Published in final edited form as:
Curr Opin Neurobiol. 2014 April ; 25: 194–200. doi:10.1016/j.conb.2014.01.012.

The role of efference copy in striatal learning

Michale S. Fee
McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge MA

Abstract

Reinforcement learning requires the convergence of signals representing context, action, and reward. While models of basal ganglia function have well-founded hypotheses about the neural origin of signals representing context and reward, the function and origin of signals representing action are less clear. Recent findings suggest that exploratory or variable behaviors are initiated by a wide array of ‘action-generating’ circuits in the midbrain, brainstem, and cortex. Thus, in order to learn, the striatum must incorporate an efference copy of action decisions made in these action-generating circuits. Here we review several recent neural models of reinforcement learning that emphasize the role of efference copy signals, as well as ideas about how these signals might be integrated with inputs signaling context and reward.

Actions that produce a satisfying effect in a particular situation become more likely to occur again in that situation [1]. This simple statement, known as Thorndike's Law of Effect, is one of the central tenets of animal behavior, and forms the basis of instrumental learning, or operant conditioning [2,3]. It is also at the core of reinforcement learning, a computational framework that formalizes the process of determining the best course of action in any situation in order to maximize a quantifiable reward signal [4]. The Law of Effect embodies the simple intuition that in order to learn from our past actions, we need to have the convergence of three distinct pieces of information: signals representing the situation (or context) in which an action takes place; a signal representing the action that is being taken; and, finally, a signal representing the outcome of that action. While the neural basis of context and reward signals in biological models of reinforcement learning are well founded, the neural basis of action signals is less apparent. Several recent neural models of reinforcement learning have emphasized the role of efference copy signals, and incorporated ideas about how such signals might be integrated with inputs signaling context and reward.

Neural circuitry in the basal ganglia (BG) is well known to be involved in the control of learned behaviors [5,6], and the striatum, the input structure of the BG, is well established as a key structure in the neural implementation of reinforcement learning [7-10]. Some of the most compelling support for this view come from work demonstrating the role of basal

Address for correspondence: fee@mit.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ganglia circuitry in oculomotor learning, in which animals are trained, using rewards, to make saccades in a particular direction depending on which visual stimulus is presented [11-13].

In one simple and elegant model for the role of BG circuitry in these behaviors [14], cortical neurons representing the appearance of the rewarded stimulus are thought to activate medium spiny neurons (MSNs) in the ‘direct pathway’ of the caudate nucleus (the oculomotor part of the striatum), which, through a process of disinhibition, activates saccade-generating neurons of the superior colliculus to cause a robust saccade in the rewarded direction. Importantly, different MSNs in this pathway project to different parts of the superior colliculus, driving saccades to different parts of visual space. More generally, one can view the striatum as a massive switchboard, capable of connecting cortical neurons signaling a vast array of different contexts, to MSNs in a large number of different motor ‘channels’, including BG outputs to midbrain and brainstem structures [15], as well as the thalamus, which can in turn activate circuits in motor and premotor cortex [16,17]. In the simple oculomotor learning model shown in Figure 1, the context and motor channels have been reduced to a minimal representation of two visual stimuli and two saccade directions, and the switchboard has only four possible connections.

The key problem of reinforcement learning, then, is to determine which connections in the switchboard to strengthen. Before learning, the association between context and action that leads to a favorable outcome is unknown. Thus, we imagine that all possible connections between context inputs and the MSNs of each motor channel exist, but they are initially weak. Thorndike's Law of Effect suggests that if any particular pairing of a context and an action taken consistently leads to reward, we would like to strengthen synapses between the cortical input representing that context and the MSNs driving that action. After learning, then, any time the context neuron becomes active, it will activate the MSNs that generate the rewarded behavior.

But how does a corticostriatal context synapse know what action was taken? Some models of basal ganglia function [18-20] assume that the ‘actor’ that generates exploratory actions during learning is in the striatum itself. In this case learning is simple: If the decision to saccade to the left or right is generated by spontaneous activity in MSNs, then all three signals important for learning are available at each synapse: the context signal is carried by the presynaptic inputs to a corticostriatal synapse, the action signal is carried by postsynaptic spiking, and the reward signal could be carried by a phasic release of dopamine [21-24]. Indeed, it has been suggested that the corticostriatal learning rule that underlies reinforcement learning is a form of gated spike-timing dependent plasticity [25-28]. This idea is consistent with recent findings on the role of dopamine in corticostriatal plasticity [29-31].

We now arrive at the crux of the problem. While the BG appears to play a powerful role in driving specific actions *after* learning, several lines of evidence suggest that it may not be the origin of exploratory behaviors *before or during* oculomotor learning (reviewed in [32]). Thus, ‘exploratory’ saccades early in learning may be initiated, not in the striatum, but in one of the many brain circuits that project to the superior colliculus and are capable of triggering or influencing saccade generation [33]. More generally, spontaneous behaviors produced by a naïve, untrained animal may be initiated by the myriad behavior-generating circuits distributed throughout motor cortex and the brainstem [34]. These circuits could be activated by external sensory stimuli, or even by intrinsic ‘noisy’ mechanisms that promote spontaneous behaviors. For example, exploratory vocal babbling and song variability in juvenile songbirds does not require basal ganglia circuitry [35]; rather, the variability that underlies vocal learning is driven by—and possibly generated within—a specialized circuit in the avian cortex known as LMAN [36-39]. Furthermore, behavior-generating circuits likely incorporate competitive mechanisms, such as mutual inhibition, that select (i.e. decide) which of many possible actions animal will take [40,41]. This view is distinct from the early ‘action-selection’ hypothesis of basal ganglia function [42,43] in which competitive processes within the BG make decisions about which action is taken. It suggests, instead, that the BG act to bias decision processes already built into behavior-generating circuits outside the BG, the purpose of which is to tilt these decisions in favor of actions that previous experience has shown lead to a better outcome [10,44,45].

So if the striatum does not decide which motor actions to take, but simply biases decisions made elsewhere, how does the striatum know which actions were actually taken, and how might this information be used to control plasticity at corticostriatal context synapses? Several recent models of basal ganglia function have highlighted the potential importance of efference copy signals to shape plasticity in the striatum [32,44,46,47]. Frank [44] describes a model in which decision-making circuits in premotor cortex send a collateral to the striatum representing a set of potential actions under consideration. The activity of premotor cortical neurons representing different actions build up until, through a competitive interaction, one of these actions is selected, silencing the neurons representing the actions not taken. The feedback signal from the winning neuron to the striatum then determines which striatal neurons can undergo plasticity. After learning, the output of the basal ganglia circuit then feeds back to cortical circuits to bias the cortical circuits toward decisions with a favorable outcome.

Redgrave and Gurney [46] have taken a related approach to address a different question: how does an animal learn agency—the consequences of its own actions? They suggest that novel actions are reinforced, not by reward, but by the appearance of any unpredicted salient stimulus, signaled by short-latency phasic dopaminergic responses [23,48,49]. They proposed a model in which the convergence of these salience signals with context and efference copy signals could be used to discover the actions that, within a particular context, led to a novel outcome.

More recently, a model of basal ganglia function that utilizes an efference copy of motor actions [47] was developed in the context of a reinforcement-learning framework of songbird vocal learning [50]. Neurons in the cortical variability-generating nucleus LMAN project to the song motor pathway but also send a collateral to the song-related basal ganglia circuit Area X [51]. It was proposed that MSNs in Area X measure the correlation between LMAN ‘variation’ commands and a measure of vocal performance [47], potentially transmitted to Area X by dopaminergic input from VTA/SNc [52]. Thus, MSNs in Area X could discover which LMAN variation commands lead to better song performance and which lead to worse song performance.

Of course, the context in which a particular variation command takes place is important. Additional tension on a muscle of the vocal organ, driven by LMAN, may make the song better at one time in the song, but might make it worse at a different time. Thus, this computation—correlating LMAN activity with song performance—should be carried out independently at each time in the song [47] by taking advantage of a context signal transmitted to Area X from neurons in the premotor cortical nucleus HVC (used as a proper name). These neurons are sparsely active [53,54], each generating a brief burst of spikes at a small number of times in the song, and, as a population, are likely active throughout the song. If each MSN in Area X receives synaptic inputs from a set of HVC neurons representing every moment in the song, together with an efference copy signal from LMAN and a song-evaluation signal, then each MSN has all the information required to carry out reinforcement learning: context, action, and reward. Specifically, at each HVC-MSN synapse, a coincidence of presynaptic input from HVC and an efference copy input from LMAN would signal the occurrence of a particular action/context pairing (i.e., a particular song variation at a particular time). According to the rule described earlier, our learning rule should simply strengthen the HVC synapse if a coincidence of HVC input and LMAN efference copy input is followed by reward (a better song). It is envisioned that a synaptic eligibility trace would maintain a ‘memory’ of the HVC-LMAN coincidence until the later arrival of the song-evaluation (reward) signal [4,27,55,56].

After learning, the strengthened HVC inputs would strongly drive the MSN at specific times in the song corresponding to times at which activity of the LMAN neuron was previously observed to make the song better. Activity in this MSN should, in turn, feed back to LMAN (through the direct pathway to the thalamus and back to LMAN) to bias the LMAN neuron to be more active during singing, thus biasing the song toward better vocal performance. Indeed, there is strong evidence that the variability generated by LMAN becomes biased during learning, pushing the song in the direction of reduced vocal errors [57,58].

Inspired by the model of song learning, the concept of using motor efference copy was recently extended to a simple model of oculomotor learning [32]. This model, which bears some conceptual similarity to that of Frank [44], envisions that the oculomotor striatum receives an efference copy of saccade commands from deep layers of the superior colliculus [59] or from oculomotor cortical areas such as the frontal eye fields (FEF), which send a topographically organized projection to the superior colliculus [60,61], and form a collateral projection to the caudate [16,62]. The idea is that early in learning, before the monkey has learned the association between visual targets and the saccades that lead to reward, FEF circuitry acts as a source of saccade variability, perhaps analogous with the songbird nucleus LMAN, that generates random ‘guesses’ in response to each visual stimulus. Figure 1 shows a simple circuit in which the function of the efference copy of saccade guesses from FEF would be to gate plasticity at context-to-MSN synapses, but not to drive activity in MSNs. Specifically, the learning rule would be as follows: strengthen context-to-MSN synapses when activation of the presynaptic context input followed by efference copy input is followed by reward. For example, if the appearance of stimulus 1 followed by a saccade to the left yields a reward, then this learning rule correctly strengthens only the connection from the stimulus 1 neuron onto the MSN in the left saccade channel. After learning, this strengthened synapses allows the neuron representing stimulus 1 to drive spiking in the left MSN, thus biasing saccade generation in the left direction, as desired. The end result of this learning rule would be that the synaptic strength of each context-to-MSN synapse resembles the state-action (Q) value envisioned in Q-learning models of reinforcement learning [4,63].

The models described above predict an asymmetry in the function and anatomy of the context and the hypothesized motor efference copy inputs to the striatum: First, reward-related plasticity occurs only at the context inputs onto MSNs, while efference copy inputs are not plastic but serve only to gate plasticity. Second, MSN spiking is driven by context inputs alone, not by motor efference copy inputs. Finally, in this model there are important differences in the required convergence of these two signals: an MSN must receive only a single neuron-wide efference copy input indicating that an action has taken place. In contrast, each MSN must receive thousands of context inputs representing the vast array of different situations or stimuli that can potentially be paired with the action represented by the MSN. How might detection of context and efference copy inputs be biophysically implemented on MSNs? One possibility is that efference copy inputs might synapse preferentially on dendritic shafts, while context inputs might synapse preferentially on dendritic spines, which could then serve as highly local detectors of coincidence between presynaptic context input and the more global postsynaptic depolarization provided by the efference copy input. Another notable asymmetry between efference copy and context inputs in this model relates to the way these inputs must project within the striatum. In particular, the projection of efference copy signals must be local within one motor channel of the basal ganglia, while the projection of context signals must be highly divergent across many motor channels.

Note that, in this model, MSNs must serve two distinct functions: during learning they measure the coincidence of context, efference copy, and reward inputs; after learning, they begin to play a motor role in which they are driven to spike by their context inputs and, in turn, act on their downstream motor targets. Early in learning, before the context inputs are sufficiently strong to drive the MSNs, there would be no conflict between these functions. However, as learning progresses, it is possible that the large context inputs required to drive spiking in the MSN could interfere with the detection of coincidences between presynaptic inputs and efference copy inputs, particularly if the latter inputs are mediated by dendritic depolarization. However, if there is a sufficient latency between the spiking of MSNs (driven by the context inputs) and the decision outcome in the downstream motor circuits it biases, then the MSN would recover from its premotor function in time to receive the resulting efference copy signaling the outcome of that decision. In this version of the model, the learning rule could be implemented as a spike-timing-dependent mechanism [25-28,31] such that potentiation would occur only when context inputs precede efference copy inputs by the appropriate latency [47].

In summary, the requirement that reinforcement learning integrates signals representing context, action, and reward, together with an emerging understanding that actions, decisions, and behavioral variability may be generated by circuitry outside the basal ganglia, have led to several models that emphasize the potential role of an efference copy of motor actions to shape striatal learning. They also suggest specific ways these signals could be incorporated into corticostriatal learning rules, and make a number of testable predictions about the way signals carrying context and efference copy signals are integrated onto MSNs at the level of axonal and synaptic anatomy and plasticity.

Acknowledgments

I thank Michael Frank, Peter Redgrave, Adrienne Fairhall, and Jesse Goldberg for helpful discussions. Funding for this work was provided by NIH grant #R01MH067105.

Literature Cited

1. Thorndike, EL. Animal Intelligence. Darien, CT: Hafner; 1911.

2. Packard MG, Knowlton BJ. Learning and memory functions of the basal ganglia. Annu Rev Neurosci. 2002; 25:563–593. [PubMed: 12052921]

3. Graybiel AM. Habits, rituals, and the evaluative brain. Annu Rev Neurosci. 2008; 31:359–387. [PubMed: 18558860]

4. Sutton, RS.; Barto, AG. Reinforcement learning: An introduction. Cambridge, MA: The MIT Press; 1998.

5. Graybiel AM, Aosaki T, Flaherty AW, Kimura M. The basal ganglia and adaptive motor control. Science. 1994; 265:1826–1831. [PubMed: 8091209]

6. Graybiel AM. The basal ganglia and chunking of action repertoires. Neurobiol Learn Mem. 1998; 70:119–136. [PubMed: 9753592]

7. Houk, JC. Information processing in modular circuits linking basal ganglia and cerebral cortex. In: Houk, JC.; Davis, JL.; Beiser, DG., editors. Models of Information Processing in the Basal Ganglia. Cambridge, MA: The MIT Press; 1995. p. 3-10.

8. Doya K. Complementary roles of basal ganglia and cerebellum in learning and motor control. Curr Opin Neurobiol. 2000; 10:732–739. [PubMed: 11240282]

9. Daw ND, Doya K. The computational neurobiology of learning and reward. Curr Opin Neurobiol. 2006; 16:199–204. [PubMed: 16563737]

10. Frank MJ. Computational models of motivated action selection in corticostriatal circuits. Curr Opin Neurobiol. 2011; 21:381–386. [PubMed: 21498067]

11. Kawagoe R, Takikawa Y, Hikosaka O. Reward-predicting activity of dopamine and caudate neurons--a possible mechanism of motivational control of saccadic eye movement. J Neurophysiol. 2004; 91:1013–1024. [PubMed: 14523067]

12. Pasupathy A, Miller EK. Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature. 2005; 433:873–876. [PubMed: 15729344]

13. Hikosaka O. Basal ganglia mechanisms of reward-oriented eye movement. Ann N Y Acad Sci. 2007; 1104:229–249. [PubMed: 17360800]

**14. Hikosaka O, Nakamura K, Nakahara H. Basal ganglia orient eyes to reward. J Neurophysiol. 2006; 95:567–584. This paper synthesizes results from electrophysiological recordings in the caudate nucleus of monkeys during an oculomotor learning task to construct a simple model of how the basal ganglia mediate reward-driven changes in saccade behavior. [PubMed: 16424448]

15. Grillner S, Hellgren J, Ménard A, Saitoh K, Wikström MA. Mechanisms for selection of basic motor programs--roles for the striatum and pallidum. Trends Neurosci. 2005; 28:364–370. [PubMed: 15935487]

16. Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu Rev Neurosci. 1986; 9:357–381. [PubMed: 3085570]

17. Hoover JE, Strick PL. Multiple output channels in the basal ganglia. Science. 1993; 259:819–821. [PubMed: 7679223]

18. Berns GS, Sejnowski TJ. A computational model of how the basal ganglia produce sequences. J Cogn Neurosci. 1998; 10:108–121. [PubMed: 9526086]

19. Hikosaka O, Nakahara H, Rand MK, Sakai K, Lu X, Nakamura K, Miyachi S, Doya K. Parallel neural networks for learning sequential procedures. Trends Neurosci. 1999; 22:464–471. [PubMed: 10481194]

20. Ito M, Doya K. Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit. Curr Opin Neurobiol. 2011; 21:368–373. [PubMed: 21531544]

21. Mirenowicz J, Schultz W. Importance of unpredictability for reward responses in primate dopamine neurons. J Neurophysiol. 1994; 72:1024–1027. [PubMed: 7983508]

22. Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997; 275:1593–1599. [PubMed: 9054347]

23. Schultz W. Predictive reward signal of dopamine neurons. J Neurophysiol. 1998; 80:1. [PubMed: 9658025]

24. Hollerman JR, Schultz W. Dopamine neurons report an error in the temporal prediction of reward during learning. Nat Neurosci. 1998; 1:304–309. [PubMed: 10195164]

25. Farries MA, Fairhall AL. Reinforcement learning with modulated spike timing dependent synaptic plasticity. J Neurophysiol. 2007; 98:3648–3665. [PubMed: 17928565]

26. Florian RV. Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. Neural Comput. 2007; 19:1468–1502. [PubMed: 17444757]

*27. Izhikevich EM. Solving the distal reward problem through linkage of STDP and dopamine signaling. Cereb Cortex. 2007; 17:2443–2452. A nice review of the problem of temporal credit assignment, and the potential role of spike timing, eligibility traces, and reward in reinforcement learning. [PubMed: 17220510]

28. Roberts PD, Santiago RA, Lafferriere G. An implementation of reinforcement learning based on spike timing dependent plasticity. Biol Cybern. 2008; 99:517–523. [PubMed: 18941775]

29. Reynolds JNJ, Hyland BI, Wickens JR. A cellular mechanism of reward-related learning. Nature. 2001; 413:67–70. [PubMed: 11544526]

30. Pawlak V, Kerr JN. Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. J Neurosci. 2008; 28:2435–2446. [PubMed: 18322089]

31. Shen W, Flajolet M, Greengard P, Surmeier DJ. Dichotomous dopaminergic control of striatal synaptic plasticity. Science. 2008; 321:848–851. [PubMed: 18687967]

**32. Fee MS. Oculomotor learning revisited: a model of reinforcement learning in the basal ganglia incorporating an efference copy of motor actions. Front Neural Circuits. 2012; 6:38. A model of songbird vocal learning incorporating motor efference copy in striatal learning is extended to the problem of oculomotor learning. The paper speculates on mechanisms by which efference copy signals could act on striatal learning, and on different possible pathways that could carry efference copy signals to the striatum. [PubMed: 22754501]

33. Wurtz RH, Albano JE. Visual-motor function of the primate superior colliculus. Annu Rev Neurosci. 1980; 3:189–226. [PubMed: 6774653]

34. Swanson LW. Cerebral hemisphere regulation of motivated behavior. Brain Res. 2000; 886:113–164. [PubMed: 11119693]

*35. Goldberg JH, Fee MS. Vocal babbling in songbirds requires the basal ganglia-recipient motor thalamus but not the basal ganglia. J Neurophysiol. 2011; 105:2729–2739. Complete bilateral lesions of the song-related basal ganglia Area X in juvenile birds have little effect on vocal variability. These findings support the idea that the BG must receive and integrate an efference copy of variation commands from LMAN in order to learn from them. [PubMed: 21430276]

36. Ölveczky BP, Andalman AS, Fee MS. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biol. 2005; 3:e153. [PubMed: 15826219]

37. Kao MH, Doupe AJ, Brainard MS. Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. Nature. 2005; 433:638–643. [PubMed: 15703748]

38. Aronov D, Andalman AS, Fee MS. A specialized forebrain circuit for vocal babbling in the juvenile songbird. Science. 2008; 320:630–634. [PubMed: 18451295]

39. Aronov D, Veit L, Goldberg JH, Fee MS. Two distinct modes of forebrain circuit dynamics underlie temporal patterning in the vocalizations of young songbirds. J Neurosci. 2011; 31:16353–16368. [PubMed: 22072687]

40. Choi GB, Dong HW, Murphy AJ, Valenzuela DM, Yancopoulos GD, Swanson LW, Anderson DJ. Lhx6 delineates a pathway mediating innate reproductive behaviors from the amygdala to the hypothalamus. Neuron. 2005; 46:647–660. [PubMed: 15944132]

**41. Mysore SP, Knudsen EI. A shared inhibitory circuit for both exogenous and endogenous control of stimulus selection. Nat Neurosci. 2013; 16:473–478. Identifies an inhibitory circuit in the tectum (i.e. superior colliculus) of the bird that mediates a winner-take-all interaction underlying visual target selection. These findings suggest that competitive interactions underlying action selection do not necessarily reside within the striatum. [PubMed: 23475112]

42. Mink JW. The basal ganglia: focused selection and inhibition of competing motor programs. Prog Neurobiol. 1996; 50:381–425. [PubMed: 9004351]

43. Redgrave P, Prescott TJ, Gurney K. The basal ganglia: a vertebrate solution to the selection problem? Neuroscience. 1999; 89:1009–1023. [PubMed: 10362291]

**44. Frank MJ. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J Cogn Neurosci. 2005; 17:51–72. Although not highlighted explicitly in this paper, the model of BG function described here was one of the first to incorporate an efference copy of cortical action signals into striatal learning. It also assigns action decisions (e.g. action selection) to a winner-take-all mechanism within cortical circuits, upon which the BG acts to bias the outcome. [PubMed: 15701239]

45. Samson RD, Frank MJ, Fellous JM. Computational models of reinforcement learning: the role of dopamine as a reward signal. Cogn Neurodyn. 2010; 4:91–105. [PubMed: 21629583]

**46. Redgrave P, Gurney K. The short-latency dopamine signal: a role in discovering novel actions? Nat Rev Neurosci. 2006; 7:967–975. Describes an interesting alternative view that dopaminergic neurons signal the presence of a surprising outcome, not a rewarding outcome. Combined with context and efference copy signals, this would allow the BG to learn which, if any, of an animals own actions was responsible for this outcome, and to acquire a sense of ‘agency’. [PubMed: 17115078]

**47. Fee MS, Goldberg JH. A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. Neuroscience. 2011; 198:152–170. This paper describes a model of vocal learning in which the song-related basal ganglia Area X uses an efference copy of song variability ‘commands’ to learn which commands make the song better and which make the song worse. [PubMed: 22015923]

48. Horvitz JC. Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. Neuroscience. 2000; 96:651–656. [PubMed: 10727783]

49. Bromberg-Martin ES, Matsumoto M, Hikosaka O. Dopamine in motivational control: rewarding, aversive, and alerting. Neuron. 2010; 68:815–834. [PubMed: 21144997]

50. Doya K, Sejnowski T. A novel reinforcement model of birdsong vocalization learning. Advances in Neural Information Processing Systems. 1995; 7:101–108.

51. Vates GE, Nottebohm F. Feedback circuitry within a song-learning pathway. Proc Natl Acad Sci U S A. 1995; 92:5139–5143. [PubMed: 7761463]

52. Gale SD, Perkel DJ. A basal ganglia pathway drives selective auditory responses in songbird dopaminergic neurons via disinhibition. J Neurosci. 2010; 30:1027–1037. [PubMed: 20089911]

53. Kozhevnikov AA, Fee MS. Singing-related activity of identified HVC neurons in the zebra finch. J Neurophysiol. 2007; 97:4271–4283. [PubMed: 17182906]

54. Prather JF, Peters S, Nowicki S, Mooney R. Precise auditory-vocal mirroring in neurons for learned vocal communication. Nature. 2008; 451:305–310. [PubMed: 18202651]

55. Frey U, Morris RGM. Synaptic tagging and long-term potentiation. Nature. 1997; 385:533–536. [PubMed: 9020359]

56. Fiete IR, Fee MS, Seung HS. Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. J Neurophysiol. 2007; 98:2038–2057. [PubMed: 17652414]

*57. Andalman AS, Fee MS. A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. PNAS. 2009; 106:12518–12523. See Warren et al (2011), below. [PubMed: 19597157]

*58. Warren TL, Tumer EC, Charlesworth JD, Brainard MS. Mechanisms and time course of vocal learning and consolidation in the adult songbird. J Neurophysiol. 2011; 106:1806–1821. Together with Andalman et al (2009), this paper provides compelling evidence that a basal ganglia-cortical circuit in the songbird biases vocal variability in the direction of improved vocal performance, which then directs long-term plastic changes in the cortical song motor pathway. [PubMed: 21734110]

59. Sommer MA, Wurtz RH. Brain circuits for the internal monitoring of movements. Annu Rev Neurosci. 2008; 31:317–338. [PubMed: 18558858]

60. Bruce CJ, Goldberg ME. Primate frontal eye fields. I. Single neurons discharging before saccades. J Neurophysiol. 1985; 53:603–635. [PubMed: 3981231]

61. Komatsu H, Suzuki H. Projections from the functional subdivisions of the frontal eye field to the superior colliculus in the monkey. Brain Res. 1985; 327:324–327. [PubMed: 2985177]

62. Künzle H, Akert K. Efferent connections of cortical, area 8 (frontal eye field) in Macaca fascicularis. A reinvestigation using the autoradiographic technique. J Comp Neurol. 1977; 173:147–163. [PubMed: 403205]

63. Watkins CJ, Dayan P. Q-learning. Machine learning. 1992; 8:279–292.

Highlights

1. Reinforcement learning requires signals representing context, action, and reward.
2. Actions can be generated by many different midbrain and cortical circuits.
3. We review recent models that incorporate an efference copy of action commands.
4. Also discussed are ideas about how these signals might be integrated in striatal circuits.

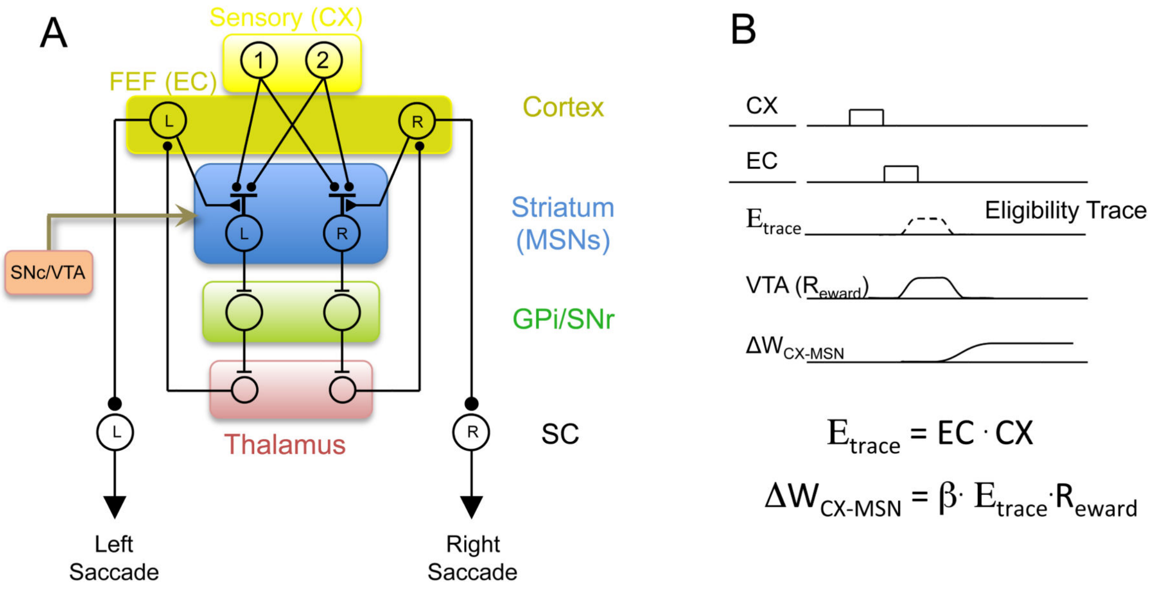

Figure 1.

A model of basal ganglia function incorporating efference copy of motor actions. Shown is the schematic of a network to implement reinforcement learning of an association between stimulus and saccade direction. In this hypothetical model of oculomotor learning, leftward or rightward saccades are driven by neurons in the frontal eye fields (FEF). The FEF neurons are configured with mutual inhibition such that saccade direction is decided by winner-take-all interaction. A) Schematic diagram of the direct pathway of the striatum (blue) and SNr (green). In this model, the BG outputs feed back to the FEF through the thalamus (red), rather than projecting directly to the superior colliculus. After learning, cortical neurons representing sensory, or context (CX), input to the striatal MSNs can bias saccade decisions by disinhibition of the thalamic neurons. Early in learning, however, saccade 'guesses' to the left or right, are generated by noisy mechanisms in the FEF. The left and right FEF neurons send an efference copy to the left and right MSNs, respectively. The EC inputs (triangular synapse) do not drive spiking in the MSN, but serve to gate plasticity at the CX-to-MSN synapse (solid circle). B) Depiction of the proposed learning rule that drives potentiation of a CX-to-MSN synapse. A context input, followed by an efference copy input activates an eligibility trace. If a dopaminergic reward signal arrives at the synapse and temporally overlaps with the eligibility trace, the CX-to-MSN synapse is potentiated.

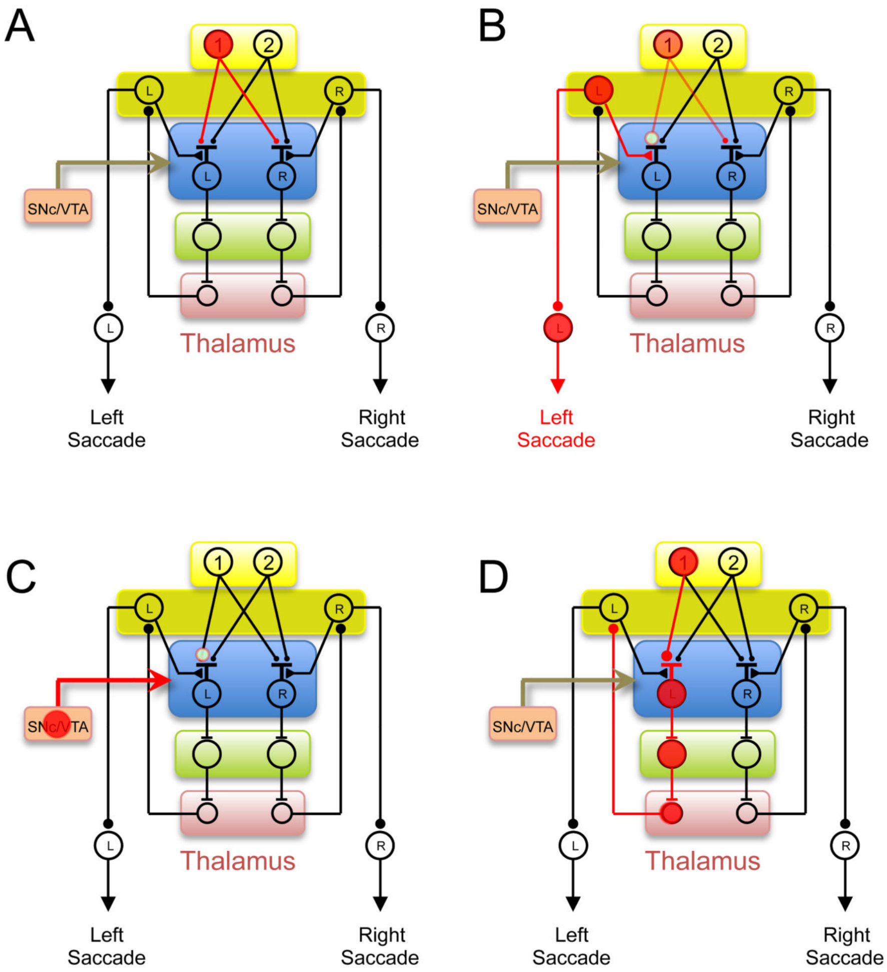

Figure 2.

The sequence of events that implements the hypothesized synaptic learning rule. One training trial is represented in which a leftward saccade was the correct response to Stimulus 1. A) Stimulus 1 is presented, activating context neuron 1 (CX1); B) The left FEF neuron turns on randomly, driving a leftward saccade and activating the left EC input; C) The presynaptic input at the CX1-to-MSN input, followed by the EC input (depicted as terminating on the dendritic shaft of the MSN), activates an eligibility trace only in the CX1-to-MSN synapse. Subsequent arrival of the dopaminergic reward signal, temporally overlapped with the eligibility trace, produces LTP only at the CX1-MSN synapse. D) On subsequent trials, presentation of Stimulus 1 will activate the left MSN, disinhibiting the left thalamic neuron, thus biasing the competitive decision in the FEF toward leftward saccades. After a short latency, this decision occurs when one of the FEF neurons becomes active, and learning continues from step (B). In this instantiation of the model, the latency of the bias signal (through the direct pathway back to the FEF) is sufficiently long that the CX1 input that drives bias is temporally separated from the efference copy input that signals the outcome of the saccade direction decided by the FEF network. Thus learning can continue without interference, even after context inputs begin to drive bias in the MSN.