Neural Networks 67 (2015) 1–13

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

A spiking neural network based on the basal ganglia functional anatomy

Javier Baladron, Fred H. Hamker *

Department of Computer Science, Chemnitz University of Technology, Strasse der Nationen 62, 09107 Chemnitz, Germany

ARTICLE INFO

Article history:
Received 5 May 2014
Received in revised form 29 January 2015
Accepted 3 March 2015
Available online 24 March 2015

Keywords:
Spiking neurons
Basal ganglia
Reinforcement learning
Cognitive modeling
Action selection

ABSTRACT

We introduce a spiking neural network of the basal ganglia capable of learning stimulus–action associations. We model learning in the three major basal ganglia pathways, direct, indirect and hyperdirect, by spike time dependent learning and considering the amount of dopamine available (reward). Moreover, we allow to learn a cortico-thalamic pathway that bypasses the basal ganglia. As a result the system develops new functionalities for the different basal ganglia pathways: The direct pathway selects actions by disinhibiting the thalamus, the hyperdirect one suppresses alternatives and the indirect pathway learns to inhibit common mistakes. Numerical experiments show that the system is capable of learning sets of either deterministic or stochastic rules.

1. Introduction

The basal ganglia are a set of nuclei located in the forebrain. Several experiments have associated this brain area to action selection and reinforcement learning (Grillner, Hellgren, Menard, Saitoh, & Wikstrom, 2005; Packard & Knowlton, 2002; Wickens, Reynolds, & Hyland, 2003). The reinforcement signal, i.e. a reward prediction error (Schultz, 2010), is transferred to basal ganglia in form of the neurotransmitter dopamine originating in the substantia nigra pars compacta and the ventral tegmental area.

The basal ganglia are composed of several cortico thalamic loops that start in the cortex and via different pathways converge in the internal globus pallidus, an output structure which projects through the thalamus back to the cortex. All loops include either the striatum or the subthalamic nucleus, areas that are considered as input stages of the basal ganglia. In the basal ganglia each loop is composed of typically three different pathways, a direct pathway, an indirect pathway and a hyperdirect pathway (Schroll & Hamker, 2013).

To allow the simulation of behavioral experiments, several models of the complete basal ganglia have been proposed that do also include synaptic plasticity. Most of them are based on mean rate neurons (see for example: Frank, 2005; Gurney, Prescott, & Redgrave, 2001; Schroll, Vitay, & Hamker, 2012, 2014). Only recently, some models using spiking neurons have appeared (Chersi, Mirolli, Pezzulo, & Baldassarre, 2013; Stewart, Bekolay, & Eliasmith, 2012). The main difference between both relates to the different ways of modifying the synaptic weights. While rate-based models typically adjust their weights based on a three factor rule, pre- and post-synaptic firing rate plus dopamine, spiking models can also consider the exact timing of pre- and post-synaptic spikes by spike timing dependent plasticity (STDP) learning rules (Markram, Gerstner, & Sjöström, 2011; Morrison, Diesmann, & Gerstner, 2008).

Recently, Schroll et al. (2014) presented a set of learning rules for rate-based models that allow to determine the function of each pathway while minimizing hard-wired connections. However, none of the published spiking neural models of the basal ganglia allows for synaptic plasticity in all three pathways. Thus, in the present work we propose a new spiking network, inspired by the rate model presented by Schroll et al. (2014), that allows for learning by means of STDP in all three pathways. As a result of learning novel interpretations in the function associated to each of the cortico-thalamic pathways that include the basal ganglia emerge.

2. Basal ganglia anatomy

The main input structure of the basal ganglia is the striatum. It is composed mainly of medium spiny neurons (MSNs) and of a small amount of interneurons. The input from the cortex and thalamus to the striatum is mediated by glutamatergic synapses (Leh, Ptito, Chakravarty, & Strafella, 2007; Smith, Raju, Pare, & Sidibe, 2004; Wiesendanger, Clarke, Kraftsik, & Tardif, 2004). MSNs

* Corresponding author.
E-mail address: fred.hamker@informatik.tu-chemnitz.de (F.H. Hamker).

http://dx.doi.org/10.1016/j.neunet.2015.03.002
0893-6080/

are quiet at rest and require a strong correlated input to activate (Nisenbaum & Wilson, 1995). Once active they inhibit through GABAergic connections the neurons in the globus pallidus.

The striatum is also the destination of many projections from the dopaminergic cells of the substantia nigra pars compacta. The dopamine signal produced by these connections provides the basal ganglia with information about the performance of the task by means of a reward prediction error (Schultz, 2007). When more reward than expected is obtained (for example, in the form of juice in animal experiments) the level of dopamine is enhanced and when less reward than expected is received the dopamine level is reduced. This dopamine signal modulates learning in the basal ganglia connections. In neurons expressing the type-1 receptor (D1) a rise in the level of dopamine produces long term potentiation, while in neurons expressing the type-2 receptor (D2), it produces long term depression (Shen, Flajolet, Greengard, & Surmeier, 2008). A reduced level of dopamine reverses this effect, producing long term depression in D1 cells and long term potentiation in D2 cells. This difference suggests that both cell types have a different function. D1 expressing cells directly project to the internal globus pallidus (GPi), the main output nucleus of the basal ganglia, while D2 expressing cells first project to neurons in the external globus pallidus (GPe), which then project to the GPi. All these connections are inhibitory.

The pathway comprising striatal D1 cells and its direct connection to GPi is usually called the direct pathway and the one including D2 cells and GPe is usually called the indirect pathway. Both pathways converge in GPi which projects to the thalamus via GABAergic connections. The direct pathway, through the projections from striatum D1 cells, reduces the tonic activity of GPi and thus reduces the level of inhibition from GPi to the thalamus. The indirect pathway, through the inhibitory connections between striatum D2 cells and GPe, is able to remove the continuous inhibition that the tonic firing of GPe provides to GPi. The absence of this inhibition increases the level of activity in GPi. Thus, standard theories associate the direct pathway with a GO-function, i.e., initiating the correct action, and the indirect pathway to a NO-GO-function, i.e., inhibiting the incorrect actions (Braak & Del Tredici, 2008; O'Reilly & Frank, 2006; Schroll & Hamker, 2013).

The Subthalamic Nucleus (STN) is another input structure of the basal ganglia which receives connections from the cortex, the thalamus and the GPe and projects to GPi and GPe (Temel, Blokland, Steinbusch, & Visser-Vandewalle, 2005). The projections from the STN to GPi are excitatory, so this pathway, usually called hyperdirect, is supposed to have a different function than the inhibitory connection from the striatum. Anatomical evidence suggests a center–surround structure where this pathway inhibits competing motor programs while the direct pathway excites the correct one (Nambu, Tokuno, & Takada, 2002). For a recent review of the computational function of the different BG pathways see Schroll and Hamker (2013).

3. Previous spiking models of the basal ganglia

Recently, spiking models of the complete basal ganglia have been proposed which are able to learn stimulus–response associations. Chersi et al. (2013) developed a model to simulate a behavioral task in which a monkey must learn a stimulus–action association. The monkey sits in front of a table with three lights and three buttons. At the beginning of a trial a light is flashed and then the animal must discover which button he has to press to turn the light on again. Only one button is correct for each light. An interesting characteristic of this model is that to reach a successful response in one trial a set of consecutive actions must be performed. The simulated monkey must first look at the flashed light, then reach the button and finally press it.

In their network each nucleus of the basal ganglia, the motor cortex and the prefrontal cortex, are represented by a layer composed of a set of populations of leaky integrate and fire neurons, one for each possible action.

The projections arriving to the striatum, but not to STN, are plastic and adapt according to a three factor rule which depends on both the level of dopamine, the timing between spikes and a trace to solve the temporal credit assignment problem. However, the learning rule is identical for all connections, independent of the pathway or dopamine receptor. All remaining connections of the direct, indirect and hyperdirect pathway and their weights are determined by an optimization routine which assures the fulfillment of a set of biological and functional restrictions. In the resulting network the direct pathway is in charge of selecting the proper action by disinhibiting the corresponding population of the thalamus, while the indirect only maintains the activity of GPi within working limits, avoiding undesired behaviors like oscillations. The hyperdirect pathway basically switches off the BG selection mechanism allowing direct connections between the prefrontal cortex and the motor areas to determine the action to be executed.

Another spiking model has been recently proposed by Stewart et al. (2012) based on the firing rate model of Gurney et al. (2001). The model of Gurney et al. (2001) is basically an action selection model that, by its implemented BG connections, ensures that a single action is determined, the one with output zero, given utility values Q for each action as input. The utility value is the estimate of reward given a particular state and a chosen action. The model of Stewart et al. (2012), uses learning between cortex and striatum but not between cortex and STN to map a sensory state to utility values represented by populations of spiking neurons while the other connections are set to replicate the same exact computations performed by the original rate model. The learning rule does not depend on the timing between the presynaptic and postsynaptic spikes but only on the amount of activity of both cells and on an error signal. The activity of a neuron is estimated using the amount of neurotransmitters released in the synapse. The error signal is computed by subtracting the vector of utility values (one value for each possible action in a particular state) from the received reward. Thus, the error signal is not uniform across striatal populations. The network is capable of learning a probabilistic bandit task, in which an agent must choose between two or more actions. Reward is non deterministic, depending on a probability distribution.

Humphries, Stewart, and Gurney (2006) proposed a different spiking model which was used to study the appearance of slow oscillations in the STN and the GPe. The dopamine signal in this model is not used to modify the synapses, which are all fixed, but only as a modulator of synaptic efficiency. This model is able to reproduce several biological datasets but its lack of plasticity makes it incomparable with this work where one of the main objectives is to test the learning capabilities of the basal ganglia.

The learning process in the previous spiking models does not differentiate between cells with a different type of dopamine receptor. However, recent experiments have shown that the effect of dopamine in plasticity varies according to the type of receptor expressed by the cell (Calabresi, Picconi, Tozzi, & Filippo, 2007; Shen et al., 2008), a factor that none of the spiking models have taken into account.

Also, none of the previous models can explain the effects of pallidotomy, a common treatment for Parkinson's disease in which the GPi is lesioned, reducing the influence of the BG in the thalamus. After the surgery, patients can perform everyday movements (Lozano et al., 1995) but are impaired at learning (Sage et al., 2003). This suggests that there exists another mechanism for action selection that can guide behavior independent of the basal ganglia. We propose that this can be performed by direct cortico thalamic connections which are trained by the BG. However, in order to change

a learned (habitual) behavior established by the cortico-thalamic interactions, the basal ganglia would require a strong inhibitory mechanism. We believe that this is possible due to the existence of asymmetrical synaptic plasticity in the distinct cortico striatal pathways.

Other modeling approaches have used a different, top-down, approach to construct spiking models of reinforcement learning inspired by classical machine learning techniques (Sutton & Barto, 1998). Potjans, Morrison, and Diesmann (2009) proposed a method in which any implementation of the TD algorithm could be directly transformed into a spiking network, including a complete correspondence between parameters. Their approach has two main problems: first from the biological point of view, there is no direct concordance between the structure of their model and the known functional anatomy of the basal ganglia, and, second, their implementation is designed for discrete state descriptions while the brains operates in a continuous way. A solution to the latter problem has been proposed in Frémaux, Sprekeler, and Gerstner (2013), but this implementation is still lacking a direct association to the different basal ganglia nuclei and their projections. Morita, Morishima, Sakai, and Kawaguchi (2012) introduce a network that is consistent by anatomical facts. However, it only operates in a state-based fashion.

4. Network description

We here introduce a bottom-up modeling approach to the basal ganglia which can explain the role of the different effects of dopamine receptors in plasticity. We have also added direct cortico thalamic projections, which are capable of conducting behavior once an stimulus action association have been learned by the basal ganglia pathways. These two features introduce a new mechanism for action selection, different than previous computational models. The model is inspired by a recent rate model presented by Schroll et al. (2014) where through different, realistic learning rules the pathways learned to perform a behavioral task.

The model of Schroll et al. (2014) includes the direct, hyperdirect, indirect and cortico-thalamic pathways of the basal ganglia. The connectivity is not pre-specified but defined by a set of plasticity rules. At the beginning the pathways do not implement any function but acquire one through experience. Learning is modulated by dopamine (except in the direct connections between the cortex and the thalamus).

Schroll et al. (2014) also induced a Parkinsonian state in their rate model by lowering the dopaminergic cell output to the striatum, which alters synaptic plasticity and as a result dysfunctional pathways functionalities emerge. The simulations showed that the changes in the direct and indirect pathways agreed with early theories about the effects of Parkinson's disease and predicted an increase in the output of both the hyperdirect pathway and the direct cortico-thalamic pathway. Dopamine replacement was simulated by increasing the tonic dopamine level. Simulations showed that initial learning performance can be alleviated by low dopamine doses while automatic execution required high doses which impede learning. In networks that received pallidotomy, a common therapeutical lesioning of GPi for Parkinsonian patients, the performance on well known tasks was restored but learning was impaired, in agreement with data.

We here show that the functions that emerge through learning do not depend on the particular kind of Hebbian and three factor learning rules used by Schroll et al. (2014) but equally well occur when spiking neurons and spike time dependent learning rules are used. All model details are given next.

As we focus on the whole basal ganglia some simplifications are made with respect to computational details. We do not model in detail the function of different interneurons and omit some known connections in the BG.

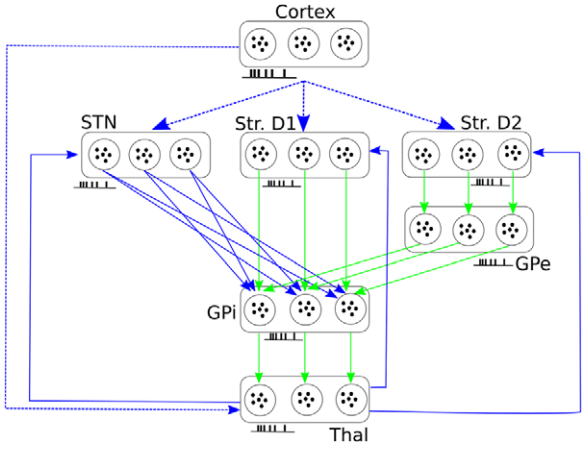

Fig. 1. Network structure. Each rectangle represents one nucleus or component of the basal ganglia. The circles inside each nucleus represent the different neuron populations in each (see text). Dashed lines represent plastic connections and solid lines fixed connections. Blue lines represent excitatory connections and green lines inhibitory. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.1. Cortico thalamic pathways

In our network, each of the basal ganglia nuclei is represented by a layer of spiking neurons (Fig. 1). The striatum is divided into two separate groups, one for D1 receptor expressing cells and one for D2 expressing cells. Each layer in the basal ganglia is composed of several populations of neurons, where each population is linked to a particular action, as implemented in most BG models. Learning takes place between cortex and striatum, cortex and STN as well as cortex and thalamus.

The other connections of basal ganglia pathways are hard coded. Both, in the direct and indirect pathway, each population is only connected to neurons encoding its same action in the next layer. In the hyperdirect pathway, each neuron is connected to all neurons encoding an action different than the one represented by its own population. All these connections are shown in Fig. 1.

Cortical input is represented using one population of neurons for each possible stimulus. Each cell of these populations is initially connected to all the neurons in the striatum and the STN.

At rest, the GPi generates spikes that keep the activity in the thalamus at a low level. Activation of a striatal D1 population inhibits the GPi, reducing the inhibition in one population of the thalamus which in turn will allow this population to increase its activity to finally select an overt response.

Activation of a striatal D2 population will inhibit one population of GPe which are otherwise firing with a baseline rate. A reduction in the activity of GPe will then reduce the amount of inhibition received by the GPi, increasing its activity and suppressing an overt response. Activation of a STN population will increase the activity of the GPi in all populations encoding a different action and therefore will inhibit alternative actions.

Both the direct and the hyperdirect pathway jointly generate patterns that will show a clear action selection in the thalamus. The direct pathway removes the inhibition of the correct action, while the hyperdirect increases the inhibition to alternative actions. This implements a center–surround inhibition pattern in GPi if the weights develop through learning. A detailed view of the action selection process is shown in Video 1 (see Appendix A) and Fig. 2, where each spike generated by the network is visualized.

The network also includes cortico-thalamic connections that allows a fast decision, without the information being processed by the basal ganglia. Biological evidence for these connections can be found in Haber (2003). Initially, each population in the

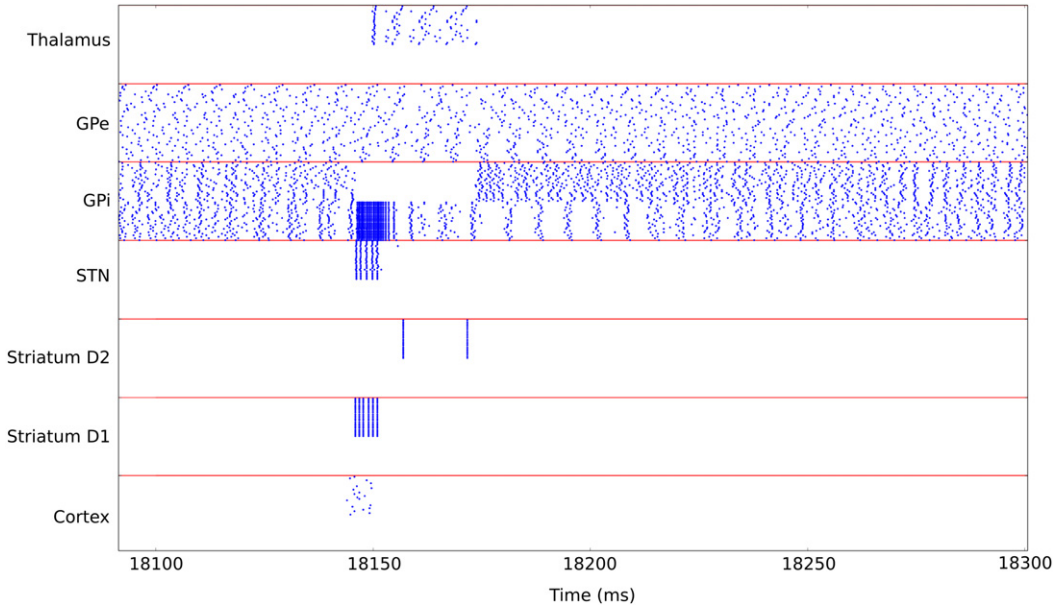

Fig. 2. Raster plot. This plot shows the activity of the network after an association has been learned. Each area is divided in two populations of 25 neurons each, both representing different possible actions. The cortex is initially silent until the stimulus is presented, this activates immediately both striatum D1 cells and STN cells. Striatum D1 cells inhibit the neurons of GPi belonging to the channel encoding the action to be selected. STN cells excite other cells to suppress the execution of alternative actions. The decrease in the inhibition that the Thalamus constantly receives from GPi activates only the population of cells that represent the correct action. The feedback projections from the Thalamus produce a small activation of D2 striatal cells, useful for learning, but incapable of canceling the effects of the striatal D1 neurons.

stimulus cortex is connected to all populations in the thalamus. Slowly, as learning progresses in the basal ganglia pathways, these connections are modified in such a way that they express the same association that has been created in the basal ganglia.

We have also implemented local inhibition between populations in GPi and striatum. Each neuron in these areas projects to cells belonging to all other populations in the same layer. This enhances competition between actions. In the striatum this effect is probably created by its complex interneuron structure (Tepper & Bolam, 2004). We also included inhibition in the thalamus and STN, but as neurons in these areas are excitatory, we had to include a small amount of inhibitory neurons. Each excitatory population in these layers are strongly connected to one small, different, group of inhibitory interneurons, which then project to all other populations.

Finally, both STN and striatal cells receive feedback connections from the population of the thalamus encoding its same action (see Table 1). This enhances the response of neurons that represent the selected action, increasing the effect of the learning rule.

4.2. Neurons and synapses

We have used the Adaptive Exponential Integrate and Fire model (Brette & Gerstner, 2005; Naud, Marcille, Clopath, & Gerstner, 2008) for all neurons in the network. The state of each neuron is described by 2 values, an adaptation variable (w) and a membrane potential (Vm). Both variables are governed by the following differential equations:

$$\begin{aligned} C\frac{dVm}{dt} &= g_L(E_L - Vm) + g_L\Delta_T \exp\left(\frac{Vm - V_T}{\Delta_T}\right) \\ &\quad + I_{ext} + g_e - g_i - w \\ \tau_w\frac{dw}{dt} &= a(Vm - E_L) - w \\ \tau_e\frac{dg_e}{dt} &= -g_e \\ \tau_i\frac{dg_i}{dt} &= -g_i. \end{aligned} \tag{1}$$

Table 1
Weight values used in all the simulations.

Pre-synaptic populations	Post-synaptic population	Weight value
Striatum D1	GPi	2.0 nA
GPi	Thalamus	0.12 nA
GPi	GPi	0.1 nA
Thalamus	D1	0.2 nA
Striatum D2	Striatum D2	2.0 nA
Striatum D2	GPe	2.0 nA
GPe	GPi	1.0 nA
Thalamus	Striatum D2	0.5 nA
STN	GPi	3.0 nA
STN	STN interneurons	2.0 nA
STN interneurons	STN	1.0 nA
Thalamus	STN	0.2 nA
Thalamus	Thal. interneurons	2.0 nA
Thal. interneurons	Thalamus	0.1 nA

The value of Vm was initialized to -65 mV and w, g_e and g_i to 0. A description of each parameter of Eq. (1) together with the value given in each of the simulations is presented in Table 2.

Once the value of Vm has reached a threshold of 30 mV we consider a spike has been emitted and reset Vm to -65 mV and increase w by 0.08 nA. The input to each neuron is composed of the sum of 3 different values, $I = I_{ext} + g_e - g_i$. Both g_e and g_i represent input coming from other neurons in the network, while I_{ext} represents the sum of all external currents. Excitatory synapses increase the value of g_e by an amount which depends on the strength of the connection between the pre-synaptic and post-synaptic neuron (the weight) each time the pre-synaptic neuron spikes. Inhibitory synapses increase g_i instead. Both variables decay exponentially with time constants τ_e and τ_i in the absence of any activity.

Only the GPe, GPi and the thalamus have a I_{ext} baseline rate greater than 0, modeled by a Gaussian random variable with different mean and variance depending on the nucleus.

4.3. Learning rule

All the projections from the cortex to the basal ganglia are learnable by dopamine modulated STDP. We have used the rule

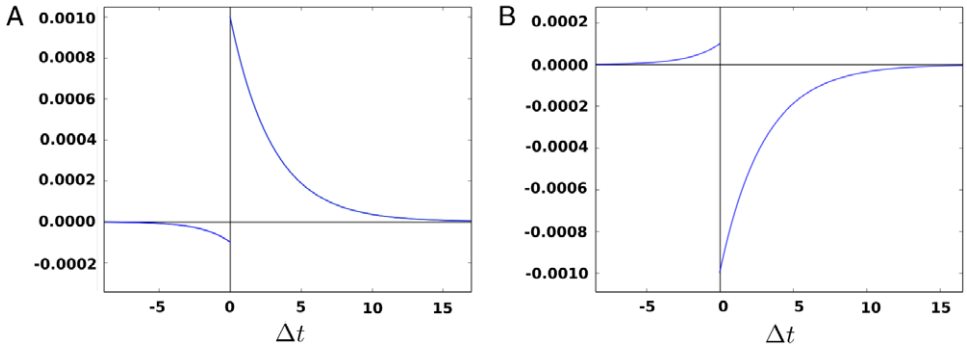

Fig. 3. $STDP(\Delta t)$ function for synaptic plasticity between the stimulus cortex and the basal ganglia. A: function for the synapses projecting to D1 expressing cells when dopamine is above baseline and for the D2 expressing cells when dopamine is below baseline. B: function for the synapses projecting to D1 expressing cells when dopamine is below baseline and for the D2 expressing cells when dopamine is above baseline.

Table 2
Neuron parameters used in all the simulations, taken from Brette and Gerstner (2005).

Parameter name	Description	Value
C	Membrane capacitance	281 pF
g_L	Leaky conductance	30 nS
E_L	Leak reversal potential	−70.6 nV
Δ_T	Slope factor	2 mV
τ_w	Adaptation time constant	144 ms
a	Subthreshold adaptation	4 nS
I_{ext} mean GPe	Mean value of GPe	3 nA
I_{ext} s.d. GPe	Standard deviation of GPe	0.1 nA
I_{ext} mean GPi	Mean value of GPi	10 nA
I_{ext} s.d. GPi	Standard deviation of GPi	0.5 nA
I_{ext} mean Thal.	Mean value of thalamus	1.5 nA
I_{ext} s.d. Thal.	Standard deviation of thalamus	0.1 nA

Table 3
Learning rule parameters for the connections between the stimulus cortex and the basal ganglia. These values were used in all the simulations.

Parameter name	Description	Value
τ_+	Pre–post time constant	3 mS
τ_-	Post–pre time constant	2 mS
A_+	Pre–post maximum STDP value	0.001
A_-	Post–pre maximum STDP value	0.0001
τ_E	Eligibility trace time constant	3 ms
τ_e	Excitatory synapses time constant	1 mS
τ_i	Inhibitory synapses time constant	1 mS

proposed by Izhikevich (2007) which is able to solve the distant reward problem by adding an eligibility trace. The change in the weights is given in the following equations:

$$\tau_E \frac{dE}{dt} = -E + STDP(\Delta t) \tag{2a}$$

$$\frac{dws}{dt} = E \cdot DA \tag{2b}$$

where E is the eligibility trace, ws is the weight, and DA is the dopamine signal. A positive DA represents a dopamine level above baseline, which happens when reward is received. A negative DA represents a dopamine level below the baseline, which in our model happens when no rewards is obtained at the end of a trial. The value of E was initialized to 0 for all synapses.

STDP is the standard spike timing dependent plasticity function (see Morrison et al., 2008 for a review) which depends on the time difference between a post-synaptic and pre-synaptic spike (Δt). If the pre-synaptic neuron spikes before the post-synaptic neuron $STDP$ will lead to long term potentiation (LTP) while if the order is reversed it will lead to long term depression (LTD). The variable E stores a history of the effect of spike pairs as a trace that will be used to change the weights once the dopamine level changes from its baseline. The STDP function is given in Eq. (3), where A_+ and A_- are the maximum values and τ_+ and τ_- are parameters which limit the maximum distance between spikes that will cause a modification. A plot of this function with the parameters given in Table 3 is shown in Fig. 3.

$$STDP(\Delta t) = \begin{cases} A_+ \exp(-\Delta t/\tau_+) & \text{if } \Delta t > 0 \\ A_- \exp(\Delta t/\tau_-) & \text{if } \Delta t < 0. \end{cases} \tag{3}$$

In D1 cells, when DA is above baseline, LTP takes place when the presynaptic neuron fires before the postsynaptic neuron. A different temporal order of the spikes leads to a small LTD. In D2 cells, the effect of dopamine is reversed, as observed in Shen et al. (2008). This has been implemented by multiplying the right hand side of Eq. (2b) by −1, see also Fig. 3.

According to the classification proposed by Frémaux, Sprekeler, and Gerstner (2010), this learning rule corresponds to a R-STDP type, a category which Frémaux et al. (2010) show to be separable into the sum of two terms: one which depends on the covariance of the eligibility trace and the reward, and a second one, which depends only on the mean of both the reward and the trace. This second term is called the unsupervised bias as it is independent of the relationship between behavior and success. The authors showed that under certain circumstances this latter term can be dominant and impede learning. In the model, to reduce this effect, the level of long term depression in D1 MSNs is very low (see Table 3 and Fig. 3A). This condition relates well to the last experiment of Frémaux et al. (2010) where LTD has been suppressed. Also, this is consistent with Shen et al. (2008), who showed that the magnitude of LTD is much smaller than LTP in striatal cells.

The modification of the cortico-thalamic projections is not modulated by dopamine, but depends only on spike timing. For these connections, each pre-synaptic spike followed by a post synaptic spike will increase the weight and each reversed pair will reduce it, independent of the level of dopamine. The amount of weight change is given directly by the STDP function shown in Eq. (3). The learning of these connections is much slower than the cortical basal ganglia projections by using smaller values of A_+ and A_-. In fact, only a large amount of consecutive correct decisions can set the cortico-thalamic connections to its maximal level.

Homeostasis is required in the cortico-thalamic connections to limit the growth of the weights and to assure a one to one mapping between stimuli and actions. This is achieved by reducing the weights by a small amount (γ_S) each time the pre-synaptic (cortical) neuron spikes. If the post-synaptic neuron also spikes, because its action has been selected by the basal ganglia, then the rise due to the pre–post pairing will be much higher than the homeostasis reduction. If the post-synaptic neuron does not fire, then the basal ganglia are currently not associating the input to

Table 4
Learning rule parameters for the direct connections between the stimulus cortex and the thalamus. These values were used in all the simulations.

Parameter name	Description	Value
τ_+	Pre–post time constant	5 mS
τ_-	Post–pre time constant	10 mS
A_+	Pre–post maximum STDP value	2×10^{-09}
A_-	Post–pre maximum STDP value	1×10^{-11}
γ_S	Homeostasis term	0.02 nA

the encoded action and the weight is reduced. The learning rule for the weight of a synapse between neuron i of the cortex and j of the thalamus is shown in Eq. (4). In the equation, the activity of each neuron is described as a set of short pulses t_k^i, where i is the index of the neuron and k the index of the spike. A spike train for neuron i is expressed as $X_i(t) = \sum_k \delta(t - t_k^i)$.

$$\Delta w_{ij}(t) = -\gamma_S X_i(t) + X_i(t) \sum_k STDP(t_k^j - t) + X_j(t) \sum_k STDP(t - t_k^i). \quad (4)$$

The first term of the right-hand side of Eq. (4) correspond to homeostasis, the second to the effect of a pre-synaptic spike and the third to the effect of a postsynaptic spike. The effect of a single spike on the weights depends on the time difference between its emission time and that of all previous spikes, according to the function of Eq. (4).

A decision is accompanied by the increase in thalamic activity enhancing the weights of the direct projection between the active stimulus and those thalamic cells. This effect is independent of the correctness of the choice and not modulated by dopamine. For this reason, the weight increases performed during wrong trials need to be removed once the correct association is found by the basal ganglia. Thus, in each correct trial, the activity in the thalamus is focused on the correct population, and due to the homeostatic term the weights to all other cells will diminish. Over multiple repetitions all the wrong information acquired during previous trials will be forgotten and only one association will be learned.

The homeostatic term is not required in the cortico striatal connections due to the effect of dopamine. In these synapses the modification of the weights is different in correct and incorrect trials.

Both learning rules include an explicit maximum weight value. All cortico striatal synapses are limited to a range between 0 and 3 nA, while the direct cortico thalamic can only take values between 0 and 1 nA.

5. Numerical experiments

We have implemented the network using the Brian spiking neural network simulator (Goodman, 2009) and ran several numerical experiments that are detailed in this section. The common parameters used in all simulations are shown in Tables 2–4 and the weights in Table 1.

5.1. Learning and relearning of stimulus–response associations

In the first experiment we tested the capacity of the network to learn an initial map between stimuli and actions and its capacity to adapt to changes in the environment. In this simple reversal learning task, in each trial, a single stimulus is presented to the model which has to learn to associate the stimulus with a single action chosen from a finite set of possible actions. If the decision is correct reward is provided. Full success is achieved if 50 correct answers are given following each other.

Once the model has accomplished the task, the stimulus–response mapping is changed. This modification is not informed to the model.

The network used for this experiment has two populations in each layer, two possible stimuli and two possible actions. All the projections starting at the stimulus cortex are randomly initialized, except for the connections to D2 cells which are set to 0 to minimize the level of action suppression in the beginning. Although the weights are randomly selected the mean value (0.5 nA) is chosen to be equal. The weights in the direct and hyperdirect pathways balance each other to select a particular, random action for each stimulus.

The weights of the local inhibitory synapses were chosen so that just one population is strongly active at each trial even if at the beginning the cortico-striatal synapses are not explicitly biased towards any action. This ensures that each stimulus is associated to just one action and not with all of them.

All the other connections have weights defined in Table 1. The stimulus presented in each trial is selected randomly and presented for 50 ms.

The output of the network is determined by temporal integration using two accumulators, one for each possible action. The value of each accumulator starts at zero at the beginning of each trial and is increased each time a neuron in the corresponding population of the thalamus spikes. If there is no activity they exponentially decrease to zero. The first accumulator to reach a fixed threshold determines the selected action. If none has reached the threshold after 100 ms, we conclude that no decision has been made.

If the correct decision is made then the value of *DA* is increased to 10×10^{-8} and if the decision is wrong it is decreased to -10×10^{-8}. The network is run for 350 ms to provide a period in which the synapses will be updated according to the information stored in the eligibility traces. In this period the level of DA approaches 0 exponentially. No resetting takes place. A summary of the protocol is presented as a diagram in Fig. 4

The 100 ms period in which the model is allowed to make a decision cannot be directly compared to reaction time measurements from behavioral experiments, as the time for stimulus processing would take place prior to the activation of the stimulus neurons. Also, we do not include any motor processing, once a decision is made this information must be transferred from the thalamus to the motor cortex. Note that the core processes of perceptual decision making appear to be very fast. Stanford, Shankar, Massoglia, Costello, and Salinas (2010) proposed a method to estimate the core decision time independent of the time required for action and visual processing and concluded that a choice can be made in 25–50 ms.

We ran 100 simulations, each with different initial conditions and different random stimuli. All networks have successfully achieved the initial and reversal learning. Only a few trials with positive dopamine are enough to learn the initial association. The amount of trials required to succeed varied between 50 and 55, with a maximum of only 5 incorrect responses. Reversal learning requires more trials. During the first period after the change, the network keeps choosing the same action until the negative *DA* sufficiently reduces the weights in the direct and hyperdirect pathways and increases enough the weights in the indirect pathway to produce a different decision. Once this happens the new association is learned very fast. The amount of trials required for the reversal learning varied between 98 and 118.

Fig. 5 shows the development of the mean weights of the connections between each stimulus population and each of the action populations connected to them over the duration of a simulation. After a small amount of trials, because of the feedback provided by dopamine, the correct association is quickly found

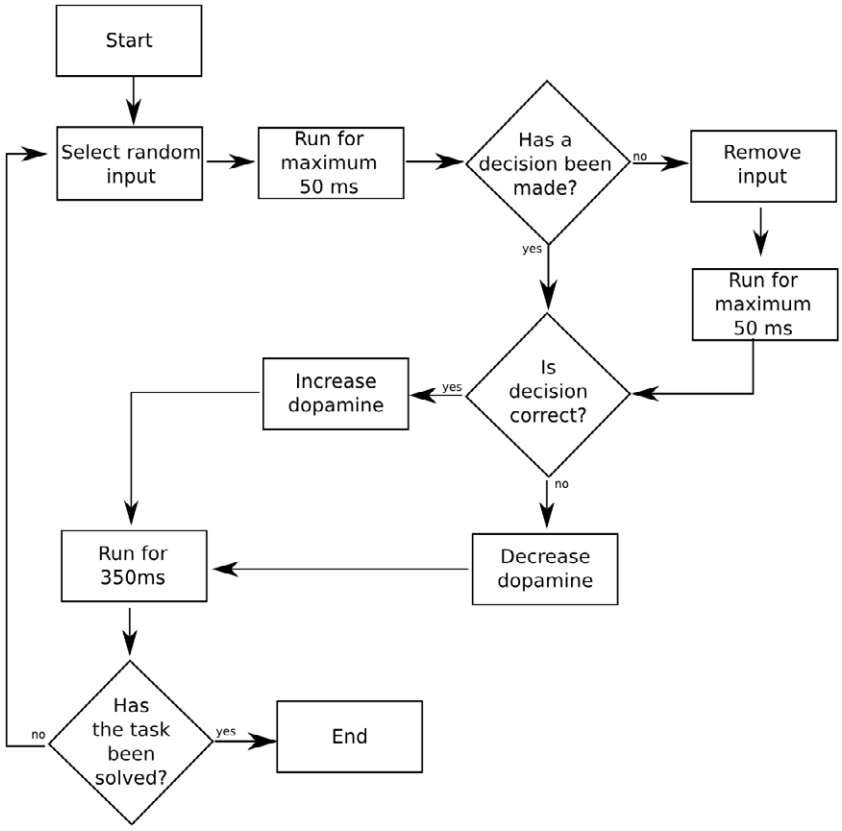

Fig. 4. The protocol performed for learning an action stimulus association task. In the case of relearning this procedure is repeated for the second mapping.

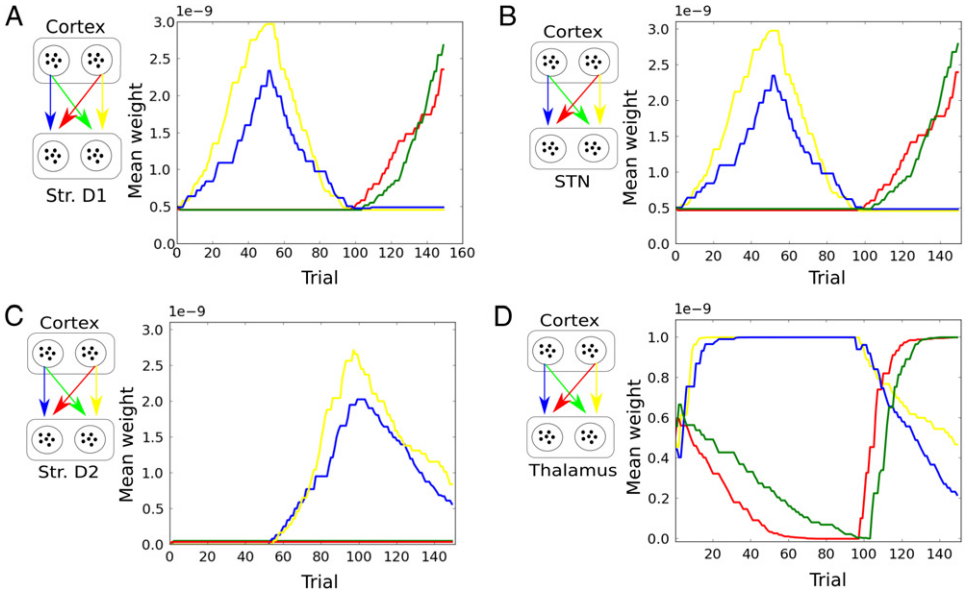

Fig. 5. Mean weight of the learnable projections from the stimulus cortex in the deterministic reversal learning experiment of one typical example model. Connections from cortex to A: Striatum D1 cells, B: STN cells, C: Striatum D2 cells, and D: Thalamus. The reversal of the mapping takes place at trial 52. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

by the weights, increasing to the correct population on the direct pathway and to the suppression of alternative actions in the hyperdirect pathway. This relationship is kept until the rules are reversed and the action is not rewarded any more. After the rule change at trial 52 the weights in the direct and hyperdirect pathway start to decrease and when they have reached the same level as the other associations the new correct one begins to rise.

The connections between cortex and the D2 cells develop differently (Fig. 5(C)). During initial learning they stay very close to zero and only rise during the re-learning process. This is because the learning of the first association is very fast and there are not enough trials with negative *DA* to increase the weights. However, once the rules have changed, the network will keep making the same decisions as initially, failing often enough to activate learning in the indirect pathway. Once the reversal rule has been learned

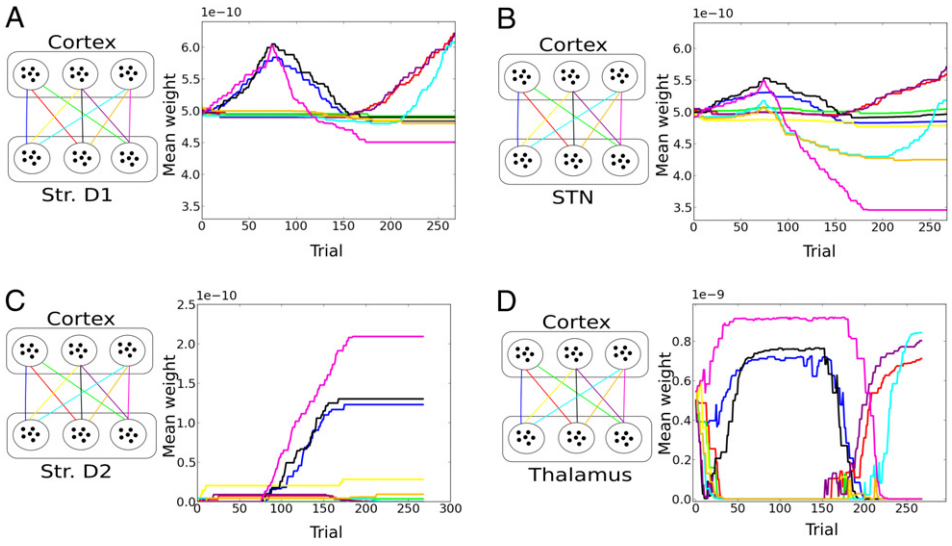

Fig. 6. Mean weight of the learnable projections from the stimulus cortex in the 3 stimuli 3 action experiment. A: stimulus cortex to striatum D1 cells. B: stimulus cortex to STN cells. C: stimulus cortex to striatum D2 cells. D: stimulus cortex to thalamus. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the projections will slowly decrease, due to the existence of only positive reward (see the yellow and blue line on Fig. 5(D)).

Thus, it is not necessary to maintain the strong inhibition of the former behavior that D2 cells had learned during the task switch. From this point on, the correct stimulus–action association can be maintained through the combination of the direct, hyperdirect and cortico-thalamic pathway without a further involvement of the indirect pathway.

Fig. 5(D) shows the evolution of the direct connections between cortex and the thalamus. These connections also learn the correct association in the initial and re-learning phase but much more slowly. Thus, they follow the rule discovered by the basal ganglia pathways.

In a second part of this experiment we incremented the number of possible stimuli and actions from 2 to 3. This was done by increasing the number of populations in the network created for the previous experiment. We have run 100 simulations using the same parameters as before. Each stimulus–action association was learned following the protocol of Fig. 4.

The amount of trials required for learning the first stimulus–action association varied between 64 and 156, increasing in comparison to the previous experiment. Now, the amount of possible combinations is higher and the network requires more exploration to find the appropriate association. For the stimulus–action association after rule reversal the number of required trials varied between 85 and 149, with a mean of 104.93 trials. Video 1 shows the activation of the different nuclei during a late trial once the correct association has been learned.

Fig. 6 shows the mean weights of one simulation in a similar way as it was shown in Fig. 5. As before, in the projections to both STN and D1, only the weights corresponding to the correct association rise until the task is solved and the rules change. Then, the negative dopamine produced by wrong decisions reduces these weights to the same level as the others and allows a new, different, association to develop.

The indirect pathway shows a mild suppression in the early period of learning, but primarily activates when the task changes due to the larger number of mistakes produced by the network during the first group of trials after reversal. The negative dopamine level produces a pattern which inhibits the previously correct association enhancing the exploration of new alternatives.

The main difference with the previous experiment (Fig. 5) is that the weights of the projections from the cortex to the D2 MSNs do not decrease once the exploration phase is finished. The reason for this is that the switch to the new rule has been already learned while the weights from cortex to striatal D2 cells being small. This happens because each stimulus is presented less often as the 50 trials must be divided between three categories and not two. Thus, the synapses in Fig. 6 do not reach the same levels as in Fig. 5. Once the new rule is discovered the pattern of activity in the thalamus changes. The cells which encode the previously correct action are not active any more, providing no feedback signal to the D2 population to inhibit the previous action. As a result, the D2 cells become not active any more because the weights of the cortico-striatal connections are not high enough to produce postsynaptic activity by themselves. Due to the lack of postsynaptic activity, the learning rule does not reduce the weight as indicated in Fig. 6.

The cortico-thalamic connections also learn the correct associations, as can be seen in Fig. 6D. The initial, longer, exploration phase is reflected in this plot during early trials where all associations reduce their weights except for the correct one.

A critical element of our model is the existence of inhibitory lateral connections in the striatum that allow a competition between populations. This winner-takes-all dynamic has been widely used to explain action selection in the basal ganglia but has been questioned by a group of more recent biological experiments (Plenz, 2003). The experiments performed by Czubayko and Plenz (2002) and Tunstall, Oorschot, Kean, and Wickens (2002) have shown that the interaction between spiny neurons in the striatum is sparse and it is dominated by unidirectional projections. Theoretical studies (Ponzi & Wickens, 2003, 2012) have proposed that through this sparse connectivity the activation of neurons in the striatum become locked at different times from stimulus onset, creating complex temporal patterns. Although this new approach is still in discussion it may contradict the assumptions we made in this neurocomputational model. For this reason, we implemented a second version of the network, in which we study the effects of a reduction in the levels of lateral inhibition.

For this second version of the model we reduced the inhibitory weights between populations of the striatum and STN to 1×10^{-11} nA. Also, the inhibition between thalamic populations was removed to further weaken the competition between populations. The resulting network was still able to learn 2 different stimulus–action associations following the protocol shown in Fig. 4. This was achieved only after reducing the amount of dopamine delivered after each trial. A slower learning was required because the

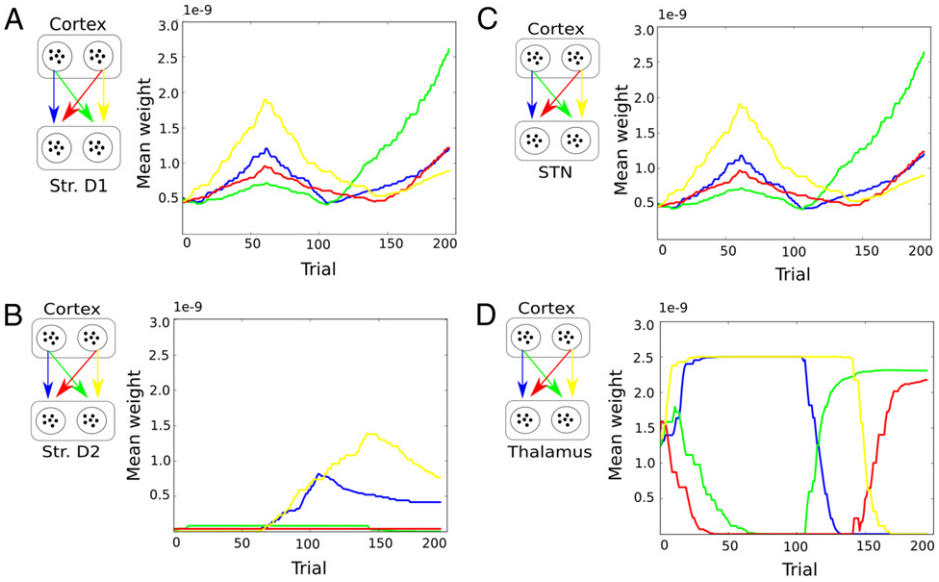

Fig. 7. Mean weight of the learnable projections from the stimulus cortex in one example simulation of the model with low lateral inhibition. Connections from cortex to A: Striatum D1 cells, B: STN cells, C: Striatum D2 cells, and D: Thalamus. The reversal of mapping takes place at trial 62. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Probability of finding reward in each side of the maze in the probabilistic learning experiment.

Block	Right path	Left path
1	0.63	0.21
2	0.21	0.63
3	0.72	0.12
4	0.12	0.72

lack of inhibition increased the amount of activity in both STN and striatum D1.

The amount of trials required to succeed slightly increased with this new version of the model. To learn the first rule, the network required between 50 and 73 trials (mean 56.5) and for the reversal learning the network required between 94 and 251 trials (mean 123.2). This is a consequence of both the slower learning rate and the reduction of lateral inhibition.

Fig. 7 shows the evolution of the weights in one example simulation in a similar way as it was shown in Fig. 5. In this new version of the model, on both STN and striatum (see Fig. 7A and B), all the weights rise together, not as in the previous network where only one association increased its value. This is produced because the reduced inhibition allows the activation of both populations of STN and striatum on every trial. Both the cortico-thalamic connections and the indirect pathway work in a similar way as before.

The correct behavior is only produced because the weights corresponding to the proper association, increased faster than the others. Although the difference between the projections is smaller than in the previous network, it is still big enough to make a correct decision. This weight pattern is now achieved mainly through the effect of the thalamic feedback signal, which increases the activity of the selected action, activating the learning rule and enhancing the current association (or decreasing it in an incorrect trial).

5.2. Probabilistic learning

In a second experiment we simulated a task used by Kim, Sul, Huh, Lee, and Jung (2009) in which a rat is situated in a maze where, after crossing a bridge, it has to choose between a left and a right path. Food is placed at the end of each path according to a probability which is changed every 40 trials.

The main difference between this task and the previous one is the probabilistic nature of the reward such that it can also be received for a wrong selection or not received although the decision has been correct.

To simulate this task we use only a single input stimulus (that may mimic the activation of a place cell) which indicates that a decision between two actions, left or right must be made. At the beginning of each trial, this population was activated for 50 ms in a similar way as in the previous task. The output decision was determined by the same accumulators.

Each simulation consisted of 4 blocks containing each 40 trials. The reward probabilities in each block are given in Table 5 identical to the ones used by Kim et al. (2009). These values assure that the path with the maximum probability changes after every block. The network does not know when this change will occur and it will only adapt based on the feedback via the dopamine level, which is positive (20×10^{-8}) if food was found in this trial and negative (-20×10^{-8}) if nothing was encountered. The probabilities of Table 5 are used only to determine the existence of reward but not to compute the amount of dopamine, which is the same on all trials in which food was encountered.

There are two main differences with respect to the protocol presented in Fig. 4. First, at the beginning of the trial, there is no input selection, as there is only one population in the first layer of the network. On every trial the same neurons are activated. Second, after a decision is made, it is not necessary to determine if it was correct or not, but, instead, the probabilities of Table 5 are used to resolve if reward is found in this trial.

Fig. 8 shows the probability of choosing the left path on each trial, computed from 100 models with different initial weights. The plot shows how the network adapts on each block and slowly discovers the new association and profits from it. On both types of blocks the network finally learns to select most of the time the pathway that produces food with a higher probability.

The plot of Fig. 8 is very similar to the experimental data of Kim et al. (2009). They reported that the animals began to choose the path associated with the highest reward probability after 10–20 trials. This behavior is replicated in our experiment where after a block change the probability of choosing the previous path gets below 0.5 in a similar amount of trials. Moreover, similar as the rats, the model does not decide for the most rewarded action in all trials.

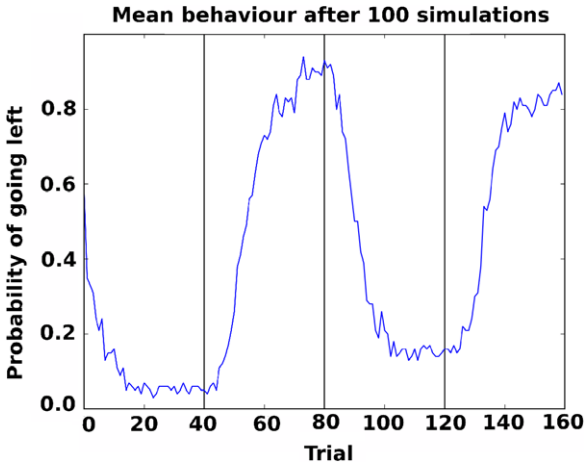

Fig. 8. Probability of choosing the second path on each trial, computed from 100 models with different initial conditions.

5.3. GPi lesioning

Parkinson's disease is a basal ganglia dysfunction that produces motor symptoms like difficulties of movement initiation or tremors. It is associated with a decrease in the production of dopamine in the substantia nigra pars compacta which affects the normal behavior of the basal ganglia. One common surgical treatment for severe cases had been the use of pallidotomy, a surgery where small areas of the internal globus pallidus are destroyed.

It has been shown by Lozano et al. (1995) that the GPi lesion caused by pallidotomy improves performance in well learned everyday movements, reducing the bradykinesia, rigidity and tremors common to parkinsonian patients. Another study showed that the same type of lesion impairs feedback learning in a categorization task (Sage et al., 2003). Both results together suggest that the basal ganglia are strongly involved in learning but less in the execution of well learned tasks.

We simulated a GPi lesion in our network by setting the weights of the connections between the GPi and the thalamus to zero. This removes completely the effect of the basal ganglia in action selection and delivers the control of behavior to the direct connections between the cortex and the thalamus. We tested if these connections alone can still produce a correct selection.

All simulations were done with the same network as in the first part of Section 5.1. In each experiment, we first let the network learn an initial task and then introduce the lesion. Then, we ran 50 more trials and determined the number of correct answers. From 100 networks with different random initial conditions and stimuli, 85 gave always correct answers after lesioning, 14 gave 98% and only 1 gave 96% correct answers. These results indicate that the cortico-thalamic connections are able to maintain an already learned behavior.

5.4. Decision time in conflicting situations

Several experimental studies have shown that the STN may be involved in suppressing premature responding. Baunez et al. (2001) tested rats where they were first trained to associate each of two possible lights with a movement into a particular direction. The rat learned that each time it heard a tone it should choose either left or right depending on the activated light, which was turned on slightly before the sound. Then, in a second set of tests, the two possible lights were both activated for a limited period of time in which the animal should keep his nose in a central hole. Then, one light is turned off and at the same time the tone is produced.

In this second experiment the initial information is not provided to the rat, so it does not know beforehand to which side it should turn when the tone is heard. Now, the rat cannot, immediately after the tone, shift to the correct side, as it was doing before, but must wait for the stimulus to be processed. In order to make a correct decision, the animal has to look at the light before starting to run. To successfully learn this task the animal must suppress or delay the initial learned association between light and the movement, triggered by the tone.

Normal rats successfully learned the task and sufficiently delayed their decision, but in surgically STN impaired rats, both the probability that the animal made a decision before the tone and that a fast and wrong decision was made were increased. Impaired rats tended to quickly move towards a random direction as soon as the tone was heard (exactly as the animal was trained to do in the first experiment), and could not wait enough to process the change in the lights. Sage et al. (2003) suggest that one function of the STN is to suppress premature responses, an effect that was removed with the surgery and that made rats make a decision without the complete information.

Desbonnet et al. (2004) used a similar task to show that deep brain stimulation of the STN can reduce the amount of premature responses in non-lesioned rats. The rat had to insert its nose in a hole and then wait until a tone is produced. The frequency of the tone informed the animal to which side it should move to receive reward. A premature response was produced when rats took out the nose from the hole and moved to one side before hearing the tone. When electrodes were inserted in the rats brain through which the STN was stimulated at different frequencies, the number of premature response decayed linearly with an increase in the frequency of the stimulation As in the previous case, the authors suggest that one function of STN is to delay the response, and that this effect is increased by the stimulation

Both results suggest that the function of the hyperdirect pathway is to modulate the response and delay the decision. This seems to be against the result of our previous numerical experiments where we showed that the STN plays an important role in selecting the correct action. To clarify the difference between the two perspectives we defined a numerical experiment where we measured the reaction time under a different experimental condition.

To recreate the conditions of the experiments in Baunez et al. (2001), we took the same network as in Section 5.1 and first used the same protocol as in Fig. 4 to learn an initial stimulus–action association. This is equivalent to the initial training done by Baunez et al. (2001) where the rats learned to move in a specific direction with each light. Then, we deactivated the learning rules, so that in each of the following trials the connections were kept fixed and the stimulus–response association was not forgotten. After this, we ran 50 new trials but changed the way in which we presented the stimuli: initially we activated both populations of the stimulus cortex instead of just one. This is equivalent to the second part of the task of Baunez et al. (2001) where the two lights are turned on before the tone. A summary of the procedure for this second group of trials (after the initial learning) is shown in Fig. 9.

Fig. 10 compares the reaction time when one stimulus is presented and the conflicting case, when both stimuli are presented. When both stimulus populations are activated, all neurons in the STN start emitting action potentials. Then, the excitatory projections to the GPi enhance the activity of the complete area, increasing thereafter the inhibition over the thalamus. The accumulators then take more time to reach the threshold because of the reduced amount of spikes they are receiving.

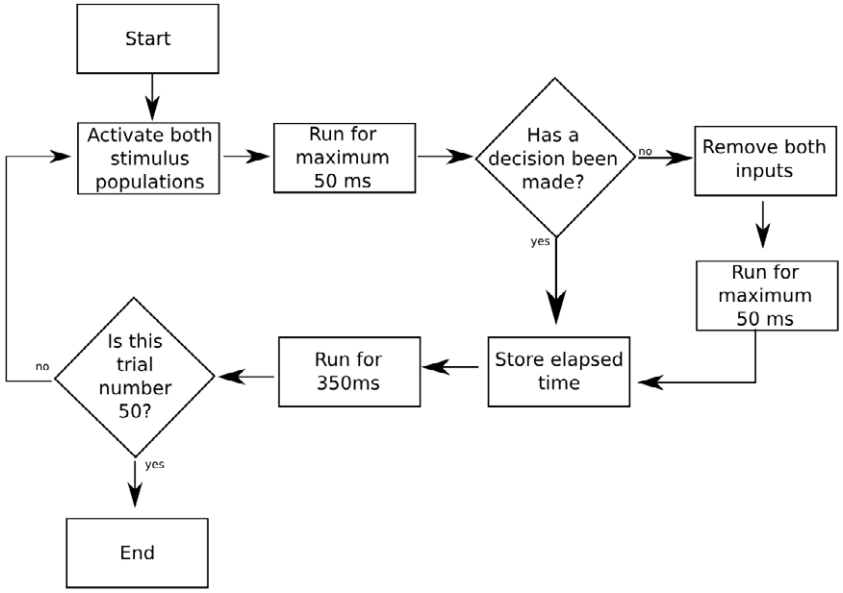

Fig. 9. Simulation protocol for measure the capacity of STN to delay the decision in a conflict situation. This process is applied after an initial association is learned using the protocol of Fig. 4 and the learning is removed (the weights are fixed).

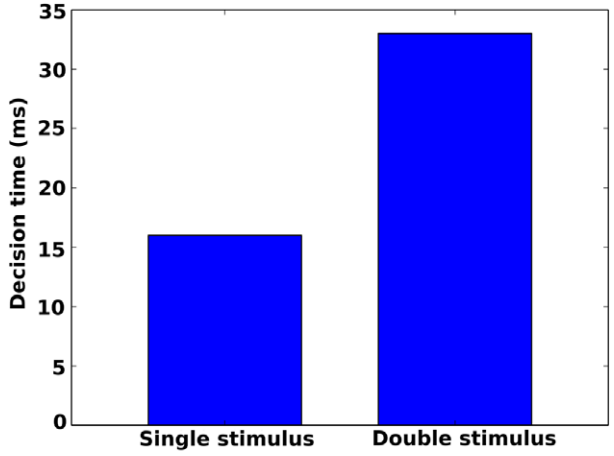

Fig. 10. Decision time for the experiment where two conflicting stimuli are presented. The decision time for the double stimulus is a mean over 50 trials and 50 simulations. The decision time for the single stimulus is the mean of the last 10 trials before the initial association is learned of the same 50 simulations.

6. Discussion

Our spiking neural network proposes new basal ganglia functionalities due to the newly introduced cortico-thalamic pathway. While learning in the direct pathway is still most essential, learning several stimulus response association tasks is explained by the cooperation between the different pathways as demonstrated in several behavioral experiments involving learning. The first one is the effect of pallidotomy in Parkinsonian patients which after surgery are able to maintain previous activities while being impaired learning new ones. This is achieved in the model by the effect of direct cortico thalamic connections which slowly learn the associations acquired by the BG.

Also, the network is capable of explaining a set of experiments where it was shown that the STN delays the response until a correct decision can be made. This effect is produced in our hyperdirect pathway by a complete activation of the STN, which indirectly increases the inhibition in the thalamus, which then holds the decision.

While several pathway functions have been already proposed in our previous rate coded model (Schroll et al., 2014), learning in the spiking neural network depends on STDP and is therefore very different than learning in the rate model.

The main difference between our network and previous spiking models of the basal ganglia is the function of the indirect pathway. Based on the observation of Shen et al. (2008), the learning rule we have used differentiates the effect of dopamine on cells with different receptor types. In neurons expressing type-2 receptor only reduced levels of dopamine produce long term potentiation. This, together with a strong thalamic feedback, activates the indirect pathway only when there are several consecutive wrong decisions. Thus, our model proposes that the indirect pathway learns to inhibit a repetitive mistake or a previously correct action that is still encoded in the cortico-thalamic pathway. This new function of the indirect pathway can be seen as a natural consequence of the additional cortico-thalamic pathway. In the model of Chersi et al. (2013) the main objective of the indirect pathway is to control the activity of the direct pathway by avoiding oscillations and keeping activity at low, normal, levels. A similar approach is used in the model of Stewart et al. (2012) which is a spiking version of the model of Gurney et al. (2001) where the indirect pathway is assumed to just regulate the activity of the GPi.

The hyperdirect pathway in our network also has a different function than in the model of Chersi et al. (2013). In their network the connections to the STN are not plastic and the objective of this area is also to keep the GPi activation within working limits. A similar effect is produced by the STN in the model of Stewart et al. (2012) where the projections do not depend on the channel, i.e., each population of this brain area excites all actions in the GPi. This provides the output nucleus with a global excitation which is proportional to the total activity afferent to the basal ganglia. Instead, in our network, this pathway learns to introduce a reasonable level of surround inhibition by exciting the incorrect actions. The final selection is then performed through a cooperation of both the direct and hyperdirect pathways which is mediated via common feedback signals sent from the thalamus.

Because of the reasons described above, our model provides a different theoretical framework to understand the basal ganglia computations, based on novel pathway functionalities. In our network all pathways are required for action selection and no

external control of activity is required. Behavioral decisions do not only depend on the direct pathway, as in previous similar models, but on a real combination of the output of all areas.

The effect of STN is critical in the probabilistic learning experiment (Section 5.2). This task is difficult for the model because of the fast change between blocks and of the stochastic nature of the reward. The inhibition provided by the hyperdirect pathway reduces the sensibility of the model to the fluctuations created by unrewarded correct or rewarded incorrect trials, hence stabilizing the network. In fact, in the model, removal of STN (ablation), produces a complete random behavior (probability of 0.5) after a few trials.

The inhibition of incorrect actions provided by STN also allows the model to solve a stimulus–action association task even when local competition between populations is almost completely removed (see Fig. 7). In this case, the incorrect striatal population is not always silent and the network requires the hyperdirect pathway to reduce the effect of its activity in the GPi. Additional simulations showed that the network with reduced lateral inhibition is not capable of learning when STN is removed.

The special connectivity we have chosen for the hyperdirect pathway implements the same functionality that was proposed in Schroll et al. (2014), and obtained through Hebbian learning. This topography disagrees with previous theoretical studies in which the STN was proposed to send a global stop signal, increasing the overall levels of activity in the GPi. In both the model of Frank (2006) and Gurney et al. (2001) this behavior is implemented by a full connection between the STN and GPi. Instead, in our model, each population does not excite all actions but only the alternatives from the one it encodes. If the missing connections were added to the network, the hyperdirect pathway would loose its ability to produce the surround inhibition required to impede the execution of alternatives. Instead, due to the learning rule of the cortico–STN synapses, with enough training, the model would just learn to inhibit everything and present no response.

The new function we propose for the STN does not impede its role in action suppression. We have shown in Section 5.4 that a strong activation of STN may increase the reaction time. A global stop signal could be produced by a higher cognitive process and not necessarily at the level of individual decisions. In fact, it has been shown that both the frontal cortex and the STN are strongly involved in stopping an already initiated response (Aron & Poldrack, 2006) and that the activation of the STN is higher in stop trials than go trials (Li, Yan, Sinha, & Lee, 2008).

Our network also agrees with the abstract model proposed by Mink (2003), in which a center–surround inhibition is used to select actions. However, in the model of Mink (2003) the activation of both the indirect and direct pathway is required for action selection, suggesting that the weights to both D1 and D2 cells should increase with reward. In our model the surround inhibition is provided by the hyperdirect pathway, through its direct cortical synapses, and not through the indirect pathway, for which we proposed a new function. For this reason, both D1 and STN cells increase their weights with reward and not D2 neurons.

Although we included a realistic learning rule of direct and indirect pathway inspired by the observations by Shen et al. (2008), our model is still much simplified. In our model synaptic plasticity only occurs when dopamine is above or below baseline. However, the data indicates that dopamine is just one player among others such as A2a and NMDA receptors reported by Shen et al. (2008) and a deviation of dopamine from baseline is not essential for synaptic plasticity to occur. However, as the in vivo baseline level of dopamine is not known the exact degree of LTP in D2 MSNs cannot presently be specified.

A further simplification is that the neuron model and parameters are the same in each layer. Experiments have shown distinctive behavior of cells in different areas of the basal ganglia. Moreover, the connections between the STN and the GPe were not included in our model. This reduction has also been made in other models (O'Reilly & Frank, 2006; Schroll et al., 2014; Stocco, Lebiere, & Anderson, 2010) mainly to keep the pathways segregated and to be able to analyze their function and effect on the output separately. We hypothesize that excitatory connections from the STN to the GPe could be useful to increase the overall activity of this last area and enhance the selection done by the inhibitory connections from the striatum whereas the projections from GPe to STN could be useful to improve competition among populations. The relationship between this two nuclei has been studied as a cause of the abnormal oscillations associated with Parkinson's disease (Kumar, Cardanobile, Rotter, & Aertsen, 2011; Rubin & Terman, 2004) but its function during learning is unknown.

However the present level of abstraction is suitable as the model is still able to successfully learn a set of cognitive tasks and to explain some biological phenomena. This is due to our new approach towards the function of the different cortico-thalamic pathways, which develops through the introduction of realistic learning rules. Through a set of numerical experiments we have shown that this new mechanism for action selection is a powerful tool that in the future could even be used to inspire a new generation of brain inspired cognitive agents or neuro-robotics. Also, our network is well suited for current neuromorphic hardware (see Schemmel et al., 2010 for an example system), a combination that should be exploited in the future for the development of large scale simulations.

Acknowledgments

This work has been supported by the German Research Foundation grant DFG HA2630/8-1 on the German-Japanese Collaboration Computational Neuroscience and the German Research Foundation grant DFG HA2630/7-2 (KFO 247).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.neunet.2015.03.002.

References

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *The Journal of Neuroscience*, *26*, 2424–2433.

Baunez, C., Humby, T., Eagle, D. M., Ryan, L. J., Dunnett, S. B., & Robbins, T. W. (2001). Effects of STN lesions on simple vs choice reaction time tasks in the rat: preserved motor readiness, but impaired response selection. *European Journal of Neuroscience*, *13*, 1609–1616.

Braak, H., & Del Tredici, K. (2008). Cortico-basal ganglia-cortical circuitry in parkinson's disease reconsidered. *Experimental Neurology*, *212*, 226–229.

Brette, R., & Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, *94*, 3637–3642.

Calabresi, P., Picconi, B., Tozzi, A., & Filippo, M. D. (2007). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends in Neuroscience*, *30*, 211–219.

Chersi, F., Mirolli, M., Pezzulo, G., & Baldassarre, G. (2013). A spiking neuron model of the cortico-basal ganglia circuits for goal-directed and habitual action learning. *Neural Networks*, *41*, 212–224.

Czubayko, U., & Plenz, D. (2002). Fast synaptic transmission between striatal spiny projection neurons.. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 15764–15769.

Desbonnet, L., Temel, Y., Visser-Vandewalle, V., Blokland, A., Hornikx, V., & Steinbusch, H. W. (2004). Premature responding following bilateral stimulation of the rat subthalamic nucleus is amplitude and frequency dependent. *Brain Research*, *1008*, 198–204.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. *Journal of Cognitive Neuroscience*, *17*, 51–72.

Frank, M. J. (2006). Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks*, *19*, 1120–1136.

Frémaux, N., Sprekeler, H., & Gerstner, W. (2010). Functional requirements for reward-modulated spike-timing-dependent plasticity. *The Journal of Neuroscience*, *30*, 13326–23337.
Frémaux, N., Sprekeler, H., & Gerstner, W. (2013). Reinforcement learning using a continuous time actor-critic framework with spiking neurons. *PLOS Computational Biology*,.
Goodman, D. F. M. (2009). The brian simulator. *Frontiers in Neuroscience*, *3*, 192–197.
Grillner, S., Hellgren, J., Menard, A., Saitoh, K., & Wikstrom, M. (2005). Mechanisms for selection of basic motor programs roles for the striatum and pallidum. *Trends in Neurosciences*, *28*, 364–370.
Gurney, K., Prescott, T. J., & Redgrave, P. (2001). A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biological cybernetics*, *84*, 401–410.
Haber, S. N. (2003). The primate basal ganglia: parallel and integrative networks. *Journal of Chemical Neuroanatomy*, *26*, 317–330.
Humphries, M. D., Stewart, R. D., & Gurney, K. N. (2006). A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. *Journal of Neuroscience*, *26*, 12921–12942.
Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, *17*, 2443–2452.
Kim, H., Sul, J. H., Huh, N., Lee, D., & Jung, M. W. (2009). Role of striatum in updating values of chosen actions. *Journal of Neuroscience*, *29*, 14701–14712.
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*.
Leh, S. E., Ptito, A., Chakravarty, M. M., & Strafella, A. P. (2007). Fronto-striatal connections in the human brain: A probabilistic diffusion tractography study. *Neuroscience Letters*, *419*, 113–118.
Li, C.-S. R., Yan, P., Sinha, R., & Lee, T.-W. (2008). Subcortical processes of motor response inhibition during a stop signal task. *NeuroImage*, *41*, 1352–1363.
Lozano, A. M., Lang, A. E., Galvez-Jimenez, N., Miyasaki, J., Duff, J., Hutchinson, W. D., & Dostrovsky, J. O. (1995). Effect of gpi pallidotomy on motor function in parkinsons disease. *The Lancet*, *346*, 1383–1387.
Markram, H., Gerstner, W., & Sjöström, P. J. (2011). A history of spike-timing-dependent plasticity. *Frontiers in Synaptic Neuroscience*, *3*.
Mink, J. W. (2003). The basal ganglia and involuntary movements: impaired inhibition of competing motor patterns. *Neurological Review*, 1365–1368.
Morita, K., Morishima, M., Sakai, K., & Kawaguchi, Y. (2012). Reinforcement learning: computing the temporal difference of values via distinct corticostriatal pathways. *Trends in Neuroscience*, *35*, 457–467.
Morrison, A., Diesmann, M., & Gerstner, W. (2008). Phenomenological models of synaptic plasticity based on spike timing. *Biological Cybernetics*, *98*, 459–478.
Nambu, A., Tokuno, H., & Takada, M. (2002). Functional significance of the cortico-subthalamo-pallidal 'hyperdirect' pathway. *Neuroscience Research*, *43*, 111–117.
Naud, R., Marcille, N., Clopath, C., & Gerstner, W. (2008). Firing patterns in the adaptive exponential integrate-and-fire model. *Biological Cybernetics*, *99*, 335–347.
Nisenbaum, E. S., & Wilson, C. J. (1995). Potassium currents responsible for inward and outward rectification in rat neostriatal spiny projection neurons. *The Journal of Neuroscience*, *15*, 4449–4463.
O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*, 283–328.
Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience*, *25*, 563–593.
Plenz, D. (2003). When inhibition goes incognito: feedback interaction between spiny projection neurons in striatal function. *Trends in Neuroscience*, *26*, 436–443.
Ponzi, A., & Wickens, J. R. (2003). Optimal balance of the striatal medium spiny neuron network. *PLOS Computational Biology*, *9*.
Ponzi, A., & Wickens, J. (2012). Input dependent cell assembly dynamics in a model of the striatal medium spiny neuron network.. *Frontiers in Systems Neuroscience*, *6*.
Potjans, W., Morrison, A., & Diesmann, M. (2009). A spiking neural network model of an actor-critic learning agent. *Neural Computation*, *21*, 301–339.
Rubin, J. E., & Terman, D. (2004). High frequency stimulation of the subthalamic nucleus eliminates pathological thalamic rhythmicity in a computational model. *Journal of Computational Neuroscience*, *16*, 211–235.
Sage, J. R., Anagnostaras, S. G., Mitchell, S., Bronstein, J. M., Salles, A. D., Masterman, D., & Knowlton, B. J. (2003). Analysis of probabilistic classification learning in patients with parkinson's disease before and after pallidotomy surgery. *Learning and Memory*, *10*, 226–236.
Schemmel, J., Brüderle, D., Grübl, A., Hock, M., Meier, K., & Millner, S. (2010). A wafer-scale neuromorphic hardware system for large-scale neural modeling. In *Proceedings of the 2010 IEEE international symposium on circuits and systems* (pp. 1947–1950).
Schroll, H., & Hamker, F. H. (2013). Computational models of basal-ganglia pathway functions: focus on functional neuroanatomy. *Frontiers in Systems Neuroscience*, *7*.
Schroll, H., Vitay, J., & Hamker, F. H. (2012). Working memory and response selection: a computational account of interactions among cortico-basalganglio-thalamic loops. *Neural Networks*, *26*, 59–74.
Schroll, H., Vitay, J., & Hamker, F. H. (2014). Dysfunctional and compensatory synaptic plasticity in parkinsons disease. *European Journal of Neuroscience*, *39*, 688–702.
Schultz, W. (2007). Behavioral dopamine signals. *Trends in Neurosciences*, *30*, 203–210.
Schultz, W. (2010). Dopamine signals for reward value and risk: basic and recent data. *Behavioral Brain Function*, *6*.
Shen, W., Flajolet, M., Greengard, P., & Surmeier, D. J. (2008). Dichotomous dopaminergic control of striatal synaptic plasticity. *Science*, *321*, 848–851.
Smith, Y., Raju, D. V., Pare, J.-F., & Sidibe, M. (2004). The thalamostriatal system: a highly specific network of the basal ganglia circuitry. *Trends in Neurosciences*, *27*, 520–527.
Stanford, T. R., Shankar, S., Massoglia, D. P., Costello, M. G., & Salinas, E. (2010). Perceptual decision making in less than 30 ms. *Nature Neuroscience*, *13*, 379–386.
Stewart, T. C., Bekolay, T., & Eliasmith, C. (2012). Learning to select actions with spiking neurons in the basal ganglia. *Frontiers in Neuroscience*, *6*.
Stocco, A., Lebiere, C., & Anderson, J. R. (2010). Conditional routing of information to the cortex: a model of the basal ganglias role in cognitive coordination.. *Psychological Review*, *117*, 541–574.
Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. MIT Press.
Temel, Y., Blokland, A., Steinbusch, H. W., & Visser-Vandewalle, V. (2005). The functional role of the subthalamic nucleus in cognitive and limbic circuits. *Progress in Neurobiology*, *76*, 393–413.
Tepper, J. M., & Bolam, J. P. (2004). Functional diversity and specificity of neostriatal interneurons. *Current Opinion in Neurobiology*, *14*, 685–692.
Tunstall, M. J., Oorschot, D. E., Kean, A., & Wickens, J. R. (2002). Inhibitory interactions between spiny projection neurons in the rat striatum. *Journal of Neurophysiology*, *88*, 1263–1269.
Wickens, J. R., Reynolds, J. N., & Hyland, B. I. (2003). Neural mechanisms of reward-related motor learning. *Current Opinion in Neurobiology*, *13*, 685–690.
Wiesendanger, E., Clarke, S., Kraftsik, R., & Tardif, E. (2004). Topography of cortico-striatal connections in man: anatomical evidence for parallel organization. *European Journal of Neuroscience*, *20*, 1915–1922.