

# Chapter 6

## Synchronization and Exploration in Basal Ganglia—A Spiking Network Model



Alekhya Mandali and V. Srinivasa Chakravarthy

**Abstract** Making an optimal decision could be to either ‘Explore’ or ‘exploit’ or ‘not to take any action,’ and basal ganglia (BG) are considered to be a key neural substrate in decision making. In earlier chapters, we had hypothesized earlier that the indirect pathway (IP) of the BG could be the subcortical substrate for exploration. Here, we build a spiking network model to relate exploration to synchrony levels in the BG (which are a neural marker for tremor in Parkinson’s disease). Key BG nuclei such as the subthalamic nucleus (STN), Globus Pallidus externus (GPe), and Globus Pallidus internus (GPi) were modeled as Izhikevich spiking neurons, whereas the striatal output was modeled as Poisson spikes. We have applied reinforcement learning framework with the dopamine signal representing the reward prediction error used for cortico-striatal weight update. We apply the model to two decision-making tasks: a binary action selection task and an n-armed bandit task. The model shows that exploration levels could be controlled by STN’s lateral connection strength which also influenced the synchrony levels in the STN–GPe circuit. An increase in STN’s lateral strength led to a decrease in exploration which can be thought as the possible explanation for reduced exploratory levels in Parkinson’s patients.

### 6.1 Introduction

Chakravarthy, Joseph, and Bapi (2010) suggested that STN–GPe loop, a coupled excitatory–inhibitory network in the IP, might be the substrate for exploration (Chakravarthy et al., 2010). It is well known that coupled excitatory–inhibitory pools of neurons can exhibit rich dynamic behavior like oscillations and chaos (Borisjuk, Borisjuk, Khibnik, & Roose, 1995; Sinha, 1999). This hypothesis has inspired models simulating various BG functions ranging from action selection in continuous spaces (Krishnan, Ratnadurai, Subramanian, Chakravarthy, & Rengaswamy, 2011), reaching movements (Magdoom et al., 2011), spatial navigation (Sukumar, Rengaswamy, & Chakravarthy, 2012), precision grip (Gupta, Balasubramani, & Chakravarthy, 2013), and gait (Muralidharan, Balasubramani, Chakravarthy, Lewis, & Moustafa, 2013) in normal and Parkinsonian conditions.

Using a network of rate-coding neurons, Kalva, Rengaswamy, Chakravarthy, and Gupte (2012) showed that exploration emerges out of the chaotic dynamics of the STN–GPe system (Kalva et al., 2012). Most rate-coded models, by design, fail to capture dynamic phenomena like synchronization found in more realistic spiking neuron models (Bevan, Magill, Terman, Bolam, & Wilson, 2002; Park, Worth, & Rubchinsky, 2010; Park, Worth, & Rubchinsky, 2011). Synchronization within BG nuclei had gained attention since the discovery that STN, GPe, and GPi neurons show high levels of synchrony in Parkinsonian conditions (Bergman, Wichmann, Karmon, & DeLong, 1994; Bevan et al., 2002; Hammond, Bergman, & Brown, 2007; Tachibana, Iwamuro, Kita, Takada, & Nambu, 2011; Weinberger & Dostrovsky, 2011). This oscillatory activity was found to be present in two frequency bands, one around the tremor frequency [2–4 Hz] and another in [10–30 Hz] frequency (Weinberger & Dostrovsky, 2011). Park et al. (2011) report the presence of intermittent synchrony between STN neurons and its local field potentials (LFP), recorded using multiunit activity electrodes from PD patients undergoing deep brain stimulation (DBS) surgery (Park et al., 2011) which is absent in healthy controls.

One of the key objectives of the current study is to use a 2D spiking neuron model to understand and correlate STN–GPe’s synchrony levels to exploration. As the second objective, we apply the above-mentioned model to the n-armed bandit problem of Daw, O’Doherty, Dayan, Seymour, and Dolan (2006) and Bourdaud, Chavarriaga, Galán, and del R Millan (2008) (Bourdaud et al., 2008; Daw et al., 2006) with the specific aim of studying the contributions of STN–GPe dynamics to exploration. The proposed model shares some aspects of classical RL-based approach to BG modeling. For example, dopamine signal is compared to reward prediction error (Schultz, 1998). Furthermore, DA is allowed to control cortico-striatal plasticity (Reynolds and Wickens 2002), modulate the gains of striatal neurons (Hadipour-Niktarash, Rommelfanger, Masilamoni, Smith, & Wichmann, 2012; Kliem, Maidment, Ackerson, Chen, Smith, & Wichmann, 2007), and influence the dynamics of STN–GPe by modulating the connections (Fan, Baufreton, Surmeier, Chan, & Bevan, 2012; Kreiss, Mastropietro, Rawji, & Walters, 1997).

## 6.2 Methods

### 6.2.1 Spiking Neuron Model of the Basal Ganglia

The network model of BG (Mandali, Rengaswamy, Chakravarthy, & Moustafa, 2015) described earlier was used to simulate the binary action selection and n-arm bandit task. For details of the model and its related equations, refer to earlier sections. The details of the tasks and the related measures are explained below.

### 6.2.2 Binary Action Selection Task

The first task we simulated was the simple binary action selection similar to Humphries, Stewart, and Gurney (2006), where two competing stimuli were presented to the model (Humphries et al., 2006). The input firing frequency is thought to represent ‘saliency,’ with higher frequencies representing higher salience (Humphries et al., 2006). The response of striatal output to cortical input falls in the range of a few tens of Hz (Sharott, Doig, Mallet, & Magill, 2012). Therefore, the frequencies that represent the 2 actions were assumed to be around 4 Hz (stimulus #1) and 8 Hz (stimulus #2). Spontaneous output firing rate of the striatal neurons (without input) is assumed to be around 1 Hz (Plenz & Kitai, 1998; Sharott et al., 2012). Selection of higher salient stimulus among the available choices could be considered as ‘exploitation’ while selecting the less salient one as ‘exploration’ (Sutton & Barto, 1998). So, the action selected is defined as ‘Go’ if stimulus #2 (more salient) is selected, ‘Explore’ if stimulus #1 (less salient) is selected, and ‘NoGo’ if none of them is selected.

The inputs were given spatially such that the neurons in the upper half of the lattice receive stimulus #1 and lower half the other (Fig. 6.1). The striatal outputs from D1 and D2 neurons of the striatum are given as input to GPi and GPe modules, respectively, with the projection pattern as shown in Fig. 6.1. Poisson spike trains corresponding to stimulus #1 were presented as input to neurons (1–1250) and were fully correlated among themselves. Similarly, Poisson spike trains corresponding to stimulus #2 were presented as input to neurons (1251–2500) and were fully correlated among themselves. Stimulus #1 and #2 are presented for an interval of 100 ms between 100 and 200 ms; at other times, uncorrelated spike trains at 1 Hz are presented to all the striatal neurons.

### 6.2.3 The $N$ -Armed Bandit Task

We now describe the four-armed bandit task (Bourdaud et al., 2008; Daw et al., 2006) used to study exploratory and exploitative behavior. In this experimental task, subjects were presented with four arms where one among them is to be selected in every trial for a total of 300 trials. The reward/payoff for each of these slots was obtained from a Gaussian distribution whose mean changes from trial to trial with payoff ranging from 0 to 100. The payoff,  $r_{i,k}$  associated with the  $i$ th machine at the  $k$ th trial, was drawn from a Gaussian distribution of mean  $\mu_{i,k}$  and standard deviation (SD)  $\sigma_0$ . The payoff was rounded to the nearest integer, in the range  $[0, 100]$ . At each trial, the mean is diffused according to a decaying Gaussian random walk. The trial was defined as an ‘exploitative’ trial if highest reward giving arm was selected else defined as an ‘exploratory’ trial.

The payoffs generated by the slot machines are computed as follows,

$$\mu_{i,k+1} = \lambda_m \mu_{i,k} + (1 - \lambda_m) \theta_m + e \quad (6.1)$$

$$r'_{i,k} \approx N(\mu_{i,k}, \sigma_0^2) \quad (6.2)$$

$$r_{i,k} = \text{round}(r'_{i,k}) \quad (6.3)$$

where

$\mu_{i,k}$  is the mean of the Gaussian distribution with standard deviation ( $\sigma_0$ ) for  $i$ th machine during  $k^{\text{th}}$  trial.  $\lambda_m$  and  $\theta_m$  control the random walk of mean ( $\mu_{i,k}$ ), and  $e \sim N(0, \sigma_d^2)$  is obtained from Gaussian distribution of mean 0 and standard deviation  $\sigma_d$ .  $r_{i,k}$  and  $r'_{i,k}$  are the payoffs before and after rounding to nearest integer, respectively. The initial value of mean payoff,  $\mu_{i,0}$ , is set to a value of 50. All the values for the parameters  $\lambda_m$ ,  $\theta_m$ ,  $\sigma_d$ ,  $\sigma_0$  were adapted from (Bourdaud et al., 2008).

To make an optimal decision, the subjects need to keep track of rewards associated with each of the four arms. The subject's decision to either Explore or exploit would depend on this internal representation which would closely resemble the actual payoff that is being obtained. It is quite difficult to identify whether the subject made an exploratory decision or an exploitative one just by observing the EEG and selected slot data. A subject-specific model is required to classify their decisions and identify the strategy (Bourdaud et al., 2008; Daw et al., 2006). Keeping this in mind, Bourdaud et al. (2008) used a 'behavioral model' that uses the softmax principle of RL to fit the selection pattern of human subjects. The parameter ' $\beta$ ' of the behavioral model was adjusted such that the final selection pattern matches that of individual subjects in the experiment (given below). The parameter ' $\beta$ ' which controls the exploration level in the behavioral model is tuned to match % exploitation obtained for each of the eight subjects (one subject's data were discarded because of artifacts); two out of the eight subjects had similar exploration levels. Hence, a total of six subjects' data are taken into account to check the performance of the proposed spiking BG model.

### 6.2.3.1 Behavioral Model (Adapted from Bourdaud et al. (2008))

The behavioral model labels each trial as corresponding to either an exploratory or exploitative decision. The model assumes that the user estimates the mean payoff of each machine using a Bayesian linear Gaussian rule (i.e., a Kalman filter). Using these estimations, he/she selects a machine according to a softmax rule. All the subjects are assumed to share the same model for tracking the payoff means, and thus, parameters are computed using the entire available data. The parameters of the model (for both mean tracking and machine selection) are estimated by maximizing the model likelihood with respect to the subject's choices.

At any given trial, the behavioral model provides the mean payoff for all machines considering previous observations (i.e., the payoff obtained at previous trials). Comparison between the model's estimated payoffs for all machines is used to label that trial as either exploration or exploitation. Those trials in which the user selects the machine with the highest estimated mean are labeled as corresponding to exploitative decisions.

The subject strategy for tracking the payoff of each machine is modeled by a Kalman filter, whose parameters are assumed to remain constant over trials. Once the  $j$ th machine is selected, at the  $k$ th trial, the estimated payoff distribution is updated from its preselection values  $\left(\hat{\mu}_{j,k}^{\text{pre}}, \left(\hat{\sigma}_{j,k}^{\text{pre}}\right)^2\right)$  to its post-selection values  $\left(\hat{\mu}_{j,k}^{\text{post}}, \left(\hat{\sigma}_{j,k}^{\text{post}}\right)^2\right)$  as follows

$$\hat{\mu}_{j,k}^{\text{post}} = \hat{\mu}_{j,k}^{\text{pre}} + K_k \left( r_k - \hat{\mu}_{j,k}^{\text{pre}} \right) \quad (6.4)$$

$$\left(\hat{\sigma}_{j,k}^{\text{post}}\right)^2 = (1 - K_k) \left(\hat{\sigma}_{j,k}^{\text{pre}}\right)^2 \quad (6.5)$$

where

$$(K_k) = \frac{\left(\hat{\sigma}_{j,k}^{\text{pre}}\right)^2}{\left(\hat{\sigma}_{j,k}^{\text{pre}}\right)^2 + \left(\hat{\sigma}_0\right)^2} \quad (6.6)$$

The mean estimation for the remaining machines does not change as result of the choice since the user cannot observe the payoff of these machines. That is,

$$\forall i \neq j$$

$$\hat{\mu}_{j,k}^{\text{post}} = \hat{\mu}_{j,k}^{\text{pre}} \quad (6.7)$$

$$\hat{\sigma}_{j,k}^{\text{post}} = \hat{\sigma}_{j,k}^{\text{pre}} \quad (6.8)$$

Then, the estimations are also evolved according to the diffusion rule:

$$\hat{\mu}_{j,k+1}^{\text{pre}} = \hat{\lambda} \hat{\mu}_{j,k}^{\text{post}} + (1 - \hat{\lambda}) \hat{\theta} \quad (6.9)$$

$$\left(\hat{\mu}_{j,k+1}^{\text{pre}}\right)^2 = \hat{\lambda}^2 \left(\hat{\sigma}_{j,k}^{\text{post}}\right)^2 + \sigma_d^2 \quad (6.10)$$

The choice of subjects is modeled by a softmax rule; i.e., at each trial  $k$ , the probability of choosing the machine is

**Table 6.1** Estimation of parameters of the behavioral model (Bourdaud et al., 2008)

	$\lambda$	$\theta$	$\sigma_d$	$\sigma_o$
Real values	0.9836	50	2.8	4
Estimated values	0.92	51.37	8.12	N/A
Subject	1 2	3 4	5 6	7 8
$\beta$	0.37	0.19	0.19	0.29
	0.28	0.21	0.29	0.23

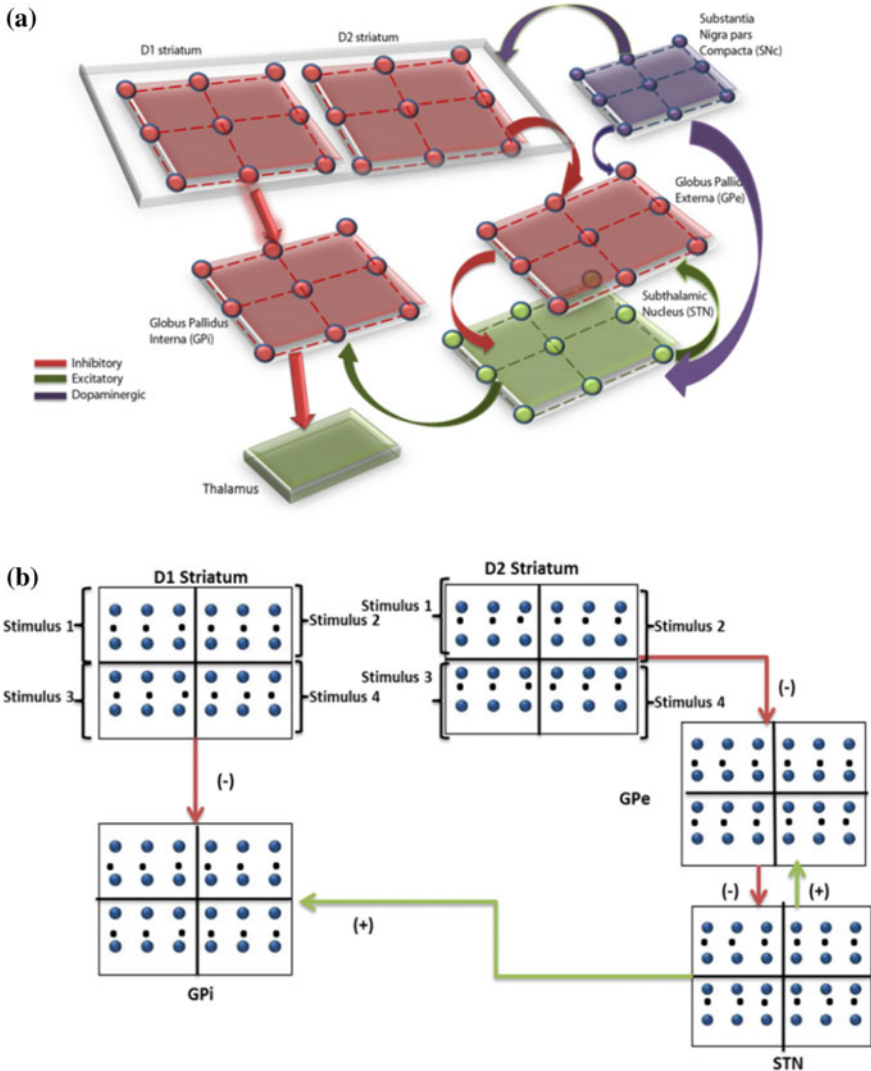
$$P_{i,k} = \frac{\exp(\beta \hat{\mu}_{i,k}^{\text{pre}})}{\sum_j \exp(\beta \hat{\mu}_{j,k}^{\text{pre}})} \quad (6.11)$$

where ‘ $\beta$ ’ is a scaling parameter. Higher values of  $\beta$  drive the system to exploitative behavior and vice versa. The parameters of the behavioral model ( $\sigma_o, \hat{\theta}, \hat{\lambda}, \hat{\sigma}_d$ ) are estimated by maximizing the log likelihood under the following constraints. To speed up convergence, estimated parameters ( $\sigma, \hat{\mu}_{j,0}^{\text{pre}} \& \hat{\sigma}_{j,0}^{\text{pre}}$ ) are initialized to the parameters of the original model ( $\sigma_o, \mu_{j,0} \& \sigma_{j,0}$ ), respectively. Fixing the last two parameters does not significantly affect the estimation of the others, because their influence vanishes quickly within a few trials. Table 6.1 shows the estimated values of the model, which are consistent with the real values of the machines.

### 6.2.3.2 Strategy for Slot Machine Selection

To simulate the experiment, we utilized the concepts of RL and combined the dynamics of BG model to select an optimally rewarding slot in each trial. Experimental data show that BG receives reward-related information in the form of dopaminergic input to striatum (Chakravarthy et al., 2010; Niv, 2009). Cortico-striatal plasticity changes due to dopamine (Reynolds & Wickens, 2002) were incorporated in the model by allowing DA signals to modulate the Hebb-like plasticity of cortico-striatal synapses (Surmeier, Ding, Day, Wang, & Shen, 2007).

The architecture of the proposed network model is depicted in Fig. 6.1. The output of striatum (both D1 and D2 parts) was divided equally into four quadrants which receive input from corresponding stimulus. The stimuli are associated with 2 weights ( $w_{i,0}^{\text{D1}}, w_{i,0}^{\text{D2}}$ ) initialized with equal value of 50 which represent the cortico-striatal weights of D1 and D2 MSNs in the striatum. Each of the cortico-striatal weights represents the saliency (in terms of striatal spike rate) for that corresponding arm. These output spikes generated from each of the D1 and D2 striatum project to GPi and GPe, respectively. The final selection of an arm is made as in Sect. 6.2.4. The reward  $r_{i,k}$  received for the selected slot was sampled from Gaussian distribution with mean  $\mu_{i,k}$  and SD ( $\sigma_o$ ) (Eq. 6.3).



**Fig. 6.1** **a** Computational spiking basal ganglia model with key nuclei such as striatum (D1, D2), STN, GPe, GPi, and thalamus. Excitatory/inhibitory/modulatory glutamatergic/GABAergic/dopaminergic projections are shown by green/red/violet arrows. **b** The BG model and the regions within each nuclei corresponding to the four decks are indicated

Utilizing the reward obtained for the input ‘i’ and trial ‘k’, the expected value of the slots, inputs to D1 and D2 striatum are updated using the following equations,

$$\Delta w_{i,k+1}^{D1} = \eta \delta_k x_{i,k}^{\text{inp}} \quad (6.12)$$

$$\Delta w_{i,k+1}^{D2} = -\eta \delta_k x_{i,k}^{\text{inp}} \quad (6.13)$$

The expected value ( $V_k$ ) for  $k$ th trial is calculated as

$$V_k = \sum_{i=1}^4 w_{i,k}^{D1} * x_{i,k}^{\text{inp}} \quad (6.14)$$

The received payoff ( $\text{Re}_k$ ) for  $k$ th trial is calculated as

$$\text{Re}_k = \sum_{i=1}^4 r_{i,k} * x_{i,k}^{\text{inp}} \quad (6.15)$$

The error ( $\delta$ ) for  $k$ th trial is defined as

$$\delta_k = \text{Re}_k - V_k \quad (6.16)$$

where  $w_{i,k}^{D1}$  are the cortico-striatal weights of D1 striatum for  $i$ th machine in  $k$ th trial,  $w_{i,k}^{D2}$  are the cortico-striatal weights of D2 striatum for  $i$ th machine for  $k$ th trial,  $r_{i,k}$  is the reward obtained for the selected  $i$ th machine for  $k$ th trial,  $x_{i,k}^{\text{inp}}$  is the binary input vector representing the four slot machines, e.g., if the first slot machine is selected  $x_{i,k}^{\text{inp}} = [1 \ 0 \ 0 \ 0]$ ,  $\eta$  ( $=0.3$ ) is the learning rate of D1 and D2 striatal MSNs,  $\text{Re}_k$  is the received payoff for selected slot for  $k$ th trial, and  $V_k$  is the expected value for selected slot for  $k$ th trial.

The cortico-striatal weights are updated (Eqs. 6.12 and 6.13) using the error term ‘ $\delta$ ’ (Eq. 6.16). The reward-related information in the form of dopaminergic input to striatum has been correlated to the error ( $\delta$ ) (Chakravarthy et al., 2010; Niv, 2009). The  $\delta$  calculated from Eq. (6.16) has both positive and negative values with no upper and lower boundaries but the working DA range in the model was limited to small positive values (0.1–0.9). Hence, a mapping from  $\delta$  to DA is defined as follows:

$$\text{DA} = \text{sig}(\lambda * \delta_k) \quad (6.17)$$

where

DA is the dopamine signal within range of 0.1–0.9,  $\lambda$  is the slope of sigmoid ( $=0.2$ ),  $\delta_k$  is the error obtained for  $k$ th trial (Eq. 6.16), and  $\text{sig}()$  is the sigmoid function.



## 6.2.4 Measures

### 6.2.4.1 Synchronization

The phenomenon of neural synchrony has attracted the attention of many computational and experimental neuroscientists in the recent decades (Hauptmann & Tass, 2007; Kumar, Cardanobile, Rotter, & Aertsen, 2011; Park et al., 2011; Pinsky & Rinzel, 1995; Plenz & Kital, 1999). It is believed that partial synchrony helps in the generation of various EEG rhythms such as alpha and beta (Izhikevich, 2007). Studying synchrony in neural networks has been gaining importance due to its presence in normal functioning (coordinated movement of the limbs) and in pathological states (e.g., synchronized activity of CA3 neurons in the hippocampus during an epileptic seizure) (Pinsky & Rinzel, 1995). Plenz and Kital (1998) proposed that STN–GPe might act as a pacemaker (Plenz & Kital, 1999), a source for generating oscillations in pathological conditions such as Parkinson’s disease. Park et al. (2011) report the presence of intermittent synchrony between STN neurons and its local field potentials (LFP), recorded using multiunit activity electrodes from PD patients undergoing DBS surgery (Park et al., 2011). They also calculated the duration of synchronized and desynchronized events in neuronal activity by estimating transition rates, which were obtained with the help of first return maps plotted using phase of neurons (Park et al., 2010, 2011). To observe how dopamine changes synchrony in STN–GPe, we calculated the phases of individual neurons as defined in (Pinsky & Rinzel, 1995).

The phase of  $j$ th neuron was calculated as follows:

$$\emptyset_j(t) = 2 * \pi * \frac{(T_{j,k} - t_{j,k})}{(t_{j,k+1} - t_{j,k})} \quad (6.18)$$

$$R^{\text{sync}}(t) * e^{i\theta(t)} = \frac{1}{N} \sum_{j=1}^N e^{i\emptyset_j(t)} \quad (6.19)$$

where

$t_{j,k}$  and  $t_{j,k+1}$  are the onset times of  $k$ th and  $k + 1$ th spike of the  $j$ th neuron  $T_{j,k} \in [t_{j,k}, t_{j,k+1}]$ ,  $\emptyset_j(t)$  = phase of  $j$ th neuron at time ‘ $t$ ’,  $R^{\text{sync}}$  is the synchronization measure  $0 \leq R^{\text{sync}} \leq 1$ ,  $\theta$  = average phase of neurons,  $N$  = total number of neurons in the network.

## 6.2.5 Action Selection Using the Race Model

Action selection is modulated by BG output nucleus GPi which projects back to the cortex via the thalamus. We have used the race model (Vickers, 1970) for the final

action selection where an action is selected when temporally integrated neuronal activity of the output neurons crosses a threshold (Frank, 2006; Frank, Samanta, Moustafa, & Sherman, 2007; Humphries, Khamassi, & Gurney, 2012).

The dynamics of the thalamic neurons is as follows:

$$\frac{dz_k(t)}{dt} = -z_k(t) + f_{\text{GPik}}(t) \quad (6.20)$$

$$f'_{\text{GPik}} = \frac{1}{(N * N)/k} \sum_{t=1}^T \left( \sum_{i=1}^N \sum_{j=1}^{N/k} S_{ij}^{\text{GPik}}(t) \right) \quad (6.21)$$

$$f_{\text{GPik}} = \frac{f_{\text{GPi}}^{\max} - f'_{\text{GPik}}}{f_{\text{GPi}}^{\max}}$$

where

$z_k(t)$  = integrating variable for  $k$ th stimulus,  $f_{\text{GPik}}(t)$  = normalized and reversed average firing frequency of GPi neurons receiving  $k$ th stimulus from striatum,  $f_{\text{GPi}}^{\max}$  = highest firing rate among the GPi neurons,  $S_{ij}^{\text{GPik}}$  = neuronal spikes of GPi neurons receiving  $k$ th stimulus,  $N$  = number of neurons in a single row/column of GPi array (=50), and  $T$  = duration of simulation.

The first neuron ( $z_k$ ) among  $k$  stimuli to cross the threshold (=0.15) represents the action selected. All the variables representing neuron activity are reset immediately after each action selection.

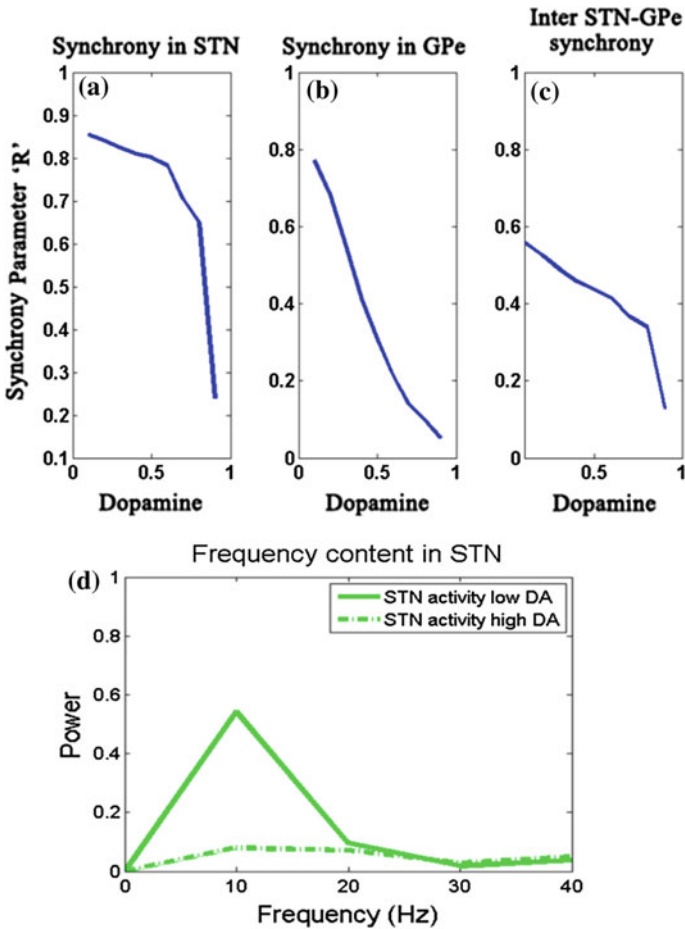
## 6.3 Results

We start with results of neural dynamics (STN–GPe) as a function of DA and then present with decision-making results.

### 6.3.1 Neural Dynamics

Pathological oscillations of STN and GP have been associated with various PD symptoms (Brown, 2003; Plenz & Kital, 1999). Correlated neural firing patterns in STN and GPi can be seen in both experimental conditions of dopamine depletion and in Parkinsonian conditions. In the present model, we show increased synchronized behavior under conditions of reduced dopamine, resembling the situation in dopamine-deficient conditions of Parkinson's disease. The effect of DA on the synchronization of STN and GPe neurons was studied by estimating the values of  $R_{\text{STN}}^{\text{sync}}$ ,  $R_{\text{GPe}}^{\text{sync}}$ ,  $R_{\text{STNGPe}}^{\text{sync}}$  for increasing values of DA (0.1–0.9).

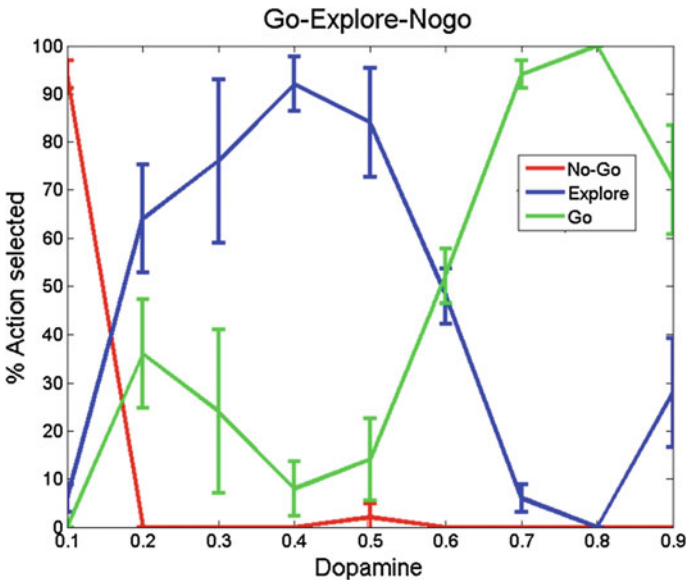
The three ' $R^{\text{sync}}$ ' (Eq. 6.19) values showed a decrease in amplitude with an increase in DA level (Fig. 6.2a–c). Under low DA conditions, GPe activity follows STN activity (Plenz & Kital, 1999), thus forming a pacemaker kind of circuit, which could be the source of STN–GPe oscillations Fig. 6.2d. One of the suspected reasons of bursting activity in STN is the decreased inhibition from GPe neurons (Plenz & Kital, 1999) at low DA levels. This feature is captured by the model since GPe firing rates are smaller for lower DA levels. The STN neurons showed oscillations around the frequency of 10 Hz at low DA but were absent at high DA level (Kang & Lowery, 2013).



**Fig. 6.2** Change in the three synchronization values  $R_{STN}^{\text{sync}}$  (a),  $R_{GPe}^{\text{sync}}$  (b) and  $R_{STNGPe}^{\text{sync}}$  (c) oscillatory activity in STN neurons (d) frequency content with the value of DA (0.1–0.9). Simulations show reduced synchronization within STN and GPe networks, and also between STN and GPe networks, as DA is increased

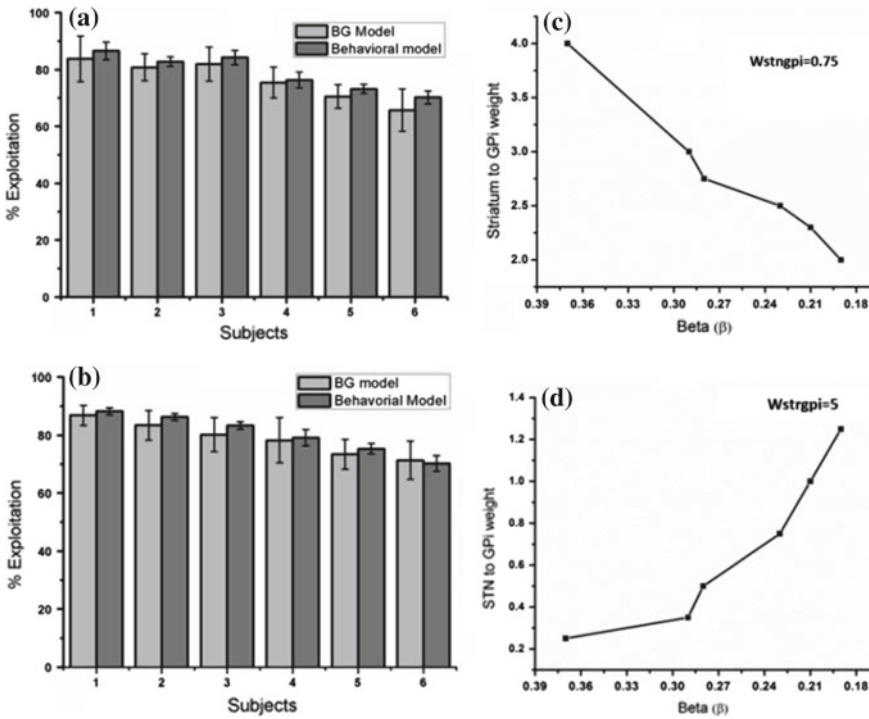
### 6.3.2 Decision Making

After the model's performance was quantified at neural level, we studied the role of BG in decision making using two tasks especially in explorative and exploitative dynamics. This work is in continuation to our earlier hypothesis that the source for exploration comes from STN–GPe dynamics (Kalva et al., 2012). The first task was a simple binary action selection similar to Humphries et al., (2006), where two competing stimuli were presented to the model. The input firing frequency is thought to represent ‘saliency,’ with higher frequencies representing higher saliency. Selection of stimulus with the higher saliency between the two available choices could be considered as ‘exploitation’ while selecting the less salient one as ‘exploration’ (Sutton & Barto, 1998). So the action selected is defined as ‘Go’ if stimulus #2 (more salient) is selected, ‘Explore’ if stimulus #1 (less salient) is selected, and ‘NoGo’ if none of them is selected. Simulations were run for 100 trials, and the percentage of actions selected under each regime (Go, Explore, and NoGo) was calculated for dopamine levels ranging from low (0.1) to high (0.9) (Fig. 6.3). We may note that the probability of NoGo, where no action is selected, decreases with increase in dopamine; probability of Go increases with dopamine; the peak of exploration is found at intermediate levels of dopamine (Fig. 6.3). The range of DA where a peak in exploration was observed is the same where STN and GPe network showed chaotic activity.



**Fig. 6.3** Percentage of action selection observed in the Go, NoGo, and Explore regimes averaged over 200 trials with DP and IP weight values at  $w_{\text{STN-GPe}} = 1.15$  &  $w_{\text{STN-GPi}} = 0.8$ . We ran the simulation for 100 trials and segmented into 4 bins (25 trials each). We then calculated the variance of each regime across all DA levels

The second task was a four-armed bandit task (Bourdaud et al., 2008; Daw et al., 2006) which is similar to a real-world decision-making scenario. In this task, the subjects are presented with four arms where one among them is to be selected in every trial for a total of 300 trials. The reward/payoff for each of these slots was obtained from a Gaussian distribution whose mean changes from trial to trial with payoff ranging from 0 to 100. The model’s performance (% exploitation) was compared with behavioral model, which represents the experimental data in the n-armed bandit task (Fig. 6.4). The parameter ‘ $\beta$ ’ of the behavioral model which controls the Exploit–Explore balance was adjusted to match the performance of individual subjects in the experiment. Exploration in the model can be obtained by either increasing the IP weight (influence from STN) or decreasing DP weight (influence from striatum).



**Fig. 6.4** Compares the performance of BG model with the behavioral model. **a** The percentage exploitation obtained for each of the six subjects from BG and behavioral model. The relationship between betas ( $\beta$ ) of the behavioral model and DP weights ( $w_{STN \rightarrow GPI}$ ) with a constant  $w_{STN \rightarrow GPI}$  value ( $=0.75$ ) used to attain (a) are shown in (c). **b** The relationship between betas ( $\beta$ ) of the behavioral model and IP weights ( $w_{STN \rightarrow GPI}$ ) of BG model with a constant  $w_{STN \rightarrow GPI}$  value of ( $=5$ ) used to attain (b) are shown in (d). Y-axis represents percentage exploitation, and X-axis represents a subject which is a specific beta value ( $\beta$ ) in behavioral model and the IP or DP weight in the BG model

## 6.4 Discussion

The synchrony results tally with the general observation from electrophysiology that at higher levels of dopamine, the STN–GPe system shows desynchronized activity and under dopamine-deficient conditions of PD exhibits synchronized bursts (Bergman et al., 1994; Gillies, Willshaw, Gillies, & Willshaw, 1998; Park et al., 2011). We observed that STN activity showed oscillatory activity with a frequency ( $\approx 10$  Hz) which falls under the beta frequency range observed in experimental PD study (Weinberger & Dostrovsky, 2011). One of the aims of the present work is also to show that the complex dynamics of STN–GPe system contributes to exploration. To this end, we first simulated the binary action selection task [similar to Humphries et al., (2006)] where saliency was coded in the firing rate. The selection of higher one was defined as ‘exploitation/Go’ and lesser one as ‘exploration/Explore’ and not selecting any of the inputs as ‘NoGo’. The model showed NoGo at low DA levels (0.1–0.3) and Go at high DA levels (0.7–0.9) consistent with the classical picture of BG function. Along with this, a peak in ‘Explore’ at intermediate levels of DA (0.4–0.6) was also observed (Fig. 6.3). To check whether any other module in the network is influencing exploration in the system, we removed the STN to GPi connection (which effectively eliminated the IP). This omission rendered the system to display only Go and NoGo regimes (no exploration, results not included). We then moved to simulating the n-armed bandit task, where the performance of model was compared with experimental result. The results obtained from BG model closely match with the behavioral model (Fig. 6.4) reinforcing the idea that STN–GPe could be a source for exploration at subcortical level.

## References

- Bergman, H., Wichmann, T., Karmon, B., & DeLong, M. (1994). The primate subthalamic nucleus. II. Neuronal activity in the MPTP model of parkinsonism. *Journal of Neurophysiology*, 72(2), 507–520.
- Bevan, M. D., Magill, P. J., Terman, D., Bolam, J. P., & Wilson, C. J. (2002). Move to the rhythm: Oscillations in the subthalamic nucleus–external globus pallidus network. *Trends in Neurosciences*, 25(10), 525–531.
- Borisyuk, G. N., Borisyuk, R. M., Khibnik, A. I., & Roose, D. (1995). Dynamics and bifurcations of two coupled neural oscillators with different connection types. *Bulletin of Mathematical Biology*, 57(6), 809–840.
- Bourdaud, N., Chavarriaga, R., Galán, F., & del R Millan, J. (2008). Characterizing the EEG correlates of exploratory behavior. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 6(6), 549–556.
- Brown, P. (2003). Oscillatory nature of human basal ganglia activity: Relationship to the pathophysiology of Parkinson’s disease. *Movement Disorders*, 18(4), 357–363.
- Chakravarthy, V., Joseph, D., & Bapi, R. S. (2010). What do the basal ganglia do? A modeling perspective. *Biological Cybernetics*, 103(3), 237–253.

- Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879.
- Fan, K. Y., Baufreton, J., Surmeier, D. J., Chan, C. S., & Bevan, M. D. (2012). Proliferation of external globus pallidus-subthalamic nucleus synapses following degeneration of midbrain dopamine neurons. *The Journal of Neuroscience*, *32*(40), 13718–13728.
- Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks*, *19*(8), 1120–1136.
- Frank, M. J., Samanta, J., Moustafa, A. A., & Sherman, S. J. (2007). Hold your horses: Impulsivity, deep brain stimulation, and medication in parkinsonism. *Science*, *318*(5854), 1309–1312.
- Gillies, A., Willshaw, D., Gillies, A., & Willshaw, D. (1998). A massively connected subthalamic nucleus leads to the generation of widespread pulses. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *265*(1410), 2101–2109.
- Gupta, A., Balasubramani, P. P., & Chakravarthy, V. S. (2013). Computational model of precision grip in Parkinson's disease: A utility based approach. *Frontiers in Computational Neuroscience*, *7*.
- Hadipour-Niktarash, A., Rommelfanger, K. S., Masilamoni, G. J., Smith, Y., & Wichmann, T. (2012). Extrastriatal D2-like receptors modulate basal ganglia pathways in normal and parkinsonian monkeys. *Journal of Neurophysiology*, *107*(5), 1500–1512.
- Hammond, C., Bergman, H., & Brown, P. (2007). Pathological synchronization in Parkinson's disease: Networks, models and treatments. *Trends in Neurosciences*, *30*(7), 357–364.
- Hauptmann, C., & Tass, P. A. (2007). Therapeutic rewiring by means of desynchronizing brain stimulation. *Biosystems*, *89*(1), 173–181.
- Humphries, M. D., Khamassi, M., & Gurney, K. (2012). Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. *Frontiers in Neuroscience*, *6*.
- Humphries, M. D., Stewart, R. D., & Gurney, K. N. (2006). A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. *The Journal of Neuroscience*, *26*(50), 12921–12942.
- Izhikevich, E. M. (2007). *Dynamical systems in neuroscience*. Cambridge: The MIT press.
- Kalva, S. K., Rengaswamy, M., Chakravarthy, V. S., & Gupte, N. (2012). On the neural substrates for exploratory dynamics in basal ganglia: A model. *Neural Networks*, *32*, 65–73. <https://doi.org/10.1016/j.neunet.2012.02.031>.
- Kang, G., & Lowery, M. M. (2013). Interaction of oscillations, and their suppression via deep brain stimulation, in a model of the cortico-basal ganglia network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), 244–253.
- Kliem, M. A., Maidment, N. T., Ackerson, L. C., Chen, S., Smith, Y., & Wichmann, T. (2007). Activation of nigral and pallidal dopamine D1-like receptors modulates basal ganglia outflow in monkeys. *Journal of Neurophysiology*, *98*(3), 1489–1500.
- Kreiss, D. S., Mastropietro, C. W., Rawji, S. S., & Walters, J. R. (1997). The response of subthalamic nucleus neurons to dopamine receptor stimulation in a rodent model of Parkinson's disease. *The Journal of Neuroscience*, *17*(17), 6807–6819.
- Krishnan, R., Ratnadurai, S., Subramanian, D., Chakravarthy, V. S., & Rengaswamy, M. (2011). Modeling the role of basal ganglia in saccade generation: Is the indirect pathway the explorer? *Neural Networks*, *24*(8), 801–813.
- Kumar, A., Cardanobile, S., Rotter, S., & Aertsen, A. (2011). The role of inhibition in generating and controlling Parkinson's disease oscillations in the basal ganglia. *Frontiers in Systems Neuroscience*, *5*.
- Magdoom, K., Subramanian, D., Chakravarthy, V. S., Ravindran, B., Amari, S.-I., & Meenakshisundaram, N. (2011). Modeling basal ganglia for understanding Parkinsonian reaching movements. *Neural Computation*, *23*(2), 477–516.
- Mandali, A., Rengaswamy, M., Chakravarthy, S., & Moustafa, A. A. (2015). A spiking Basal Ganglia model of synchrony, exploration and decision making. *Frontiers in Neuroscience*, *9*, 191.

- Muralidharan, V., Balasubramani, P. P., Chakravarthy, V. S., Lewis, S. J., & Moustafa, A. A. (2013). A computational model of altered gait patterns in parkinson's disease patients negotiating narrow doorways. *Frontiers in Computational Neuroscience*, 7.
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3), 139–154.
- Park, C., Worth, R. M., & Rubchinsky, L. L. (2010). Fine temporal structure of beta oscillations synchronization in subthalamic nucleus in Parkinson's disease. *Journal of Neurophysiology*, 103(5), 2707–2716.
- Park, C., Worth, R. M., & Rubchinsky, L. L. (2011). Neural dynamics in parkinsonian brain: the boundary between synchronized and nonsynchronized dynamics. *Physical Review E*, 83(4), 042901.
- Pinsky, P. F., & Rinzel, J. (1995). Synchrony measures for biological neural networks. *Biological Cybernetics*, 73(2), 129–137.
- Plenz, D., & Kitai, S. T. (1998). Up and down states in striatal medium spiny neurons simultaneously recorded with spontaneous activity in fast-spiking interneurons studied in cortex–striatum–substantia nigra organotypic cultures. *The Journal of Neuroscience*, 18(1), 266–283.
- Plenz, D., & Kitai, S. T. (1999). A basal ganglia pacemaker formed by the subthalamic nucleus and external globus pallidus. *Nature*, 400(6745), 677–682.
- Reynolds, J. N. J., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, 15(4), 507–521.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80(1), 1–27.
- Sharott, A., Doig, N. M., Mallet, N., & Magill, P. J. (2012). Relationships between the firing of identified striatal interneurons and spontaneous and driven cortical activities in vivo. *The Journal of Neuroscience*, 32(38), 13221–13236.
- Sinha, S. (1999). Noise-free stochastic resonance in simple chaotic systems. *Physica A: Statistical Mechanics and its Applications*, 270(1), 204–214.
- Sukumar, D., Rengaswamy, M., & Chakravarthy, V. S. (2012). Modeling the contributions of Basal ganglia and Hippocampus to spatial navigation using reinforcement learning. *PLoS ONE*, 7(10), e47467.
- Surmeier, D. J., Ding, J., Day, M., Wang, Z., & Shen, W. (2007). D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. *Trends in Neurosciences*, 30(5), 228–235.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1). UK: Cambridge University Press.
- Tachibana, Y., Iwamuro, H., Kita, H., Takada, M., & Nambu, A. (2011). Subthalamo-pallidal interactions underlying parkinsonian neuronal oscillations in the primate basal ganglia. *European Journal of Neuroscience*, 34(9), 1470–1484.
- Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, 13(1), 37–58.
- Weinberger, M., & Dostrovsky, J. O. (2011). A basis for the pathological oscillations in basal ganglia: The crucial role of dopamine. *NeuroReport*, 22(4), 151.