Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

The role of efference copy in striatal learning

Michale S Fee

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. Also described are ideas about how these signals might be integrated with inputs signaling context and reward.

Addresses
McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, United States

Corresponding author: Fee, Michale S (fee@mit.edu)

Current Opinion in Neurobiology 2014, **25**:194–200

This review comes from a themed issue on **Theoretical and computational neuroscience**

Edited by **Adrienne Fairhall** and **Haim Sompolinsky**

For a complete overview see the Issue and the Editorial

Available online 21st February 2014

0959-4388/$ – see front matter,

http://dx.doi.org/10.1016/j.conb.2014.01.012

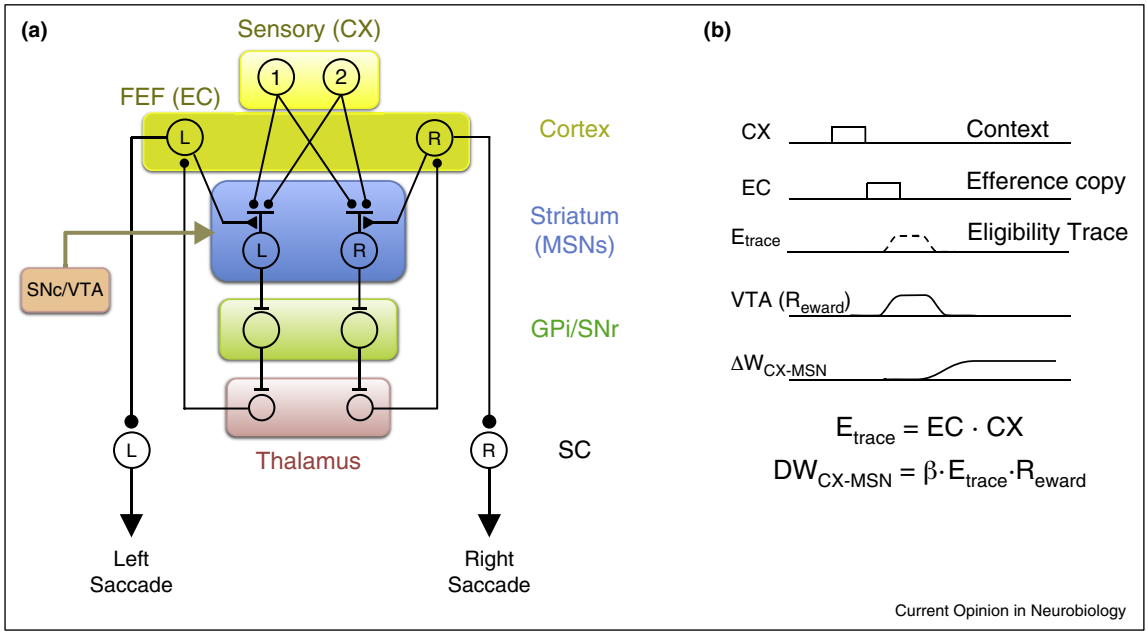
Actions that produce a satisfying effect in a particular situation become more likely to occur again in that situation [1]. This simple statement, known as Thorndike's Law of Effect, is one of the central tenets of animal behavior, and forms the basis of instrumental learning and operant conditioning [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 described, 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 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 Fig. 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,

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 (not shown) such that saccade direction is decided by a cortical 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 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 (EC) 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.

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 [25,26,27•,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