lower frequency bands and by moving a 0.3 s window in 0.02 s steps for the 30–50 Hz band. The power components were then summed for the frequency interval between the upper and lower limits of each band. The resulting time series was then linearly interpolated at the sample rate used for the initial LFP recordings (1 KHz).

Scatter plots of the band-limited power versus the speed and acceleration of locomotion were computed by the procedure described above for calculating band-limited power, except that the spectrogram window was 1 s wide and was moved in 0.1 s steps. Speed and acceleration were calculated from video tracker data that were linearly interpolated at the original LFP sample rate and smoothed using a Hanning window 2001 samples wide. The acceleration trace was smoothed again with the 2001-sample Hanning window before further analysis. Finally, every one-hundredth sample was selected for plotting in the scatter plot, and Pearson's linear correlation coefficient and two-tailed statistical significance levels were computed with Matlab's corr function.

TABLE 3.2.1. *Characteristics of Taper Sets Used in the Figures*

Window Width (s)	Time-bandwidth Product	Number of Tapers	Half-Power Bandwidth (Hz)	Power at First Side Lobe Relative to Center Lobe (dB)
0.75	1.8	1	1.8	-40
0.75	1.8	2	3.2	-26
0.75	3	2	4	-53
1	2	3	3.3	-20

Conversion between phase and time. The equations we used for converting between time in milliseconds and phase in degrees were:

$$t(ms) = \frac{1}{f(kHz)} \bullet \frac{\mathrm{mod}(\phi(degrees),360)}{360}$$

$$\phi(degrees) = \mathrm{mod}(t(ms) \bullet f(kHz) \bullet 360,360)$$

Two key features of these equations are that the computation depends on the frequency of the oscillation whose phase is being measured, and that one can freely add or subtract any integral multiple of 360 degrees to any phase value without changing its meaning. We have observed the convention $0 \leq \Phi < 360$.

Histology. At the end of the recording sessions, rats were deeply anaesthetized with Nembutal (150 mg/kg) and were perfused transcardially with 4% paraformaldehyde in 0.1 M $NaKPO_4$ buffer. Frozen 24 μm-thick sections stained for Nissl substance were examined to locate tetrodes tracks. For some animals, the final positions of the tetrodes were marked by electrolytic lesions made 2 days before perfusion (25 μA, 10 s).

Figures

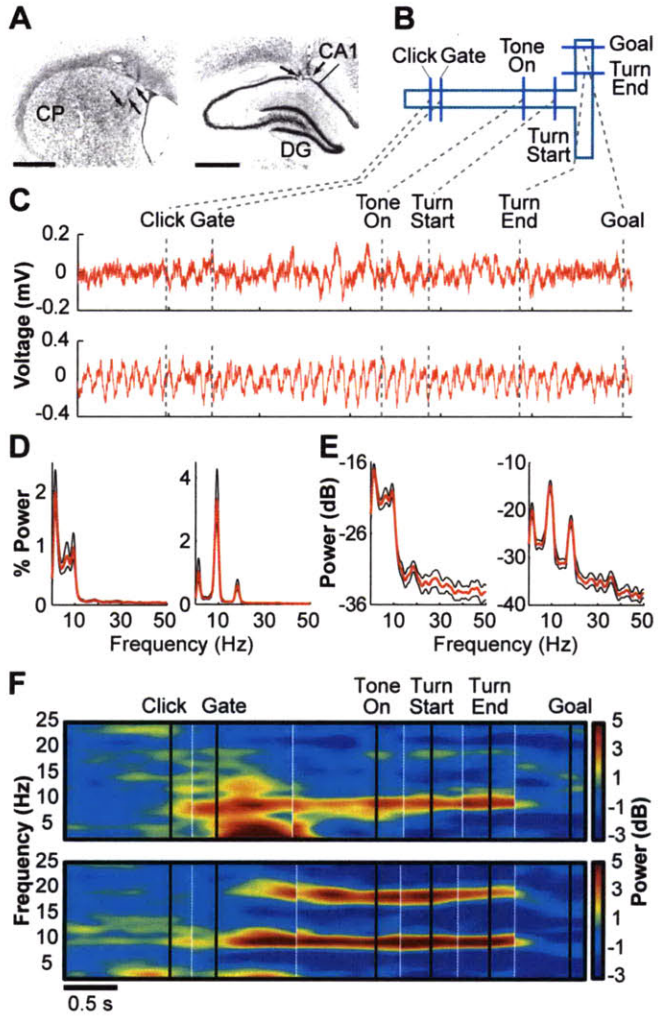

Figure 3.2.1. Simultaneously recorded LFP oscillations in the caudoputamen and the CA1 field of the dorsal hippocampus exhibit distinguishable task-related modulation during instructed running in a T-maze task. (A) Nissl-stained transverse sections illustrating, at arrows, the tracks of tetrodes in the medial caudoputamen (left) and the CA1 pyramidal cell layer (right). Scale bars represent 1 mm. CP: caudoputamen. CA1: hippocampal CA1 field. DG: dentate gyrus. (B) T-maze with task events. (C) Raw striatal LFP trace recorded during a single representative trial. (D and E) Mean power (red) with 95% confidence limits (black) of LFP activity in the striatum (left) and hippocampus (right) during a 0.75 s epoch after tone onset, plotted on linear (D) and log (E) scales. Data were averaged across values for 3 rats (S23, acq. session 7, S31, acq. session 5, S36, acq. session 10) during the session in which each reached running-time asymptote. (F) Reconstructed spectrograms of LFP activity in the medial striatum (top) and in the dorsal hippocampus (bottom) averaged for data from the 3 rats at their running-time asymptotes, as in D. The task-time was reconstructed by abutting individual peri-event windows (bracketed by white vertical lines) with widths reflecting median inter-event intervals. Data are plotted as normalized power relative to pre-trial baseline activity on a pseudocolor log scales (right). Labeled task event-times are indicated by black vertical lines.

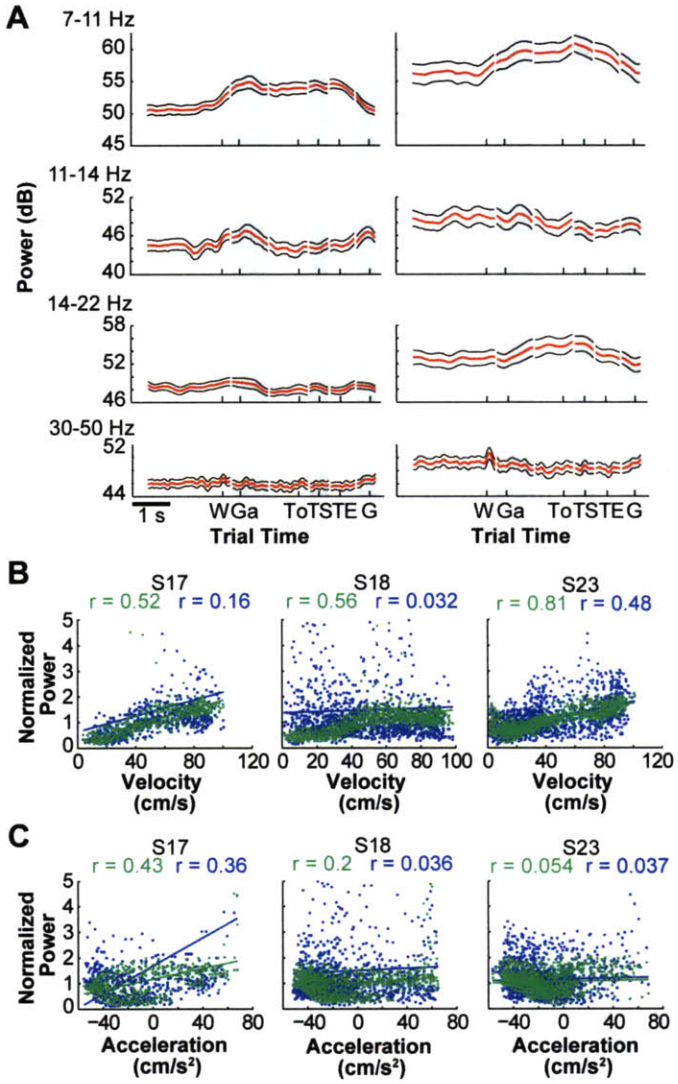

Figure 3.2.2. LFP oscillations in the striatum and the hippocampus exhibit different task-dependent modulation.

(A) Average spectral power in 4 frequency bands of LFPs recorded in the medial striatum (left) and the hippocampus (right) averaged across values for the 3 rats (S17, S31 and S36). Black lines indicate upper and lower 95% confidence limits. Alternating white and shaded zones indicate time-windows around task events (W: warning click, Ga: gate opening, To: instruction tone onset, TS: turn start, TE: turn end, and G: goal reaching). (B and C) Correlations of broad-band theta power (5-12 Hz) in the medial striatum (dark blue) and hippocampus (green) with movement velocity (B) and acceleration (C) of 3 individual rats (left: S17 acq. session 8, middle: S18 acq. session 6, right: S23 acq. session 7) sampled at 101 ms intervals during the 2.5 s before and 0.5 s after goal reaching in each trial. Each dot represents one such sample. Power is normalized to the median of all points within each recording site and session.

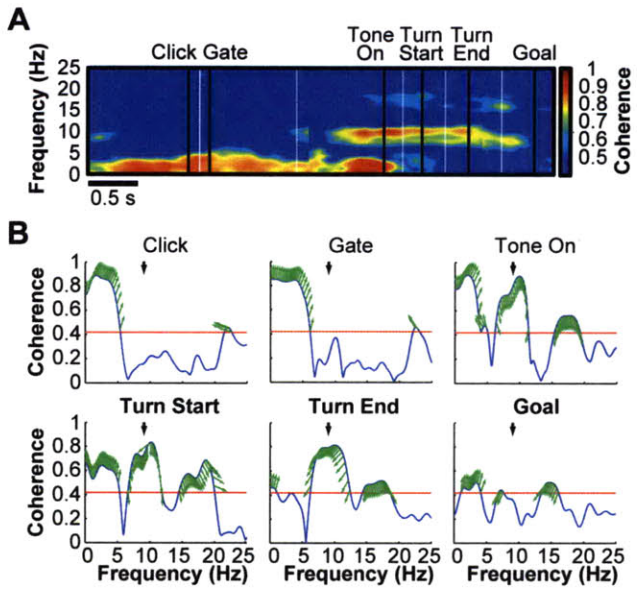

Figure 3.2.3. The coherence between striatal and hippocampal theta-band LFP oscillations is the strongest at the decision period of the maze runs.

(A) Single-session average coherogram (S17, acq. session 8), assembled by abutting 6 peri-event striatal-hippocampal coherograms, smoothed with 2 tapers (width = 3). Window widths reflect median inter-event intervals. The average coherence values are indicated in pseudocolor according to scale at right. (B) Plots of session-averaged coherence magnitude (black lines) and phase (green arrows) showing the dynamics of the synchrony between the striatal and hippocampal signals. The phase angle of significantly coherent signals is indicated by the direction of green arrows (up: 0°, down: 180°, left: 90° lead or 270° lag of hippocampus relative to striatum). Horizontal red lines indicate the level of significant coherence. Black arrows mark 9 Hz.

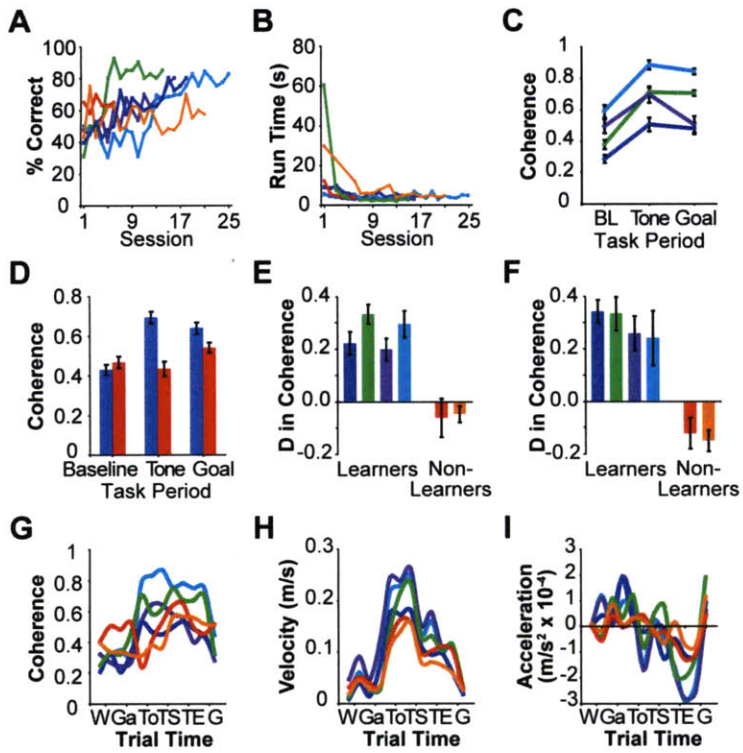

Figure 3.2.4. Coherence of striatal and hippocampal theta-band LFP oscillations increases in rats that successfully learn the T-maze task. (A and B) Performance accuracy
(A) and running times (B) of each rat during training on the procedural T-maze task. Four rats (S17: light blue, S18: purple, S23: dark blue, and S36: green) reached the acquisition criterion, but two rats (S31: red, S35: orange) did not. (C) Average magnitude of peak coherence in 7–11 Hz band during 0.75-s pre-trial baseline (BL), post-tone and pre-goal periods. Each line represents coherence values for a single rat that learned the task, averaged over all sessions for each rat (color coded as in A). Error bars indicate standard errors of the mean. (D) Average magnitude of coherence in theta-band oscillations during 0.75-s pre-trial baseline, tone and goal periods for the 4 learners (blue) and the 2 non-learners (red). (E and F) Changes in coherence during the post-tone period relative to pre-trial baseline. Data from individual rats are color-coded as in A. Significant increases in coherence magnitude of striatal-hippocampal theta were found for all learners but not for non-learners in averages across all sessions (E, ANOVA, $F = 54.42$, $P < 0.0001$) and also in averages across the first 5 available training sessions (F, ANOVA, $F = 45.21$, $P < 0.0001$), during which behavioral performance of the learners and non-learners was comparable (Fig. 10A and B). (G–I) Values for coherence magnitude at 9 Hz (G), running velocity (H) and acceleration (I) calculated for pre- and post-event periods for 6 task events (as described for Fig. 2A) averaged over all sessions for each rat (color-coded as in A).

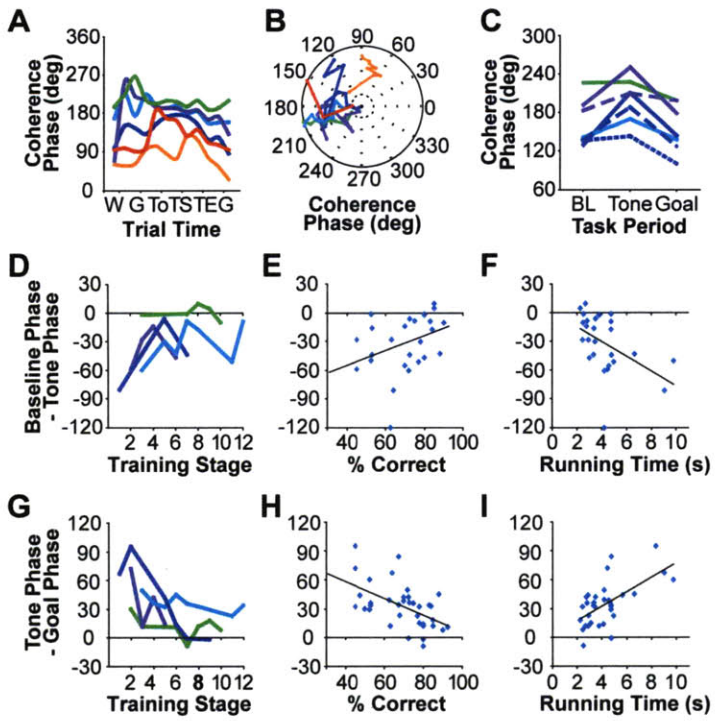

Figure 3.2.5. The phase of striatal-hippocampal theta coherence is modulated during learning.
(A) Average phase angles plotted for the 6 rats whose percent correct and running times are shown in the same color codes as in Fig. 4. Coherence phase was calculated by subtracting the striatal phase from the hippocampal phase and converting the angles to a 0–360° range. (B) Phase angles during the post-tone period for the first to last training sessions for individual rat are shown from center to periphery of the polar plots. (C) Coherence phase angles measured at three task periods during individual sessions up to asymptote of running speed for rats S17 (1 session), S18 (2 sessions), S23 (3 sessions) and S36 (1 session). (D–I) Amounts of change in coherence phase angles from pre-trial baseline period to post-tone period (D–F) and from post-tone period to pre-goal period (G–I) during T-maze training for the 4 learners. Changes in coherence phase angles from pre-trial baseline to post-tone period were significantly correlated with learning stage (D, $R = 0.46$, $P < 0.05$) and with running time (F, $R = -0.48$, $P < 0.02$) but not with percent correct response (E, $R = 0.35$, $P = 0.079$). Tone-to-goal changes in coherence phase angles were significantly correlated with all behavioral measures: stage (G, $R = -0.53$, $P < 0.01$), percent correct response (H, $R = -0.55$, $P < 0.001$), and running time (I, $R = 0.61$, $P < 0.001$).

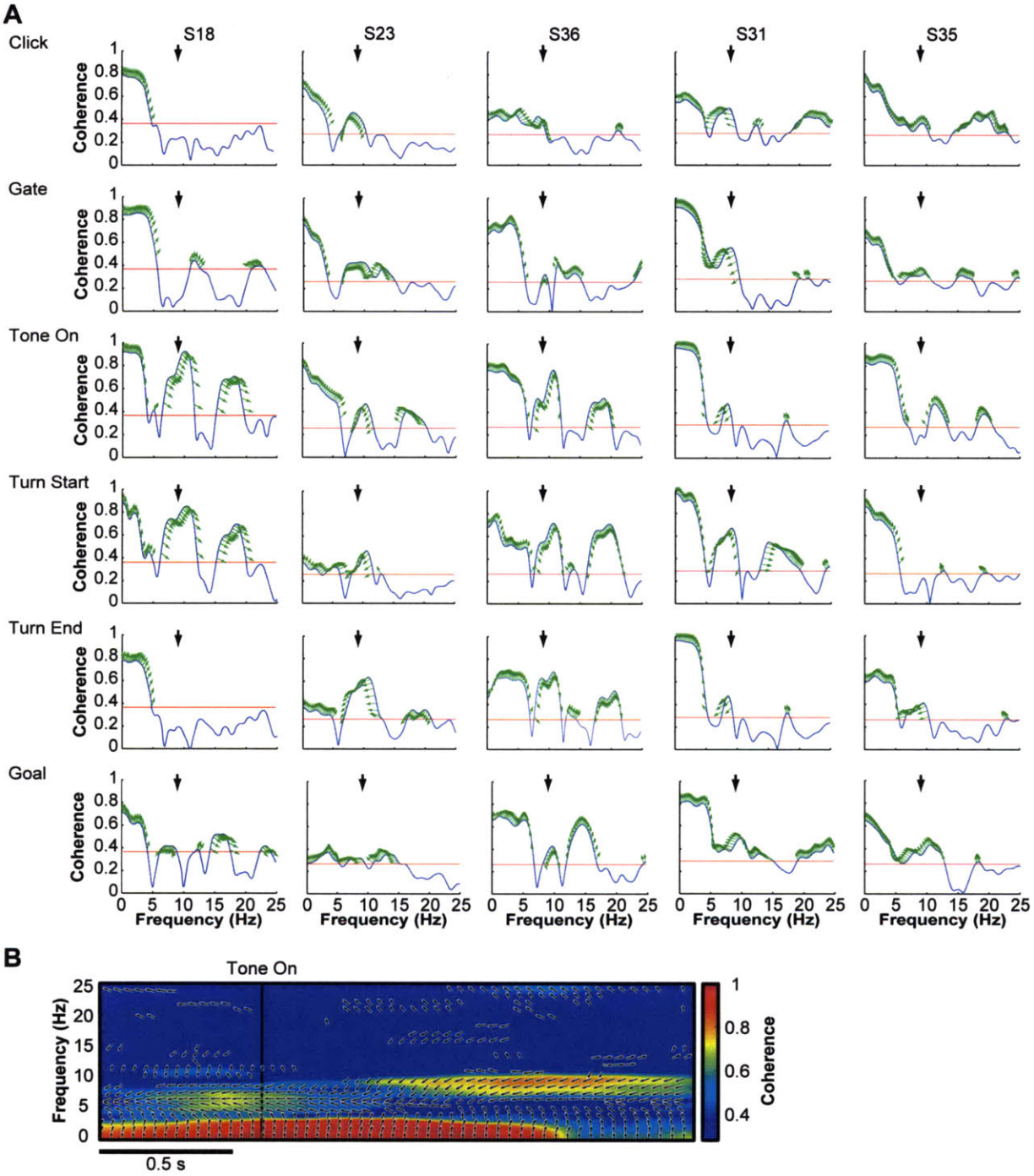

Figure 3.2.6. Average coherence between medial striatal and hippocampal LFPs during peri-event intervals. (A) Average coherence magnitude and phase in 0.75-s peri-event windows around each task event for three rats that learned the task (S18, S23, and S36) and two that did not (S31 and S35). Coherence phase is indicated by the angle of the green arrows (up, 0°; down, 180°; left, 90° lead or 270 lag of hippocampus relative to striatum). Red horizontal lines represent the level of significant coherence. Black arrows mark 9 Hz. Compare with Fig. 3. (B) Peri-event coherogram (S31, acquisition session 6) between the striatal and hippocampal theta, constructed by using a two-taper spectrum with smoothing width = 1.8. Arrows indicate phase angles (right, 0°; left, 180°; up, 90 lead or 270° lag of hippocampus relative to striatum). Note differences in phase angles for coherent signals at different frequencies.

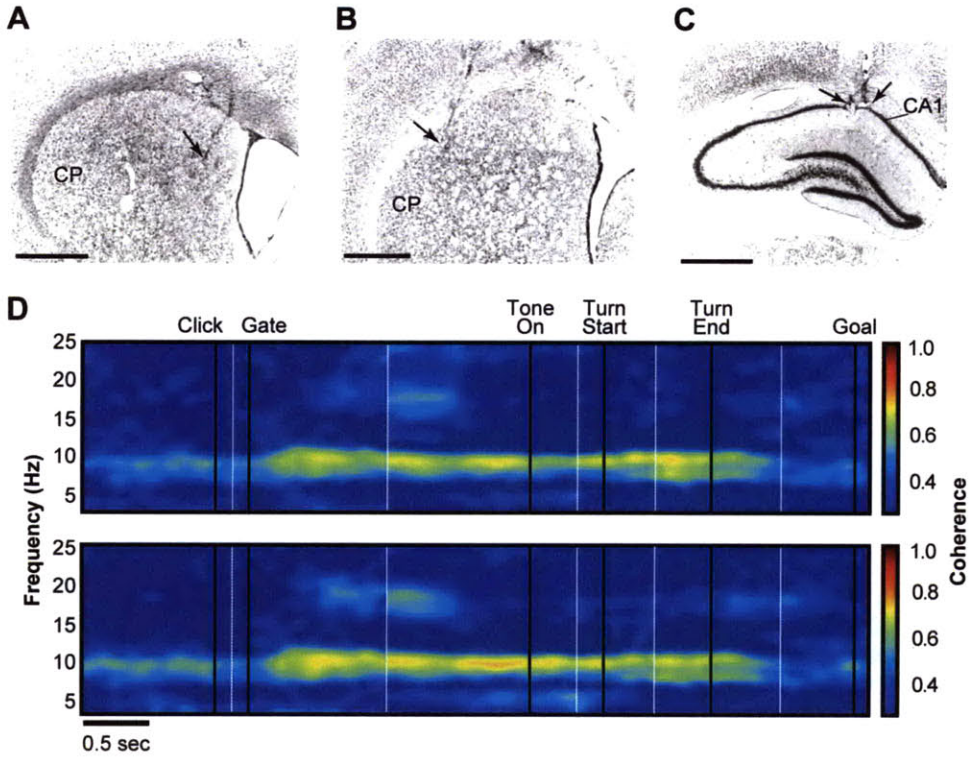

Figure 3.2.7. Coherence of LFP activity recorded in two areas of the striatum and in the hippocampus. (A-C) Photographs of Nissl-stained sections illustrating, at arrows, the tracks of tetrodes in the medial caudoputamen (A), the dorsolateral caudoputamen (B), and the CA1 of the dorsal hippocampus (C). Horizontal scale bars represent 1 mm. CP, caudoputamen; CA1, CA1 field of the hippocampus. (D) Pseudocontinuous single-session average coherograms showing coherence between LFPs recorded simultaneously from the medial striatum and the hippocampus (Upper) and the dorsalateral striatum and the hippocampus (Lower) for rat S25 (acquisition session 8).

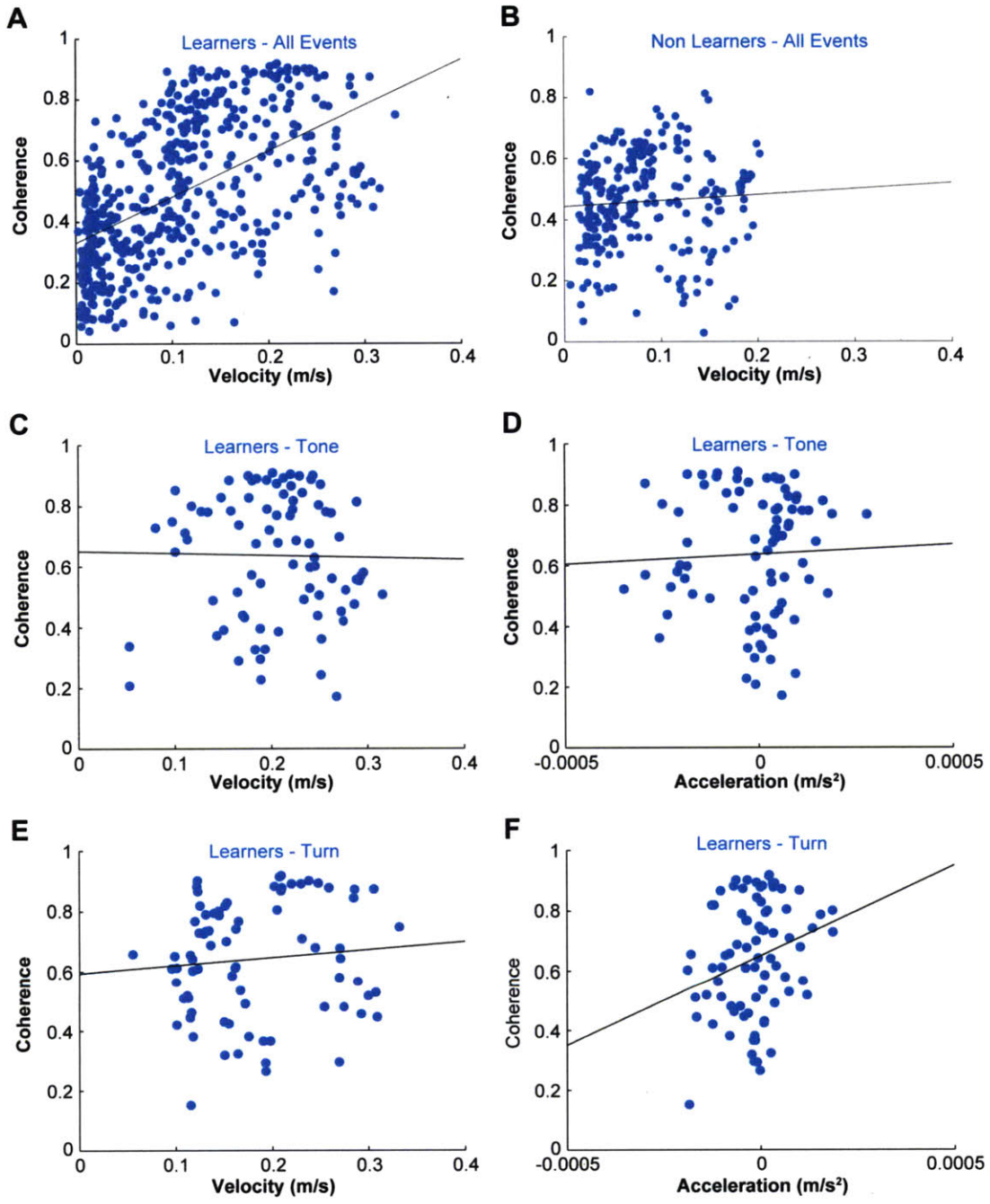

Figure 3.2.8. Striatal-hippocampal theta coherence magnitude is directly related to neither the velocity nor the acceleration of rats performing the T-maze task.

Significant correlations were found between amounts of coherence and velocity for four rats that learned the task ($R = 0.5283$, $P < 0.001$) (A) but not for two rats that did not ($R = 0.0632$, $P = 0.3326$) (B) based on all peri-event intervals (click, gate opening, tone onset, turn beginning, turn end, and goal reaching). However, for the learners, correlations calculated individually for the posttone and turn windows were not significant between coherence and velocity [at tone, $R = -0.0195$, $P = 0.8626$ (C); at turn, $R = 0.0923$, $P = 0.4163$ (E)] or between coherence and acceleration at tone ($R = 0.04$, $P = 0.2495$) (D). (F) Coherence and acceleration were significantly correlated at turn ($R = 0.253$, $P = 0.0235$).

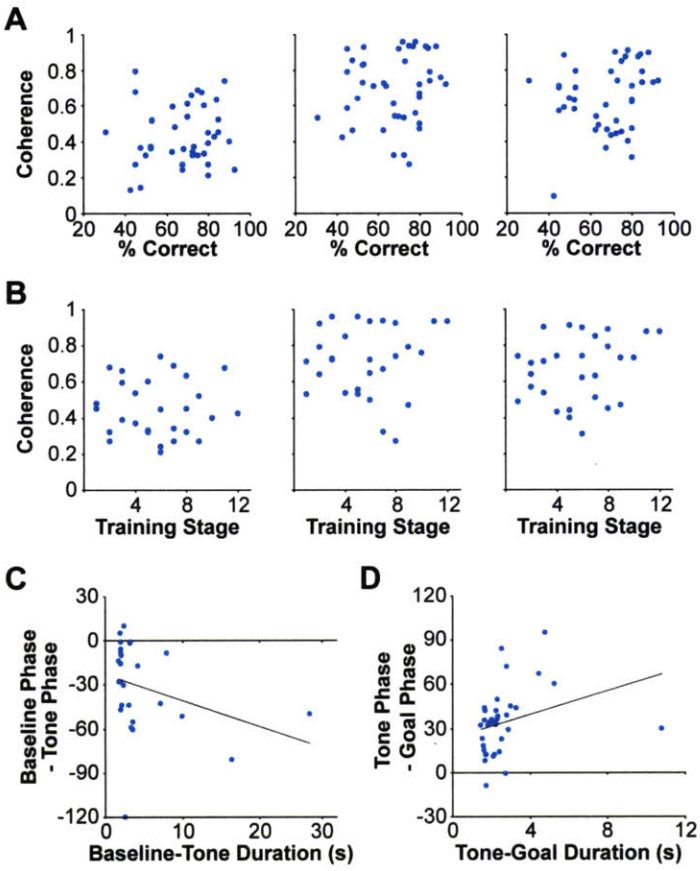

Figure 3.2.9. Nonsignificant correlations between striatal-hippocampal theta-band coherence and behavioral measures.

(A and B) Magnitude of coherence between the dorsomedial striatum and hippocampal CA1 area during the baseline (Left), tone (Center), and goal (Right) periods for the four rats that learned the T-maze task. Coherence magnitude was not correlated either with performance accuracy (P = 0.22 - 0.62) (A), with running time (P = 0.30 - 0.98) (data not shown) or with training stages (P = 0.23 - 0.97) (B). (C and D) Changes in coherence phase between striatal and hippocampal theta-band LFP signals and running-time duration. Amounts of changes in coherence phase were not significantly correlated with duration for the maze run intervals from baseline to tone onset (R = -0.3233, P = 0.1072) (C) or from tone onset to goal reaching (R = 0.3016, P = 0.0738) (D).

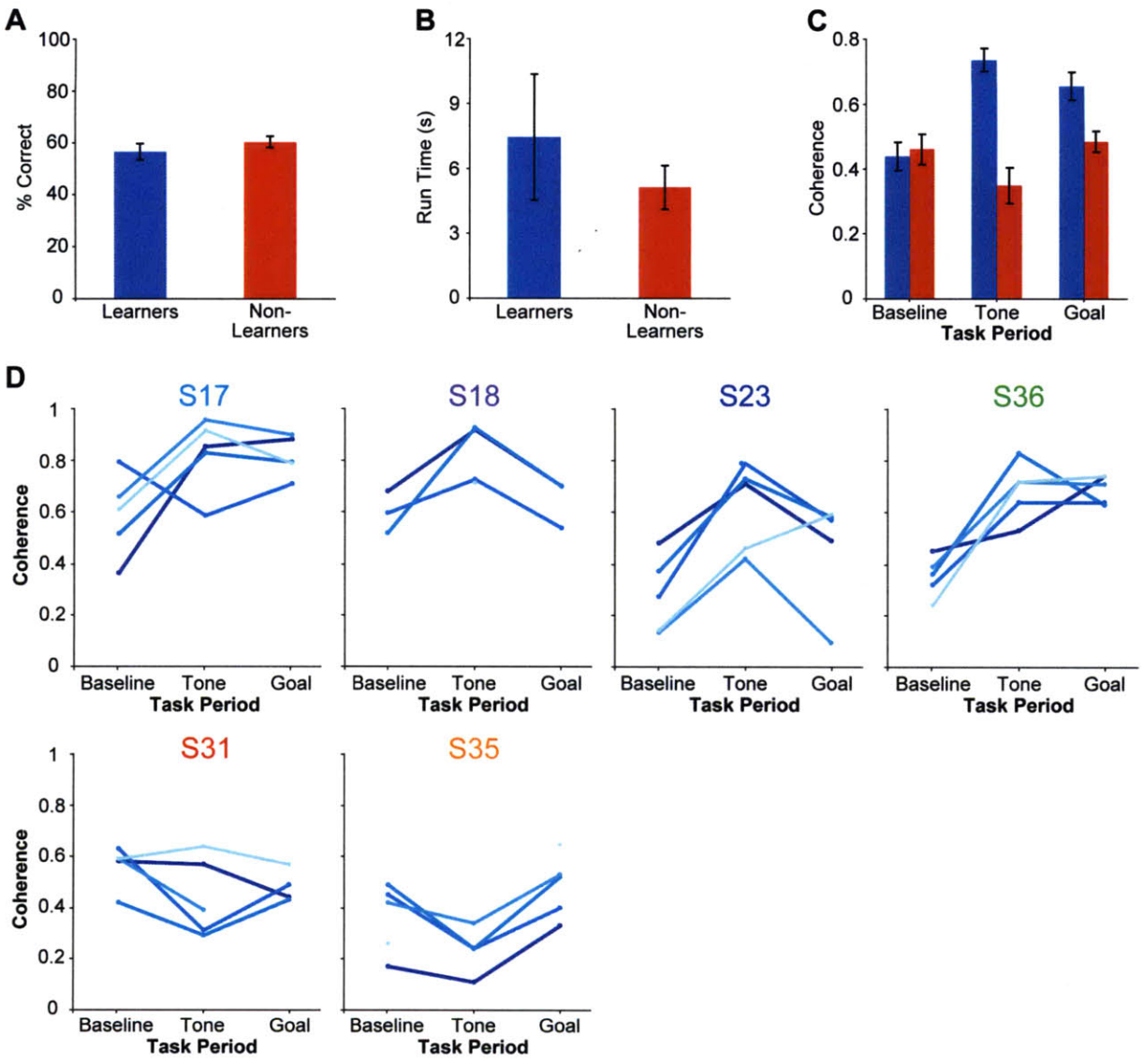

Figure 3.2.10. Coherence magnitude and behavioral data during first five sessions acceptable for analysis. (A and B) Average percentages of correct responses (A) and running times (B) for learners (blue, n = 4) and nonlearners (red, n = 2) during the first five sessions. (C) Coherence magnitude during the pretrial baseline, posttone, and pregoal periods (0.75 s) for learners and nonlearners. (D) Coherence magnitude during the three task periods for each of the first five sessions (from dark blue to light blue) plotted for each rat, as labeled. Note increases from the baseline to tone periods in learners (Upper), but not in nonlearners (Lower), even from the first session.

References

Baker, S. N., Chiu, M., and Fetz, E. E. (2006). Afferent encoding of central oscillations in the monkey arm. J Neurophysiol *95*, 3904-3910.

Baker, S. N., Kilner, J. M., Pinches, E. M., and Lemon, R. N. (1999). The role of synchrony and oscillations in the motor output. Exp Brain Res *128*, 109-117.

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., and Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature *437*, 1158-1161.

Buhusi, C. V., and Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. Nat Rev Neurosci *6*, 755-765.

Buzsaki, G. (2005). Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. Hippocampus *15*, 827-840.

Courtemanche, R., Fujii, N., and Graybiel, A. M. (2003). Synchronous, focally modulated beta-band oscillations characterize local field potential activity in the striatum of awake behaving monkeys. J Neurosci *23*, 11741-11752.

DeCoteau, W. E., and Kesner, R. P. (2000). A double dissociation between the rat hippocampus and medial caudoputamen in processing two forms of knowledge. Behav Neurosci *114*, 1096-1108.

DeCoteau, W. E., Thorn, C., Gibson, D. J., Courtemanche, R., Mitra, P., Kubota, Y., and Graybiel, A. M. (2007). Oscillations of local field potentials in the rat dorsal striatum during spontaneous and instructed behaviors. J Neurophysiol *97*, 3800-3805.

Devan, B. D., and White, N. M. (1999). Parallel information processing in the dorsal striatum: relation to hippocampal function. J Neurosci *19*, 2789-2798.

Dragoi, G., and Buzsaki, G. (2006). Temporal encoding of place sequences by hippocampal cell assemblies. Neuron *50*, 145-157.

Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. Nat Rev Neurosci *1*, 41-50.

Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci *2*, 704-716.

Ergorul, C., and Eichenbaum, H. (2006). Essential role of the hippocampal formation in rapid learning of higher-order sequential associations. J Neurosci *26*, 4111-4117.

Fell, J., Klaver, P., Elfadil, H., Schaller, C., Elger, C. E., and Fernandez, G. (2003). Rhinal-hippocampal theta coherence during declarative memory formation: interaction with gamma synchronization? Eur J Neurosci *17*, 1082-1088.

Gervasoni, D., Lin, S. C., Ribeiro, S., Soares, E. S., Pantoja, J., and Nicolelis, M. A. (2004). Global forebrain dynamics predict rat behavioral states and their transitions. J Neurosci *24*, 11137-11147.

Gray, C. M. (1994). Synchronous oscillations in neuronal systems: mechanisms and functions. J Comput Neurosci *1*, 11-38.

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. Neurobiol Learn Mem *70*, 119-136.

Graybiel, A. M. (2005). The basal ganglia: learning new tricks and loving it. Curr Opin Neurobiol *15*, 638-644.

Hasselmo, M. E. (2005). What is the function of hippocampal theta rhythm?--Linking behavioral data to phasic properties of field potential and unit recording data. Hippocampus *15*, 936-949.

Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, X., Nakamura, K., Miyachi, S., and Doya, K. (1999). Parallel neural networks for learning sequential procedures. Trends Neurosci *22*, 464-471.

Huxter, J., Burgess, N., and O'Keefe, J. (2003). Independent rate and temporal coding in hippocampal pyramidal cells. Nature *425*, 828-832.

Hyman, J. M., Zilli, E. A., Paley, A. M., and Hasselmo, M. E. (2005). Medial prefrontal cortex cells show dynamic modulation with the hippocampal theta rhythm dependent on behavior. Hippocampus *15*, 739-749.

Jones, M. W., and Wilson, M. A. (2005). Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. PLoS Biol *3*, e402.

Laurent, G., Stopfer, M., Friedrich, R. W., Rabinovich, M. I., Volkovskii, A., and Abarbanel, H. D. (2001). Odor encoding as an active, dynamical process: experiments, computation, and theory. Annu Rev Neurosci *24*, 263-297.

Lisman, J. E. (1999). Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions. Neuron *22*, 233-242.

Mauk, M. D., and Buonomano, D. V. (2004). The neural basis of temporal processing. Annu Rev Neurosci *27*, 307-340.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., and Moser, M. B. (2006). Path integration and the neural basis of the 'cognitive map'. Nat Rev Neurosci *7*, 663-678.

Mehta, M. R., Lee, A. K., and Wilson, M. A. (2002). Role of experience and oscillations in transforming a rate code into a temporal code. Nature *417*, 741-746.

Moore, R. A., Gale, A., Morris, P. H., and Forrester, D. (2006). Theta phase locking across the neocortex reflects cortico-hippocampal recursive communication during goal conflict resolution. Int J Psychophysiol *60*, 260-273.

Nerad, L., and Bilkey, D. K. (2005). Ten- to 12-Hz EEG oscillation in the rat hippocampus and rhinal cortex that is modulated by environmental familiarity. J Neurophysiol *93*, 1246-1254.

O'Keefe, J., and Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. Hippocampus *3*, 317-330.

Packard, M. G., and Knowlton, B. J. (2002). Learning and memory functions of the Basal Ganglia. Annu Rev Neurosci *25*, 563-593.

Packard, M. G., and McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiol Learn Mem *65*, 65-72.

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., and Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. Nat Neurosci *5*, 805-811.

Rauch, S. L., Wedig, M. M., Wright, C. I., Martis, B., McMullin, K. G., Shin, L. M., Cannistraro, P. A., and Wilhelm, S. (2007). Functional magnetic resonance imaging study of regional brain activation during implicit sequence learning in obsessive-compulsive disorder. Biol Psychiatry *61*, 330-336.

Rizzuto, D. S., Madsen, J. R., Bromfield, E. B., Schulze-Bonhage, A., Seelig, D., Aschenbrenner-Scheibe, R., and Kahana, M. J. (2003). Reset of human neocortical oscillations during a working memory task. Proc Natl Acad Sci U S A *100*, 7931-7936.

Seidenbecher, T., Laxmi, T. R., Stork, O., and Pape, H. C. (2003). Amygdalar and hippocampal theta rhythm synchronization during fear memory retrieval. Science *301*, 846-850.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., and Woldorff, M. G. (2007). Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. Neuropsychologia *45*, 561-571.

Siapas, A. G., Lubenov, E. V., and Wilson, M. A. (2005). Prefrontal phase locking to hippocampal theta oscillations. Neuron *46*, 141-151.

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., and Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. Hippocampus *6*, 149-172.

White, N. M., and McDonald, R. J. (2002). Multiple parallel memory systems in the brain of the rat. Neurobiol Learn Mem *77*, 125-184.

Yin, H. H., and Knowlton, B. J. (2004). Contributions of striatal subregions to place and response learning. Learn Mem *11*, 459-463.

4. Reinforcement learning approaches to basal ganglia function

Reinforcement learning has garnered much attention among basal ganglia researchers in recent years, in large part due to the discovery that the dopaminergic neurons of the VTA and SNc fire in a manner consistent with the computation of reward prediction errors. This error signal can then provide a "teaching signal" to downstream structures, including the prefrontal cortex and the basal ganglia. The question remains, however: what information is being learned, and how? Ideas from the field of artificial intelligence, in particular from reinforcement learning, have helped basal ganglia researchers frame these questions. In this chapter, a brief and general background on reinforcement learning is provided in Section 4.1, followed in Section 4.2 by a summary of the applications of some of these RL ideas to basal ganglia research. Finally, in Section 4.3, these ideas are extended to the experiments described in Chapter 2, and two hypotheses are suggested by which the dorsolateral and dorsomedial ensemble patterns observed during learning on the T-maze might result from two systems engaged in RL-based processes.

4.1. Reinforcement learning

For a complete introduction to the field of reinforcement learning, *Reinforcement Learning* by Sutton and Barto (1998) is the classic and highly recommended text. In the following section, we briefly summarize the basic concepts and those most relevant to basal ganglia research.

4.1.1. Introduction to reinforcement learning

The basic idea of reinforcement learning (RL) is that an agent interacts with its environment with the goal of maximizing its reward. Even this seemingly simple statement raises a number of questions. What is reward? What is meant by 'maximum'? What strategy or strategies should be used to accomplish this goal? Finally, how can an agent learn all of these things in a new environment? In this section, we unpack these concepts.

First, reward is a scalar function given by the agent's environment. At each point in time, a signal indicating the reward currently being received is provided by the environment to the agent. In biology experiments, for example, the delivery of food or water is often considered to be the reward that the animal is interested in obtaining more of. By learning to predict where, when, and/or how to behave in order to receive more food/water, the animal is solving a typical reinforcement learning problem. While the animal/agent can learn to perform actions that result in the receipt of more reward, it is important to realize that the reward is nonetheless external to the agent. The agent cannot manufacture a reward signal independent of its external environment - if it could, there would be no need for interacting with or learning about its world.

We could imagine several schemes for achieving maximum reward. For example, the agent could try to maximize the average reward over some time period, the total reward over the same period, or it could try to maximize its immediate reward. Each of these maximization schemes may result in the selection of a different behavioral strategy to attain them. Different schemes may be best in different environments, and thus the choice of which to use may be dependent on the parameters of the task to be performed.

Reinforcement learning formalizes these ideas and provides methods by which an agent may learn which actions will lead it to maximum reward. This formalization leads us to define certain terms. Sutton and Barto enumerate four components of any RL system: 1) a reward function, 2) a value function, 3) a policy, and 4) a model of the environment. The *reward function*, as we have already discussed, represents the goal of any reinforcement learner – defining what is good in the environment. The reward function is not alterable by the agent, and may be stochastic. The *value function* is the agent's estimate of how good an environmental state is in terms of generating future rewards; in other words, it is an estimate of the total reward that an agent can expect to receive in the future after having seen that state. This estimate must be acquired through the process of interacting with the environment, and may change over time. The *policy* defines the way in which the agent behaves when a given state is encountered, i.e. it maps the perceived states to the action to be performed. A policy may be learned and may change over time. Finally, a reinforcement learner may explicitly store a *model* of the environment, defined as the probabilities of transitioning between states. A number of RL algorithms do not explicitly learn or store a model of the environment, and there is thus a distinction between *model-based* and *model-free* RL. Models provide several benefits, but require additional memory and computational resources. They can be used to simulate experience, such that learning of value functions and policies progresses faster. Or, they can be used for *planning*, in which possible future situations are considered in determining a course of action.

Much of reinforcement learning is concerned with how to efficiently learn value functions, and how to select an optimal policy. Two central problems can be defined in these terms. The first is how to determine which states or actions led to the ultimate delivery of reward, as these may be separated by a long delay and several intervening steps. This is the "credit assignment problem" and relates to how value functions can be learned. The second is whether to perform actions that have the highest values, given that the value estimates may be inaccurate or unknown (either from a lack of experience, or because the environment has changed), or whether to branch out and try a new action that may result in higher reward. This is the "explore/exploit tradeoff" and relates to the issue of choosing a policy. Different reinforcement learning solutions address these two issues in different ways, and several general approaches are described in Section 4.1.2.

The reinforcement learning problem can be considered in the following generic terms. At time t, an agent encounters a state s_t, and must decide what action a_t to perform. Based on the amount of reward the agent receives at $t+1$, and the new state it encounters s_{t+1}, the agent determines whether the action performed led to an outcome that was better or worse than expected. Based on this feedback, the agent then updates its estimates of the value of state s_t and the probability that it will perform action a_t the next time s_t is encountered. The agent is thus tasked with perceiving the correct state s_t, acquiring an accurate estimate of the value of that state $V(s_t)$, using the estimate of $V(s_t)$ to select the best action a_t, evaluating the success of action a_t and using that evaluative information to improve its future performance. Throughout the following discussion, $V(s)$ is used to denote the agent's estimate of the value of state s. $V^*(s)$ is the optimal value function, which the agent is trying to approximate with $V(s)$. A state is denoted as s, and the state encountered at time t is denoted as s_t. We further assume that the accurate perception of state s is successfully accomplished somehow and focus on the remaining steps.

4.1.2. Basic approaches to solving RL problems

In this section, we consider some approaches that have been developed to solve reinforcement learning problems. In general, the solution to this class of problems involves an iterative process of updating a value function given a current policy, and then using the improved value function to improve the current policy. Three approaches are discussed: dynamic programming, Monte Carlo methods, and temporal difference learning. Each of these approaches has advantages and disadvantages. The first set of solutions, dynamic programming methods, use a full model of the external environment, including the transition probabilities between states, to update values of all states following feedback. Dynamic programming solutions can be computationally expensive, but generally do not require as much exploration to arrive at optimal value functions and policies. The second set of solutions, Monte Carlo methods, learn value estimates for each state from experience with those states, following feedback given at the end of an episode. These have the advantage that a full model is not required, and state values are not tied to other state values. These methods require extensive exploration and may suffer if reward is very delayed from the actions which produced it. The third class of solutions, temporal difference methods, also learn from experience with the environment, but also update the value of the current state according to the values learned for other states - thus combining the advantages of both dynamic programming and Monte Carlo methods. This class of algorithms has been extended to bridge long temporal delays, and has garnered particular interest among basal ganglia researchers.

4.1.2.1. Dynamic programming

As described above, RL algorithms involve an iterative process of improving the value estimates given the current policy (policy evaluation), combined with the process of improving the policy based on the updated value functions (policy improvement). This process is termed *generalized policy iteration.*

In the simplest case, the agent observes the state at time t and decides on an action based on the current policy, π. After observing the reward at the next time step, r_{t+1}, and the next state, s_{t+1}, the agent updates the value of s_{t+1} and the values of all of the other states are subsequently updated according to their known or assumed transition probabilities and the current state values.

If we define S to be the set of all states in the discretized state space, and A to be the set of all possible actions, then for all states:

$$V(s) = \sum_{a \in A} \left[\pi(s,a) \sum_{s' \in S} P^{a}_{ss'} \left(R^{a}_{ss'} + \gamma V(s') \right) \right]$$

Above, $\pi(s,a)$ denotes the probability of taking action a in state s, $P^{a}_{ss'}$ is the probability that performing action a in state s will result in next state s'. $R^{a}_{ss'}$ is the reward obtained in transitioning from s to s', and $V(s')$ is the value of the future state s'. The value of state s is then the expected value of transitioning out of that state, considering the probability of landing in all possible next states and the values of those states. The parameter γ is a *discount factor*, which can take a value between 0 and 1, and is used to discount the values of potential states encountered in the more distant future compared to those encountered sooner.

We can imagine that the agent continues to select actions until it finds a *termination state* ending the trial or episode, receives a reward for transitioning to that state, and updates the values of all the

states one final time. With this updated value function, we now find that certain states have better or worse actions that can be chosen - those that move the agent closer or farther from the goal state. To improve its performance on the next trial, the agent should update its policy to increase the probability of performing the action at each state that will take it to the highest-valued next state.

The key point here is that by exploring its environment, an agent can estimate the value of each state it visits and determine an appropriate action to take from each state in order to maximize its reward over time. Dynamic programming methods provide intuitive procedures for estimating state values and determining a policy. Further, they are able to leverage full knowledge of the state transition probabilities to update the values of all states at each time step – not just the value of the current state. This *model-based* approach has the advantage of reducing the amount of exploration required to achieve optimal performance (though if the model must be learned in addition to the policy, this may no longer be true). It comes, however, at the cost of requiring significant memory and computational processing capabilities to store the model and value functions and perform the update at each time step. We consider methods that reduce these demands in the next section.

4.1.2.2. Monte Carlo Methods

Rather than assuming a full model of the state transition probabilities is known, Monte Carlo methods acquire estimates of state-values by experiencing different states and keeping track of their overall probability of resulting in future reward. These model-free methods are generally less computationally intense than dynamic programming methods. These methods, however, come at the cost of required exploration, which ensures that the value estimates, $V(s)$, converge to the optimal value function $V^*(s)$, and that the agent does not settle too soon into a suboptimal solution. In the simple idealized Monte Carlo case, several episodes, or trials, are experienced, each of which consists of a finite number of state transitions and ends when a termination state is encountered. State-values are computed by averaging the rewards obtained after visiting state s. Unlike the dynamic programming methods, the value estimate for a state does not depend on the values computed for any other state, and is learned only after having experienced that state. A typical algorithm is shown in **Box 4.1**.

For each episode k:

Choose actions according to policy π until a termination state is encountered.

For each state s visited at least once in episode k, update the number of trials, n_s, in which that state was observed:

$$n_s = n_s + 1$$

Update the cumulative reward tally associated with each s in k:

$$R_s = R_s + R^k$$

where R_s denotes the cumulative reward associated with state s, and R^k denotes the reward delivered in the current episode k.

Update the value of all states observed in episode k:

$$V(s) = R_s / n_s$$

Box 4.1 *Monte Carlo algorithm for estimating state-values*

Because state value estimates are acquired through experience, the problem arises of how much to exploit the previous experience acquired, versus how much to explore unknown or underexplored options. Only policies that continue to explore are guaranteed to converge on the optimal value functions, though exploratory policies can only approach (not equal) optimal performance. Two policies are commonly used in RL algorithms that select actions based on their relative values, but also guarantee continued exploration. These are ε-greedy and softmax policies. The ε-greedy policy behaves greedily most of the time (choosing the action associated with the highest value), but occasionally, with probability ε, chooses a random action with uniform probability over all available actions. Whereas the ε-greedy policy chooses among all policies with equal probability when not behaving greedily, the softmax policy distributes the selection probabilities according to the action values. The most common softmax method uses a Boltzmann distribution, which provides a mathematically convenient way in which to distribute the selection probabilities:

$$p(s,a) = \frac{e^{Q_t(s,a)/\tau}}{\sum_{b=1}^{n} e^{Q_t(s,b)/\tau}}$$

Above, $p(s,a)$ is the probability of performing action a when in state s. The parameter τ is the temperature parameter, and determines to what degree differently-valued actions are selected with different frequencies. A lower τ results in the highest-valued actions assigned higher probabilities of selection; a higher τ results in a more uniform probability distribution. We have also introduced a *Q function*, or action-value function, which denotes the value of performing action a in state s. In Q-learning, the value of each state-action pair is acquired instead of the state values $V(s)$. Q-values are acquired in a similar manner to that described above for acquiring state-value functions (**Box 4.2**).

For each episode k:

Choose actions according to policy π until a termination state is encountered.

For each state-action pair (s,a) observed at least once in episode k, update the number of trials, $n_{(s,a)}$, in which that pair was observed:

$$n_{(s,a)} = n_{(s,a)} + 1$$

Update the cumulative reward tally associated with each (s,a) in k:

$$R_{(s,a)} = R_{(s,a)} + R^k$$

where $R_{(s,a)}$ denotes the cumulative reward associated with each state-action pair (s,a), and R^k denotes the reward delivered in the current episode k.

Update the value of all states-action pairs observed in episode k:

$$Q(s,a) = R_{(s,a)} / n_{(s,a)}$$

Box 4.2 *Monte Carlo algorithm for estimating Q-values*

4.1.2.3. Temporal difference learning

As we will see in Section 4.2, temporal difference learning is perhaps the single most influential idea from reinforcement learning theory to impact brain research. Temporal difference (TD) learning is a model-free approach, meaning that like Monte Carlo methods, TD learners acquire state- and/or

action-value functions from experience with their environments, and require a certain amount of exploration to guarantee that the $V(s)$ and/or $Q(s,a)$ estimates converge to $V^*(s)$ and/or $Q^*(s,a)$, repectively. Unlike Monte Carlo methods, TD approaches also use the already-acquired value estimates for other states to estimate the value of the current state (or state-action pair).

Whereas Monte Carlo methods must wait until the end of an episode to update the values of all states encountered in that episode, temporal difference algorithms update the value estimate for the state encountered at time t, denoted $V_t(s_t)$, on the next time step. The updated value estimate is then denoted $V_{t+1}(s_t)$ This is accomplished by comparing the current estimate of the state value $V_t(s_t)$ to the combined total of the reward received after performing action a_t, denoted R_{t+1}, and the estimated value of future rewards to be received from the new state, denoted $V_t(s_{t+1})$. This comparison results in a *reward prediction error*, or δ-function, which is used to determine by how much $V(s_t)$ should be incremented at time $t+1$.

$$\delta_t = R_{t+1} + \gamma V_t(s_{t+1}) - V_t(s_t)$$
$$V_{t+1}(s_t) = V_t(s_t) + \alpha \delta_t$$

Similarly, for Q-values:

$$\delta_t = R_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q(s_t, a_t)$$
$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha \delta_t$$

We have included two additional parameters above, the learning rate α and the discount factor γ. The learning rate, α, has a value between 0 and 1, and determines by how much the value estimates are incremented on each time step. The TD algorithm is guaranteed to converge for small enough values of α, but small incremental updates lengthen the training time required to obtain accurate value estimates. The discount factor, γ, also has a value between 0 and 1, and determines how much value to give to expected future rewards compared to those received in the current time step. Animals and humans generally exhibit such discounting behavior, valuing smaller immediate rewards over larger delayed rewards. The discount factor helps capture this effect in the TD framework.

As discussed above for Monte Carlo methods, a policy such as ε-greedy or softmax should be used that both exploits value information as it is acquired, and continues to explore so as to guarantee convergence.

4.1.2.3.1. Actor-critic architecture

In the actor-critic framework, the agent explicitly maintains both a state-value function $V(s)$, and separate policy $\pi(s,a)$. The temporal prediction error δ is then used to update both the state values and the probability of choosing an action in a given state.

After taking action a_t and observing s_{t+1}, and R_{t+1}:

$$\delta_t = R_{t+1} + \gamma V_t(s_{t+1}) - V_t(s_t)$$
$$p_{t+1}(s_t, a_t) = p_t(s_t, a_t) + \beta \delta_t \quad \text{(incremental policy improvement)}$$
$$V_{t+1}(s_t) = V_t(s_t) + \alpha \delta_t \quad \text{(incremental state-value update)}$$

where α and β are both learning rate parameters, which need not be the same for actor and critic.

The actor-critic architecture has been especially linked to basal ganglia substrates thought to perform different reinforcement learning functions. In particular, the ventral striatum is thought to maintain state-value estimates $V(s)$, the dopamine neurons of the VTA and SNc are thought to compute the TD error δ, and the dorsolateral striatum is proposed to maintain the policy, $\pi(s,a)$.

4.1.2.3.2. Eligibility traces

So far we have discussed one-step temporal difference learning methods, which encounter some difficulty when an action leading to reward is separated in time from the actual reward delivery by a probabilistic sequence of intervening events. One way to address this "temporal credit assignment" problem is with eligibility traces. Temporal difference learning with eligibility traces is called TD(λ), and this algorithm employs the weighting of future returns to obtain an average return for state s_t:

$$R(\lambda) = (1-\lambda) \sum_{n=1}^{\infty} \lambda^{n-1} R_t^{(n)}$$

As this "forward view" of TD(λ) is acausal, we consider a computationally implementable "backward view." For each state an eligibility trace in maintained, which is set to a value of 1 when the state is encountered, and decays with successive time steps:

$$e_t(s) = \begin{cases} 1 & \text{if } s = s_t, \\ \gamma \lambda e_{t-1}(s) & \text{otherwise} \end{cases}$$

This describes how "eligible" each state is to receive credit or blame if a return is better or worse than expected. Note that in addition to storing a value for each state, we must now also store an additional parameter for each state corresponding to its eligibility. The update rule for state-values using eligibility traces is modified as follows:

$$\delta_t = R_{t+1} + \gamma V_t(s_{t+1}) - V_t(s_t)$$

$$V_{t+1}(s) = V_t(s) + \alpha \delta_t e_t(s)$$

Note that all states are now updated at time $t+1$ according to their eligibility, not just the state encountered at time t. For action-values instead of state-values, the equations become:

$$\delta_t = R_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)$$
$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha \delta_t e_t(s,a)$$

When using an ε-greedy or softmax policy, the update rule must be adjusted, as the values of actions that occurred prior to a randomly-chosen 'exploratory' action should not be updated to the same extent as those that occurred after. In these cases, eligibility traces are generally set to 0 when such an exploratory action occurs.

For actor-critic implementations of TD(λ), the actor and the critic must maintain separate eligibility traces. The critic uses $e_t(s)$ to update the state-value function V, as described above. The actor must maintain $e_t(s,a)$ and update the probability that an action will be chosen in state s accordingly:

$$p_{t+1}(s,a) = \begin{cases} p_t(s,a) + \alpha\, \delta_t\, e_t(s,a) & \text{if } a = a_t \text{ and } s = s_t \\ p_t(s,a) & \textit{otherwise} \end{cases}$$

4.1.3. More advanced solutions to RL problems

A number of methods have been proposed for combining and/or extending the basic concepts of reinforcement learning summarized in the previous section. In this section, we summarize some of these topics that are particularly relevant to basal ganglia researchers.

4.1.3.1. Neural networks and function approximation

A number of practical issues arise in implementing the algorithms described above. Perhaps the most obvious of these issues is that the memory required to explicitly represent the value of each state or state-action pair quickly becomes impractical in solving complicated real-world problems. A typical way of dealing with this issue is with *function approximation*, where rather than explicitly maintaining a value estimate for each state or state-action pair, the state space is parameterized and state values are estimated as a combination of these parameters. A common way of implementing function approximation is with neural networks.

Neural networks, as the name implies, are thought to be particularly analogous to the implementation adopted by the brain. Here, units in an input layer are connected through varying weights to units in an output layer. In the simplest case of estimating a value, the first layer is vector of parameters, $x^1 \ldots x^n$, chosen to represent the state space. The value function is the product of this input vector and the associated weights, w^i, for each x^i.

$$V_t(s_t) = \sum_i w_t^i x_t^i$$

Updating the value function is accomplished by *gradient descent*. Here, the estimated value is compared to a reference value, and the vector of weights is adjusted in the direction that minimizes this error. In practical terms, this means that a change to a particular weight w^i depends not only on the learning rate α, the calculated prediction error δ and an eligibility trace e, but also on the activation of the units connected by w^i. This dependence addresses the credit assignment problem - only those weights that are active leading up to reward delivery are adjusted. Generally, function approximation methods cannot be guaranteed to converge to the global optimal policy or value functions. Rather, these methods are expected to converge to a local optimum, or to within a small window around this value. The iterative process by which the best possible value function or policy is learned is otherwise the same as those described previously.

$$\delta_t = r_{t+1} + \gamma\, V_t(s_{t+1}) - V_t(s_t)$$
$$\Delta w_t^i = \alpha\, \delta_t\, e_t(w_t^i)\, x_t^i$$

The question of how to best parameterize the state space in order to represent a large or continuous set of states with a limited number of neural "units" is another difficult issue. A related issue is how to scale these representations when more complex problems are encountered or new sensors are added to an agent. The computational concept of *generalization* approaches these issues, but is beyond the scope of our discussion here. A number of specific representations have been proposed to ensure good coverage of the possible states by the parameter space, to limit the dimensionality of the

parameter space, and to ensure that the dimensionality of the parameter space scales in a reasonable way.

4.1.3.2. Using models for learning and planning

As mentioned previously, model-based approaches provide advantages over model-free RL methods, though at the expense of requiring significant additional computational resources. The first of these advantages enables the agent to learn through simulated experience, in addition to actual experience, thus speeding the learning process and limiting the time required for potentially dangerous exploration. The second advantage is *planning*, which allows the agent to search through the model of its environment to find the optimal next action, rather than relying on the current potentially inaccurate value estimates. Models can be acquired using any of the basic RL approaches described in Section 4.2, using an iterative improvement process to update not only the value estimates and policies, but also the state transition probabilities.

For illustration of the use of models in the service of generating virtual experience, the Dyna-Q algorithm is described. Dyna-Q integrates the three processes of model building, model evaluation and decision making into a single algorithm, using the same tabular computation of Q-values in both the model building (learning) and model evaluation (planning) phases. The Dyna-Q model provides *indirect RL*, allowing simulated experience to be used to build the value functions and the policy. For example, previous experiences can be regenerated in addition to using current experience to build value functions or policies. For each iteration, the agent chooses an action and proceeds to state s_{t+1}, updates its value estimates $Q(s,a)$ according to the experienced error calculated, performs some number of simulated actions at $t+1$ and updates its Q-value estimates according to the simulated errors (planning), and again chooses an action based on the new value functions after both direct and indirect RL steps have completed. With this type of algorithm, which learns from regenerated experience, there is a danger that the agent will settle on a suboptimal solution before exploring all states. This is addressed by ensuring that the policy guarantees some exploration, such as ε-greedy or softmax. Another common solution is to add a "reward bonus" to the values of actions that have not been experienced recently to ensure that all states are continually explored.

To illustrate the use of models in planning functions, we discuss the heuristic search, or "tree search" approach. In heuristic search, the planning phase is used directly to find the best action, rather than to update the current value estimates. At each state, a tree of possibilities is generated according to the current model, and the values of each state-action pair are computed according to the transition probabilities and the current value estimates. After generating the tree and the values for the leaves, the best action is chosen and the values are discarded. These methods consider all future next steps in determining how best to optimize reward, and generally have an advantage in situations in which the Q-function is not yet accurate but a good model of the environment has been obtained. Obviously, building a full tree at each state in a complicated environment is impractical, and a number of methods have been developed to address this issue, the most intuitive of which is simply to truncate the tree after n steps.

4.1.3.3. Uncertainty, value estimation, and action selection

Kaelbling et al. (1996) summarized the next set of issues succinctly when they stated that "exploration is often more efficient when it is based on second-order information about the certainty or variance of the estimated values of actions." Such estimates of variance (or uncertainty) provide

an idea of how confident the agent is that its current value estimates are accurate. It has been shown that incorporating variance estimates can improve the selection of optimal actions. It has been further suggested that variance estimates can be used to arbitrate between model-based versus model-free reinforcement learners such that each is used when it is likely to be most accurate.

In using variance to improve action selection, a number of approaches have been suggested which serve to bias action selection toward choosing actions that have the highest level of uncertainty associated with their value estimates. One approach, used by Kaelbling (1993), is to compute a confidence limit around the estimated state- or action-values and choose the action associated with the highest upper bound. A similar solution would be to adjust the estimated value by adding its estimated variance, and again choose the action associated with the highest adjusted value. An alternative conceptualization of these ideas is that they place a value on the additional information that can be gained by taking exploratory actions to reduce uncertainty. These algorithms guarantee efficient exploration by biasing action selection toward the most uncertain actions.

The idea that uncertainty can be used to arbitrate between model-based and model-free learners was explored by Daw et al. (2005) and is discussed in more detail in Section 4.3.

4.1.3.4. Hierarchical reinforcement learning

If the same sequence of actions is performed from a given state, it may be advantageous to "chunk" these primitive actions into a higher-level meta-action, or *option*. This is the general idea behind hierarchical reinforcement learning (HRL). One way to think of HRL is as a master-slave architecture, where the high-level option is chosen by the master, and the slave executes primitive actions in order to achieve a subgoal state defined by the current option. In other words, the high level option "gates" the low-level policies. One conceptually straightforward method of implementing such a system is with memory registers that maintain a representation of the current high-level option. The values contained can then modulate which actions are chosen by the low-level policy. Reinforcement learning techniques can be used to learn not only the low-level actions, but also when to gate the high-level options. In such systems, the low-level policy must be reinforced not only for finding external rewards in the environment, but also for achieving the subgoals defined by the high-level option. Thus, in addition to requiring extra memory and computation to represent more complicated task architectures, they also must include a mechanism for producing internally-generated intermediate rewards for successfully accomplishing subgoals. The acquisition of such subgoals through trial-and-error learning is a non-trivial problem, but specific solutions are beyond the scope of this brief summary.

4.1.4. Summary

In this section, a number of general solutions to RL problems were described. These basic solutions all include 1) a method for acquiring state-value and/or action-value functions, 2) a policy by which the estimates of value may be used to select specific actions, and 3) a method for improving value estimates and action selection over time and with increasing experience. We distinguished model-free approaches from those that are model-based, where the latter include an additional explicit representation of each state in the environment and the transition probabilities between states. While model-based approaches require more computational resources to implement, they have the advantage of being able to improve performance based on simulated experience and can be used to evaluate future states and outcomes prior to experiencing them.

While it is easiest to think in terms of deterministic state values and state transitions, it is important to realize that most problems of interest are best described probabilistically, and may be non-stationary. An agent may be most successful in such complex environments if a number of different approaches are implemented and integrated in some way. Combining multiple learning approaches is of particular interest to brain researchers, as animal and human decision making is thought to be governed by a number of factors, including model-free trial-and-error learning and model-based forward planning, as well as other types of learning not reviewed here, such as supervised learning, or learning from instruction. In the following section, we summarize some of the approaches inspired by reinforcement learning that have been applied to the study of basal ganglia function.

4.2. Reinforcement learning in basal ganglia research

Following the discovery that dopamine may encode reward prediction errors and thus could be used as a teaching signal in a neural learning algorithm, a number of neuroscientists have turned their attention to RL. The applications of RL in brain research range from providing a quantitative description of behavior, to the development of neural network models with components mapped onto specific brain regions. In this section, we review some of this work.

4.2.1. RL and classical conditioning

Prior to the characterization of dopamine neurons according to reward prediction error, Rescorla and Wagner (1972) developed a mathematical model to explain the strength of the association between a conditioned stimulus and an unconditioned stimulus in classical, or Pavlovian, conditioning experiments. In this type of conditioning, a stimulus is repeatedly paired with reward and eventually the stimulus itself acquires a value predictive of the to-be-delivered reward. The Rescorla-Wagner model uses a δ-function to update the values associated with presented reward-predicting stimuli, similar to the RL algorithms discussed above. The δ-function is computed according to the difference between the delivered reward and the current value of the context, which includes the presented stimulus. The value of stimulus x is then updated on trial k according to:

$$\delta_x^{k+1} = \alpha_x \beta \left[R - \sum_x V_x \right]$$
$$V_x^{k+1} = V_x^k + \delta_x^{k+1}$$

The Rescorla-Wagner model successfully accounts for a number of phenomena observed in classical conditioning experiments, including blocking, overshadowing, and conditioned inhibition. In *blocking*, if A is repeatedly paired with reward and acquires a predictive value, then later pairing A+X with reward will "block" the acquisition of any value associated with X because the stimulus A already accounts for the full value of the predicted reward. In *overshadowing*, the compound stimulus A+X is repeatedly paired with reward, and subsequent pairing of A alone or X alone reveals that the values associated with the individual stimuli are lower than that acquired for A+X together. In *conditioned inhibition*, the stimulus A predicts reward, but compound stimulus A+X predicts no reward, and the animal thus acquires a negative value for stimulus X. The Rescorla-Wagner model fails to account for detailed temporal relationships between stimuli and rewards, or explain trace conditioning (in which the stimulus is not present at the time of reward presentation).

These failings motivated Sutton and Barto (1981) to develop a modified model that uses stimulus eligibility traces to account for animals' ability to predict the timing of reward delivery. Their model incrementally updates the synaptic weights according to the difference between expected and actual reward, and assigns credit to the encountered stimuli based on their eligibility. This model accurately learns to respond at the time of the earliest reliable predictor of reward, even if the stimulus presentation never overlaps the time of reward delivery.

4.2.2. Dopamine and reward prediction error

In the late 1980s and early 1990s, Schultz and colleagues conducted a number of experiments in which they recorded from the dopamine-containing neurons of the VTA and SNc. They found that these neurons fired phasically in response to the delivery of an unexpected reward (but not to delivery of an expected reward), and with training developed responses to stimuli that predicted reward delivery. They also observed that if no reward was delivered at the time when it was expected, the dopamine neurons exhibited a pause in firing. Montague et al. (1996) first formalized these findings in terms of temporal difference (TD) learning, assigning to these neurons a role in calculating the reward prediction error, or δ_t, used to update state-value estimates. Schultz et al. (1997) provides a review of the experimental and modeling results.

The ability of dopamine to serve as a teaching signal rests not only in the firing of these neurons according to a reward prediction error, but also in the ability of this signal to modulate activity of target neurons. As discussed in Section 1.4.3, dopamine has been shown to influence the synaptic plasticity of D1- and D2-receptor expressing medium spiny neurons in the striatum, providing the mechanisms by which a DA teaching signal could be utilized in this region.

While the TD framework is intuitive, easy to implement, and captures a number of features of Pavlovian learning, several issues remain. Bullock et al. (2009) outline a number of these, including the fact that dopamine responses are not limited to reward prediction errors, and that negative prediction errors may not be well captured by DA neuron firing. Neurons in the VTA and SNc have been shown to respond to novel stimuli, as well as salient stimuli unassociated with reward, and even to salient stimuli associated with negative outcomes (for review, see Horvitz, 2000). Dopamine neurons have also been shown to exhibit uncertainty-related firing (Fiorillo et al., 2005), adding another dimension by which DA signaling may influence behavior. Uncertainty-related DA firing may be represented by changes in the tonic, rather than phasic, firing of these neurons, and the tonic actions of DA are not considered in the standard TD learning framework. Another interesting point is that the TD framework generally predicts a "backpropagation through time" of reward prediction error, which has never been seen experimentally. It has been shown that this backpropagation need not occur for sufficiently large values of λ in a TD(λ) implementation, so this is perhaps not a critical shortcoming. Finally, Dayan & Balleine (2002) as well as Berridge and colleagues (see (Berridge, 2007) also emphasize the failure of TD algorithms to explain the "incentive salience" of a reward-related stimulus, which critically depends on dopamine, and can modulate the overall value of a stimulus on the fly according to physiological states such as hunger or satiety. Recently, Zhang et al. (2009) have provided a computational model by which incentive salience can be incorporated into a reinforcement learning framework.

O'Reilly et al. (O'Reilly et al., 2007) developed a "Pavlovian-Value Learned-Value" reinforcement paradigm to address some of the issues associated with the TD algorithm. Their algorithm, unlike TD(0), does not rely on a predictable chain of intervening events (or predictable timing) between a

presented stimulus and reward presentation. The authors also highlight TD's lack of a clear biological mapping, which is at least two-fold: 1) it is unclear how the prediction error signals might arise in the DA neurons and 2) the trace conditioning under investigation is known to depend on neural structures implicated in memory maintenance, and TD does not predict that this should be so. Instead of a TD-like chaining mechanism, their algorithm involves two systems - one to learn about "primary values" at the time of reward delivery, and one that learns about stimulus values ("learned values"), a representation of which must be available at the time of reward delivery. Of course, modifications can be made to the standard TD algorithm, such as the inclusion of eligibility traces, which make it more capable of coping with the timing issues addressed by O'Reilly et al., so further work is required to show whether their conceptualization is more biologically accurate.

In summary, the TD framework provides a convenient and concise way to conceptualize animal behavior via dopamine-mediated reinforcement learning. This framework has proven useful in interpreting a number of scientific results and has provided testable hypotheses for DA activity. While it has been shown to capture key features of animal learning, the various components of TD implementations may not map precisely onto known neuroanatomy. The best-supported mapping is that of dopamine neurons of the VTA and SNc in computing a reward prediction error, which can be used as a teaching signal in target structures, especially the striatum and prefrontal cortex. Even this mapping has limitations, however, as VTA/SNc neurons are known to exhibit a number of additional responses unrelated to prediction errors (including novelty, salience and tonic responses). Further, state-value estimation according to TD reinforcement learning cannot capture all the known features of how animals assign value to a stimulus.

4.2.3. Stimulus-response learning and actor-critic architectures

Perhaps the most influential extension of temporal difference learning in basal ganglia research is the mapping of striatal anatomy onto an actor-critic architecture for reinforcement learning. Takahashi et al. (2008) review the current conceptualization of the actor-critic mapping onto basal ganglia substrates. Here, the ventral striatum and dopamine neurons play the role of the "critic." The ventral striatum computes a state-value function *V(s)*, and sends projections to the dopamine neurons of the VTA/SNc. Using the current estimate of the state value and information about current rewards being received, the VTA/SNc can then compute a TD reward prediction error. Feedback from the dopamine neurons to the ventral striatum is used to update the state-values such that they come to accurately reflect the real values of each state. Importantly, the dopamine neurons in the SNc also send substantial projections to the dorsolateral striatum, which plays the role of the "actor." The dorsolateral striatum receives information regarding the current state from motor and somatosensory regions of cortex, and is thought to be involved in action selection. The same dopamine reinforcement signal computed by the critic and used to update its state-value estimates, can be used to update the strength of the state-action associations in the DLS/actor to improve the animal/agent's policy.

Atallah et al. (2007) found that activation of the dorsal striatum was not necessary for learning about the value of a stimulus, but was necessary for performing the appropriate action based on that value. They performed temporary inactivation of the dorsal striatum and trained rats for three training sessions on an odor discrimination task. They found that rats were unable to improve their performance when the dorsal striatum was inactive during training. However, when they tested the animals during a 4^{th} training session without inactivation, the animals' performance immediately recovered - suggesting they were able to acquire the state-values and an appropriate policy, but were

unable to implement that policy when the dorsal striatum was inactivated. The authors suggest an actor-director-critic conceptualization to explain their results. The dopamine neurons play the role of the "critic" and the dorsal striatum is the "actor" as above. Crucially, in this conceptualization the dorsal striatum plays no part in learning which actions to perform, rather it is biased ("directed") toward the correct action by the ventral striatum. The authors suggest that the role of the ventral striatum as "director" may be mediated through the dopamine neurons, or through other brain regions known to connect to dorsal striatal neurons (e.g., OFC), which could then bias action selection in the dorsal striatum.

The idea that the dorsal striatum (specifically, the dorsolateral striatum) may be involved in implementing a policy raises the questions of what type of policy is preferred and how this information may be stored. The policy implemented in most reinforcement learning models is generally softmax. Daw et al. (2006) specifically compared subjects' behavior on a 4-arm bandit task to that predicted by the most common RL policies (ε-greedy, softmax, and softmax with an uncertainty bonus) and found that the standard softmax policy provided the best fit to subjects' performance. Interestingly, Daw et al. further observed that during "exploratory" actions versus "exploitative" actions, activity was high in the anterior frontopolar cortex. They interpret this finding as suggesting that exploratory activity may "override" exploitation - a suggestion with some similarities to the idea proposed in Chapter 2 that dorsomedial activation might serve to modulate the access of dorsolateral loops to the control of action.

It remains unclear precisely how the dorsal striatum learns, stores and/or implements a policy. The most common suggestion is a somewhat literal computation of Q-values by the cortico-striatal network, with some probabilistic winner-take-all mechanism in the striatum to determine which action is chosen. As mentioned above, the dorsolateral striatum receives converging input from sensory and motor areas of cortex, and is known to be involved in action selection and stimulus-response learning and habit formation. The basic idea put forward is that converging state information from the cortex activates a subset of striatal neurons, which select desired actions and/or inhibit undesired actions through activation of the direct and indirect pathways as described in Section 1.3. If the selected action results in a better-than-expected state, a positive prediction error causes a strengthening of the connections between the active cortical and striatal neurons. Conversely, if the result is worse than expected, the negative prediction error causes a weakening of these connections.

A number of authors have reported action-value (or Q-value) correlated activity in the dorsal striatum during task performance (Histed et al., 2009; Lau and Glimcher, 2008; Pasquereau et al., 2007; Samejima and Doya, 2007), but this conceptualization still raises some concerns. First, the firing of most striatal neurons follows movement, suggesting that action value-correlated firing in the striatum plays a larger role in action evaluation than in action selection (Lau and Glimcher, 2008). By contrast, most RL-based models of striatal function predict that activity related to the values of upcoming actions should be prominent in this region during behavior. Current theories of striatal function suggest that the differing functional roles of different striatal regions result from the differing inputs received from connected regions of cortex, consistent with the parallel loop architecture of these structures. However, as discussed further in Section 4.2.5, the most widely accepted RL mapping onto striatal substrates, the actor-critic architecture, fails to capture the role of the dorsomedial striatum in goal-directed behaviors. Specific neural network implementations extending the basic model to address some of these issues are discussed in the next section.

A complementary explanation of how the dorsal striatum may be involved in the selection of actions comes from Lo and Wang (2006). They present a model of cortico-basal ganglia-superior colliculus interaction in which the basal ganglia pathway sets a threshold, which when a crossing is detected by SC neurons, triggers a motor response. The authors suggest that the strength of the cortico-striatal synapses can adjust the threshold level, providing a way to tune behavior to achieve an optimal speed-accuracy tradeoff. This model is related to a class of random-walk-to-threshold models that capture important behavioral aspects of decision-making. Further experimental and modeling work are needed to determine whether and how this hypothesized function of the basal ganglia in threshold setting may relate to its hypothesized role in action-value encoding. It will be interesting to see if future work supports the model of Lo and Wang, and whether their model will be able to account for a broad range of observed clinical and experimental results, or whether it exists alongside other roles of the basal ganglia in action selection.

4.2.4. Specific neural network models of motor control and action selection

While the algorithms described above provide convenient mathematical descriptions of behavior, they generally make only superficial attempts to map these functions onto neural substrates. A few authors have developed specific neural network models that attempt to capture the major functions of the basal ganglia while remaining true to the known anatomical connectivity of the structures involved. Cohen and Frank (2009) provide an excellent review of the strengths and issues surrounding both the abstract mathematical and the neural network modeling approaches. Here, we review two recently developed neural network models.

Building on previous conceptualizations of basal ganglia function, O'Reilly, Frank and colleagues have developed perhaps the most extensive model of cortico-basal ganglia interaction to date. We therefore discuss this model in some detail. The O'Reilly and Frank model proposes that the primary function of the basal ganglia is to gate frontal cortical registers (Hazy et al., 2006; O'Reilly and Frank, 2006). Their implementation considers a number of independent PFC-BG sets of registers ("stripes"), each toggled by an associated basal ganglia loop. They suggest that through DA-mediated reinforcement, D1 ("GO") and D2 ("NO-GO") pathways within each loop can learn to update or maintain, respectively, the current representation held in the PFC. The authors developed an extended version of the commonly used AX task, which they call the 1-2-AX task. In the standard AX version of the task, the subject is presented with a series of characters, and asked to respond with a button press when an X is presented immediately following presentation of an A. In the modified version, the subject is asked to press the button following the A-X presentation only if a '1' was the most recently presented number. If a '2' was presented most recently, the target sequence is then B-Y. This more difficult task was used to verify that their model could learn when to update/maintain the "outer-loop" representations of the number most recently encountered, which then bias the "inner-loop" representations of each presented character such that a button push occurs only during the correct condition.

To support their model, O'Reilly, Frank and colleagues have developed a "Pavlovian Value - Learned Value" algorithm for DA-mediated reinforcement of the D1 and D2 pathways (O'Reilly et al., 2007). Critically, each stripe in their model receives its own reinforcement signal, such that activity is reinforced based on a combination of the prediction error and the activation of the BG neurons in that stripe. A positive prediction error results in the strengthening of active "GO" or direct-pathway weights, whereas a negative prediction error results in a decrease in D2 weights.

As discussed briefly in Chapter 1, Hikosaka and colleagues found that the D1 and D2 pathways are differentially implicated in speeding reaction times when a large reward is predicted versus the slowing of responses to a small reward (Nakamura and Hikosaka, 2006). The O'Reilly et al. model suggests that in the former case, activation of the D1/direct pathway results in rapid selection of action, whereas in the latter case, disinhibition of the D2/indirect pathways occurs more slowly. Interestingly, the model has also generated specific predictions about the differential involvement of D1 and D2 pathways in learning from positive versus negative feedback. The model predicts that under conditions of DA-depletion in which a positive phasic prediction error cannot be generated, learning from positive feedback should be impaired but learning from negative feedback relatively spared. These results have been verified in PD patients and in normal subjects given D1 antagonists. Similarly, O'Reilly and colleagues have shown that individual variations in D2 receptor genetics correlate with a subject's propensity to learn from negative reinforcement.

The O'Reilly et al. model has been used by Reynolds and O'Reilly (2009) to investigate how a hierarchical arrangement of cortico-basal ganglia loops may contribute to the development of representations of different levels of hierarchical abstraction in different cortical regions (e.g. the outer-loop versus inner-loop representations required in the 1-2-AX task). They show that hierarchical architecture is sufficient to bias the network toward developing such hierarchical representations. Interestingly, they find that their model does not develop stable maintained representations of the outer-loop information. Rather the outer-loop representations are implemented via conjunctive representations of the outer- and inner-loop information that vary at each time step. In a further extension to this framework, Doll et al. (2009) used the O'Reilly et al. model to investigate how RL learning in the PFC-BG network might be influenced by inaccurate prior instruction on a task. They conclude that the updating of information acquired via instruction is governed by "special rules" and not just overridden by accumulated contradictory experience.

A somewhat similar conceptualization is proposed by Massaquoi and Mao in their MIMOAS model (Steve Massaquoi, *personal communication*), who like O'Reilly et al. propose a gating function for the basal ganglia. Significant differences exist between the two models in the anatomical assumptions used to construct the models, however. Critical to the MIMOAS model is the architecture of the cortico-basal ganglia loop, which differs from that envisioned by O'Reilly et al. The activation of a pattern of cortical units in the MIMOAS model excites essentially one striatal unit, which in turn projects into *either* the direct pathway or the indirect pathway. By contrast, the "stripes" of O'Reilly et al. contain both direct and indirect projections. Dopamine reinforcement acts in the MIMOAS model in same direction at D1 and D2 cortico-striatal synapses, with D1 synapses updating more rapidly than D2 synapses. O'Reilly et al. assume opposite actions of dopamine at D1 and D2 synapses. The MIMOAS model thus predicts extremely sparse encoding in the striatum during skilled motor performance, and provides a novel suggestion for the interaction of the direct and indirect pathways.

The MIMOAS model further presumes that the primary function of the basal ganglia is to enable the recreation of specific patterns of cortical activation. Consequently, they suggest that continuous basal ganglia activation is needed to enable/disable frontal cortical registers, whereas O'Reilly et al. suggest that phasic activation in the striatum results in the toggling (updating) of cortical registers. Both the MIMOAS and the O'Reilly et al. models provide a single architecture that can support low-level action selection and sequencing as well as high-level working memory. However, the MIMOAS model also includes an abstraction of muscle activation. By combining abstractions of both the central nervous system and the periphery into one unifying framework, the MIMOAS model

is able to suggest specific mechanisms by which DA-depletion in the basal ganglia can result in the movement deficiencies (tremor and rigidity) observed in Parkinson's Disease.

As both Mao and Massaquoi and O'Reilly et al. review, there is substantial evidence supporting both of the proposed models of the cortico-basal ganglia network, and significant future work is needed to determine which may be the more accurate simplification of the anatomy. A number of prior models propose one or more of the ideas that inspired the two models described above (Beiser and Houk, 1998; Berns and Sejnowski, 1998; Houk and Wise, 1995). Without diminishing the importance of this previous work, the reader is referred to the specific papers for the details of these implementations.

4.2.5. Stimulus-response (S-R) versus Action-Outcome (A-O) learning and model-free versus model-based RL

As discussed above, in the most common RL framework for understanding basal ganglia function, the ventral striatum is thought to compute state-values V(s), whereas the dorsolateral striatum is thought to implement a policy, perhaps through the computation of Q-values and a winner-take-all selection mechanism. This conceptualization leaves open the question of how the dorsomedial striatum may contribute to exploratory behavior and decision-making. As reviewed in Section 1.5, lesions to the dorsomedial striatum result in behavior that is insensitive to outcome devaluation, suggesting that the dorsomedial striatum is critical for "goal-directed" behavior that depends on a representation of the outcome and its value. A number of authors have suggested how this might come about.

The most intuitive conceptualization comes from Horvitz (2009), in which the same architecture is proposed for dorsolateral and dorsomedial striatal circuits. In the dorsolateral striatal loop, cortical areas that represent context and movement parameters are thought to map onto action representations in the striatum, and acquire stimulus-response associations through trial-and-error and dopamine-driven reinforcement. In the dorsomedial striatum, cortical areas that maintain "outcome value" representations map onto striatal action representations, and over time, the appropriate action-outcome associations are formed. The model can learn sequences of actions if feedback from the striatum to the cortex is provided, and can provide a mechanism for sustaining cortical activation. Several problems exist with this simple approach, however. The author notes that it is unclear what cortical region would map on to the "outcome" representations in the model, nor is it clear that the action-outcome associations developed by the model could drive the goal-directed behavior exhibited by animals. For example, after an outcome-response mapping is learned, subsequent devaluation (which reduces the outcome value) results in an immediate adjustment in behavioral responding, without requiring further exploration or incremental learning. The Horvitz model can predict a lack of the learned lever-press responding in a relatively simple instrumental conditioning paradigm, but it is unclear that it would respond appropriately in a more complex navigation task, for example.

This last issue has led to the idea that the systems that implement goal-directed behavior use a model of the environment and forward planning to direct decision-making and action selection. Thus, the distinction between dorsolateral striatum-based stimulus-response learning, and dorsomedial striatum-based action-outcome learning, is mapped onto the dichotomy between model-free RL and model-based RL (Daw et al., 2005; Matsumoto et al., 2006; Redish et al., 2008; Samejima and Doya, 2007). While this provides a formal framework for thinking about animal behavior, the biological mapping of this approach remains unclear. Wickens et al. (2007) point out that a fundamental

contradiction must be resolved in any model of striatal function: different striatal regions perform different functions, but at the same time there is relative consistency throughout the striatum in chemical architecture, microcircuitry, and physiology. In mapping model-free and model-based systems onto brain function, authors generally envision the dorsolateral striatum as implementing a model-free system, and the prefrontal cortex as implementing a model-based system. The specific role of the dorsomedial striatum, which is both intimately connected to prefrontal cortical areas and likely to perform computations similar to those performed in the dorsolateral striatum, is left undefined.

The idea that multiple systems may compete to direct behavior also raises the question of how arbitration between these systems may occur (Daw et al., 2005; Matsumoto et al., 2006). Daw et al. (2005) suggest that uncertainty in the estimates made by each system can be used to determine which will gain control of behavior, such that each system is used when it is most certain. They use a "tree search" algorithm to implement a model-based "goal-directed" performer, and a "cache value" implementation of a model-free "habitual" performer. Variances in the estimates for each controller are also computed, and the system with the lowest variance (least uncertainty) is chosen to direct action selection. The biological mapping of this approach is again unclear. The authors favor the view that the model based "tree-search" may be localized in the prefrontal cortex, whereas the model-free "cache" system may be localized in the dorsolateral striatum. Arbitration, they suggest, may take place via modulation of these systems by cholinergic or noradrenergic tone (ACh and NE expression have been observed to correlate with uncertainty arising from different sources), or in specific brain regions observed to exhibit uncertainty-related firing (e.g., the anterior cingulate or infralimbic cortices). Again, the role of the dorsomedial striatum is undefined, though the authors speculate that it may be engaged with the PFC in model-based control.

4.2.6. Hierarchical RL

Neuroscientists have long recognized the hierarchical nature of anatomical organization and functional involvement in the brain. The incorporation of reinforcement learning approaches into brain research has thus recently brought attention to the field of hierarchical reinforcement learning (HRL). Botvinick et al. (2009) offer a review of hierarchical reinforcement learning theory and how it may map on to different neural substrates. The general idea that they review is that hierarchical reinforcement learning provides a formal mechanism by which a sequence of low-level actions may be "chunked" into a single higher-order *option*, and then applied as a whole in various contexts.

The computational issues associated with this framework include how to acquire a useful set of options and how to learn when to use them. Options may be acquired through trial-and-error, as in standard RL. Here though, it is necessary to provide a "pseudo-reward" for accomplishing a "subgoal" - i.e., the agent must be rewarded for reaching a desired option-termination state, even if no external reward is available in the environment. Once a set of options has been acquired, it has been shown that learning can proceed faster if both low-level actions and high-level options are available during subsequent RL learning.

The PFC-BG model of O'Reilly, Frank and colleagues operates under essentially these principles - a representation of the high-level "outer loop" rule must be maintained and bias the low-level "inner loop" target representations for correct performance in the 1-2-AX task, as discussed in Section 4.2.4. In an extension of their model, Reynolds and O'Reilly (2009) show that hierarchical architectural constraints are sufficient to encourage separate, hierarchical representations of outer-

loop and inner-loop information. A dopamine reinforcement signal provides both "pseudo" and traditional reward prediction error, enabling the model to learn when to gate the high-level inputs.

Whether the brain implements a version of HRL may hinge on whether internally-generated "pseudo-rewards" are in fact available as animals acquire sequences of primitive actions. However, other features make HRL a particularly attractive approach for understanding behavior and brain function. As noted above, anatomical and experimental findings have highlighted a hierarchical structuring of different brain areas, with increasingly more frontal regions providing increasingly more abstract representations. Moreover, HRL provides a framework for understanding the behavioral observation of "positive transfer" and "negative transfer." Animals may transfer the options acquired on one task to a new task, which can result in faster learning in the new environment if these options are useful (positive transfer), or slower learning if these options are sub-optimal (negative transfer). The basal ganglia have also long been associated with grouping a sequence of movements into a single efficient "chunk" (Graybiel, 1998), and in acquiring if-then rules such as in stimulus-response learning. Thus, HRL may provide insight into the requirements of such a system, and shed light on the ways in which the basal ganglia may implement these processes. Botvinick et al. (2009) propose an extended actor-critic architecture which incorporates structures for implementing HRL.

Other authors suggest hierarchical or semi-hierarchical architectures for different cortico-basal ganglia loops, though these are less specifically tied to HRL theory. Samejima and Doya (2007) suggest that cortical networks may calculate *belief states* according to Bayesian inference, implementing a model-based approach for learning and planning. The basal ganglia may then implement model-free Q-learning, with the different functions of different regions assigned according to the type of cortical input received. For example, high-level orbitofrontal areas connected to ventral striatum may operate on "context/goal" information, while motor cortical regions connected to the putamen may operate on "motor/stimulus" information. Wickens et al. (2007) make a similar proposal, without specifically drawing on the reinforcement learning theory. These authors suggest that not only do cortical projections to different striatal regions differ, but there may exist a gradient in dopamine reinforcement from ventromedial to dorsolateral striatum such that the dopamine signals in ventromedial striatum are more temporally and spatially diffuse than those in the dorsolateral striatum. Haruno and Kawato (2006) build on this idea further by proposing a "heterarchical" model by which coarse state-value representations are formed quickly by ventral striatal circuits, which then train specific fine-grained Q-value representations in the dorsolateral striatum via DA-mediated reinforcement.

In short, many proposals exist to explain how hierarchically arranged cortico-basal ganglia loops may cooperate and/or compete to guide decision-making. Substantial experimental work will be required to validate or disprove any of these theories.

4.2.7. Summary

Reinforcement learning has been making inroads in the field of neuroscience for decades, beginning with the development of the Rescorla-Wagner δ rule to model animal learning in classical conditioning experiments. With the discovery that dopamine neurons fire phasically in a manner consistent with the encoding of a reward prediction error, RL has been increasingly applied in models of striatum-based trial-and-error learning with DA-mediated reinforcement. Particularly popular is the biologically-removed, but mathematically convenient framework of temporal difference learning. This framework has been used to make predictions regarding the magnitude of

phasic responses by DA neurons, and has proven especially useful in investigating individual variation in behavioral performance in terms of variations in model parameters. In mapping TD learning onto neural architecture, the actor-critic framework has achieved the most success. In the current actor-critic conceptualization, the ventral striatum learns state values, which are used by the DA neurons of the SNc to compute a prediction error signal. This error signal or "critic" is then used to update both the state values in the ventral striatum and the separately maintained policy stored by the "actor" in the dorsolateral striatum. This conceptual framework has provided the inspiration for a number of neural network implementations.

One of the drawbacks of these models is that they hypothesize a clear distinction between ventromedial and dorsolateral striatum, but the role of the dorsomedial striatum remains undefined. A series of experiments by Yin, Knowlton, Balleine and colleagues revealed that the dorsolateral striatum is critical for the expression of habitual stimulus-response behavior, whereas the dorsomedial striatum is critical for the expression of goal-directed action-outcome behavior. A number of authors have equated this stimulus-response versus action-outcome behavior with model-free versus model-based reinforcement learning systems. However, the mechanisms by which these regions mediate stimulus-response versus action-outcome behavioral control, and the biological mapping of model-free and model-based reinforcement learning algorithms is unknown. Generally, the model-free system is localized to the dorsolateral striatum, and the model-based system is localized to the prefrontal cortex, again leaving the role of the dorsomedial striatum undefined.

An alternate version of dorsolateral versus dorsomedial engagement may be related to hierarchical reinforcement learning, which describes a mechanism by which sequences of low-level actions may be grouped into higher-level "chunks." A hierarchical arrangement of dorsolateral to dorsomedial-based corticobasal ganglia loops has been noted, with progressively higher-level loops representing more abstract information. It may be that both model-based and model-free computations are performed at each level of hierarchy, but on different types of information. In any of these proposed models, one of the major issues to be resolved is how similar structural architecture and microcircuitry can used for neurocomputation throughout different striatal regions, despite the different functional contributions of these different regions to animal behavior.

In the interest of clarity and space, the discussion above has entirely omitted Bayesian approaches to decision-making, which overlap extensively with reinforcement learning approaches and have been commonly applied in brain research. For a review of how Bayesian and RL approaches have been applied to animal decision-making, see Doya (2008) and references therein.

4.3. Two RL-based hypotheses on medial-lateral interactions during learning

As was reviewed in Chapter 1, the dorsolateral striatum is thought to be involved in stimulus-response learning and habitual behavioral performance, whereas the dorsomedial striatum is thought to be involved in goal-directed action-outcome learning and flexible behavioral performance. Reinforcement learning approaches have generally attributed these functions to a model-free dorsolateral striatum-based system and a model-based dorsomedial striatum-centered planning system (Redish et al., 2008). The experimental findings summarized in Section 1.5 are consistent with either a direct role for the dorsomedial striatum in the goal-directed action selection process, or

with a role for this region in arbitrating between a model-free (habit) system and a model-based (goal-directed) controller.

In Chapter 2, it was shown that the dorsolateral striatum develops patterned activity in conjunction with the improved motor performance and increasing amount of reward received across training. The medial striatum, by contrast, developed patterned activity in conjunction with the difference in performance between the auditory and tactile task versions. In this section, we explore two reinforcement learning based models that may account for the dorsomedial pattern development. The first is based on the assumption that the dorsomedial striatum may be directly involved in action selection according to a model-based planning system. The second is based on the idea that the dorsomedial striatum may be part of a system involved in arbitrating between multiple controllers, one of which is the dorsolateral striatum-based model-free habit system.

These two ideas were explored computationally by Daw et al. (2005), though their approach was purely theoretical and not constrained to any specific biological implementation. Daw et al. suggest that the prefrontal cortex may implement the model-based controller, whereas the dorsolateral striatum may implement the model-free controller. In this conceptualization, the role of the dorsomedial striatum is undefined. The authors speculate that it may be engaged with the PFC in implementing model-based planning. Here, we additionally explore the idea that the dorsomedial striatum could be activated with the anterior cingulate cortex in arbitrating between multiple memory systems. We thus extend the work of Daw et al. (2005) by mapping their theoretical approach onto a biologically-inspired implementation, supported by the experimental results presented in Chapter 2.

4.3.1. Implementation

The model-free RL system used a TD update rule to incrementally update state-action values and was implemented alongside a model-based RL system that estimated its own state-action values based on the transition probabilities between states and values of subsequent states. Each of these controllers selected a right or left turn to perform probabilistically based on the current state-action values. Thus, an additional arbitration scheme was required to determine which controller would direct action selection. These components are illustrated in **Figure 4.1A-B** and described in detail in the following sections.

4.3.1.1. The T-maze task

The T-maze task was simplified to include only 3 time steps (start, cue onset, goal reaching), as shown in **Figure 4.1C**. At start, the agent has only one possible choice of action – move forward. After moving forward, a stimulus is presented to the rat. Each stimulus is equally likely, and for simplicity, the order of stimulus presentations was fixed for all runs. Each modality was presented in blocks of 20 trials, and within each block, the specific cue presented was alternated each trial (**Figure 4.1D**). For each stimulus, the agent has some probability of detection, which increases with the number of trials encountered,

$$p_{det,stim} = p^*_{det,stim} (1 - e^{-N/\tau_{stim}})$$

where $p^*_{det,stim}$ is the maximum probability of detection for a given stimulus. For rats that failed to acquire the tactile cues, $p^*_{det,stim} = 0$ for the tactile cues (rough and smooth textures). To reproduce the different learning rates observed for auditory and tactile cues in rats that were able to acquire both

versions, τ_{stim} was set to 500 for both auditory cues and 1500 for both tactile cues. If the stimulus is detected, the agent enters a state corresponding to the presentation of that stimulus, and if the stimulus is not detected, it enters the "Other/unknown" stimulus state. The agent then chooses one of two available actions (right turn or left turn) according to the state and action values calculated by the model-based and model-free controllers, described in the subsequent sections. The trial terminates when the agent reaches the goal state, receives reinforcement for its actions, and updates both the model-based and model-free systems. For correct trials, a reward of 1 was received following action selection; otherwise reward was equal to 0.

4.3.1.2. The model-free controller

The model-free controller computes the values of each state as well as the values of each state-action pair. As suggested by other authors, these roughly map on to a ventral striatal state-value learner and a dorsolateral striatal Q-value learner. In keeping with the idea that D1 and D2 expressing MSNs respond differentially to dopamine reinforcement signals, especially in the dorsolateral striatum where D2 receptors are differentially expressed, a modified update rule was adopted. Conceptually, for a better-than-expected outcome, an increase in dopamine representing the positive prediction error serves to strengthen the activation of D1-class MSNs while weakening the activation of D2-class MSNs. This serves to simultaneously increase "go" activity and decrease "no-go" activity for the chosen action. Conversely, a negative prediction error weakens activation of D1-class MSNs and strengthens activation of D2-class MSNs. This dual updating of D1- and D2-class MSNs serves to amplify the effects of reinforcement. This is captured in the model by the modified update rule, in which the values of both the chosen action and the non-chosen action were incremented on each time step. The state and action values for the model-free controller, $V_{MF}(s)$ and $Q_{MF}(s, a)$, respectively, were thus updated according to the equations below:

$$\delta_V = r_t + \gamma V_{MF}(s_{t+1}) - V_{MF}(s_t)$$
$$V_{MF}(s_t) \leftarrow V_{MF}(s_t) + \alpha \delta_V$$

$$\delta_Q = r_t + \gamma V_{MF}(s_{t+1}) - Q_{MF}(s_t, a_t)$$
$$Q_{MF}(s_t, a_t) \leftarrow Q_{MF}(s_t, a_t) + \alpha \delta_Q$$
$$Q_{MF}(s_t, a \neq a_t) \leftarrow Q_{MF}(s_t, a \neq a_t) - \alpha \delta_Q$$

A softmax selection rule was used to determine the probability of selecting each action based on the current Q-values for the possible stimulus-action pairs. This selection rule ensures continued exploration as discussed in Section 4.1.2, and previous studies have suggested that human behavior is best described using such a rule (Daw et al., 2006).

$$\pi_{MF}(s,a) = \frac{e^{Q_{MF}(s,a)/\tau}}{\sum_b e^{Q_{MF}(s,b)/\tau}}$$

For simplicity, $\gamma = 1$ for all runs and $\alpha = 0.005$.

4.3.1.3. The model-based controller

The model-based controller calculates the values of the current states and state-action pairs based on the state transition probabilities and the values of the successive states:

$$V_{MB}(s) = \sum_a \left[\pi(s,a) \sum_{s'} P^a_{ss'} \left(R^a_{ss'} + \gamma V_{MB}(s') \right) \right]$$
$$Q_{MB}(s,a) = \sum_{s'} P^a_{ss'} \left[R^a_{ss'} + \gamma V_{MB}(s') \right]$$

Note that $V_{MB}(s) = \sum_a \pi(s,a) Q_{MB}(s, a)$. The model-based controller works by storing the state-action probabilities, $\pi(s,a)$, along with the state transition probabilities, $P^a_{ss'}$, and calculating the values of current and future states according to those probabilities and the values of the rewards to be obtained. For all runs, $\gamma = 1$. The action and transition probabilities are updated according to:

At $t=0$ (warning click):

$$P_{stim} = \frac{n_{stim}}{\sum n_{stim}}$$

After stimulus presentation, the counts of each stimulus type, n_{stim}, are updated:

$$n_{stim} \leftarrow \begin{cases} \beta\, n_{stim} + 1 & \textit{for observed stimulus} \\ \beta\, n_{stim} & \textit{otherwise} \end{cases}$$

After an action is performed and reward (or no reward) is delivered, the state-action probabilities and probabilities of reward in each state are updated:

$$\pi(s,a) \leftarrow \begin{cases} \beta\, \pi(s,a) + (1-\beta) & \textit{for chosen action} \\ \beta\, \pi(s,a) & \textit{otherwise} \end{cases}$$

$$P(Rew \mid a,s') \leftarrow \begin{cases} \beta\, P(Rew \mid a,s') + (1-\beta) & \textit{for rewarded transition} \\ \beta\, P(Rew \mid a,s') & \textit{for unrewarded transition} \end{cases}$$

Above, $\pi(s,a)$ is a stored value indicating the probability of performing action a in state s, updated after the agent experiences a state-action pair. Selection of an ultimate action by the model-based system, by implementing a policy $\pi_{MB}(s,a)$, is discussed below. All initial stimulus counts were set to 0, as were the initial probabilities for reward in each state. The probabilities of all actions were initialized to a uniform distribution across all available actions. For all runs, all $\beta = 0.99$.

A softmax selection rule was used to determine the probability of selecting each action, based on the Q-values for the possible state-action pairs at t = 1:

$$\pi_{MB}(s,a) = \frac{e^{Q_{MB}(s,a)/\tau}}{\sum_b e^{Q_{MB}(s,b)/\tau}}$$

4.3.1.4. Arbitrating between the model-free and model-based controllers

Two schemes were developed to combine the two controllers. For model 1, in which we envision that the dorsomedial striatum is related to computing action-values according to a model-free approach, we simplified the arbitration scheme such that a final action was chosen according to the combined probabilities from the two controllers:

$$\pi_{RAT}(s,a) = \frac{e^{[\,(\pi_{MF}(s,a) + \kappa\, \pi_{MB}(s,a))\,/\tau\,]}}{\sum_b e^{[\,(\pi_{MF}(s,b) + \kappa\, \pi_{MB}(s,b))\,/\tau\,]}}$$

Reflecting the idea that the model-based controller is computationally expensive to operate, and should be decreasingly active as the model-free controller becomes more strongly activated, we also added a gain parameter, $\kappa = 1 - |Q_{MF}(s, a)|$. This allows the action selection to be increasingly determined by the model-free controller as training progresses.

For model 2, we envision that the dorsomedial striatum is involved in arbitrating between the two controllers. For this scheme, we use a softmax selection rule to determine the influence each controller exerts in the ultimate selection of an action.

$$\Pi_x(s, \pi_x) = \frac{e^{p_x(s, \pi_x)\,/\tau}}{\sum_y e^{p_y(s, \pi_y)\,/\tau}}$$

Above, $p_x(s, \pi_x)$ represents the propensities of selecting each controller. For the model-free controller, $p_{MF}(s, \pi_{MF})$ was incrementally updated following action selection toward the current absolute value of $Q_{MF}(s, a)$; for the model-based controller, $p_{MB}(s, \pi_{MB})$ was updated toward $(\kappa\, Q_{MB}(s, a))$.

$$p_{MF}(s, \pi_{MF}) \leftarrow \beta\, p_{MF}(s, \pi_{MF}) + (1 - \beta)\, |Q_{MF}(s, a)|$$

$$p_{MB}(s, \pi_{MB}) \leftarrow \beta\, p_{MB}(s, \pi_{MB}) + (1 - \beta)\, \kappa\, Q_{MB}$$

Finally, the probabilities determined above were used to bias the contribution of each controller in the final selection of an action. This has the same effect as the simplified rule of model 1, but enforces analogous architecture in the model-free, model-based and arbitration systems such that each of the three components may map onto cortico-basal ganglia loop architecture in a similar manner.

$$\pi_{RAT}(s,a) = \frac{e^{[\,(\Pi_{MF}\, \pi_{MF}(s,a) + \Pi_{MB}\, \kappa\, \pi_{MB}(s,a))\,/\tau\,]}}{\sum_b e^{[\,(\Pi_{MF}\, \pi_{MF}(s,b) + \Pi_{MB}\, \kappa\, \pi_{MB}(s,b))\,/\tau\,]}}$$

4.3.2. Simulation results

We tested the models under a variety of conditions, representing several common experimental paradigms. First, we ensured that the models could adequately reproduce the behavioral performance observed for the T-maze task used in Chapter 2. Next, we tested the models under devaluation and lesion conditions, to ensure previous behavioral results could be reproduced and to predict activation patterns in the three component systems under such conditions. The results of these experiments are described in the following sections.

4.3.2.1. The models reproduce rodent T-maze learning

Recall from Chapter 2 that the five rats in Group 1 failed to acquire the tactile cues, but began to perform above the 72.5% correct criterion on the auditory task version in an average of 13 training

sessions. By contrast, the three rats in Group 2 acquired the auditory task version in an average of 16 sessions, and the tactile version in an average of 23 sessions. We suggest that the difference in learning abilities in the two groups is the result of a failure by the Group 1 rats to detect the tactile stimuli, modeled here as a low probability of detection for the tactile cues ($p^*_{det,rough} = p^*_{det,smooth} = 0$). The slower learning of the tactile cues by the Group 2 animals was modeled by a larger τ_{stim} value for the tactile cues than for the auditory ($\tau_{8k} = \tau_{1k} = 500$; $\tau_{rough} = \tau_{smooth} = 1500$). Interestingly, the Group 1 rats that failed to acquire the tactile cues responded with similar turn probabilities during the tactile trials and during the 1 kHz tone trials, suggesting there was some confusion among the 3 stimulus types. The Group 2 rats showed a similar trend prior to acquiring good performance on the tactile cues. The model behaves similarly if the probability of detecting the 1 kHz tone is low compared to the probability of detecting the 8 kHz tone ($p^*_{det,8kHz} = 1$ and $p^*_{det,1kHz} = 0.7$ were used unless otherwise noted). **Figure 4.2** shows learning curves for a typical simulated Group 1 and Group 2 rat using these parameters for Model 1 and Model 2.

4.3.2.2. Parameter choices affect model component activation patterns

As shown in **Figure 4.2**, the models reproduce behavior for Group 1 and Group 2 rats.

Both models predict that as training progresses, the model-free controller comes to select the correct action in each state and deselect all other actions with increasing likelihood. The distribution of value across the available actions thus becomes increasingly non-uniform as the value of the correct action converges toward the expected value of reward and the value of all incorrect actions converges toward the negative of this value. The dynamics of the activation of the model-free controller across training depend critically on the selection of the learning rate α. Here, we have tuned α to replicate the behavior of the rats under a devaluation paradigm: α was set such that upon initially reaching criterial performance, the Q-values had not yet reached saturation, but after several days of overtraining, these values had saturated. As discussed further below, this results in a rapid change in behavior if one of the rewards is devalued after initial acquisition, but only a gradual change in behavior if devaluation occurs after extended training.

The activation of the model-based controller depends critically on the gain function, κ. Our selection for κ here is somewhat arbitrary, but was designed to capture the idea that as the habit system becomes increasingly activated, the computationally intense calculations of the model-based system become less likely to be performed. As we have chosen a function for κ that depends on the Q-values of the model-free system, activation of the model-based system thus also critically depends on the model-free learning rate α.

For both models, we assume that the activation of the model-free controller depends on the Q-values computed by the system for all state-action pairs, determined in part by the number of potentially activated actions and the strength of their activations. Likewise, the strength of the model-based controller depends on the Q-values computed by that system, determined by the state transition probabilities, state-action probabilities, the estimated values of future states, the strength of all these component activations, and the gain κ of the model-based system. Finally, the activation of the arbiter in model 2 depends on the number of available controllers and the strength of their activations. **Figure 4.3** shows the Q-values computed by the model-free and model-based controllers in both models, as well as the propensities associated with each controller and used by the arbiter in model 2. Note that both models recreate the general activation patterns reported in Chapter 2 for dorsolateral and dorsomedial striatum during learning of the T-maze task. In the model-free

controller, activation steadily strengthens with training. In the model-based controller, and the arbiter in model 2, activation initially strengthens, then declines for the simulated Group 2 rats that successfully acquire both auditory and tactile stimuli. For the simulated Group 1 animals, activation in the model-based controller remains elevated even late in training, resulting additionally in continuted competition and heightened activation in the model 2 arbiter.

While the models capture the average activation of dorsolateral and dorsomedial striatum, they fail to predict the equivalent activations observed during auditory and tactile trials in both dorsolateral and dorsomedial recordings, especially in the Group 1 rats that failed to learn the tactile cues (**Figure 4.4**). This is likely the result of several assumptions and simplifications, one of which may be the probability of detecting the 1 kHz stimulus. **Figure 4.4** compares the average activations across training for each component system during each of the four stimulus types. As the probability of detection of the 1 kHz tone decreases, the mean activation between auditory and tactile trials becomes more nearly equal. Psychophysical evidence suggests that this may be a plausible explanation: the 1 kHz tone lies at the boundary of audible frequencies for rats, whereas 8 kHz lies midrange (Kelly and Sally, 1988). While this is perhaps the simplest explanation for the equivalent activations, and is consistent with the behavioral performance of the rats, both models predict that even in the case of poor detection of the 1 kHz tone, stronger activation should occur during the trials in which the 8 kHz tone is presented than for all other stimuli. At an ensemble level, we failed to observe such preferential activation during 8 kHz trials, though there is perhaps some evidence that at the single unit level, neurons are more strongly activated to the 8 kHz tone (**Figures 2.5F**, **2.SF-G** and **2.SK-L**), especially in the dorsomedial striatum.

4.3.2.3. The models reproduce the results of previous lesion studies

The models were designed not only to reproduce learning and activation patterns in the T-maze task, but also to adequately capture devaluation results. **Figure 4.5** demonstrates that after only two days of overtraining on an auditory-only task, if the value of reward at one of the goals is devalued, the animal is more likely to use this information to direct behavior and reduce its tendency to make the previously-rewarded turn (**Figure 4.5A**). After extensive overtraining, in this case 25 sessions in which performance remained above 72.5% correct, the model-free controller has achieved saturated Q-values and the gain of the model-based system is low. Under these conditions, devaluation fails to elicit a change in behavior (**Figure 4.5B**). This is true for both Model 1 and Model 2, though Model 2 more robustly demonstrates this effect.

In addition to these studies, which the model was designed to replicate, the model offers an explanation for the more surprising findings of Atallah et al. (2007). This group showed that if dorsal striatum was temporarily inactivated during training by injecting muscimol into a dorso-central location, rats showed no improvement during training, but performed nearly as well as controls during a post-training test session in which no inactivation was present. We use the auditory-only version of the T-maze task to model their paradigm. We model a striatal lesion by setting the Q-values of both controllers to 0. When the model is trained under inactivation of both the model-based and model-free Q-value computations, it fails to improve its performance over time, though the state values are still adequately updated. The model reproduces the Atallah et al. results if one assumes that inactivation knocks out both the model-free and model-based controllers, but that learning of state values still occurs in the ventral striatum. During the test session, when the systems are intact, the rat is able to immediately access this stored state value information and use it to direct behavior via the model-based controller (**Figure 4.6**). Atallah et al. also showed that performing the same procedure in the ventral striatum impaired performance during both the training and test phases - a

result predicted by a failure to learn the state-values used by both the model-based and model-free controllers to correctly update or calculate their respective state-action values.

4.3.2.4. The models predict different activation patterns for the model-based system and the arbiter under lesion conditions

The two models perform similarly, and during the task paradigms explored above suggest similar activation patterns should arise within the model-based system and the arbiter. Only when the model-free system is unable to acquire task-relevant state-action values should the model-based system be activated above the model-free system without necessitating arbitration between the two controllers. This may occur when the actions required to obtain reward are not consistent or when the model-free system is impaired. Here, we investigate the latter of these situations by manipulating the models to represent various lesion conditions. We find that different activation patterns are suggested for the model-based system and the arbiter when the model-free system is impaired or absent.

Figure 4.7A-B shows the behavioral performance of the models, along with the Q-values computed by each component system, during training on the two-version T-maze paradigm under conditions of inactivation of the model-free system. The model-based controller is able to direct improving performance on the task, and the simulation attains criterial performance (>72.5% correct on both auditory and tactile task versions for 10 consecutive sessions) after 30-35 sessions, somewhat slower than for the intact case. The model-based system is increasingly activated as the calculated Q-values come to be more accurate, and without competition from or overriding by the model-free system, these values remain high throughout training. By contrast, the arbiter remains inactive as there is no competition between the two controllers. Identical results were obtained when dopamine lesions to the dorsal striatum were simulated by setting $\delta_Q = 0$ (data not shown). Similarly, when lesions to the model-based system are made, the model-free system is nonetheless able to drive improving performance on the task (**Figure 4.7C-D**).

4.3.3. Discussion

We have suggested two hypotheses regarding dorsomedial pattern activation. The first of these proposes that the dorsomedial striatum may be involved directly in the computation of the model-based value estimations. The second proposes that the dorsomedial striatum may be involved in arbitrating between the model-based and model-free controllers. Above, we have shown that either of these functions may account for the waxing and waning of dorsomedial striatal activation observed during T-maze learning. However, the two models provide differing explanations for the similar training-related patterns of activation observed in the two systems.

Model 1 assumes that the dorsomedial striatum is directly involved in the calculation of Q-values according to a model-based approach. Here, initial activity is low due to a lack of familiarity with the task construct. In the middle stages of training, a model of the world is developing and activation of the model-based controller increases. Finally, in the later stages of training, activation of the model-based system is reduced as the model-free system comes to dominate. By contrast, Model 2 assigns an arbitration role to the dorsomedial striatum. In this model, initial activation in the arbiter is low as the activation of the model-free system is low and behavior is biased toward use of the model-based controller. As training continues, both the model-free and model-based values gain strength and come to compete for behavioral control. With extended training, the model-based controller inactivates and competition between the two likewise declines as behavior becomes biased toward the model-free system.

Critical for distinguishing the two possibilities, the two models make differential predictions for activation of the model-based system versus the arbiter after lesions of the model-free system. In this case, the model-based system should be reactivated to direct behavior appropriately, but competition between the two systems should remain low. Thus, if the dorsomedial striatum computes the value of acting according to a model-based approach, it should be reactivated under lesion conditions when the model-free controller is absent. By contrast, if the dorsomedial striatum is involved in arbitrating between controllers, it should remain inactive following dorsolateral lesions, as there is no competition between the two controllers. Such experiments will thus be important for determining which (if any) model best captures the functions of the dorsomedial striatum.

4.3.3.1. Model assumptions, performance and potential biological implementations

4.3.3.1.1. Neural tuning of critical model parameters

As noted in Section 4.3.2.2, the activation of the model components depends critically on the rate at which the model-free controller increments its Q-values, determined by the learning rate parameter α. Additionally, we have assumed that the model-based controller becomes inactive according to a gain parameter κ, which decreases as training progresses. These parameters were hand-tuned for the simulations described above, but in this section we discuss more realistic approaches.

For simplicity, we selected a constant learning rate α in our simulations, and tuned this parameter such that the model-free system had not yet reached saturated Q-values when criterial performance was initially acheived, but had reached saturation after extended overtraining. A more realistic approach would be to tune α according to the uncertainty of the model-free controller, which would necessarily decrease with further experience in the task. Early in training, a larger α would enable more flexible behavior and faster devaluation, whereas later in training, the smaller α would contribute to the inflexibility of behavioral performance. This approach to tuning the α parameter was explored by Daw et al. (2006) for behavioral and neural data. They used a Kalman filtering approach to determine the uncertainty and optimal gain of a model-free RL system (for a related review, see also (Dayan et al., 2000). These authors suggest that modulation by the acetylcholine system may reflect expected uncertainty during learning, providing a potential mechanism for the modulation of a learning rate. The cholinergic neurons intrinsic to the striatum, with their strong interconnections with the dopamine system, may perform an analogous function for this region.

In our model, we used a simple inverse relationship between the Q-values computed by the model-free system to compute the gain of the model-based system: $\kappa = 1 - Q_{MF}$. We suggest several reasons why the activation of the model-based system might depend on the activation of the model-free controller in a biological-based system. The first of these is related to the difference in resource consumption by the two systems. The computations performed by the model-based system are expensive to perform. This is especially true in the more general case, when the model-based controller may look ahead through many available next states to try to determine the best course of action, or when potential states many steps into the future may be explored. From an energy consumption standpoint, there are thus substantial savings to be gained by not performing these computations when they are no longer needed. By contrast, the model-free controller is relatively inexpensive to operate, and therefore should be deployed as long as it can accurately direct behavior. From a speed perspective, the model-based computations take substantial time to compute, especially the farther ahead in time for which the search is performed. If the model-free system "knows" what

to do, it may signal the appropriate action before the model based system has time to settle on a final answer. This suggests a natural mechanism by which the model-based controller might become less active as the model-free controller gains "confidence." The second reason is related to the more practical considerations of a neural implementation of the computations performed by the two systems. Updating of the model-free Q-values according to the reward prediction errors requires reactivation of the current value in addition to the prediction error, suggesting that the model-free system cannot be shut down if it is to be properly updated. By contrast, computation of the Q-values according to the model-based approach relies on estimates of state values and transition probabilities that may be updated apart from the final computation of the Q-values. Thus, in practical terms, the model-free system must remain active to be properly updated, whereas the model-based system can be inactive at the level of Q-value calculation as long as the model and state values continue to be incrementally updated.

4.3.3.1.2. Parallel architecture of model-based, model-free and arbitration systems and implications for neural implementations

The critical issue for any model of dorsal striatal function is how regions with similar architecture can perform what appear to be substantially differing functions. In the two models presented above, we have constructed the three component systems in such a way as to emphasize the parallel nature of the respective computations. **Figure 4.8** illustrates potential biological mappings for the two models, discussed in more detail below.

The anatomical mapping of the two models onto neural substrates relies on the known functions and connections of multiple brain regions. For the model-free system representing the sensorimotor cortico-basal ganglia loop (including the dorsolateral striatum), the functionality and connections of the somatosensory and motor systems are relatively well-established. For this loop, motor and somatosensory cortical areas project onto dorsolateral striatal sites, which project via direct and indirect pathways to GPi/SNr. The pallidum then sends feedback projections through the thalamus back to the somatosensory and motor cortices, and additionally sends projections to the brainstem. Bidirectional dopamine-driven learning occurs at the cortico-striatal synapses driven by SNc projections to the striatum. Considerably more confusion surrounds the functions of various prefrontal cortical areas, and as reviewed in Chapter 1, some controversy exists regarding the mapping of even the best-understood prefrontal regions in primates onto analogous sites in the rat. We thus propose plausible mappings for each model based on what is known. For both models, we suggest that the model-based planning system engages hippocampal-prefrontal circuitry implicated in working memory. The hippocampus is interconnected with ventral striatum, and projects strongly to prelimbic cortical areas, and this network has been implicated in memory manipulation, perhaps in the service of planning according to a model-based control scheme. As discussed in Section 1.5.1.3, other cortical areas, especially the orbitofrontal cortex, may be additionally involved in the calculation of state-value information. These cortical areas have been shown to project strongly to ventral striatal regions, but project also to the dorsomedial striatum. In Model 1, we suggest that these projections from prelimbic cortex to the dorsomedial striatum may engage circuitry there in the calculation of state-action values based on input from the hippocampal-prefrontal planning system. Like the dorsolateral striatum, the dorsomedial model-based system then projects through direct and indirect pathways, and sends ultimate output projections both to the cortex and to the brainstem. The anterior cingulate cortex also projects strongly to the region of dorsomedial striatum from which we recorded. As reviewed in Chapter 1, the anterior cingulate is reciprocally connected to both sensorimotor cortex and to other prefrontal cortical regions and has been shown to activate strongly

during situations of high conflict or uncertainty. In Model 2, we suggest that the anterior cingulate cortex may be engaged in arbitrating between the competitive model-free and model-based systems. The projections from anterior cingulate cortex to dorsomedial striatum may thus engage the dorsomedial circuitry in this arbitration function. Via the feedforward brainstem projections and feedback cortical projections, we suggest that an anterior cingulate/dorsomedial striatum-based arbitration system could then bias the activation of the model-free and model-based controllers such that each dominates behavioral control when it is most strongly activated.

As in the conceptualization of Samejima and Doya (2007), we envision different learning rules are implemented by the components mapping onto cortical versus striatal regions. In the model-free system, the incremental δ learning rule dominates learning in the dorsolateral striatum. By contrast, in the model-based system, a different incremental learning rule that drives activations toward the currently experienced values, is thought to dominate in the cortex or cortico-hippocampal complex. This is meant to be analogous to the Hebbian learning commonly thought to be implemented by cortical ensembles. In the simple model presented here, we have made the additional simplifying assumption that striatal-based Q-learning dominates in the model-free system whereas cortical-based Hebbian learning dominates in the model-based system. This simplification is believed to be plausible based on the differential projections of the various neuromodulatory systems to different regions of striatum and cortex. Most notably, dopamine projections and D2 receptor expression are stronger in the dorsolateral striatum than in the dorsomedial striatum, suggesting that learning according to a δ-function may be stronger in the sensorimotor than in the associative loop. In the cortex, dopaminergic projections are targeted primarily toward prefrontal regions, suggesting a preference for Hebbian learning over δ-function based learning in the associative loop.

The implementation of this simplification was twofold. First, we included strong bidirectional updating of both selected and non-selected actions within the model-free system, enabling the selection of desired actions with higher probabilities than was possible in the model-based system (which did not update according to prediction errors). This bidirectional updating is an abstraction of the opposing actions of D1 and D2 dopamine receptors in the direct and indirect pathways, respectively. Second, we implemented Hebbian learning rules only in the model-based and arbitration systems, to update the estimates for state transition probabilities and controller activations, respectively. In a more realistic neural network implementation of these loops, the Hebbian versus TD-based learning rules need not be executed in such an all-or-none fashion. **Figure 4.9** illustrates how connections between neurons representing sensory and motor states may be simultaneously updated in the cortex according to the Hebbian update rule, and the combinations of those states may be updated according to a reward prediction error function within the striatum. Output from the basal ganglia may then prime the motor pathway via descending projections to brainstem (as was suggested by the simple abstract model presented in this chapter) and/or may modulate the updating of cortical activations via feedback projections through the thalamus. As the model-based and arbitration systems exhibit this same basic architecture, similar schemes may be imagined for these other components.

An important implication of this more complicated view relates again to the work of Atallah et al. (2007) In addition to the experiments described above, Atallah et al. also performed experiments in which rats were trained on the discrimination task without inactivation, but tested with either the ventral or dorsal striatum inactivated. Either lesion made only during the test phase had only a minor effect on performance. This result is predicted for test-session lesions to the ventral striatal state-value system, because during training, Q-values are learned by the dorsolateral striatum that can then drive performance even after ventral lesions. However, the maintenance of good performance

following lesions to the dorsal striatum is not currently predicted by the models presented in this chapter. We suggest that plasticity occurring in the cortex and/or brainstem during learning, which were not included in the simplified and abstract framework implemented here, may enable continued good performance on already-learned tasks in the absence of the dorsal striatum based model-free and model-based controllers.

Considering a neural network implementation of these systems may shed further light on the activation of the component systems across task time, which is difficult to capture using the more abstract models presented above. Activation of neurons in the dorsolateral striatum, which we envision calculates the values of actions for the model-free system, depends on the number of actions under consideration and the strength of their cortical representations, as well as the strength of the cortico-striatal connections. This suggests that the firing rate in the dorsolateral striatum should increase when more actions are possible (e.g. in the start block, or after goal reaching), or when more muscles are recruited to control an action (e.g. during turning). Across training, behaviors become more efficient, tuning the cortical representations. In conjunction with this cortical tuning, dopamine-mediated reinforcement strengthens the striatal activations that occur during the times when the reward prediction errors are the largest - presentation of warning click, and presentation of reward. This is precisely the pattern that we observe in the dorsolateral striatum as the rats acquire the T-maze task. In the dorsomedial striatum, which we suggest may calculate the value of acting according to a model-based approach, firing rates depend on the stored state transition probabilities, and the calculated values of future states. This suggests that activation in this region should be highest as the number of possible future states increases, which occurs as the rats approach cue onset and turning. Again, this is precisely what we observe experimentally. Considering the dorsomedial striatum as part of an arbiter between the two systems, similar activation should be observed, as both systems are strongly engaged mid-task as the model-based activity increases.

The consideration of the dorsomedial striatum as part of the model-based system has a number of advantages over the alternative that it is involved in arbitration between the two systems. The simplified 2-system scheme of Model 1 provides a straightforward mechanism by which the two systems may cooperate or compete in the control of behavior at the level of the brainstem and/or motor cortex, and a number of experiments have shown that both modes of interaction are possible, depending on the task paradigm (Balleine et al., 2007; Corbit and Janak, 2007; Whishaw et al., 2007; Yin and Knowlton, 2006). Model 2 provides the same functionality, at the expense of significant additional architecture. Moreover, the experimental results presented in Chapter 2 show that both the dorsomedial and dorsolateral regions of striatum exhibit similar proportions of neurons that differentiate between stimulus, action and trial outcome parameters. It is easier to map this similarity onto two systems that are both engaged in computing action value functions, as opposed to a dorsomedial system hypothesized to arbitrate between controllers. However, as hinted by devaluation results presented in **Figure 4.5**, Model 2 may more robustly capture some behavioral results, and may additionally provide more flexibility and tighter control of the interaction between the two controllers. It is important to note however, that we need not assign a single function to the dorsomedial or dorsolateral striatum, as both regions are large and in themselves heterogeneous. Yin and colleagues (Yin and Knowlton, 2004; Yin et al., 2005) as well as Corbit and Janak (2010) have shown dissociations between anterior and posterior dorsomedial striatal lesions, suggesting that a single "dorsomedial" function is a highly oversimplified view. Similar distinctions have been made for dorsal versus ventral sites within the lateral striatum, and anatomical projection patterns suggest that within dorsolateral striatum, functionality is likely to vary along the anterior-posterior axis as well.

Finally, our results show that the models fail to replicate the similarity in ensemble activations during auditory and tactile trials unless a failure to detect the 1 kHz tone is assumed. Under this assumption, however, the models predict enhanced activation during the presentation of the salient 8 kHz stimulus. We found no evidence of such enhanced ensemble activity by either the dorsolateral or dorsomedial ensembles during 8 kHz tone trials, though there is some evidence that at the single-unit level, the 8 kHz tone was strongly represented in comparison to the other stimuli. Though the experimental results in this regard are rather inconclusive, we suggest that the high variability in firing rates of striatal MSNs, combined with the low number of trials of each stimulus type, may have masked the preference for the 8 kHz tone in our analyses. It is additionally likely that the cortical representations are highly overlapping for all four of the presented stimuli, as all other sensory inputs are identical across all trials. This may additionally contribute to the similarity in ensemble activations seen in the striatum during auditory versus tactile trials.

4.3.3.2. Relation to previous work

As reviewed in Section 4.2, a number of authors have proposed that the dorsolateral striatum based loop is engaged in model-free reinforcement learning and behavioral control, whereas the dorsomedial striatum based loop is engaged in model-based reinforcement learning and behavioral control. Redish et al. (2008) provides an especially detailed outline of this general scheme, and its mapping onto neural substrates, which serve as an inspiration for the implementation developed in this chapter. However, neither this comprehensive review by Redish et al., nor the similar conceptualization presented by Horvitz (2009), includes an implementation of their model, nor do they relate their ideas to predictions about neural activity in the parallel model-free and model-based systems. The work presented here thus provides a significant extension to these previous frameworks by providing a concrete implementation which can be used to make predictions regarding the activation of each system as well as the interactions between them under various experimental paradigms.

Several authors have provided computational frameworks which account for the interaction between model-free and model-based RL systems, without mapping these systems explicitly onto any specific neural architecture (Daw et al., 2005; Matsumoto et al., 2007; Samejima and Doya, 2007). We extend this work by relating the activations predicted in the model systems to the experimental results obtained from dorsolateral and dorsomedial striatum during T-maze learning. The models were designed to capture the main features of activation in the two regions, but they also reproduce the results of Atallah et al. (2007) for dorsocentral lesions, and make an unexpected prediction regarding the salience of the 1 kHz tone used in our experiments.

Daw et al. (2005) in particular suggested that uncertainty-based arbitration could account for the pattern of behavioral results observed in devaluation experiments. Here, we used a simplified arbitration scheme based simply on the strength of activation of the model-free and model-based systems. Behavioral and fMRI experiments have shown that humans are adept at tracking uncertainty and volatility within an environment, and using these parameters to adjust behavioral performance. An interesting extension of the model would thus be to incorporate the calculation and use of these high-level parameters into the implementation.

As envisioned in the models presented by these authors, and in the specific implementation put forward in this chapter, the ability to improve behavioral performance on a number of tasks requiring calculation of state-action values should be severely impaired by lesions encompassing both model-

based and model-free controllers. Conversely, the failure of partial lesions to impair performance can be attributed in most cases to the ability of the other system to compensate for the loss of functionality in the lesioned system. It is thus interesting that dorsolateral striatal lesions often result in a failure to acquire even simple stimulus-response discriminations, suggesting that compensation by the model-based system is not possible in these domains. The source of such failures by the model-based system is unclear in the standard conceptualizations of dorsolateral striatum based model-free versus dorsomedial striatum based model-based control, nor can they be accounted for by the implementations presented in this chapter. Exploring these shortcomings is thus an interesting avenue for future research.

The understanding of striatal function according to traditional psychology and neuroscientific research has benefited from extensive neuroanatomy and behavioral studies, but has often lacked a more formal framework for understanding neural computation. On the other hand, the computational formality of reinforcement learning provides a framework for integrating model-free and model-based approaches to learning, but can be far removed from any neural implementation. Thus, despite the simplicity of the modeling work presented in this chapter, it nonetheless provides an important link between a conceptual understanding of brain function and the computational description of how these functions might arise within neural architectures.

4.3.4. Modeling summary

Both models presented above suggest that the appearance of dorsomedial striatal activation is indicative of a goal-directed behavioral strategy, and that behavior becomes habitual only once this pattern of activation has subsided. In model 1, we propose that the dorsomedial striatum is actively engaged as part of the goal-directed controller. This model suggests that activation at the choice point reflects the computational load required in conducting a search for the appropriate action, and that the engagement of this system increases initially as a model of state transitions develops, and decreases in overtraining as the model-free system takes over. In model 2, we suggest that the dorsomedial striatum may be involved in arbitrating between the model-based and model-free controllers, resulting in the biasing of action selection toward the controller that is most strongly active. This model suggests that the enhanced activity during the decision period may reflect the enhanced competition between the two systems around the choice point, when neither the values of the available actions nor the value of the potential goal states have been fully determined. Importantly, the two models make different predictions regarding the pattern of activation that should be observed in the dorsomedial striatum following lesions to the dorsolateral striatum-based habit system. Model 1 predicts that with the reinstatement of goal-directed behavior in this case, activity should reappear in the dorsomedial striatum. Model 2 suggests that the lack of competition between the model-based and model-free approaches should fail to activate a dorsomedial system responsible for arbitrating between the two controllers.

4.4. Reinforcement learning summary

In Section 4.1, we reviewed the fundamentals of reinforcement learning (RL), including Dynamic Programming (DP), Monte Carlo (MC) and Temporal Difference (TD) approaches to RL problems. We distinguished model-based approaches (e.g. DP) from those that are model-free (e.g. MC and TD), where the former include an explicit representation of each state in the environment and the transition probabilities between states. We saw that model-based approaches require more computational resources to implement, but have the advantage of being able to improve performance

based on simulated experience and can be used to evaluate future states and outcomes prior to experiencing them. TD learning in particular has attracted the interest of basal ganglia researchers as the concepts of incremental updating according to reward prediction errors, especially as implemented according to an actor-critic framework, may map well onto biological architectures.

In Section 4.2, we reviewed how reinforcement learning has been applied to the study of brain function, in particular how it may provide a framework for understanding the function of the basal ganglia. These applications began with the development of the Rescorla & Wagner δ rule to model animal learning in classical conditioning experiments, but exploded with the discovery that dopamine neurons fire phasically in a manner consistent with the encoding of a reward prediction error. Since this latter discovery, RL has been increasingly applied in models of striatum-based trial-and-error learning with DA-mediated reinforcement. The TD framework has been usefully applied to make predictions regarding the magnitude of phasic responses by DA neurons, and to model variation in behavioral performance during procedural learning. In the current conceptualization of how a TD framework might be represented neurally, an actor-critic implementation is most commonly envisioned. Here, the ventral striatum learns state-values, which are used by the DA neurons of the SNc to compute a prediction error signal. This error signal is then the "critic" used to update both the state values in the ventral striatum and the separately maintained policy stored by the "actor" in the dorsolateral striatum. This idea has provided the inspiration for a number of neural network implementations, but the role of the dorsomedial striatum remains undefined in this framework. A number of authors have hypothesized that the dorsomedial striatum may be engaged in performing model-based RL, in contrast to the model-free RL assigned to the dorsolateral and ventral striatal systems. However, confusion remains regarding the biological mapping of model-free and model-based reinforcement learning algorithms. Generally, the model-free system is thought to be localized in the dorsolateral striatum, and the model-based system is localized in the prefrontal cortex and/or hippocampus, again leaving the role of the dorsomedial striatum undefined.

In Section 4.3, we presented two reinforcement learning-based models that may explain the dorsomedial activation during training on the T-maze task presented in Chapter 2. The first of these models proposes that the dorsomedial striatum is directly involved in a model-based system implementing goal-directed behavioral control. The second model proposes that the dorsomedial striatum may be involved in the arbitration between model-based and model-free controllers by biasing action selection toward the controller that is most strongly activated. These two models make opposing predictions regarding activation of the dorsomedial striatum following lesions to the dorsolateral striatum. The first predicts that the dorsomedial striatum will be reactivated during subsequent control by the goal-directed system, as the habit-based system is no longer available. The second predicts that no reactivation should occur, as there should be no resulting competition between the two systems.

Figures

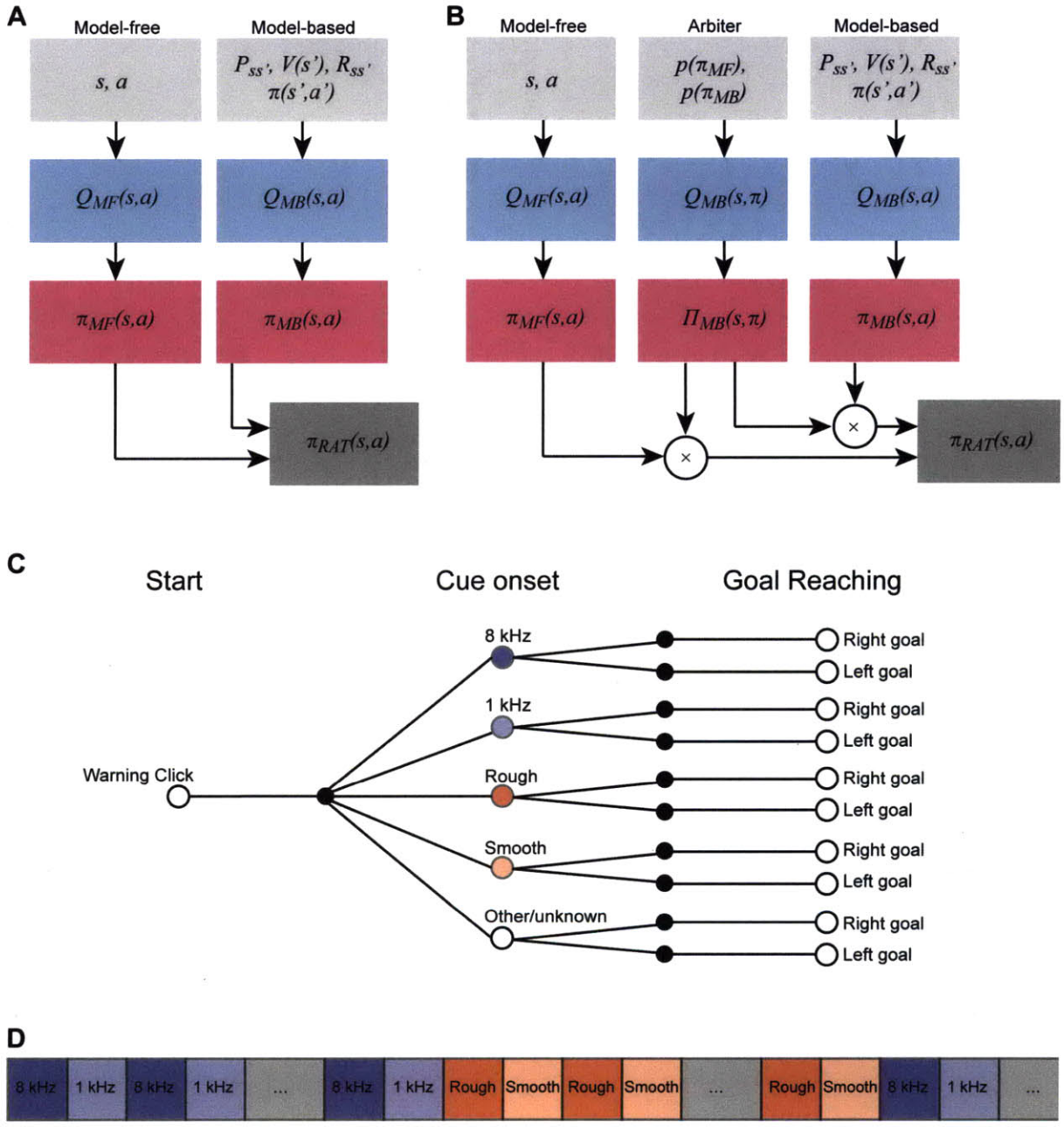

Figure 4.1. Using reinforcement learning to model learning in the T-maze task.
(A-B) Schematics of the proposed models. Model 1 (A) consists of a model-free and a model-based controller and uses a simplified interaction rule for ultimate action selection. Model 2 (B) is composed of the same model-free and model-based systems, and includes an additional arbitration system with architecture analogous to that of the model-based and model-free controllers. The arbiter then biases the two component systems in the final selection of an action.
(C) Backup diagram for simplified T-maze task.
(D) Order of cues presented to the models during training.

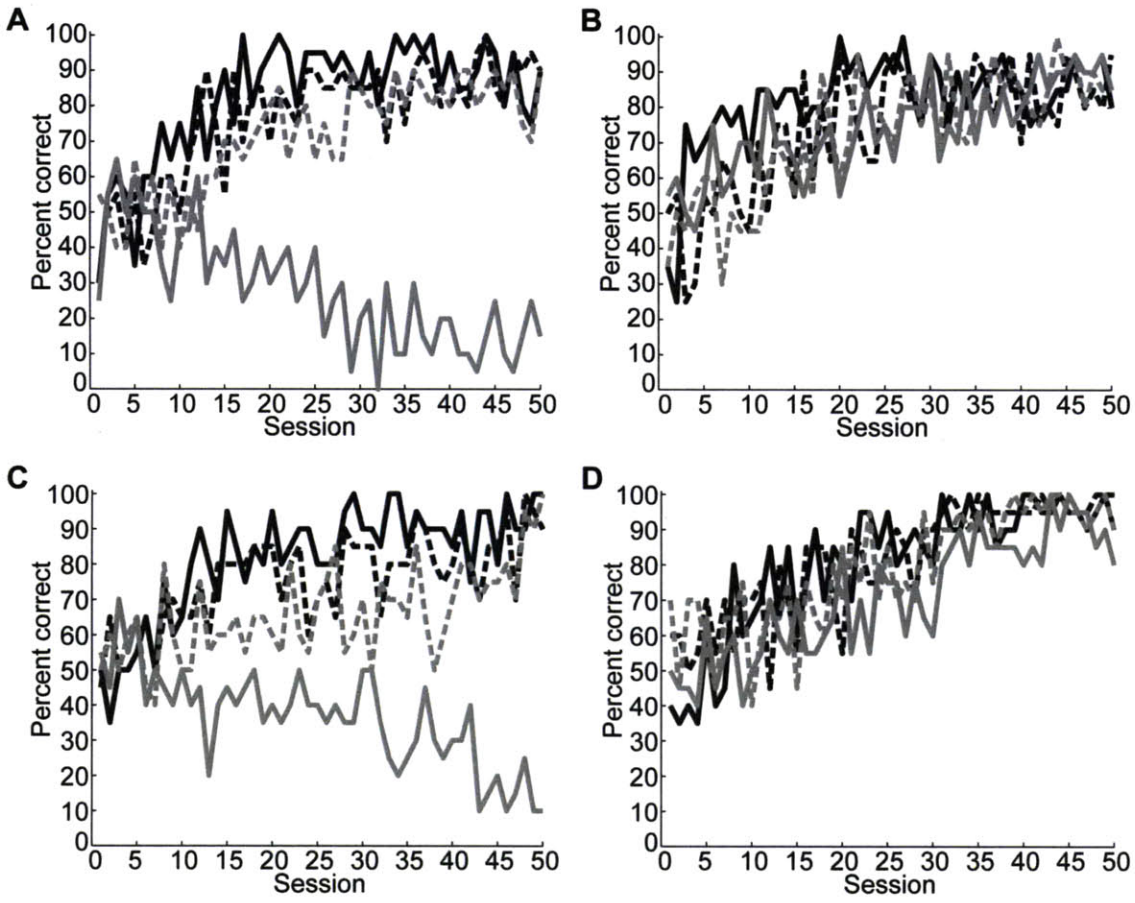

Figure 4.2. Behavioral performance of the models during T-maze learning
(A-B) Percent correct performance of Model 1 for a simulated Group 1 (A) and Group 2 (B) rat. Dark lines indicate performance during simulated 8 kHz (solid) and 1 kHz (dashed) auditory trials, light lines indicate performance during rough (solid) and smooth (dashed) tactile trials.
(C-D) Percent correct performance of Model 2 for a simulated Group 1 (C) and Group 2 (B) rat. Colors and line styles as above.

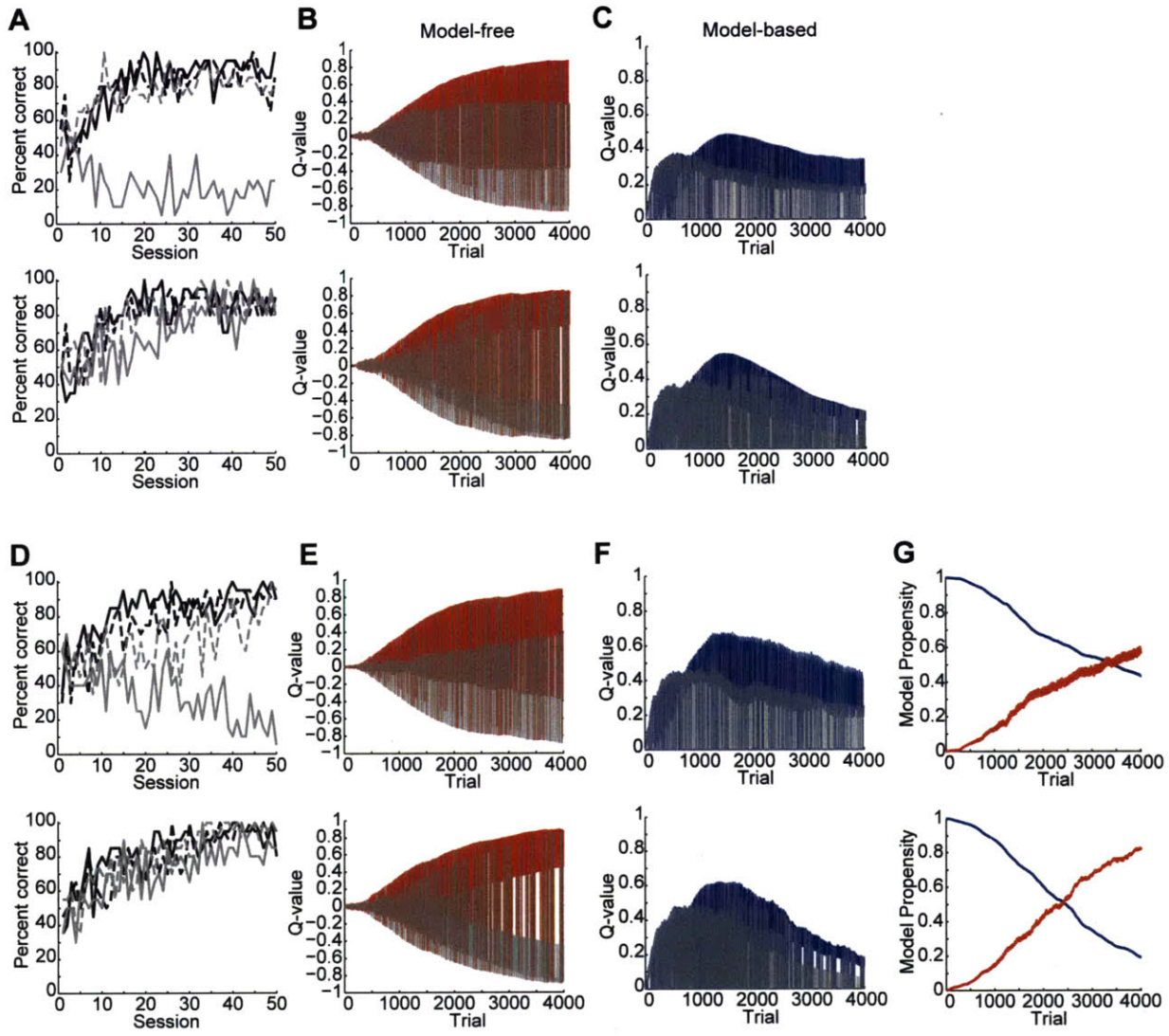

Figure 4.3. Activation of model component systems during T-maze learning.
(A-C) Model behavioral performance across training sessions (A, 80 trials/session), Q-values at stimulus onset for the model-free (B), and model-based (C) systems of model 1 during T-maze learning for Group 1 (top) and Group 2 (bottom) simulations. Q-values for the chosen action are shown in color, for the unchosen action in gray. Note that Q-values in the model-based system remain elevated late in training in the Group 1 simulations, but are reduced late in training for Group 2 simulations.
(D-G) D-F as in A-C for model 2. G shows the propensities for using the model-free (red) and model-based (blue) systems; competition is high between the model-free and model-based controllers when their propensities are nearly equal.

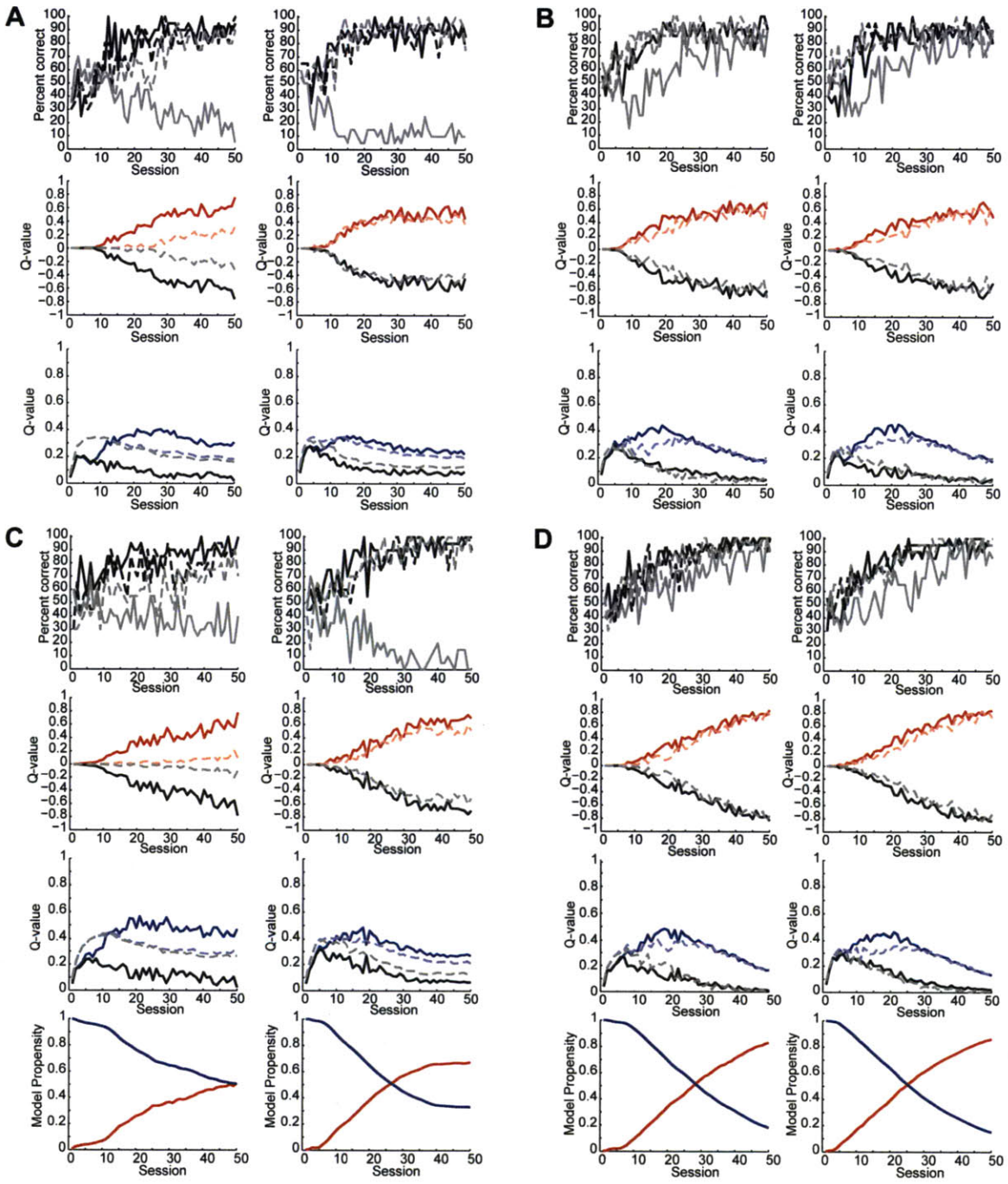

Figure 4.4. Performance of models for high and low probabilities of detection for the 1 kHz tone.
(A-B) Model 1 behavioral results (top), model-free Q-values (center), and model-based Q-values (bottom) for a simulated Group 1 rat (A) and a Group 2 rat (B) under conditions of high probability of detecting the 1 kHz tone (left, pdet,1kHz = 0.7) and low probability of 1 kHz detection (right, pdet,1kHz = 0). Solid dark lines indicate session-averaged Q-values computed during auditory trials, lighter dashed lines indicate the same for tactile trials, for chosen (color) and unchosen (grayscale) actions. Note that Q-values are approximately equal during auditory and tactile trials in the low-detection case, but not in the high-detection case.
(C-D) Top 3 rows are as in A-B for simulated Group 1 (C) and Group 2 (D) rats using model 2. Bottom row shows propensities for the model-free (red) and model-based (blue) controllers calculated by the input stage of the arbiter. Dark lines indicate propensities during auditory trials, light lines indicate propensities during tactile trials, which for all conditions are nearly identical for all trials.

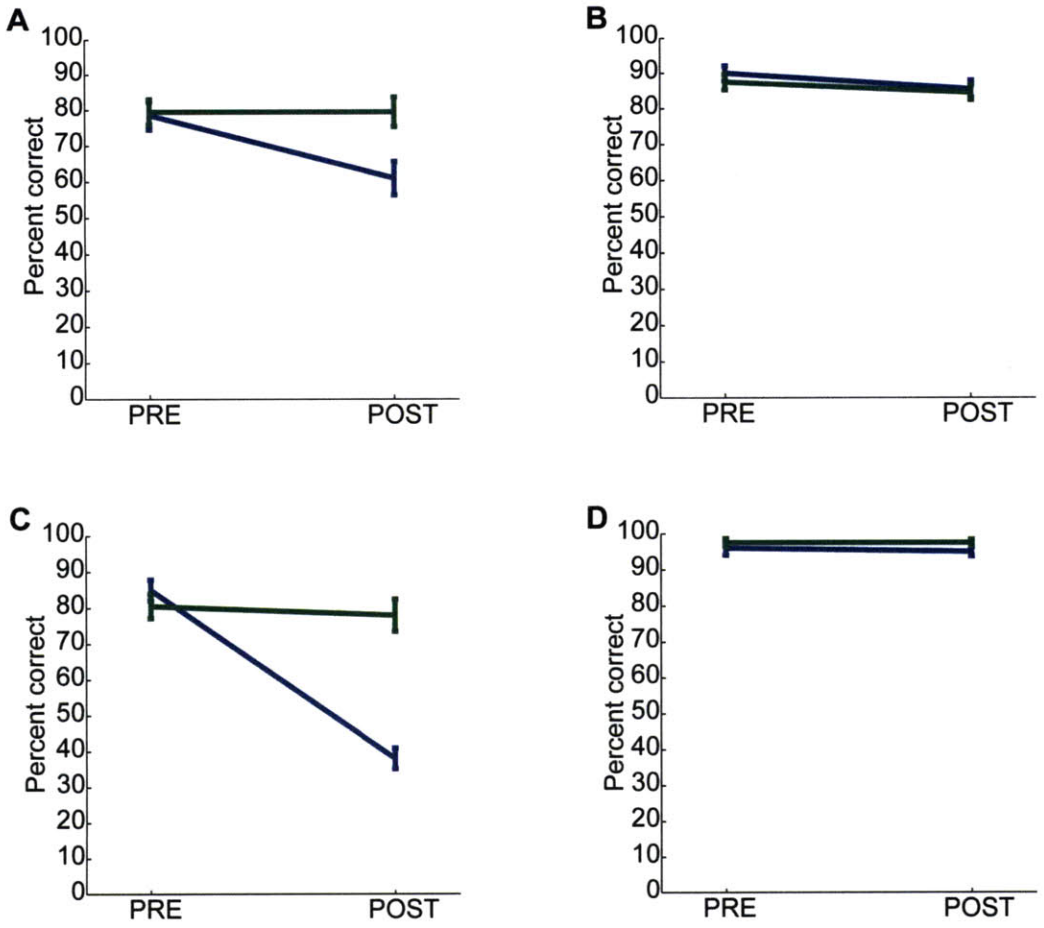

Figure 4.5. Performance of models under devaluation.
(A-B) Devaluation simulation results for model 1 for devaluations performed early (A) and late (B) in training. Percent correct performance during the session immediately preceding simulated devaluation (PRE) and the first session in which the reward associated with the 8 kHz tone was devalued (POST) for the devalued (blue) and still-valued (green) stimuli. Model was presented with only auditory stimuli and pdet,8kHz = pdet,1kHz = 1; early devaluation (A) was performed after 2 sessions of >72.5% correct performance; late devaluation (B) was performed after 25 days of >72.5% correct performance.
(C-D) As in A-B for model 2.

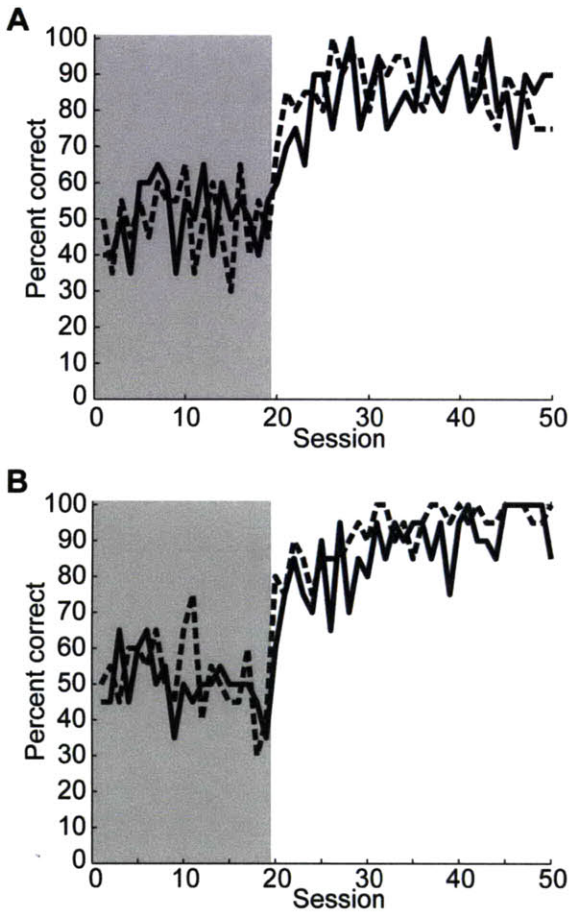

Figure 4.6. Performance of models under temporary inactivation.
(A) Model 1 behavioral results when both model-free and model-based Q-values are inactivated (set to 0) during initial 20 training sessions (gray shaded region). Gains are restored to normal on trial 20, at which point performance jumps to above chance, as observed in the dorsal striatal lesion experiments of Atallah et al (2007).
(B) As in A for Model 2.

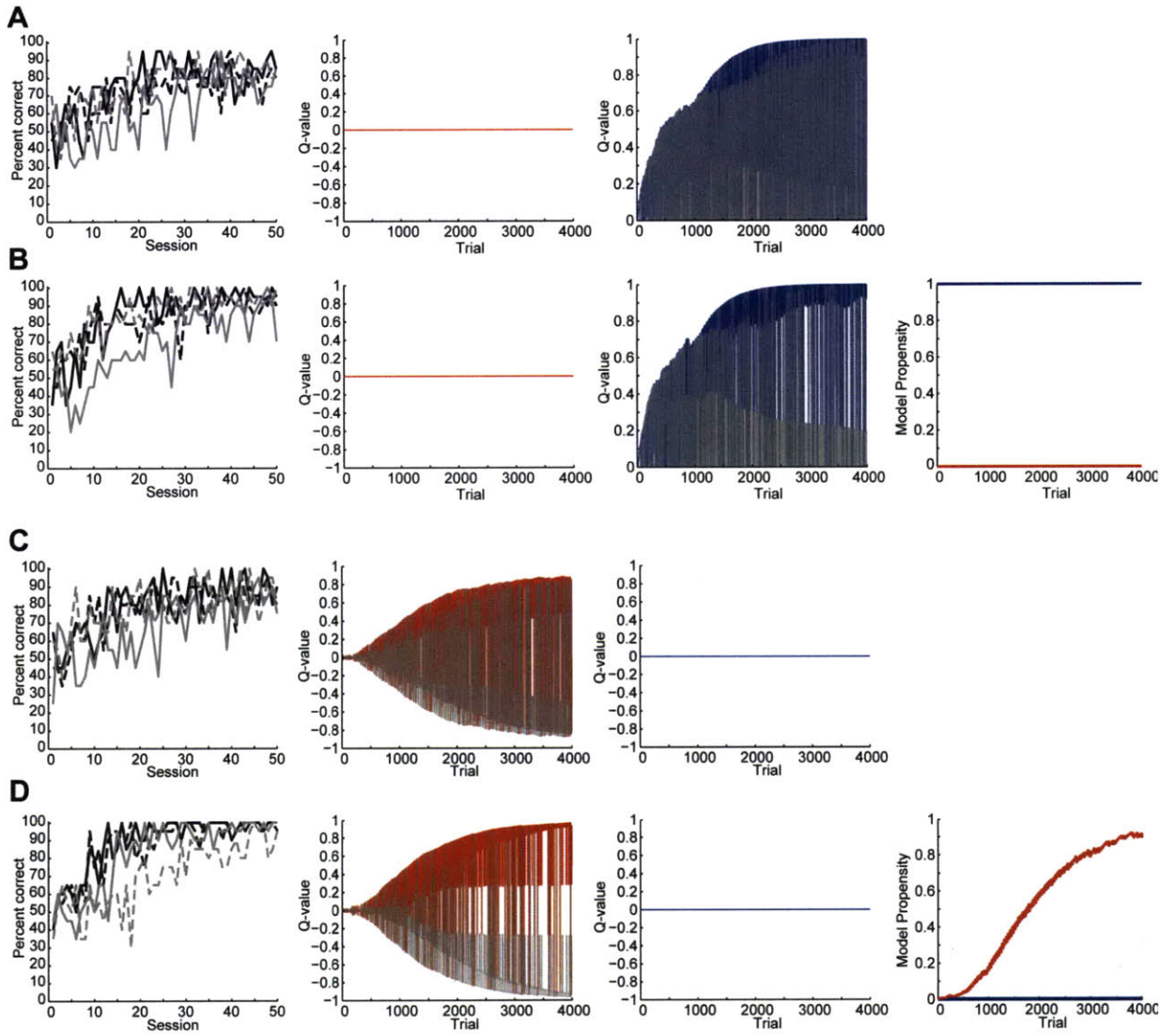

Figure 4.7. Modeled activation of model-based and arbitration systems under simulated lesions of the model-free or model-based systems.

(A-B) Model performance during simulated lesions of the model-free system for model 1 (A) and model 2 (B). Session-averaged percent correct performance during trials of each stimulus type (left), Q-values across trials computed for the chosen (color) and unchosen (gray) actions by the model-free (center left, red) and model-based (center right, blue) systems. For model 2, propensities for the model-free (red) and model-based (blue) systems are shown far right. Note that the model learns to perform correctly, and Q-values calculated by the model-based system remain high late in training. However, no competition exists between the two controllers.

(C-D) Model performance during simulated lesions of of the model-based system for model 1 (C) and model 2 (D). Conventions as in A-B.

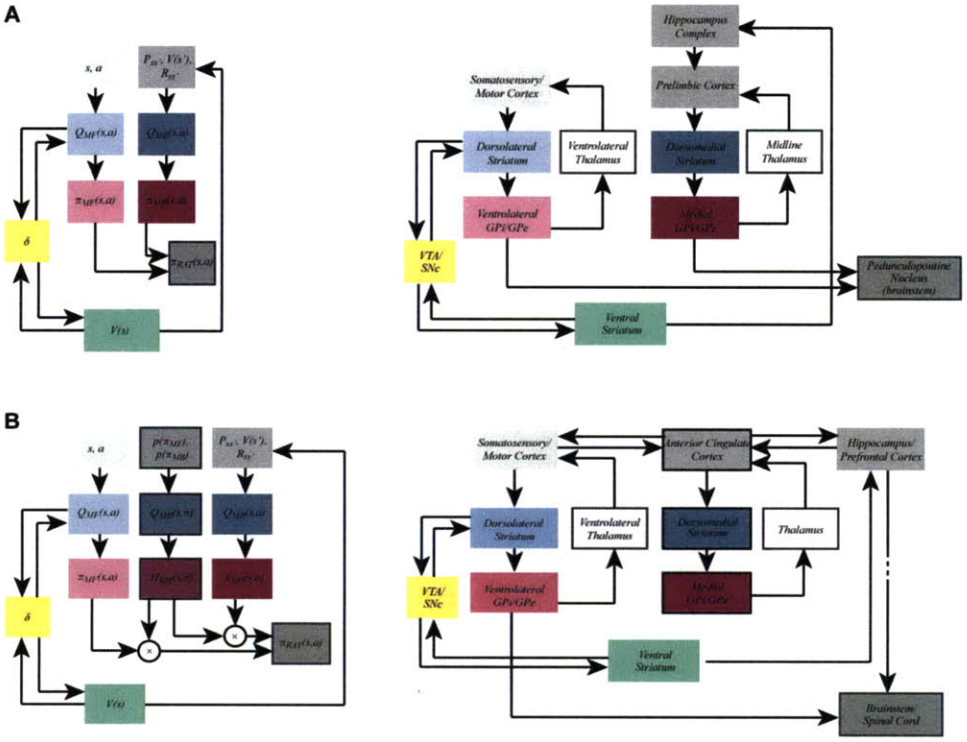

Figure 4.8. Suggested biological mapping for Models 1 and 2.
(A) Proposed mapping of the model-free and model-based systems of Model 1 onto sensorimotor neural circuitry (including the somatosensory/motor cortex and dorsolateral striatum) and associative neural circuitry (including the hippocampus, prefrontal cortex and dorsomedial striatum), respectively. Motor output can be modulated via brainstem connections from pallidal regions in both loops. Note that direct motor output from motor cortex via the pyramidal tract is not drawn, but may also be influenced by plasticity within the model-free system.
(B) Proposed biological mapping for Model 2. The model-free system is localized to the somtsensory and motor cortices and the dorsolateral striatum as in Model 1. The model-based system is localized to hippocampal-prefrontal cortex complex involved in working memory and planning. This system has its own route to influence motor output, indicated by the dashed line to brainstem and spinal cord. This system may involve yet another parallel cortico-basal ganglia loop (not drawn), including perhaps to more posterior regions of the dorsomedial striatum. Finally, both systems share reciprocal connections with anterior cingulate cortex, which could serve to bias either system such that it dominates the control of behavior according to task demands. Again, direct pyramidal tract projections to spinal cord and motor output have been omitted for clarity.

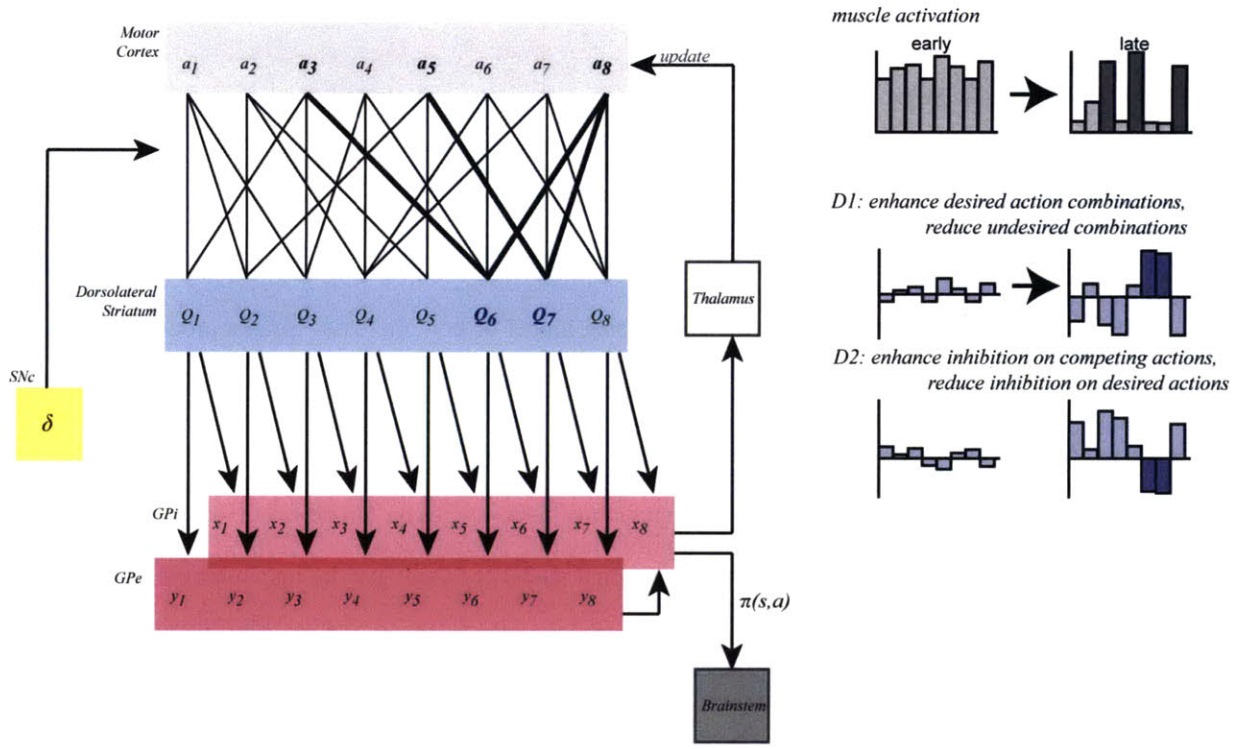

Figure 4.9. Neural network schematic for the model-free system.
Left: Motor cortical units a1..a8 represent activation patterns associated with specific muscles. These randomly converge and diverge onto striatal units Q1..Q8, such that activation of a combination of muscles may be needed to excite a given striatal unit. Each striatal unit depicted includes both D1 and D2 MSNs. Via dopamine-mediated reinforcement from the SNc, the corticostriatal synapses are modified during learning such that inputs onto D1 MSNs are enhanced for actions that result in reward, and reduced for actions that fail to result in reward. Conversely, inputs onto D2 MSNs are strengthened for competing actions such that they remain suppressed during movement. Q-values are amplified at the level of the pallidum, where enhancement of desired actions converges with suppression of undesired actions to increase the probability of selecting a single best motor program. Pallidal outputs then bias actions via output to brainstem nuclei projecting to the retrorubral spinal tract. Feedback projections through the thalamus may additionally contribute to proper updating of cortical ensembles such that over time, actions leading to reward become more efficient and automatic. Right: Early in training, muscle activation is uniform and inefficient, resulting in uniform baseline activation of striatal ensembles. Through the dual processes of Hebbian updating of the cortical ensembles and reinforcement-driven updating of striatal ensembles, neural activity is restructured with learning in both regions. Cortical ensembles become tuned, enabling more efficient production of movement. Striatal ensembles become simultaneously tuned based on the combinations of active cortical activity, reducing the probability that movement will be interrupted by competing programs.

References

Atallah, H. E., Lopez-Paniagua, D., Rudy, J. W., and O'Reilly, R. C. (2007). Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. Nat Neurosci *10*, 126-131.

Balleine, B. W., Delgado, M. R., and Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. J Neurosci *27*, 8161-8165.

Beiser, D. G., and Houk, J. C. (1998). Model of cortical-basal ganglionic processing: encoding the serial order of sensory events. J Neurophysiol *79*, 3168-3188.

Berns, G. S., and Sejnowski, T. J. (1998). A computational model of how the basal ganglia produce sequences. J Cogn Neurosci *10*, 108-121.

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

Botvinick, M. M., Niv, Y., and Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. Cognition *113*, 262-280.

Bullock, D., Tan, C. O., and John, Y. J. (2009). Computational perspectives on forebrain microcircuits implicated in reinforcement learning, action selection, and cognitive control. Neural Netw *22*, 757-765.

Cohen, M. X., and Frank, M. J. (2009). Neurocomputational models of basal ganglia function in learning, memory and choice. Behav Brain Res *199*, 141-156.

Corbit, L. H., and Janak, P. H. (2007). Inactivation of the lateral but not medial dorsal striatum eliminates the excitatory impact of Pavlovian stimuli on instrumental responding. J Neurosci *27*, 13977-13981.

Corbit, L. H., and Janak, P. H. (2010). Posterior dorsomedial striatum is critical for both selective instrumental and Pavlovian reward learning. Eur J Neurosci *31*, 1312-1321.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci *8*, 1704-1711.

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

Dayan, P., and Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. Neuron *36*, 285-298.

Dayan, P., Kakade, S., and Montague, P. R. (2000). Learning and selective attention. Nat Neurosci *3 Suppl*, 1218-1223.

Doya, K. (2008). Modulators of decision making. Nat Neurosci *11*, 410-416.

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. Neurobiol Learn Mem *70*, 119-136.

Haruno, M., and Kawato, M. (2006). Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. Neural Netw *19*, 1242-1254.

Hazy, T. E., Frank, M. J., and O'Reilly, R. C. (2006). Banishing the homunculus: making working memory work. Neuroscience *139*, 105-118.

Histed, M. H., Pasupathy, A., and Miller, E. K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. Neuron *63*, 244-253.

Horvitz, J. C. (2009). Stimulus-response and response-outcome learning mechanisms in the striatum. Behav Brain Res *199*, 129-140.

Houk, J. C., and Wise, S. P. (1995). Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. Cereb Cortex *5*, 95-110.

Kaelbling, L. P. (1993). Hierarchical learning in stochastic domains: Preliminary results. Paper presented at: Tenth International Conference of Machine Learning (San Mateo, CA).

Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research, 237-285.

Kelly, J. B., and Sally, S. L. (1988). Organization of auditory cortex in the albino rat: binaural response properties. J Neurophysiol *59*, 1756-1769.

Lau, B., and Glimcher, P. W. (2008). Value representations in the primate striatum during matching behavior. Neuron *58*, 451-463.

Lo, C. C., and Wang, X. J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. Nat Neurosci *9*, 956-963.

Matsumoto, K., Matsumoto, M., and Abe, H. (2006). Goal-based action selection and utility-based action bias. Neural Netw *19*, 1315-1320.

Matsumoto, M., Matsumoto, K., Abe, H., and Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci *10*, 647-656.

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci *16*, 1936-1947.

Nakamura, K., and Hikosaka, O. (2006). Role of dopamine in the primate caudate nucleus in reward modulation of saccades. J Neurosci *26*, 5360-5369.

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

O'Reilly, R. C., Frank, M. J., Hazy, T. E., and Watz, B. (2007). PVLV: the primary value and learned value Pavlovian learning algorithm. Behav Neurosci *121*, 31-49.

Pasquereau, B., Nadjar, A., Arkadir, D., Bezard, E., Goillandeau, M., Bioulac, B., Gross, C. E., and Boraud, T. (2007). Shaping of motor responses by incentive values through the basal ganglia. J Neurosci *27*, 1176-1183.

Redish, A. D., Jensen, S., and Johnson, A. (2008). A unified framework for addiction: vulnerabilities in the decision process. Behav Brain Sci *31*, 415-437; discussion 437-487.

Rescorla, R. A., and Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Classical Conditioning II: Current Research and Theory, A. H. Black, and W. F. Prokasy, eds. (New York, Appleton Century Crofts), pp. 64-99.

Reynolds, J. R., and O'Reilly, R. C. (2009). Developing PFC representations using reinforcement learning. Cognition *113*, 281-292.

Samejima, K., and Doya, K. (2007). Multiple representations of belief states and action values in corticobasal ganglia loops. Ann N Y Acad Sci *1104*, 213-228.

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

Sutton, R. S., and Barto, A. C. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. Psychological Review *88*, 135-140.

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

Takahashi, Y., Schoenbaum, G., and Niv, Y. (2008). Silencing the critics: understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. Front Neurosci *2*, 86-99.

Whishaw, I. Q., Zeeb, F., Erickson, C., and McDonald, R. J. (2007). Neurotoxic lesions of the caudate-putamen on a reaching for food task in the rat: acute sensorimotor neglect and chronic qualitative motor impairment follow lateral lesions and improved success follows medial lesions. Neuroscience *146*, 86-97.

Wickens, J. R., Budd, C. S., Hyland, B. I., and Arbuthnott, G. W. (2007). Striatal contributions to reward and decision making: making sense of regional variations in a reiterated processing matrix. Ann N Y Acad Sci *1104*, 192-212.

Yin, H. H., and Knowlton, B. J. (2004). Contributions of striatal subregions to place and response learning. Learn Mem *11*, 459-463.

Yin, H. H., and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. Nat Rev Neurosci *7*, 464-476.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2005). Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. Eur J Neurosci *22*, 505-512.

Zhang, J., Berridge, K. C., Tindell, A. J., Smith, K. S., and Aldridge, J. W. (2009). A neural computational model of incentive salience. PLoS Comput Biol *5*, e1000437.

CONCLUSIONS AND FUTURE WORK

The work presented in this thesis provides evidence that multiple forebrain structures are simultaneously active during the decision-making process, and that these structures likely work together to produce behavior. We have shown that the dorsolateral and dorsomedial striatum are differentially engaged during T-maze learning, and have proposed specific computational functions for these regions that can explain the patterns of activation expressed in these regions. Further, we have shown that structures specialized for different types of learning and memory, the dorsal striatum and the hippocampus, are simultaneously engaged and highly coordinated during task performance in animals that successfully learn the T-maze task. While these results provide insight into the neural control of behavior by multiple simultaneously active learning and memory systems, many questions remain.

The firing of striatal medium spiny neurons results from a complex interaction of excitatory inputs from cortex and thalamus as well as the neuromodulatory activity of various interneurons and dopaminergic inputs. It remains unknown how these different components contribute to the development of the specific patterns of neural activity observed in dorsolateral and dorsomedial striatal recording experiments presented in Chapter 2. We captured the activities of striatal interneurons during these experiments, and the analysis of the firing patterns of these cells is likely to shed light on the region-specific processing by striatal microcircuits. Moreover, dopamine is known to play a crucial role in synaptic plasticity in the dorsal striatum, and its role in oscillatory activity is currently under investigation. Ongoing experiments in the lab combining electrophysiological recording with the recording and manipulation of dopamine signaling in the striatum will be particularly important in determining the role of dopamine in the function of striatal microcircuits.

At the theoretical level, computational learning theory is increasingly providing insight into brain function. Implementation of reinforcement learning-based models of dorsolateral and dorsomedial interaction, such as those presented in Chapter 4, should shed light on the mechanisms required for these two systems to produce both normal and pathological behaviors. By providing novel predictions and testable hypotheses, such modeling work plays a critical role in increasing our understanding of basal ganglia involvement in motor control and decision-making processes. Performing the lesion experiments proposed in Chapter 4 will be crucial in determining the validity of the assumptions and simplifications made in constructing the models. At least two extensions to this framework should then be made such that the model can provide further insights into the dynamic interactions of multiple learning and memory systems. First, a neural network implementation of the model would enable the implications of simultaneously operating cortical and striatal learning mechanisms to be explored, and make specific predictions regarding the neural activity in parallel striatal loops under more complex and realistic operating regimes. Second, a network model that incorporates the temporal dynamics of neural transmission will be critical for investigating the mechanisms by which one system may dominate the other in the competition for control of behavior.

Combined, future experimental and modeling studies will be critical in providing a link between the cellular-level mechanisms that give rise to striatal firing patterns and the role of striatal firing in animal behavior.