Basal Ganglia System as an Engine for Exploration

Chapter · March 2014

DOI: 10.1007/978-1-4614-6675-8_81

Srinivasa Chakravarthy
Indian Institute of Technology Madras

Pragathi Priyadharsini Balasubramani
University of Rochester

Basal Ganglia System as an Engine for Exploration

Definition

The basal ganglia (BG) system is a deep brain circuit with wide-ranging brain functions. Exploration refers to the sampling of a variety of behaviors not firmly established within a learned repertoire. While the neural source of variability driving exploration within the subcortex has not been identified, the hypothesis that the indirect pathway of the BG is the subcortical substrate for exploration leads to explanations for how a range of putative BG functions might be performed.

Detailed Description

Reinforcement Learning and the Basal Ganglia

For nearly a century, a certain "mysteriousness" has been attributed to the function of the basal ganglia (BG) system - a deep brain circuit of multiple interconnected nuclei, with rich connections to large parts of the cortex (Kinnier Wilson in his Croonian lectures in 1925, Marsden 1982). The mystique surrounding BG has its roots perhaps in the multifarious functions of this circuit. Action selection, action gating, sequence generation, motor preparation, reinforcement learning, timing, working memory, goal-directed behavior, and exploratory behavior - the list of putative BG functions is long and perhaps, by the current state of knowledge, is also incomplete. Lesions of this circuit can show manifestations in many forms - from simple reaching movements to handwriting, balancing and gait, speech and language function, eye movements, and force generation, in addition to cognitive and affective manifestations. A long line of neurological (Parkinson's disease, Huntington's disease, athetosis, chronic fatigue syndrome) (DeLong 1990) and neuropsychiatric disorders (schizophrenia, obsessive-compulsive disorder, ADHD, apathy, abulia, insomnia) (Ring and Serra-Mestres 2002) are associated with BG impairment.

In one of the earliest forays into the mystery of the BG, Albin et al. (1989) interpreted BG anatomy to consist of two pathways (Albin et al. 1989) with complementary roles in motor manifestations (Fig. 1). BG consist of several nuclei - striatum (caudate and putamen), globus pallidus externa/interna, subthalamic nucleus, and substantia nigra pars compacta (SNc)/reticulata. The BG circuit as a whole receives extensive inputs from the cortex and projects back to cortex via thalamus. Cortical signals to BG flow along two pathways: the direct pathway (DP) and the indirect pathway (IP). The projections from the striatum, the input port to BG, to globus pallidus interna (GPi), one of the output ports, constitute the DP. The indirect route which connects the striatum to GPi via globus pallidus externa (GPe) and subthalamic nucleus (STN) is the IP (Alexander and Crutcher 1990). Lesions of DP affecting particularly the projections from the striatum to GPi are associated with hypokinetic disorders (distinguished by a paucity of movement), and lesions of IP produce hyperkinetic disorders like chorea and tremor. These findings led to the thinking that activation of the DP facilitates movement, and hence it became known as the "Go" pathway. Contrarily, IP was dubbed the "NoGo" pathway since its activation typically inhibits movement (Contreras-Vidal and Stelmach 1995).

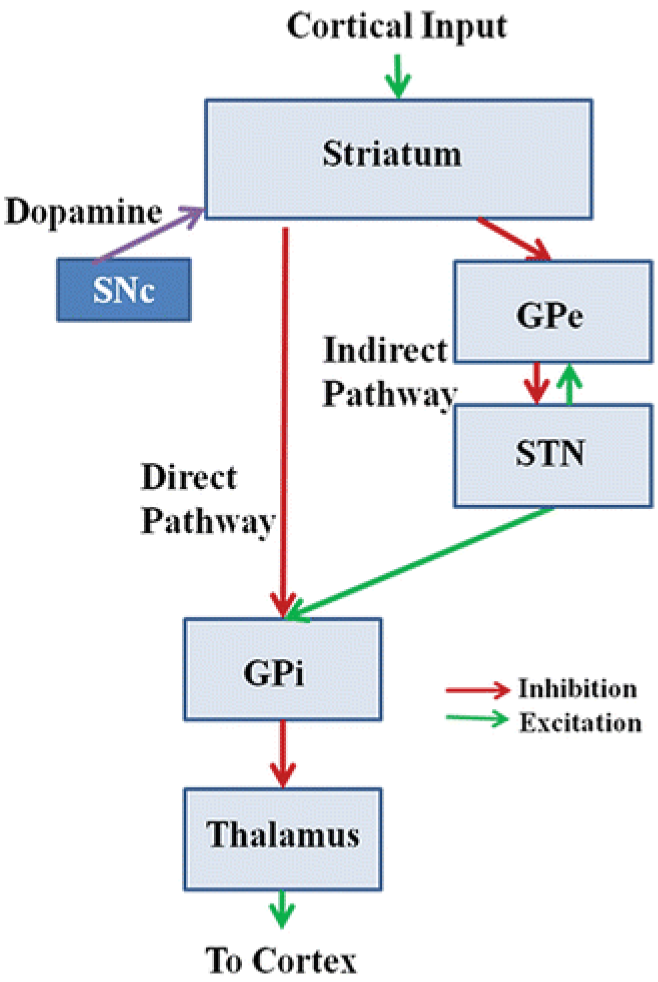

Fig. 1
The schematic showing the direct and indirect pathways of basal ganglia

A need to go beyond the simple Go/NoGo picture of BG function had its seeds in experiments (Houk et al. 1995; Schultz et al. 1997) on the firing properties of mesencephalic dopamine cells. Although activities of dopaminergic cells have been linked to reward sensing for a long time, experiments by Schultz et al. (1997) specifically showed that dopamine neurons of ventral tegmental area (VTA) respond to unconditional rewards (food or juice). When a sensory stimulus (like a sound or a light flash) consistently precedes the appearance of reward such that the stimulus is predictive of the reward, then dopamine cells fire in response to the stimulus and also the reward. Such findings led to the insight that dopamine cell activity is analogous to a quantity known as temporal difference error (TD error) that appears in reinforcement learning (RL) theory - a branch of machine learning (Sutton and Barto. 1998). Recognition of the analogy between mesencephalic dopamine signals and TD error signals of RL has inspired a much larger effort to draw parallels between other elements of RL theory and anatomical components of BG. Although the effort to explain various functions of BG using RL concepts is a story in the making, it is believed that RL holds the promise to create a comprehensive theory of BG in the long term (Chakravarthy et al. 2010).

RL theory describes how an agent can learn correct stimulus-response (SR) relationships using reward feedback from the environment. For a given stimulus, responses that yield rewards are reinforced, while those that result in punishment are attenuated. The problem is often complicated by the fact that rewards come after a delay following the response or even after a whole series of responses. The agent then needs a surrogate to reward, which guides its responses in the intervening period. RL theory proposes the value function as such a surrogate; a computational module called the critic computes value. (The module that performs actions is known as the actor.) Value is defined as the total discounted future reward that an agent expects to receive from a given state. Thus, once the value is known, at any instant, the agent chooses the response or action that brings about the greatest increase in value, a process known as exploitation. Sometimes it is desirable for the agent to try out actions that are not optimal, by way of adapting to the changing reward patterns of the real world. This selection of suboptimal actions, typically stochastically, is known as exploration. Thus, exploitation and exploration are two key complementary processes, the yin and yang, of RL theory (Sutton and Barto. 1998).

The Indirect Pathway and Exploration

In actor/critic (AC) models of BG, the emphasis is often on the respective substrates for the actor and critic components in the striatum (Joel et al. 2002), and the dopamine signal for its role in training the AC components. Thus, though exploitation and exploration are complementary processes, exploitation receives most of the attention. Such omission is perhaps not surprising as even in the AC framework, the actor and critic are recognized explicitly as modules, while exploration is a mere mechanism obtained by a "noise term" in RL equations. Since variability is ubiquitous in the brain - either arising due to thermal noise or chaotic neural dynamics - the search for a specific substrate for exploration in BG was felt unnecessary.

Experimental evidence particularly from functional neuroimaging seems to support this partial view, with a bias towards cortical substrates over subcortical ones. In fMRI studies, gamblers were asked to choose between slots that are expected to give highest rewards ("exploit") and less familiar slots that might turn out to be more profitable ("explore"). The areas in the brain that are preferentially activated during exploitation or exploration are noted (Daw et al. 2006). While substrates for value computations were found in orbitofrontal cortex (Knutson et al. 2001), substrates for exploration were found in anterior frontopolar cortex and intraparietal sulcus (Daw et al. 2006). The anterior cingulate cortex (ACC) is suggested to be involved in balancing between exploitation and exploration (Rushworth and Behrens 2008). Studies by Yoshida and Ishii (2006) found activation in prefrontal cortex and ACC when subjects were exploring a maze (Yoshida and Ishii 2006). In the subcortex, it was suggested that the ventral and dorsal striata correspond to critic and actor, respectively (O'Doherty et al. 2004). Thus, though both cortical and subcortical substrates of exploitation have been discovered, no corresponding subcortical substrates for exploration have been found.

Can there be subcortical substrates for exploration? Stein et al. (1997) showed that decorticated kittens can exhibit exploratory and goal-oriented behavior (Stein et al. 1997). Rats with damaged STN were shown to exhibit perseverative behavior or reduced exploration of new options, with persistent selection of older unrewarding ones (Baunez et al. 2001). When bicuculline, a GABA antagonist, was injected into the anterior GPe of primates, the animals exhibited stereotypic movements, and when it was injected into dorsolateral GPe, the animal produced hyperactivity that included exploratory or searching movements for food (Grabli et al. 2004). Drawing inspiration from the studies of Usher and Cohen et al. (1999), Doya (2002) suggested a link between norepinephrine levels and the "inverse temperature" parameter which controls exploration in RL literature. It is noteworthy that globus pallidus is reported to have high norepinephrine levels (Russell et al. 1992).

Thus, it appears compelling that the STN-GPe system constituting the IP of BG might be the subcortical substrate for exploratory behavior. The STN-GPe system and its intriguing oscillatory activity do not seem to occupy a prominent place in AC modeling literature. On the other hand, there is an entire line of modeling work that presents the STN-GPe system as a pacemaker in the brain, in reference to its oscillatory activity (Ring and Serra-Mestres 2002; Willshaw and Li 2002). These oscillations have also been linked to Parkinsonian tremor (Hurtado et al. 1999; Terman et al. 2002). Though the aforementioned STN-GPe models explain the behavioral effects of pathological oscillations, they attribute no role to the oscillations in the RL framework that is thought to govern the processes of BG. Under dopamine-deficient or Parkinsonian conditions, the firing patterns of STN and GPe neurons show dramatically increased correlation without significant increase in firing rate (Bergman et al. 1994; Brown et al. 2001). Since exploration is driven by noise in RL models, a brain region that drives exploration is expected to be a source of noise generated perhaps by complex neural dynamics. Considering the low correlation in STN-GPe under normal conditions, with increased correlation or loss of complexity in pathology, it is plausible that the STN-GPe is a subcortical substrate for exploration.

By an extended application of RL concepts, a comprehensive model of BG can be built in which the exploitative dynamics of the DP can be combined with STN-GPe oscillations that drive exploratory behavior (Chakravarthy et al. 2010). Thus emerges a view that while DP supports exploitation, IP subserves exploration, differing from the classical Go/NoGo view of BG. In a recent modeling study, it was shown that the exploitation (DP) versus exploration (IP) can be reconciled with Go (DP) versus NoGo (IP) view, by inserting a third regime dubbed the Explore regime. This regime would correspond to exploration and resides between the classic Go and NoGo regimes (Kalva et al. 2012). A series of BG models based on this view have been developed to account for a wide variety of BG-related motor and cognitive behaviors such as spatial navigation, saccades, reaching, and reward-punishment learning (Sridharan et al. 2006; Chakravarthy et al. 2010; Krishnan et al. 2011; Kalva et al. 2012; Priyadharsini et al. 2012; Gupta et al. 2013; Muralidharan et al. 2013).

The Basic Model

The intuitive ideas outlined before can be embodied in a simple mathematical model of BG. The framework explicitly represents striatum, STN, GPe and GPi, and DP and IP, to capture the structural aspects of BG, and also models:

- The nigrostriatal dopamine signal as TD error, using it to train corticostriatal connections
- The action of dopamine in switching between DP and IP, via its differential action on the D1 and D2 receptors of striatal medium spiny neurons
- Oscillations in the STN-GPe system
- Value computations in the striatum
- The classical Go (DP) and NoGo (IP) with the added "Explore" behavior.

Striatum

The binary action selection problem consists of choosing between two inputs based on their "salience." Thus, the input, a 2D vector, I^{ext}, representing the corticostriatal afferents, is presented to the striatum that consists of two 1D layers of sigmoidal neurons representing medium spiny neurons (MSNs) (Fig. 2). The first layer represents neurons that express D1-type dopamine receptors (R), whereas the second layer represents D2R-expressing MSNs. The D1R- and D2R-MSN layers project to GPi and GPe, respectively, and they also receive dopaminergic projections from SNc. Each component of I^{ext} is uniquely connected to one neuron each in D1 and D2 layers. The dopamine signal (δ) controls the sigmoidal gain of neurons in D1 and D2 layers. Whereas the gain of D1 neurons increases with δ, that of D2 neurons decreases with δ. With such an arrangement, it can be noticed that the DP is selected at higher dopamine levels and the IP at lower levels (Humphries and Gurney 2002; Humphries and Prescott 2010).

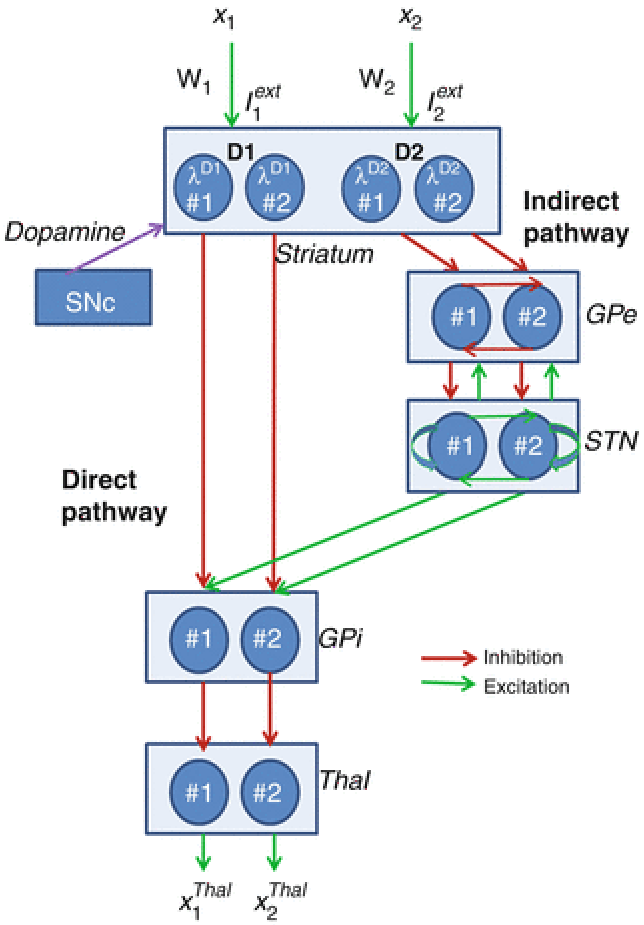

Fig. 2
The schematic of signal flow in the BG network model

The STN-GPe System

STN and GPe form a loop with excitatory projections from STN to GPe and inhibitory projections in the reverse direction. Excitatory/inhibitory pairs of neuronal pools are known to exhibit limit cycle oscillations (Gillies et al. 2002). The STN and GPe system receives two inputs: the inhibitory (GABAergic) projection from D2R MSNs of striatum to GPe and the excitatory (glutamatergic) cortical input via the hyperdirect pathway to STN. Several factors elicit oscillations in a STN-GPe neuron pair.

1. Increased striatal input to GPe: This property is corroborated by electrophysiological data from Bergman and Wichmann et al. (1994). Kravitz et al. (2010) observed that increased firing of D2R-MSNs in the striatum induce a state similar to Parkinson's, with motor symptoms like freezing, bradykinesia, and difficulty in movement initiation (Kravitz et al. 2010).
2. Increasing cortical input to STN: Electrophysiological studies show that ablation of cortical areas that project to STN largely abolished low-frequency oscillations in STN-GPe (Magill et al. 2001).
3. Reducing dopamine levels in STN-GPe: Organotypic culture studies show that the STN-GPe system exhibits low-frequency oscillations under dopamine-deficient conditions as in that of Parkinson's disease (Plenz and Kital 1999). STN and GPe oscillations seem to be triggered by the effect of dopamine loss on D2R which strengthens the STN-GPe coupling (Steiner and Tseng 2010).

Now consider a network model of the STN-GPe system in which STN and GPe layers have equal number of neurons with each STN neuron uniquely connected bidirectionally to a GPe neuron. Assume that both STN and GPe have complete internal connectivity with all connections within a nucleus having the same strength: ε_s for STN, ε_g for GPe; the one-to-one common connection strength for STN → GPe is w_{sg} and that of GPe → STN is w_{gs}. With ($\varepsilon = \varepsilon_s = -\varepsilon_g$) and ($w = w_{sg} = -w_{gs}$) and searching the (ε , w) space for finding the oscillatory behavior of the STN-GPe system, simulations show that for strong connections between STN and GPe (w) and weak lateral connections (ε), the network exhibits oscillations (Fig. 3) with poor correlation between the STN-GPe neuronal activity (Fig. 4).

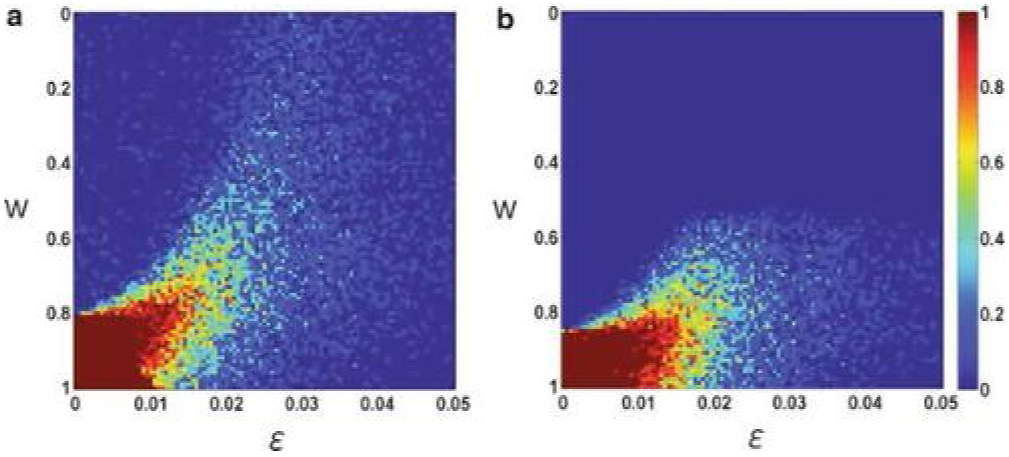

Fig. 3
The regions of parameter space (ε, w) over which probability of oscillations in (a) STN and (b) GPe is depicted (averaged over 50 trials)

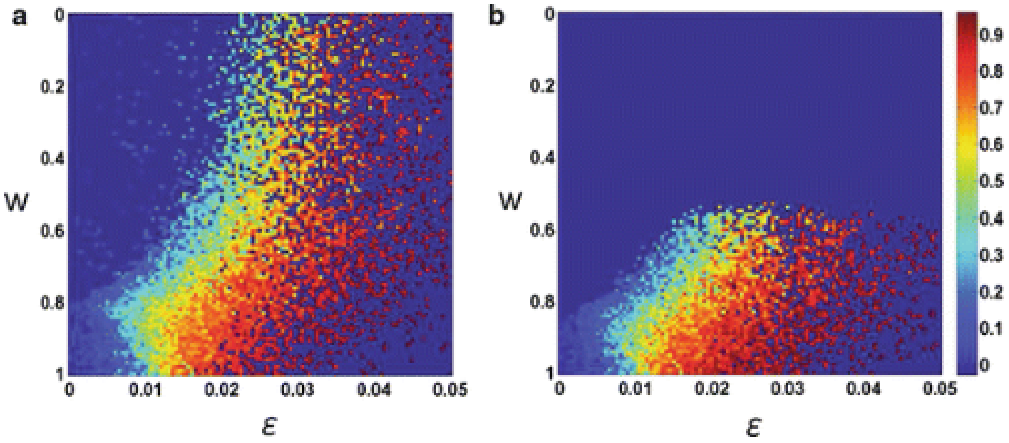

Fig. 4
Correlations within (a) STN and (b) GPe over the parameter space (ε, w) averaged for 50 instances. For given values of parameter pair (ε, w), correlation is calculated only when there is at least 1 oscillating neuron in the case of STN for (a) and GPe for (b)

If the STN-GPe system is to serve as a source of exploration, a high spatiotemporal complexity in STN activity manifested in the form of low pair-wise correlations among neurons is expected. The Figures 3 and 4 also illustrate the ability of ε_s to control correlation within STN and show that ε_s can be used to control the exploration in the BG model.

GPi

GPi combines the GABAergic striatal output via DP with glutamatergic STN output from IP. There is evidence that this combination of DP and IP outflows in GPi is modulated by dopamine projections to GPi. When D1R in GPi, primarily located on the GABAergic striato-pallidal axonal projections, is activated, firing levels of GPi neurons are reduced (Kliem et al. 2007). Since D1Rs are activated at increased dopamine levels, the facilitation of the DP outflow over IP at higher dopamine levels is consistent with the nature of switching facilitated by dopamine in the striatum.

Action Selection in Thalamus

If the primary function of the BG circuit is action selection, where in the circuit is the precise site of such selection? If action salience is computed in the striatum and STN-GPe provides exploration, action selection could be happening downstream in GPi or in the thalamic nuclei receiving afferents from GPi. The competitive dynamics of neurons of thalamic reticular complex makes them ideally suited for implementing action selection (Humphries and Gurney 2002). During binary action selection in the model, the GPi outputs to thalamus converge on two neurons that represent the two action alternatives. These two thalamic neurons integrate the GPi inputs through time: the one that crosses a preset threshold first wins the competition, while the second neuron is reset immediately. Accordingly, if $x_i^{Thal}(t) > x_{th}$ for i (= 1, 2) at time t, then the states of all the other thalamic neurons immediately reset when "i"th action being selected is expressed by $x_j^{Thal}(t) = 0; j \neq i$. If all $x_i^{Thal}(t)$ fail to reach x_{th}, no action is selected.

Simulation Experiments

Binary Action Selection

Salience-based action selection is considered to be one of the primary functions of BG (Redgrave et al. 1999; Gurney et al. 2001). Consider a binary action selection problem. The corticostriatal input, I_i^{ext} with i = 1, 2, represents two possible actions, and the components of I_i^{ext} represent action saliencies. The selected action is denoted by the winning neuron in the thalamus. Due to the complex dynamics of the STN-GPe system, it is not necessary that the winning action is always the one with greater saliency. There could be three possible outcomes:

- "Go" - winning neuron has greater salience.
- "Explore" - winning neuron has lesser salience.
- "NoGo" - no winner and therefore no action selection.

Now consider the effect of δ (dopamine) in determining the type of action selection: When the network depicted in Fig. 2 is simulated with the STN-GPe system exhibiting uncorrelated oscillations, a new Explore regime in addition to the classical Go and NoGo regimes (Fig. 5b) can be observed. This is consistent with the classical picture of selecting the Go regime with high probability for large δ and the NoGo for small δ. In addition, the Explore regime is also selected with the probability maximized for moderate δ (Fig. 5a). This Go/Explore/NoGo profile is called the GEN profile.

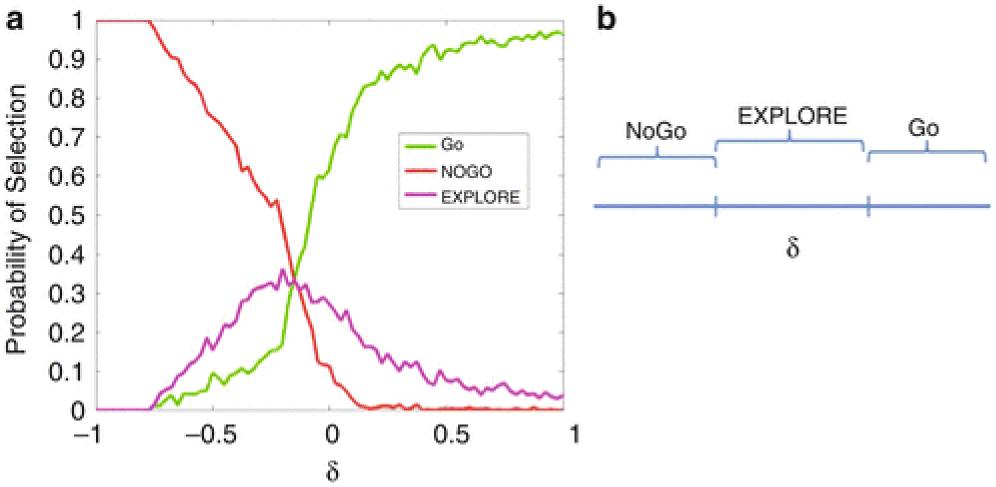

Fig. 5 (a)
Probability of selection of Go/Explore/NoGo; (b) schematic: Go-Explore-NoGo (GEN) regime selection for higher-intermediate-lower values of δ, respectively

The notion that higher correlations (caused by higher ε_s) among STN neurons result in weaker exploration is depicted by the GEN profiles of Fig. 6a, b, c. Note the progressive reduction in the Explore regime with increasing ε_s (Fig. 6).

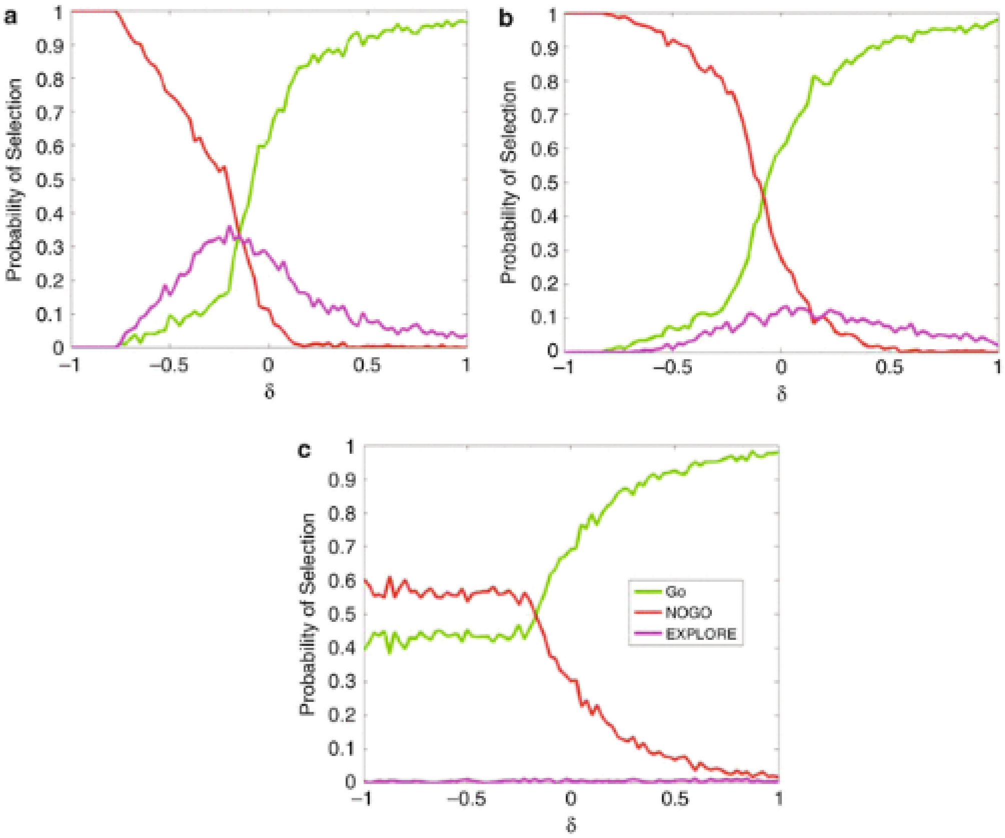

Fig. 6
Probability of selection of Go/Explore/NoGo regimes as a function of δ: (a) for ε $_s$ = 0.001, (b) for ε $_s$ = 0.05, (c) for ε $_s$ = 0.95

Modeling the N-Armed Bandit Problem

In the n-armed bandit problem, a generalization of the binary action selection problem, the setup consists of n slot machines each delivering a fixed reward (deterministically or probabilistically) on selection. The objective is to maximize the total reward received by the agent. Additional features that need to be added to the binary action selection for simulating the n-armed bandit problem are (1) value computation in the striatum, (2) feedback of previous action to the striatum, and (3) resolving the nigrostriatal signal into two dopamine signals - δ_{TD} and δ_V.

If x denotes the action selected (x_i = 1, if the ith arm is selected; x $_j$ = 0 for j ~ = i), the action value is computed as

$$V = \sum_{i=1}^{n} w_i x_i \tag{1}$$

where w $_i$ denote the corticostriatal weights. Instantaneous output error, δ $_{TD}$, is defined as

$$\delta_{TD} = r - V \tag{2}$$

where r is the reward. Note that δ $_{TD}$ in Eq. 2 above is a special case of the more general temporal difference (TD) error (Eq. 3) in RL,

$$\delta_{TD} = r(t) + \gamma V(t+1) - V(t) \tag{3}$$

that arises when γ = 0.

This δ $_{TD}$ is used to update the corticostriatal connections as

$$\Delta w_i = \eta \delta_{TD} x_i \tag{4}$$

In addition to δ $_{TD}$, δ $_{v}$ is introduced as a similar yet novel quantity representing the temporal gradient of value function and is expressed as

$$\delta_V(t) = V(t) - V(t-1) \tag{5}$$

The difference between TD error (Eqs. 2 and 3) and value gradient (Eq. 5) is as follows. While TD error controls learning of corticostriatal weights (Eq. 4), value gradient controls the gain of D1R- and D2R-MSNs. D1R-MSNs are modeled to be activated at higher δ_V and D2R-MSNs at lower levels. Thus, δ_V controls exploration by determining the relative contributions of DP or IP.

A large positive δ $_{v}$ implies a large increase in value, which therefore recommends selection of the same action next time (Go), since the contribution of DP to GPi dominates that of STN. A large negative δ $_{v}$ implies a large reduction in value, suggesting exploration for new actions. Since the IP is selected for strong negative δ $_{v}$, IP contribution dominates that of DP at GPi, thereby suppressing action (NoGo). For small magnitudes of δ $_{v}$, DP is still reduced, and driven by the complex dynamics of STN, a random action is selected next time (Explore).

Figure 7a shows the change in value function with time for the above model applied to a five-armed bandit problem. The rewards are generated by the following distribution: r $_{i}$ = i/n + A*v, where r $_{i}$ is the reward of the ith arm, v is a random variable uniformly distributed over [0,1], and A = 0.3. Note the effect of the strength of STN lateral connection, ε_s, on the growth of value function. For small ε_s (= 0.0145), value increases rapidly but settles at a lower value. For large ε_s (= 0.0475), value rises slowly, but for moderate ε_s (= 0.0305), it rises slower than the previous case but settles at a level higher than two previous cases. In this respect, the effect of ε_s on action selection seems to be analogous to that of the ε parameter in ε-greedy methods used in RL (Fig. 7b). The parameter ε, usually a small positive number, refers to the probability with which a nonoptimal action is selected (Sutton and Barto. 1998).

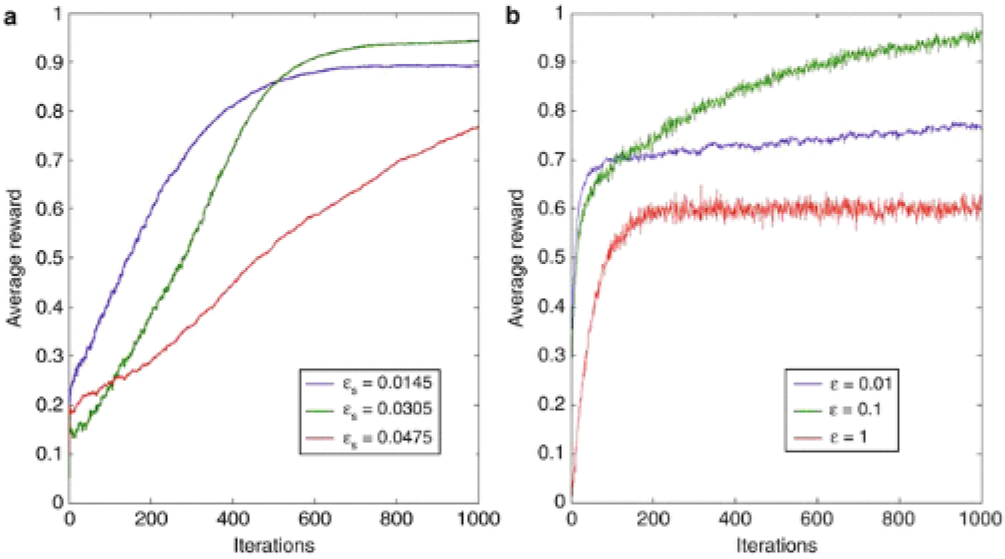

Fig. 7
The mean reward versus iterations obtained with (a) different ε $_{s}$ using the BG network model and (b) different ε values using ε-greedy policy, for a five-armed bandit task averaged for 500 trials

Climbing Value Gradient Using δV

It can be observed from the network model used for the n-armed bandit problem that the value function increases gradually when the action selected is iterated through the loop (thalamo-striatal) in Fig. 2. Thus, the network dynamics implements hill-climbing over the value function, which is a form of stochastic hill-climbing, thanks to the complex dynamics of STN-GPe system. In fact, the aforementioned effect of δ_V on value change (by selecting the previous/random/no action for large positive/moderate/large negative values of δ_V, respectively) is strongly reminiscent of simulated annealing - a form of stochastic optimization (Kirkpatrick, Jr. et al. 1983). It is noted that the BG model exhibits three behaviors depending on dopamine: (1) Go regime ("repeat the previous action") for large positive δ_V, (2) Explore regime ("try random actions") for intermediate values of δ_V, and (3) NoGo regime ("no action") for large negative values of δ_V. These regimes inspire a simple mechanism of hill-climbing in continuous action spaces as follows.

Let V(x) be value function, x be n-dimensional state vector ($x \in S$, where $S \subset R^n$), $\delta_V = V(t) - V(t-1)$, and $\Delta x = x(t) - x(t-1)$. Δx is then updated by the following equations:

$$\begin{aligned}
&\text{if } (\delta_v > D_{hi}) \\
&\quad \Delta x\,(t+1) = \Delta x(t) && -\text{"Go"} && \text{(a)} \\
&\text{else if } (\delta_v > D_{lo} \ \wedge \ \delta_v(t) \le D_{hi}) \\
&\quad \Delta x\,(t+1) = \phi && -\text{"Explore"} && \text{(b)} \\
&\text{else } \ (\delta_v \le D_{lo}) \\
&\quad \Delta x\,(t+1) = 0 && -\text{"NoGo"} && \text{(c)}
\end{aligned} \tag{6}$$

where $D_{hi} > 0$, $D_{lo} < 0$, and Φ is a random vector whose each component, Φ_i, is given by Eq. 7:

$$\phi_i = G(0,1)\exp\left(-{\delta_v}^2/\sigma^2\right) \tag{7}$$

where G(0,1) is a Gaussian random variable with mean 0 and standard deviation 1. The last Eq. 6c seems redundant as there is no state update in that case. It can be altered by substituting 0 in Eq. 6c with -Δx(t). Now in NoGo regime, the state x is updated in an opposite direction compared to the previous update. The Eq. 6a, b, c may be expressed in modified form as follows:

$$\begin{aligned}
&\text{if } (\delta_v > D_{hi}) \\
&\quad \Delta x\,(t+1) = \Delta x(t) && -\text{"Go"} && \text{(a)} \\
&\text{else if } (\delta_v > D_{lo} \ \wedge \ \delta_v(t) \le D_{hi}) \\
&\quad \Delta x\,(t+1) = \phi && -\text{"Explore"} && \text{(b)} \\
&\text{else } \ (\delta_v \le D_{lo}) \\
&\quad \Delta x\,(t+1) = -\Delta x(t) && -\text{"NoGo"} && \text{(c)}
\end{aligned} \tag{8}$$

In Eq. 8 above, the Go, Explore, and NoGo regimes are depicted as discrete, disjoint regimes demarcated by thresholds - D_{hi} and D_{lo}. But the regimes observed in the GEN profile obtained from binary action selection problem are not disjoint; their probability distributions as a function of δ_V overlap (Fig. 5). This inspires a combination of Eq. 8a, b, c into a single update equation where the three regimes smoothly overlap as follows:

$$\begin{aligned}
\Delta x\,(t+1) = {} & \log\text{sig}\left(\kappa_3(\,\delta_V(t) - D_{hi}\,)\right) \ \ \Delta x(t) \\
& + \psi \exp\left(-\delta_V^2(t)/\sigma_E^2\right) \\
& - \log\text{sig}\left(\kappa_4\,(\delta_V(t) - D_{lo})\right) \ \ \Delta x(t)
\end{aligned} \tag{9}$$

where logsig(x) = 1/(1 + exp(−x)). Thus, Eq. 9 describes a map between the current state update (Δx(t)) to the next state update (Δx(t + 1)). Δx(t + 1) is nearly in the same direction as Δx(t) for large positive δ_V and is nearly in the same direction as −Δx(t) for large negative δ_V; Δx(t + 1) is random for the intermediate values of δ_V. Note that the map of Eq. 9 is stochastic due to the middle term on the right-hand side.

Equation 9, known as the GEN rule, represents an abstract, summarized representation of how BG selects actions based on DA signals. The GEN rule or the approach that it is based on has been applied successfully to model a range of BG functions.

Discussion

Application of RL concepts to BG function (Houk et al. 1995; Schultz et al. 1997; Hollerman and Schultz 1998) sought a revision of the Go/NoGo picture as the models may not have given sufficient attention to exploration - the complementary process to exploitation. The Go/NoGo picture of BG can explain how the BG circuit can learn simple binary action selection using RL, but it is inadequate to know how the BG circuit can solve more challenging RL problems in continuous state and action spaces.

The binary Go/NoGo thinking about BG function is supported by a simplistic interpretation of the functional neurochemistry of the two BG pathways (Albin et al. 1989; Contreras-Vidal and Stelmach 1995). But presence of the feedback from STN to GPe allows the possibility of complex dynamics in the STN-GPe loop, thereby introducing an added complication in our functional understanding of BG pathways. The STN-GPe loop has also been dubbed as the "pacemaker" of BG considering its role in generating pathological oscillations associated with Parkinsonian tremor (Hurtado et al. 1999; Terman et al. 2002). This STN-GPe system is an excitatory-inhibitory pair that is capable of exhibiting oscillations and other forms of complex dynamics (Brunel 2000). The fact that neurons in this system exhibit uncorrelated firing patterns in normal conditions, and highly correlated and synchronized firing under dopamine-deficient pathological conditions, seems to offer an important clue to the possible role of this circuit in exploration. From the above studies, it can be concluded that the STN-GPe system is in the best position to serve as an explorer, thereby supplying the missing piece in the RL-machinery of BG. Behavioral strategies of reaching, saccades, spatial navigation, gait, and willed action are shown to be better modeled using the above theory as follows.

A model of reaching movements highlighting the role of BG was described in Magdoom and Subramanian et al. (2011), where a neural network representing motor cortex is trained to drive a two-joint arm to a target. The output of the BG that is predominant in early stages of learning is combined with that of motor cortex whose relative contribution grows with learning. The BG dynamics governed by the earlier described GEN rule discovers desired activations which are used by the motor cortex for learning. When the dopamine signal is clamped to reflect dopamine deficiency in Parkinsonian conditions, the model exhibited Parkinsonian features in reaching like bradykinesia, undershoot, and tremor.

The idea that the DP and IP subserve exploitation and exploration respectively was used in a model of BG on saccade generation (Krishnan et al. 2011) when applied to standard visual search tasks like feature and conjunction search, directional saccades, and sequential saccades. On simulating Parkinsonian conditions by diminished BG output, the model exhibited impaired visual search with longer reaction times - a characteristic symptom in Parkinson's disease (PD) patients.

The GEN approach (Eqs. 6, 8, and 9) was used to model the relative contributions of BG and hippocampus to spatial navigation (Sukumar et al. 2012). The model combines two navigational system: the cue-based system subserved by BG and the place-based system by hippocampus. The two navigational systems are associated with their respective value functions that are combined by the softmax policy (Sutton and Barto. 1998) to select the next move. This model describes the results of an experimental study that investigates competition between cue-based and place-based navigation (Devan and White 1999). Under dopamine-deficient conditions, the model exhibited longer escape latencies similar to PD-model rats (Miyoshi et al. 2002).

The GEN approach to BG was also applied to model impaired gait patterns in PD (Muralidharan et al. 2013). Studies by Cowie and Limousin et al. (2010) investigated gait changes, as PD patients walked through a narrow doorway and observed a strong dip in velocity a short distance from the doorway. In the model of Muralidharan and Balasubramani et al. (2013), the simulated agent passed through the doorway without any significant velocity dip under control conditions and exhibited a significant reduction in velocity close to the doorway under PD conditions.

Clinical literature shows that BG has a role in willed action, and its impairment is seen in conditions of BG lesions or diseases like Parkinson's disease that affect BG (Mink 2003). Recently it was suggested that the BG circuit amplifies will signals, presumably weak, by a stochastic resonance process (Chakravarthy 2013). This study shows that the GEN policy, which is a combination of deterministic hill-climbing process and a stochastic process, may be reinterpreted as a form of stochastic resonance. Applying the model to a simple reaching task, it was shown that the arm reaches the target with probability close to unity at optimal noise levels. The arm dynamics for subthreshold noise is reminiscent of Parkinsonian akinesia, whereas for superthreshold noise the arm shows uncontrolled movements resembling Parkinsonian dyskinesias.

A perspective that the BG circuit is an exploration engine, carved out of the popular RL approach to BG modeling, can thus be substantiated.

References

- Albin RL, Young AB et al (1989) The functional anatomy of basal ganglia disorders. Trends Neurosci 12(10):366-375
- Alexander GE, Crutcher MD (1990) Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends Neurosci 13(7):266-271
- Baunez C, Humby T et al (2001) Effects of STN lesions on simple vs choice reaction time tasks in the rat: preserved motor readiness, but impaired response selection. Eur J Neurosci 13(8):1609-1616
- Bergman H, Wichmann T et al (1994) The primate subthalamic nucleus. II. Neuronal activity in the MPTP model of parkinsonism. J Neurophysiol 72(2):507-520
- Brown P, Oliviero A et al (2001) Dopamine dependency of oscillations between subthalamic nucleus and pallidum in Parkinson's disease. J Neurosci 21(3):1033-1038
- Brunel N (2000) Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. J Comput Neurosci 8(3):183-208
- Chakravarthy VS (2013) Do basal ganglia amplify willed action by stochastic resonance? A model. PLoS One 8(11):e75657
- Chakravarthy VS, Joseph D et al (2010) What do the basal ganglia do? A modeling perspective. Biol Cybern 103(3):237-253
- Contreras-Vidal J, Stelmach GE (1995) Effects of Parkinsonism on motor control. Life Sci 58(3):165-176
- Cowie D, Limousin P et al (2010) Insights into the neural control of locomotion from walking through doorways in Parkinson's disease. Neuropsychologia 48(9):2750-2757
- Daw ND, O'Doherty JP et al (2006) Cortical substrates for exploratory decisions in humans. Nature 441(7095):876-879
- DeLong MR (1990) Primate models of movement disorders of basal ganglia origin. Trends Neurosci 13(7):281-285
- Devan BD, White NM (1999) Parallel information processing in the dorsal striatum: relation to hippocampal function. J Neurosci 19(7):2789-2798
- Doya K (2002) Metalearning and neuromodulation. Neural Netw 15(4-6):495-506
- Gillies A, Willshaw D et al (2002) Functional interactions within the subthalamic nucleus. The basal ganglia VII. Springer, New York. pp 359-368
- Grabli D, McCairn K et al (2004) Behavioural disorders induced by external globus pallidus dysfunction in primates: I. Behavioural study. Brain 127(9):2039-2054
- Gupta A, Balasubramani PP et al (2013) Computational model of precision grip in Parkinson's disease: a utility based approach. Front Comput Neurosci 7:172
- Gurney K, Prescott TJ et al (2001) A computational model of action selection in the basal ganglia. I. A new functional anatomy. Biol Cybern 84(6):401-410
- Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nat Neurosci 1(4):304-309
- Houk JC, Davis JL et al (1995) Models of information processing in the basal ganglia. The MIT press, Cambridge, MA
- Humphries M, Gurney K (2002) The role of intra-thalamic and thalamocortical circuits in action selection. Netw Comput Neural Syst 13(1):131-156
- Humphries MD, Prescott TJ (2010) The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. Prog Neurobiol 90(4):385-417
- Hurtado JM, Gray CM et al (1999) Dynamics of tremor-related oscillations in the human globus pallidus: a single case study. Proc Natl Acad Sci 96(4):1674-1679
- Joel D, Niv Y et al (2002) Actor-critic models of the basal ganglia: new anatomical and computational perspectives. Neural Netw 15(4-6):535-547
- Kalva SK, Rengaswamy M et al (2012) On the neural substrates for exploratory dynamics in basal ganglia: a

model. Neural Netw 32:65-73

- Kirkpatrick S, Gelatt CD Jr et al (1983) Optimization by simulated annealing. Science 220(4598):671-680
- Kliem MA, Maidment NT et al (2007) Activation of nigral and pallidal dopamine D1-like receptors modulates basal ganglia outflow in monkeys. J Neurophysiol 98(3):1489-1500
- Knutson B, Adams CM et al (2001) Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J Neurosci 21(16):RC159
- Kravitz AV, Freeze BS et al (2010) Regulation of Parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. Nature 466(7306):622-626
- Krishnan R, Ratnadurai S et al (2011) Modeling the role of basal ganglia in saccade generation: is the indirect pathway the explorer? Neural Netw 24(8):801-813
- Magdoom KN, Subramanian D et al (2011) Modeling basal ganglia for understanding Parkinsonian reaching movements. Neural Comput 23(2):477-516
- Magill P, Bolam J et al (2001) Dopamine regulates the impact of the cerebral cortex on the subthalamic nucleus-globus pallidus network. Neuroscience 106(2):313-330
- Marsden C (1982) The mysterious motor function of the basal ganglia: the Robert Wartenberg lecture. Neurology 32:514-539
- Mink JW (2003) The basal ganglia and involuntary movements: impaired inhibition of competing motor patterns. Arc Neurol 60(10):1365
- Miyoshi E, Wietzikoski S et al (2002) Impaired learning in a spatial working memory version and in a cued version of the water maze in rats with MPTP-induced mesencephalic dopaminergic lesions. Brain Res Bull 58(1):41-47
- Muralidharan V, Balasubramani PP et al (2013) A computational model of altered gait patterns in parkinson's disease patients negotiating narrow doorways. Front Comput Neurosci 7:190
- O'Doherty J, Dayan P et al (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304(5669):452-454
- Plenz D, Kital ST (1999) A basal ganglia pacemaker formed by the subthalamic nucleus and external globus pallidus. Nature 400(6745):677-682
- Priyadharsini BP, Ravindran B et al (2012) Understanding the role of serotonin in basal ganglia through a unified model. Artificial neural networks and machine learning-ICANN 2012, Springer, Berlin, pp 467-473
- Redgrave P, Prescott TJ et al (1999) The basal ganglia: a vertebrate solution to the selection problem? Neuroscience 89(4):1009-1023
- Ring H, Serra-Mestres J (2002) Neuropsychiatry of the basal ganglia. J Neurol Neurosurg Psychiatry 72(1):12-21
- Rushworth MF, Behrens TE (2008) Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci 11(4):389-397
- Russell V, Allin R et al (1992) Regional distribution of monoamines and dopamine D1-and D2-receptors in the striatum of the rat. Neurochem Res 17(4):387-395
- Schultz W, Dayan P et al (1997) A neural substrate of prediction and reward. Science 275(5306):1593-1599
- Sridharan D, Prashanth PS et al (2006) The role of the basal ganglia in exploration in a neural model based on reinforcement learning. Int J Neural Syst 16(2):111-124
- Stein PS, Grillner S et al (1997) Neurons, networks, and behavior. MIT Press, Cambridge, MA
- Steiner H, Tseng KY (2010) Handbook of basal ganglia structure and function: a decade of progress. Access online via Elsevier. Academic press, San Diego
- Sukumar D, Rengaswamy M et al (2012) Modeling the contributions of basal ganglia and hippocampus to spatial navigation using reinforcement learning. PLoS One 7(10):e47467
- Sutton R, Barto A (1998) Reinforcement learning: an introduction. Adaptive computations and machine learning. MIT Press/Bradford, Cambridge, MA
- Terman D, Rubin J et al (2002) Activity patterns in a model for the subthalamopallidal network of the basal ganglia. J Neurosci 22(7):2963-2976
- Usher M, Cohen JD et al (1999) The role of locus coeruleus in the regulation of cognitive performance. Science 283(5401):549-554
- Willshaw D, Li Z (2002) Subthalamic-pallidal interactions are critical in determining normal and abnormal functioning of the basal ganglia. Proc R Soc Lond Ser B Biol Sci 269(1491):545-551
- Yoshida W, Ishii S (2006) Resolution of uncertainty in prefrontal cortex. Neuron 50(5):781-789

Basal Ganglia System as an Engine for Exploration

Prof. V. Srinivas Chakravarthy — Department of Biotechnology, Indian Institute of Technology, Madras, India

Pragathi Priyadarshini — Department of Biotechnology, Indian Institute of Technology, Madras, India

DOI:	10.1007/SpringerReference_348146
URL:	http://www.springerreference.com/index/chapterdbid/348146
Part of:	Encyclopedia of Computational Neuroscience
Editors:	Prof. Dieter Jaeger and Prof. Ranu Jung
PDF created on:	April, 20, 2014 04:16