The Striatum: Where Skills and Habits Meet

Ann M. Graybiel[1,2] and Scott T. Grafton[3,4]

[1]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 20139

[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 20139

[3]Institute for Collaborative Biotechnologies, University of California, Santa Barbara, California 93106-9660

[4]Department of Psychological and Brain Sciences, University of California, Santa Barbara, California 93106-9660

Correspondence: graybiel@mit.edu

After more than a century of work concentrating on the motor functions of the basal ganglia, new ideas have emerged, suggesting that the basal ganglia also have major functions in relation to learning habits and acquiring motor skills. We review the evidence supporting the role of the striatum in optimizing behavior by refining action selection and in shaping habits and skills as a modulator of motor repertoires. These findings challenge the notion that striatal learning processes are limited to the motor domain. The learning mechanisms supported by striatal circuitry generalize to other domains, including cognitive skills and emotion-related patterns of action.

The nuclei and interconnections of the basal ganglia are widely recognized for modulating motor behavior. Whether measured at the neuronal or regional level, the activities of neurons in the basal ganglia correlate with many movement parameters, particularly those that influence the vigor of an action, such as force and velocity. Pathology within different basal ganglia circuits predictably leads to either hypokinetic or hyperkinetic movement disorders. In parallel, however, the basal ganglia, and especially the striatum, are now widely recognized as being engaged in activity related to learning. Interactions between the dopamine-containing neurons of the midbrain and their targets in the striatum are critical to this function. A fundamental question is how these two capacities—(motor behavior and reinforcement-based learning)—relate to each other and what role the striatum and other basal ganglia nuclei have in forming new behavioral repertoires. Here, we consider relevant physiological properties of the striatum by contrasting two common forms of adaptation found in all mammals: the acquisition of behavioral habits and physical skills.

Without resorting to technical definitions, we all have an intuition of what habits and skills are. Tying one's shoes after putting them on is something we consider a habit—part of a behavioral routine. The capacity to tie the laces properly is a skill. Habits and skills have many common features. Habits are consistent behav-

iors triggered by appropriate events (typically, but not always, external stimuli) occurring within particular contexts. Physical skills are changes in a physical repertoire: new combinations of movements that lead to new capacities for goal-directed action. Both habits and skills can leverage reward-based learning, particularly during their initial acquisition. In either instance, after sufficient experience, the need for reward becomes lower and lower. With sufficient practice, both lead to "automaticity" and a resilience against competing actions that might lead to unlearning.

THE DEGREES OF FREEDOM PROBLEM AND OPTIMAL PERFORMANCE

When acquiring a new habit or skill, an organism is faced with an enormous space of possibilities from which to choose. For habits, how does the organisms select from the many potential behaviors that it could perform? In the sections to follow, we review some of the evidence, indicating that the striatum has a principal function in learning-related plasticity associated with selecting one set of actions from many, resulting in the acquisition of habitual behavior. Similarly for skills, the motor system is also faced with an enormous set of possible solutions. There is an analogous problem of understanding how the organism narrows a search to find an effective solution when acquiring new motor skills. Of the many possible challenges in skill learning, two are renowned: the degrees of freedom problem (Bernstein 1967) and the problem of optimal control (Todorov and Jordan 2002).

How are behaviors or movements chosen so that the result is optimal? Optimality could be defined in a variety of ways, but there are two particularly relevant for habits and skills. First, optimality can be driven by the outcome of achieving a specific goal and receiving the reward. Monkeys, for example, will work with increased urgency to maximize the number of rewarded trials per hour. Second, optimality can also be determined by the particular ways in which motor behaviors are combined so that a cost is minimized. For example, an animal can learn to find the shortest path to a reward, or find the most efficient combination of movements leading to a reward. This is at the heart of solving the degrees of freedom problem: a process that optimizes movement to a goal in terms of some metric, such as the energetics of the movement.

Much evidence points to the cerebellum and its reliance of online feedback to shape ongoing activity from multiple cortical motor regions, along with spinobulbar pattern generators so that the dynamics of movements are smoother, faster, and more efficient (Takemura et al. 2001; Gao et al. 2012). Notably, this error-based cerebellar shaping of behavior occurs independently of any reward signal. Here, we propose that the basal ganglia, of which the striatum is a main input station, also have a profound effect on optimizing behavior by implementing reinforcement-based feedback to allow effective combination of sequential motor elements. Thus, whether we speak of habits or skills, we see the striatum as a sort of learning machine dedicated to achieving success in behavior. We view this learning capacity as not only adhering to the main challenges of motor control, but also as extending beyond these to influence cognitive and emotional control.

REINFORCEMENT-BASED LEARNING REPRESENTS A CORE MECHANISM THOUGHT TO UNDERLIE BEHAVIORAL OPTIMIZATION BY STRIATUM-BASED CIRCUITS

The behavioral literature on reinforcement learning shows that it is not the reward (or punishment) per se that reinforces (extinguishes) behaviors. Rather, it is the difference between the predicted value of future rewards or punishments and their ultimate reward or punishment. Learning theory has formalized the process by which these reinforcement contingencies, referred to as reward prediction errors (RPEs), drive behavioral change (Sutton and Barto 1998). These lead to the notion that, as an agent (actor) interacts with the environment, it develops state-specific behavioral policies. An influential idea is that these are instantiated in the

brain according to algorithms, such as those in temporal difference models. Through experience, eligibility traces are built up and models of behavioral tasks can be optimized.

The reinforcement-related learning functions of the striatum are driven by evaluative circuits interconnecting the striatum both with the brainstem and with the neocortex and noncortical regions of the forebrain, especially the thalamus. The gradual selection of particular behavioral repertoires can lead toward optimal behavioral control by means of reducing the degrees of freedom normally used in navigating through daily behavior. As we note below, this evidence is mainly derived by recording from multiple neurons in the striatum and interconnected circuits, and in studies in which optogenetic methods are used to manipulate corticostriatal circuits.

Electrophysiological recordings in humans are rare, but it is possible to use functional magnetic resonance imaging (fMRI) to relate the activity of dopaminergic cell groups to metabolic signals recorded from the striatum. Results from a large number of fMRI studies suggest that the human ventral striatum changes its activity in relation to many different kinds of rewards, ranging from juice rewards to abstract social or esthetic qualities. In effect, evidence from these fMRI studies suggests that the ventral striatum is involved in learning by trial-and-error irrespective of the specific nature of the rewards (Daniel and Pollmann 2014). In tasks involving decision-making and economic games, there is overwhelming evidence that the ventral striatum and the ventral tegmental area (VTA) (Diuk et al. 2013) form a key circuit for encoding for RPE (for meta-analysis of 779 fMRI articles, see Garrison et al. 2013). Studies on experimental animals concur, but point to striking changes over the course of learning in signals related to correct or incorrect behavior (Atallah et al. 2014; K Smith and A Graybiel, unpubl.).

Following classic work suggesting that reward was the main driver the nigrostriatal system (Schultz 2002), studies began to show that nonrewarding, aversive drive also could be applied through this system. Experiments in rodents have suggested that nondopamine neurons ex-pressing γ-aminobutyric acid (GABA) in the nigroventral tegmental region are sensitive to aversive stimuli (Bevan et al. 1996; Brown et al. 2012). In macaque monkeys, divisions of the substantia nigra pars compacta region have been identified as having differential positive or negative reinforcement sensitivities (Matsumoto and Hikosaka 2009). Thus, neurons in the midbrain dopamine-containing cell groups can show responses corresponding to the positive and negative RPEs of computational models. Within the striatum, as well, many neurons have spike responses that are related more to rewarding versus risky or aversive contexts (Yamada et al. 2013; Yanike and Ferrera 2014). New data suggest that some striatal neurons perform an integration of cost and benefit, which predicts natural behavioral learning (T Desrochers, K Amemori, and AM Graybiel, unpubl.).

Hence, the striatum is poised to be a hub for neuroplasticity, as it receives major input from aminergic fiber systems, including the dopamine-containing nigral innervation, and input from nearly every region of the neocortex. Not only nigrostriatal synapses, but also corticostriatal synapses are thought to be sites of neuroplasticity. Many neurons in the striatum fire in a given context as though encoding expectancy signals and priors—signals crucial for smooth behavioral performance with advanced planning (Hikosaka et al. 1989). These signals are likely generated as a result of experience-dependent plasticity in the circuits that form the input–output networks of this large region of the basal ganglia. This includes cholinergic interneurons (Doig et al. 2014). It is likely that through these and other network activities, including processing through the thalamus and neocortex and their projections to the striatum, neurons in the striatum build up selective responses to particular environmental events and certain behavioral actions and contexts.

HABIT LEARNING: A MODEL FOR STUDYING BEHAVIORAL PLASTICITY INFLUENCED BY STRIATAL CIRCUITS

Two main lines of evidence have linked the striatum and its associated neural circuits with the

development of habitual behaviors. First, a long line of lesion studies in rodents has shown that the striatum is necessary for habit formation (Balleine and Dickinson 1998; Yin and Knowlton 2006; Belin et al. 2009). Further, once acquired, habitual behaviors can be blocked or blunted by lesions of the striatum made after the habits are learned. These studies have shown that different districts within the striatum operate during habit formation. The ventral striatum is necessary for initial learning of motivated behaviors that could become habitual (Atallah et al. 2014). The dorsal striatum then becomes critical. First, behaviors are driven largely by the anticipated outcome of the behavior itself; this process, according to rodent lesion studies, requires the dorsomedial striatum ([DMS] in rodents). But then, according to these lesion studies, as the behaviors are repeated and bring about a positive outcome, the DMS is no longer required, but the dorsolateral parts of the striatum ([DLS] in rodents) is required for habitual performance.

A striking parallel to these behavioral findings on transitions occurring during habit learning has come from studies in which multiple simultaneous recordings have been made within the striatum on a daily basis as the acquisition of the habits occurs (Jog et al. 1999; Barnes et al. 2005; Thorn et al. 2010; Smith and Graybiel 2013; Atallah et al. 2014). Remarkable plasticity is seen in these recordings. For example, in T-maze learning studies, in which rodents learn to navigate a maze according to cues given midrun that instruct them about which side food reward will be given, recordings have been made in the DLS, the part of the striatum thought to be essential for postlearning habit performance, the DMS, the part of the striatum thought to be essential for initial goal-directed behavior during early acquisition, and in the ventromedial striatum, the region thought, along with the VTA, to be critical for initial acquisition. Striatal projection neuron ensembles in each of these regions develop different response patterns but, in all of the regions, the ensembles reflect the entire behavioral time.

These behavioral and physiological findings are important in supporting the view that there is a critical transition period during habit formation. Before this transition, a given behavior being learned remains sensitive to outcome (usually tested as sensitivity to reward value). But, after this period, the same behavior becomes independent of the reward value. This distinction, introduced formally by Dickinson and his colleagues (Dickinson 1985) with reward devaluation paradigms, has been influential in models of habit formation and the shifts between goal-dependent and semiautomatic performance characteristics of habits. As we note below, this transition is marked by changes in the activity patterns of striatal neurons, and also of neurons in the prefrontal cortex.

Such transitions have not been tested directly in humans. However, brain imaging has suggested that, after conditioning, activity in the dorsal striatum is sensitive to the relative value of an action choice compared with other actions rather than to the relative value of rewards per se (Li and Daw 2011). Evidence suggests that the capacity to undergo this transition is influenced by the human *FOXP2* gene, a gene implicated in speech and language function in humans (Schreiweis et al. 2014). This influence is particularly intriguing because mutants of *FOXP2* do not alter motor performance or skill acquisition as tested, for example, by rotorod.

A second major line of work implicating the striatum in habitual behaviors comes from work on the neural origins of addictive behaviors. Much evidence suggests that the midbrain dopamine system, particularly the VTA, is influential in the initial stages of generation of these behaviors, and that the striatal target of the VTA system, which largely lies in the ventral striatum, is essential for the neural changes, leading to addiction. Remarkably, here too, with the progression of the addictive behavior, the dorsal striatum becomes more and more involved with time (Everitt and Robbins 2013). Thus, across very different domains of what in common parlance we call habits, there appear to be progressive stages for the ingraining of stereotyped action patterns into an individual's repertoire of behaviors, and a corresponding change in emphasis of striatal regions predominantly implicated (Graybiel 2008).

Although not focusing in depth on addictive behavior, we emphasize that an important unresolved question is the degree to which addictions are extreme forms of habits, developed by trial-and-error learning and leading to distorted RPEs. In this context, addictions reveal fragility in this otherwise robust habit-learning mechanism, exposed by abnormal activation of the reward circuitry by exogenous chemicals, including cocaine and other psychomotor stimulants, ethanol, and nicotine. Positron emission tomography (PET) studies in humans have shown that drugs can induce rapid increases of striatal dopamine but that, once a person is addicted, these drug-induced dopamine increases (as well as their subjective effects on behavior) are blunted. By contrast, addicts experiencing craving to any one of a number of drugs can show a significant dopamine increase in striatum in response to drug-conditioned cues (such as thoughts leading to craving) with response levels that can be greater than those to the drug themselves (Volkow et al. 2011, 2014a,b). These cue-induced responses are particularly prominent in the dorsal striatum, including in the putamen, consistent with evidence that the dorsal striatum is heavily involved in "normal" habit learning. The cue responses likely reflect a profoundly distorted RPE.

Taken together, the findings from these two lines of work have sometimes been interpreted as suggesting that the striatum is "the seat of habitual behavior"—that the "habit," or its neural representation, is stored within the striatum itself. Classic methods, however, did not allow adequate testing of this notion.

An important detail of the anatomy of the striatum makes a definitive answer difficult to achieve. The projection neurons of the striatum, that is, the neurons that project to the pallido-nigral output nuclei of the basal ganglia, also are the main striatal neurons receiving inputs to the striatum. Thus, any procedure, genetic or otherwise, affects circuits, and not just the striatum. We suggest that this is a critical distinction. It becomes impossible to say that habit representations are stored in the striatum because the striatum is only a node in larger networks. Neurobiological experimental techniques, such as optogenetics, now open the possibility of testing these assumptions directly.

TIME SCALES OF LEARNING AND THE FORMATION OF MOTOR–MOTOR ASSOCIATIONS

Just as there are remarkable changes within striatal circuits occurring on multiple time scales during habit learning, of which we point out only some examples, during physical skill acquisition, there are analogous temporal shifts across circuits. For example, as nonhuman primates learn a novel behavior, such as a unique sequence of arm or finger movements, neural recordings typically show shifts in major activity from associative to sensorimotor districts of the striatum (Miyachi et al. 1997, 2002; Hikosaka et al. 1999), regions to which the DMS and DLS of rodents are thought to correspond (Graybiel 2008). Detailed modeling of reaction time behavior as the animal makes sequential reaches shows that they switch between two modes of control, consistent with the use of multiple motor control or prefrontal circuits to generate actions.

In rodents, profound changes in spike activity patterns of neurons occur simultaneously in the associative and sensorimotor striatum, and, as activity in the associative striatum declines, activity in the sensorimotor striatum becomes strong. These dynamics suggest that potentially competing circuits are organized to favor habit formation (Thorn et al. 2010). There is much evidence from classic rodent studies that different corticostriatal loops can compete with one another during the learning process and subsequent performance. This now has been found in humans. fMRI evidence shows that individuals who reduce activity in prefrontal regions sooner are those who acquire sequential skills more rapidly (Bassett et al. 2015).

In human brain-imaging experiments (Lehéricy et al. 2005; Grol et al. 2006; Doyon et al. 2009), a shift from anterior associative to sensorimotor striatum is also observed as people practice sequential finger movements. Initially, there is a broad recruitment of prefrontal, pre-

motor, and sensorimotor cortex (along with the underlying corticostriatal target regions). With time, there is a progressive reduction of activity in prefrontal regions and associative striatum. These different cortical regions could potentially acquire, represent, or forget sequential information differently. For example, prefrontal regions that support working memory are invaluable for explicitly remembering a sequence of external cues that could guide movement. There are not yet sufficient studies of the striatum to relate all of these findings for the neocortex to selective changes in spike activities of different striatal regions, but we emphasize, again, that these dynamic changes occur across corticostriatal and other circuits.

A key advance in recent human brain-imaging methods is the ability to map activity that corresponds to a specific sequence or skill using either machine learning or repetition suppression methods (Wiestler and Diedrichsen 2013; Wymbs and Grafton 2014). These show that the degree to which different cortical areas represent a specific skill depends in large part on the depth of training experience, and not simply on time (Wymbs and Grafton 2014). Over a longer training horizon, there is less reliance on premotor areas to represent a sequence. Ultimately, a central feature of motor skill is the ability to guide actions without explicit memory or external stimuli through the creation of direct motor–motor associations. Not surprisingly, skill-specific changes also emerge within motor cortex (Karni et al. 1995; Wymbs and Grafton 2014). Thus, across habits and skills, different sorts of automaticity are gained. There are compelling parallels between these recording studies in animals during habit-learning and human-imaging experiments of skill learning.

Although the acquisition of physical skills is commonly attributed to online feedback-based error learning mediated by the cerebellum, allowing for powerful tuning of complex musculoskeletal dynamics, it is important to note that all of the associative and sensorimotor cortical areas that have been implicated in motor-skill acquisition and performance project directly to the striatum as parts of corticostriatal circuits. As shown by the recording work in rodents,

multiple corticostriatal loops, as judged by striatal projection neuron activity, are simultaneously active as reward-based learning occurs. Dopamine can modulate each of these loops. Dopamine-receptor blockade in nonhuman primates impairs skill acquisition (Tremblay et al. 2009). Deficits of dopamine signaling stemming from Parkinson's disease or dopamine-receptor blockade are both determinants of the rate of skill acquisition (Weickert et al. 2013). Genetically determined reductions of striatal dopamine function can also influence response to rewards and impact skill learning as well (Frank and Fossella 2011; Stice et al. 2012).

There is evidence from experiments in monkeys that well-practiced behaviors do not require the pallidal output nuclei of the basal ganglia for their expression (Desmurget and Turner 2010). This finding does not necessarily hold for habits, but it accords well with the proposal that the basal ganglia, and here we emphasize the striatum, is critical in the acquisition of action repertoires. Reinforcement-related signals reaching nonstriatal regions are very likely also important for influencing the formation of cortical or other connections that mediate motor–motor associations (in which the commands for one movement directly trigger the next command). It is important to note that there are significant striatal output connections that do not include the classical pallidal output nuclei of the basal ganglia. These connections could be part of the habits–skill performance circuitry even when the classic pallidal pathways are not required.

BRACKETING: A READOUT THAT FRAMES AN ACTION

In the part of the rodent striatum thought to be necessary for habitual performance (i.e., DLS), a striking pattern of neuronal activity emerges. As animals learn to run in maze tasks, the striatal activity, at first, marks the full run time, but later begins to bracket the entire run. Activity becomes more and more prominent at the beginning and end of the runs, or beginning and end of the action through the turns. Surprisingly, at the same time, activity during the rest of the run time declines and may even be below

prerun baseline levels. The kinematics of the runs are changing as the rats become more and more repetitive in their navigational routes, but the "end" activity can occur even after the rats are no longer running and, thus, cannot simply be attributed to velocity or acceleration signals. Similarly, the "beginning" or "start" activity can occur before the runs have begun.

Such beginning and end activity has been seen repeatedly in different experiments in rodents (Jog et al. 1999; Barnes et al. 2005; Thorn et al. 2010; Smith and Graybiel 2013), is long lasting, and has also been found in lever-pressing tasks (Jin and Costa 2010). Task-bracketing activities have also been recorded in the striatum (and prefrontal cortex) of macaque monkeys performing well-learned motor skills, including oculomotor sequential saccade tasks (Fujii and Graybiel 2003) and sequential arm-reaching tasks (J Feingold, D Gibson, AM Graybiel, et al., unpubl.; R. Turner, pers. comm.). Concurrent phasic episodes of oscillatory local field potential (LFP) activity also occur in relationship to action boundaries (Howe et al. 2013). In patients with Parkinson's disease, LFP recordings within the subthalamic nucleus show anticipatory suppression of β-band activity at sequence boundaries that is linked to better performance (Herrojo Ruiz et al. 2014). This feature is also seen in monkeys performing arm-reaching tasks (J Feingold, D Gibson, AM Graybiel, et al., unpubl.).

Remarkably, a nearly inverse pattern of spike activity has been shown to develop in the associative, dorsomedial part of the rodent striatum, a region critical for goal-directed behavior. The projection neuron ensembles gradually develop increased firing during the runs, especially around the decision period of the task. There is much less activity at the beginning and end of the runs. Moreover, this decision-period activity then subsides during late learning—the very time that the "beginning-and-end" activity in the DLS is especially strong (Thorn et al. 2010).

Finally, in the ventromedial striatum, ensembles in the aggregate also show beginning-and-end responses. Yet, others fire throughout the runs, ramping up to the time of reward receipt (Atallah et al. 2014). These findings provide unequivocal evidence that striatal projection neurons have highly dynamic ensemble response patterns during habit learning. Especially notable is the fact that these patterns reflect entire behavioral sequences from beginning to end, which initially are goal directed, but after long training can become nearly autonomous except for being triggered by a start cue.

Protocols have now been used in rodents to determine how fixed the striatal task-bracketing patterns are. Extinction protocols, in which rewards were either removed or rarely given after acquisition and prolonged overtraining, nearly abolish the beginning-and-end pattern, but if the rewards are returned, the beginning-and-end pattern reappears almost immediately. Thus, the form of action boundary representation in the sensorimotor striatum somehow can be suppressed but cannot be erased by removal of rewards. After prolonged training, application of the classic reward devaluation procedure of Dickinson (1985), in which the reward is maintained, but is made unpalatable, hardly changes the striatal beginning-and-end pattern (Smith and Graybiel 2013). Thus, the task-bracketing pattern in the DLS is extremely resistant to degradation—it takes wholesale removal of rewards to block it fully, and even then, the pattern is latent, but not gone, and is rapidly retrievable.

Although the function of this bracketing activity remains unknown, a strong case can be made from the studies on habit learning that it is tied to feedback about how successful a sequence of actions within the bracket has been in gaining a desired outcome. This function is at the heart of the trial-and-error learning that leads to the forming of habits composed of multiple sequential actions. The end patterns found in these sequences could provide such outcome signals. These phasic end responses are themselves dynamic, changing or coming and going during the course of training (Smith and Graybiel 2013; Atallah et al. 2014). The mechanisms underlying these habit-related activity patterns are not understood, but evidence from the rodent experiments suggests that the activity of striatal interneurons is strongly modulated dur-

ing habit learning. This result is notable because it indicates that intrastriatal networks undergo profound reorganizational changes. Thus, the ensemble patterns, such as task bracketing, cannot solely be attributed to the direct input connections of the projection neurons themselves; changes in local intrastriatal microcircuit occur as well. Below, we suggest that the bracketing could be a neural sign of the chunking of behaviors that have proven successful enough to the organism to merit prolonged expression.

FROM BRACKETING TO CHUNKS: SHAPING THE ELEMENTS OF ACTION IN RELATION TO COSTS AND BENEFITS

The bracketing activity surrounding a habit formed from multiple behaviors is defined by both start- and end-related changes of neuronal activity. Although the end activity could clearly be used to predict a subsequent reward, the purpose of the start activity is less obvious. One possibility is that it serves as the opening of a bracketed behavioral unit. Interestingly, the beginning-and-end activities often are built and changed together, and sometimes appear in the responses of single striatal neurons. In monkeys, it has been shown that experimentally triggered changes in the end activity can induce changes in the accompanying start activity.

A key idea in control theory is that optimal behavior is determined not only by the reward obtained but also by the minimization of some cost function related to the set of actions needed to accomplish the action. If a behavior needs to be optimized over a particular time interval, then there needs to be an indication of when an action actually begins and ends. This estimate of a distinct time interval becomes increasingly important during transitions from habits to motor skills for which there is a further refinement of behavior at the level of kinematics and limb dynamics based on optimal control principles. Ultimately, optimal control requires an estimate of the physical or neural cost of performing a particular action. Together, the start-and-end activities could contribute to this estimation by providing a reading frame for labeling a given action.

A direct investigation of optimization within a habitual sequence of eye movements was examined in oculomotor scanning patterns generated by naïve, untrained monkeys (Desrochers et al. 2010). In this study, monkeys who were never trained experimentally were placed in a booth with a computer screen in front of them, on which colored discs appeared. The monkeys naturally looked around the display and, without experimental training, tended to acquire particular favored scan patterns that gradually changed over months of experience. The bit-by-bit changes in the monkeys' scanning resulted in a succession of favored spatio-temporal chunking patterns of the untrained habitual saccade sequences, and these habitual scanning patterns eventually became optimal or nearly optimal, as judged by models of the task. Remarkably, nearly all of these adjustments in the scanning patterns took place long after maximum reward had been obtained. Analysis showed that the behavior was driven by small trial-by-trial differences in the cost of the scans: the distance required. This result has led to the conclusion that extremely fine-grain, trial-by-trial monitoring of least cost can be a driver of habit learning, as well as a driver of skill learning. These results are in line with the notion that the brain has a natural tendency to reduce cost across many cognitive domains and that this tendency, in addition to sensitivity to reward, can drive the character of habits (Desrochers et al. 2010; Kool et al. 2010; Gepshtein et al. 2014). Recent electrophysiological recordings suggest that striatal projection neurons encode these outcome and cost signals in their end activity (T Desrochers, K Amemori, and AM Graybiel, unpubl.).

A set of behaviors that is reliably combined and expressed as a habit can ultimately be viewed as a "chunk," framed by the neuronal bracketing activity in the striatum. This notion of chunking, introduced by George Miller (1956) in reference to helping deal with memory load, invokes, for the motor system, the binding together of multiple behaviors into a single behavioral unit. There are different aspects to such packaging up of behaviors. First, there is a form of chunking (concatenation), bundling the in-

dividual elements into a whole. As we note below, there is reason to think that the striatum and its circuits could be critical to this function (Graybiel 2008). Alongside this is another well-known phenomenon in cognitive science wherein adjacent elements of a long sequential stimulus or behavior are temporally divided up (parsing), leading to detectable pauses between groups of adjacent elements. Chunking is often used strategically by humans to parse long strings of stimuli into smaller sets to facilitate memory, as proposed by Miller, by analogy to remembering sequences of numbers. For example, a U.S. phone number is strategically divided into a 3-3-4 pattern.

Pauses within a sequence of movements provide a useful way to identify chunks embedded within long sequences. New computational tools are emerging to identify chunks for habits or motor skills based on concatenation rather than parsing. Concatenation can be spotted by examining the covariation or timing of movements within a suspected chunk or by assessing the frequency with which errors are made at the boundary of adjacent chunks rather than within chunks (Acuna et al. 2014). Thus, it is becoming possible to assess the strength by which successive elements of a complex action are combined. Whether it is habits or skills, the animal is seeking to minimize some cost function over the time interval of the complete action. This development of smooth kinematics is universally observed as animals combine fragmented movements together. By concatenating muscle synergies or movements together in specific groups, it might be possible to improve efficiency. Interestingly, the learning of the kinematics can progress more rapidly than the progression of the behavior becoming a habit (Jog et al. 1999; Barnes et al. 2005; Smith and Graybiel 2013).

Based on the bracketing patterns that form in the striatum and elsewhere, and the behavioral changes that occur alongside them, it has been suggested that one function of the striatum (and, hence, of the basal ganglia) could be to facilitate such chunking as habits and routines form (Graybiel 1998, 2008). The key property of this process is the selection of behaviors that are

successful, either through optimizing reward or cost or their integral. This notion fits well with our suggestion that the striatum and associated circuits could be important for achieving optimality in the performance of both habits and skills. Marking action-sequence boundaries allows the sequences to be represented as units that could then be released more readily than unconcatenated chains of elements. Importantly, this function should serve cognitive as well as motor packaging up of beneficial (or, in pathology, nonbeneficial) behaviors (Graybiel 2008).

Evidence for a role of the striatum in chunking movements stems, in part, from pharmaceutical blockade of dopamine receptors in monkeys with the dopamine D2-type receptor antagonist, raclopride. This manipulation does not interfere with well-learned sequences, but disrupts the formation of new chunks (Levesque et al. 2007). Chronic dopamine denervation in patients with Parkinson's disease can also lead to an impairment of chunking for new sequences of movement (Tremblay et al. 2010). There is also emerging evidence from human neuroimaging that the strength of activity in the associative striatum varies on a trial-by-trial basis with the degree to which subjects put together elements of motor sequences into chunks, the concatenation aspect of chunking (Wymbs et al. 2012).

Ultimately, if a string of behavioral elements is represented as a single unit, then there should be concomitant neuronal activity reflecting this. Not only is the representation of an action boundary present in the spiking of ensembles of striatal neurons, it is also observed in striatal neurons with maintained activity (Kubota et al. 2009; Barnes et al. 2011; Hernandez et al. 2013; Howe et al. 2013). Here, the action-related activity can be identified throughout the duration of the set of actions forming a chunk. This could represent a prolonged form of dopamine signaling, which, in the ventromedial striatum, can span the entire habitual behavior (Howe et al. 2013). Thus, the neuronal populations within rodent striatum associated with the bracketing and task-on activity could serve as neural signatures of the concatenation and pars-

ing functions identified in human studies of chunking.

Collectively, these studies across multiple species and tasks suggest that striatal circuits can help to bring together advantageous behavioral segments into sequences that help to achieve behavioral goals. The result, in this view, is that the behaviors could be released readily as a complete "set" when the appropriate context calls for this release. We now know that these beginning-and-end patterns can develop in cortical regions and elsewhere, suggesting that this process is a network property, strongly evidenced in parts of the striatum and corresponding corticostriatal loops. This idea finds strong resonance in clinical observations, for example, in the problems that Parkinson's patients have in starting a sequential set of actions, such as walking, and then in ending the sequence once underway.

WHERE ARE HABITS "STORED"? CIRCUIT DYNAMICS ARE CRITICAL TO HABIT LEARNING AND PERFORMANCE

Causal evidence for circuit-level control of habits is just emerging. Lesion studies show that the medial prefrontal cortical region, called infralimbic (IL) cortex, in rodents, like the DLS, is necessary for habits to be performed. New optogenetic studies have shown that the IL exerts online control of the performance of well-ingrained habits (Smith et al. 2012) and is necessary for their formation (Smith and Graybiel 2013). This work is critical to any account of the role of the striatum in habit formation, as it suggests a form of cortical control that can, on a moment-by-moment basis, determine whether a behavior is performed habitually or not.

Strikingly, IL develops, in its upper layers, a strong task-bracketing pattern during habit learning, one similar to that in the DLS. But, the cortical bracketing pattern, unlike the DLS bracketing pattern, is sensitive to reward devaluation; it is nearly lost. As IL does not project directly to the DLS, the results suggest that the online control is a circuit-level effect. This online control suggests that we need to rethink our ideas about the control of habits—both the

learning of these behaviors and their expression. At the very least, there are dual operators, cortical and subcortical, acting as habits become crystallized, and these act simultaneously with the cortical control being online (Smith and Graybiel 2013).

There is not yet comparable optogenetic evidence for the effects of perturbations within the striatum itself. This is key missing information. What can be said is that multiple circuits are simultaneously active as habits form, and these circuits have differential sensitivities and patterns of connectivity. There is no evidence for habits being "stored" in one site, such as the striatum, although local networks within the striatum acquire new activity patterns during habit learning and could be local controllers. These considerations deflate controversies pitting different individual regions as being most important for habit learning.

CHUNKS, THE DEGREES-OF-FREEDOM PROBLEM, AND OPTIMAL CONTROL

We return to the key issues of the degrees-of-freedom problem and how to achieve optimal control. We suggest that a major advantage of habit formation is that this process allows many possible degrees of freedom to be essentially dropped from the animal's normal, habitual repertoire so long as the conditions surrounding the habit are not at odds with its performance. Change in these conditions could lead to a return to behavior typical of early acquisition, a kind of trial-and-error behavior akin to early language learning in children and to song learning in passerine birds. The degradation of the task-bracketing patterns with removal of positive outcomes fits with this reverse plasticity.

How does chunking relate to the problem of having unmanageable numbers of degrees of freedom and the need for optimality in motor control? In this instance, finding optimal solutions becomes increasingly difficult as the sequence of movements is lengthened. It is possible that by grouping fine-grained movement elements into chunks, the solutions for optimality become easier to compute. An alternative and intriguing possibility is that the pauses

observed in complex movements are a result of optimal control. Preliminary evidence has been gathered in both humans and nonhuman primates as they perform five element sequential reaching tasks. For a given sequence, different monkeys will converge on the same pattern of chunking. Kinematic analysis suggests that the chunks lead to a more global pattern of movement efficiency than what is obtained otherwise (Ramkumar et al. 2014). However, these observations need to be tempered in light of the fact that, in many situations, optimality is not essential. Instead, muscle synergies seem to be built on habits that are "good enough" rather than optimal (de Rugy et al. 2012).

BEYOND ACTIONS: THE UTILITY OF STRIATAL CIRCUITS FOR EMOTIONAL AND COGNITIVE HABITS AND SKILLS

As habitual actions become ingrained, the kinematics of the habitual actions, that is, the physical skills enabling these habits, become standardized. But there is a further connotation of the term "habit" to consider, one that invokes motivational processes that shape the expression of cognitive processes in particular contexts, irrespective of kinematics and physical skill. Although these kinds of habits are not restricted to motor acts or sequences of motor actions, they are likely to also rely on corticostriatal circuits that use contextual information to shape behavior. Further, these habits of thought can be powerfully shaped by complex social cues (Graybiel 2008). For example, striatal neurons are able to distinguish reward predictions intended for the monkey undergoing neuronal recordings from those destined for another animal (Báez-Mendoza et al. 2013). Habits of thought, at least in humans, are probably as common as motor habits and, like motor habits, are vulnerable to pathologic distortion. Our view is that such habits of mind can be created by cognitive pattern generators much as habits of action are generated (Graybiel 1997). Understanding these wider implications of learning repertoires of thought and action is an important goal for future work.

REFERENCES

Acuna DE, Wymbs NF, Reynolds CA, Picard N, Turner RS, Strick PL, Grafton ST, Körding K. 2014. Multi-faceted aspects of chunking enable robust algorithms. *J Neurophysiol* **112:** 1849–1856.

Atallah HE, McCool AD, Howe MW, Graybiel AM. 2014. Neurons in the ventral striatum exhibit cell-type-specific representations of outcome during learning. *Neuron* **82:** 1145–1156.

Báez-Mendoza R, Harris CJ, Schultz W. 2013. Activity of striatal neurons reflects social action and own reward. *Proc Natl Acad Sci* **110:** 16634–16639.

Balleine BW, Dickinson A. 1998. Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology* **37:** 407–419.

Barnes TD, Kubota Y, Hu D, Jin DZ, Graybiel AM. 2005. Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* **437:** 1158–1161.

Barnes TD, Mao J-B, Hu D, Kubota Y, Dreyer AA, Stamoulis C, Brown EN, Graybiel AM. 2011. Advance cueing produces enhanced action-boundary patterns of spike activity in the sensorimotor striatum. *J Neurophysiol* **105:** 1861–1878.

Bassett DS, Yang M, Wymbs NF, Grafton ST. 2015. Learning-induced autonomy of sensorimotor systems. *Nat Neurosci* **18:** 744–751.

Belin D, Jonkman S, Dickinson A, Robbins TW, Everitt BJ. 2009. Parallel and interactive learning processes within the basal ganglia: Relevance for the understanding of addiction. *Behav Brain Res* **199:** 89–102.

Bernstein NA. 1967. *The coordination and regulation of movements.* Pergamon, New York.

Bevan MD, Smith AD, Bolam JP. 1996. The substantia nigra as a site of synaptic integration of functionally diverse information arising from the ventral pallidum and the globus pallidus in the rat. *Neuroscience* **75:** 5–12.

Brown MT, Tan KR, O'Connor EC, Nikonenko I, Muller D, Luscher C. 2012. Ventral tegmental area GABA projections pause accumbal cholinergic interneurons to enhance associative learning. *Nature* **492:** 452–456.

Daniel R, Pollmann S. 2014. A universal role of the ventral striatum in reward-based learning: Evidence from human studies. *Neurobiol Learn Mem* **114:** 90–100.

de Rugy A, Loeb GE, Carroll TJ. 2012. Muscle coordination is habitual rather than optimal. *J Neurosci* **32:** 7384–7391.

Desmurget M, Turner RS. 2010. Motor sequences and the basal ganglia: Kinematics, not habits. *J Neurosci* **30:** 7685–7690.

Desrochers TM, Jin DZ, Goodman ND, Graybiel AM. 2010. Optimal habits can develop spontaneously through sensitivity to local cost. *Proc Natl Acad Sci* **107:** 20512–20517.

Dickinson A. 1985. Actions and habits: The development of behavioural autonomy. *Philos Trans R Soc Lond B Biol Sci* **308:** 67–78.

Diuk C, Tsai K, Wallis J, Botvinick M, Niv Y. 2013. Hierarchical learning induces two simultaneous, but separable, prediction errors in human basal ganglia. *J Neurosci* **33:** 5797–5805.

Doig NM, Magill PJ, Apicella P, Bolam JP, Sharott A. 2014. Cortical and thalamic excitation mediate the multiphasic responses of striatal cholinergic interneurons to motivationally salient stimuli. *J Neurosci* **34:** 3101–3117.

Doyon J, Bellec, Amsel, Penhune, Monchi, Carrier, Lehéricy S, Benali H. 2009. Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behav Brain Res* **199:** 61–75.

Everitt BJ, Robbins TW. 2013. From the ventral to the dorsal striatum: Devolving views of their roles in drug addiction. *Neurosci Biobehav Rev* **37:** 1946–1954.

Frank MJ, Fossella JA. 2011. Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology* **36:** 133–152.

Fujii N, Graybiel AM. 2003. Representation of action sequence boundaries by macaque prefrontal cortical neurons. *Science* **301:** 1246–1249.

Gao Z, van Beugen BJ, De Zeeuw CI. 2012. Distributed synergistic plasticity and cerebellar learning. *Nat Rev Neurosci* **13:** 619–635.

Garrison J, Erdeniz B, Done J. 2013. Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. *Neurosci Biobehav Rev* **37:** 1297–1310.

Gepshtein S, Li X, Snider J, Plank M, Lee D, Poizner H. 2014. Dopamine function and the efficiency of human movement. *J Cogn Neurosci* **26:** 645–657.

Graybiel AM. 1997. The basal ganglia and cognitive pattern generators. *Schizophr Bull* **23:** 459–469.

Graybiel A. 1998. The basal ganglia and chunking of action repertoires. *Neurobiol Learn Mem* **70:** 119–136.

Graybiel AM. 2008. Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* **31:** 359–387.

Grol MJ, de Lange FP, Verstraten FA, Passingham RE, Toni I. 2006. Cerebral changes during performance of overlearned arbitrary visuomotor associations. *J Neurosci* **26:** 117–125.

Hernandez LF, Kubota Y, Hu D, Howe MW, Lemaire N, Graybiel AM. 2013. Selective effects of dopamine depletion and L-DOPA therapy on learning-related firing dynamics of striatal neurons. *J Neurosci* **33:** 4782–4795.

Herrojo Ruiz M, Rusconi M, Brücke C, Haynes J-D, Schönecker T, Kühn AA. 2014. Encoding of sequence boundaries in the subthalamic nucleus of patients with Parkinson's disease. *Brain* **137:** 2715–2730.

Hikosaka O, Sakamoto M, Usui S. 1989. Functional properties of monkey caudate neurons. III: Activities related to expectation of target and reward. *J Neurophysiol* **61:** 814–832.

Hikosaka O, Nakahara H, Rand MK, Sakai K, Lu X, Nakamura K, Miyachi S, Doya K. 1999. Parallel neural networks for learning sequential procedures. *Trends Neurosci* **22:** 464–471.

Howe MW, Tierney PL, Sandberg SG, Phillips PE, Graybiel AM. 2013. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* **500:** 575–579.

Jin X, Costa RM. 2010. Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature* **466:** 457–462.

Jog MS, Kubota Y, Connolly CI, Hillegaart V, Graybiel AM. 1999. Building neural representations of habits. *Science* **286:** 1745–1749.

Karni A, Meyer G, Jezzard P, Adams MM, Turner R, Ungerleider LG. 1995. Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature* **377:** 155–157.

Kool W, McGuire JT, Rosen ZB, Botvinick MM. 2010. Decision making and the avoidance of cognitive demand. *J Exp Psychol Gen* **139:** 665–682.

Kubota Y, Liu J, Hu D, DeCoteau WE, Eden UT, Smith AC, Graybiel AM. 2009. Stable encoding of task structure coexists with flexible coding of task events in sensorimotor striatum. *J Neurophysiol* **102:** 2142–2160.

Lehéricy S, Benali H, Van de Moortele PF, Pelegrini-Issac M, Waechter T, Ugurbil K, Doyon J. 2005. Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proc Natl Acad Sci* **102:** 12566–12571.

Levesque M, Bedard MA, Courtemanche R, Tremblay PL, Scherzer P, Blanchet PJ. 2007. Raclopride-induced motor consolidation impairment in primates: Role of the dopamine type-2 receptor in movement chunking into integrated sequences. *Exp Brain Res* **182:** 499–508.

Li J, Daw ND. 2011. Signals in human striatum are appropriate for policy update rather than value prediction. *J Neurosci* **31:** 5504–5511.

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

Miller GA. 1956. The magic number seven plus or minus two: Some limits on our automatization of cognitive skills. *Psychol Rev* **63:** 81–97.

Miyachi S, Hikosaka O, Miyashita K, Karadi Z, Rand MK. 1997. Differential roles of monkey striatum in learning of sequential hand movement. *Exp Brain Res* **115:** 1–5.

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

Ramkumar P, Acuna DE, Berniker M, Grafton S, Turner RS, Körding KP. 2014. Movement chunking as locally optimal control. In *Translational and Computational Motor Control (TCMC) 2014*. Washington, DC, November 14.

Schreiweis C, Bornschein U, Burguiere E, Kerimoglu C, Schreiter S, Dannemann M, Goyal S, Rea E, French CA, Puliyadi R, et al. 2014. Humanized Foxp2 accelerates learning by enhancing transitions from declarative to

procedural performance. *Proc Natl Acad Sci* **111:** 14253–14258.

Schultz W. 2002. Getting formal with dopamine and reward. *Neuron* **36:** 241–263.

Smith KS, Graybiel AM. 2013. A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron* **79:** 361–374.

Smith KS, Virkud A, Deisseroth K, Graybiel AM. 2012. Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. *Proc Natl Acad Sci* **109:** 18932–18937.

Stice E, Yokum S, Burger K, Epstein L, Smolen A. 2012. Multilocus genetic composite reflecting dopamine signaling capacity predicts reward circuitry responsivity. *J Neurosci* **32:** 10093–10100.

Sutton RS, Barto AG. 1998. *Reinforcement learning*. MIT Press, Cambridge, MA.

Takemura A, Inoue Y, Gomi H, Kawato M, Kawano K. 2001. Change in neuronal firing patterns in the process of motor command generation for the ocular following response. *J Neurophysiol* **86:** 1750–1763.

Thorn CA, Atallah H, Howe M, Graybiel AM. 2010. Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron* **66:** 781–795.

Todorov E, Jordan MI. 2002. Optimal feedback control as a theory of motor coordination. *Nat Neurosci* **5:** 1226–1235.

Tremblay P-L, Bedard M-A, Levesque M, Chebli M, Parent M, Courtemanche R, Blanchet PJ. 2009. Motor sequence learning in primate: Role of the D2 receptor in movement chunking during consolidation. *Behav Brain Res* **198:** 231–239.

Tremblay PL, Bedard MA, Langlois D, Blanchet PJ, Lemay M, Parent M. 2010. Movement chunking during sequence learning is a dopamine-dependent process: A study conducted in Parkinson's disease. *Exp Brain Res* **205:** 375–385.

Volkow ND, Wang G-J, Fowler JS, Tomasi D, Telang F. 2011. Addiction: Beyond dopamine reward circuitry. *Proc Natl Acad Sci* **108:** 15037–15042.

Volkow ND, Tomasi D, Wang GJ, Logan J, Alexoff DL, Jayne M, Fowler JS, Wong C, Yin P, Du C. 2014a. Stimulant-induced dopamine increases are markedly blunted in active cocaine abusers. *Mol Psychiatry* **19:** 1037–1043.

Volkow ND, Wang GJ, Telang F, Fowler JS, Alexoff D, Logan J, Jayne M, Wong C, Tomasi D. 2014b. Decreased dopamine brain reactivity in marijuana abusers is associated with negative emotionality and addiction severity. *Proc Natl Acad Sci* **111:** E3149–E3156.

Weickert TW, Mattay VS, Das S, Bigelow LB, Apud JA, Egan MF, Weinberger DR, Goldberg TE. 2013. Dopaminergic therapy removal differentially effects learning in schizophrenia and Parkinson's disease. *Schizophr Res* **149:** 162–166.

Wiestler T, Diedrichsen J. 2013. Skill learning strengthens cortical representations of motor sequences. *eLife* **2:** e00801.

Wymbs NF, Grafton ST. 2014. The human motor system supports sequence-specific representations over multiple training dependent time scales. *Cereb Cortex* doi: 10.1093/cercor/bhu144.

Wymbs NF, Bassett DS, Mucha PJ, Porter MA, Grafton ST. 2012. Differential recruitment of the sensorimotor putamen and frontoparietal cortex during motor chunking in humans. *Neuron* **74:** 936–946.

Yamada H, Inokawa H, Matsumoto N, Ueda Y, Enomoto K, Kimura M. 2013. Coding of the long-term value of multiple future rewards in the primate striatum. *J Neurophysiol* **109:** 1140–1151.

Yanike M, Ferrera VP. 2014. Representation of outcome risk and action in the anterior caudate nucleus. *J Neurosci* **34:** 3279–3290.

Yin HH, Knowlton BJ. 2006. The role of the basal ganglia in habit formation. *Nat Rev Neurosci* **7:** 464–476.