Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 71 (2008) 2782–2793
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Differences in prefrontal and motor structures learning dynamics depend on task complexity: A neural network model

S.E. Lew[a,b,*], H.G. Rey[a,b], D.A. Gutnisky[c], B.S. Zanutto[a,b]

[a]*Instituto de Ingeniería Biomédica, Facultad de Ingeniería, Universidad de Buenos Aires, Paseo Colon 850, C1063ACV Capital Federal, Buenos Aires, Argentina*
[b]*IBYME-CONICET, Buenos Aires, Argentina*
[c]*Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, Houston, TX 77225, USA*

Received 8 May 2006; received in revised form 6 September 2007; accepted 9 September 2007
Communicated by T. Heskes
Available online 22 October 2007

Abstract

Neurons in the basal ganglia (BG) of monkeys learning a simple visual discrimination (VD) task show faster changes in activity than those in the prefrontal cortex (PFC). This motivated the hypothesis that changes in the BG activity can "lead" those in the PFC. Given that the PFC is a key player in the learning of complex tasks, we tested the former hypothesis by using a neural network model that learns simple and complex contingencies as VD and delayed matching to sample (DMTS) tasks. Even though the model accounted for the results in the VD task no such "lead" was observed in the DMTS task. We propose that when the task requires learning high-order contingencies, such as in the DMTS case, motor structures quickly select the subset of responses allowing the subject to obtain reward, but learning in the cortico-BG loop progresses in a concurrent way in order to maximize reward.

Keywords: Prefrontal cortex; Premotor cortex; Basal ganglia; Visuomotor learning; Dopamine; Computational neuroscience

1. Introduction

A general agreement concerning the role of the prefrontal cortex (PFC) has been stated in the past two decades. The PFC has the ability to orchestrate and drive information from and to other less cognitive structures [26,45]. It can actively maintain task relevant stimuli and internal goals [40]. It has a strategic neuroanatomical position in the primate brain [12]. All these facts and the behavioral deficits produced by its partial or total damage [34] sustain the hypothesis of the PFC being involved in the learning of behaviors where simple input–output mapping are not sufficient. On the other hand, simple behaviors, as avoiding aversive stimuli or learning to respond in the presence of visual cues in order to get appetitive reward can be achieved by animals whose PFC is not developed enough [47] and also in prefrontal damaged primates unable to solve more complex rules [34]. Computational models have been widely used to analyze different roles of the PFC, ranging from working memory processes [7–9] to the interaction of different structures and neurotransmitters conveying to the PFC [14,29]. In this work we developed a neurobiological plausible neural network model to test recent interpretation of some neurophysiological experiments in awake behaving monkeys [33].

In previous works, we developed a neural network model that learns different tasks as the matching law, partial reinforcement extinction, blocking, learned helplessness, response selection, successive negative contrast effect, modulation of the avoidance response, transfer of control between conditioned stimuli and spontaneous recovery [24], as well as all the experiments explained by the model of Schmajuk and Zanutto [41]. Further refinement was required in order to acquire other tasks like matching to

*Corresponding author. Instituto de Ingeniería Biomédica, Facultad de Ingeniería, Universidad de Buenos Aires, Paseo Colon 850, C1063ACV Capital Federal, Buenos Aires, Argentina. Tel.: +54 11 43430891; fax: +54 11 43432775.
E-mail address: slew@fi.uba.ar (S.E. Lew).

0925-2312/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2007.09.010

sample and perceptual categorization, specially regarding the role of lateral PFC (lPFC) [22]. This is in agreement with the ideas presented in O'Reilly et al. [29]. The authors suggested that mPFC encodes specific featural information while lPFC is engaged in the representation of featural dimension. Here, we further hypothesize that lPFC is the key element for encoding higher-order contingencies and solve more complex tasks.

Despite the general agreement concerning the importance of the PFC in learning high-order contingencies, recently, strong experimental evidence has shown that changes in the basal ganglia (BG) activity precede those in the PFC during conditional visuomotor learning [33] (and references therein). This evidence induced some authors to hypothesize that anticipation in the BG activity can be a "training" signal for the PFC. In order to explore the "training" hypothesis, and to evaluate whether this hypothesis is paradigm dependent (i.e. it depends on the task complexity), we tested it in a refined version of our previous model [24]. The model takes into consideration: (a) the effects of dopamine (DA) on neuron excitability and synaptic changes in the PFC and also in cortical and subcortical motor structures such as the premotor cortex and the basal ganglia (BG–PMC), (b) the effects of norepinephrine (NE) modulating the energy flow of visual information conveying to the BG–PMC, (c) the neuronal mechanisms of competition and lateral inhibition in deep layers of the PFC and the BG–PMC.

We run simulations of our neural network model in two different paradigms. When our model was trained in a visual discrimination (VD) task we obtained similar results as in [33]. In contrast, when the model was trained in a delayed matching to sample (DMTS) paradigm (a more cognitively demanding task), the model predicted that there is no leading between the BG–PMC and the PFC, thus, contradicting the interpretation of [33].

2. Experimental paradigms

Subjects performing a VD task associate two different cues (conditioned stimuli CS^1 and CS^2) with saccadic movements in opposite directions (R^1 and R^2). A conditioned stimulus (CS^i) is presented for 500 ms followed by a delay period of 1 s [1]. Then the subject is allowed to respond and, if it is correct, an appetitive unconditioned stimulus (US) is delivered 2 s after the cue onset.

In the DMTS task, the subject has to release (R^1 = Go) or hold (R^2 = No–Go) a lever depending on the matching or non-matching of two previous sequentially presented stimuli: the sample and the comparison [36]. After presentation of the sample stimulus for 500 ms, a delay lasting 1 s is interposed and then the comparison stimulus is shown. When the matching rule is rewarded, if the sample and comparison stimuli match/non-match, the subject must release/hold the lever to obtain reward, which is delivered 2.4 s after the onset of the sample stimulus.

In Fig. 1 temporal sequences of stimuli presentation, response execution, and reward delivery in the VD and the DMTS tasks are shown.

3. Model

The activity of each neuron in the model represents the activity of a certain functional cluster of neurons. The time is discretized in steps representing 100 ms each. As shown in Fig. 2, the input layer is constituted by a set of cue selective neurons. Each time the CS^i is present, $CS^i = 1$, otherwise $CS^i = 0$. Each time the reward is delivered, $US_t = 1$ for the following 10 time steps, otherwise $US_t = 0$. In behaving monkeys, short-term memory activity (STM) is observed in later stages of the visual ventral pathway, that is, in V4 and inferotemporal cortices [11,21], as well as in orbitofrontal cortex, where reward-related stimuli are processed [10]. As neuroanatomical data reveals [6], those STMs would be inputs to the BG–PMC and the PFC. The STMs of those stimuli are computed in the input layer of the model as

$$\tau_t(S^i) = (1-\alpha)\tau_{t-1}(S^i) + \alpha S_t^i \quad \text{if} \quad S_t^i > 0,$$
$$\tau_t(S^i) = (1-\beta)\tau_{t-1}(S^i) \quad \text{if} \quad S_t^i = 0, \tag{1}$$

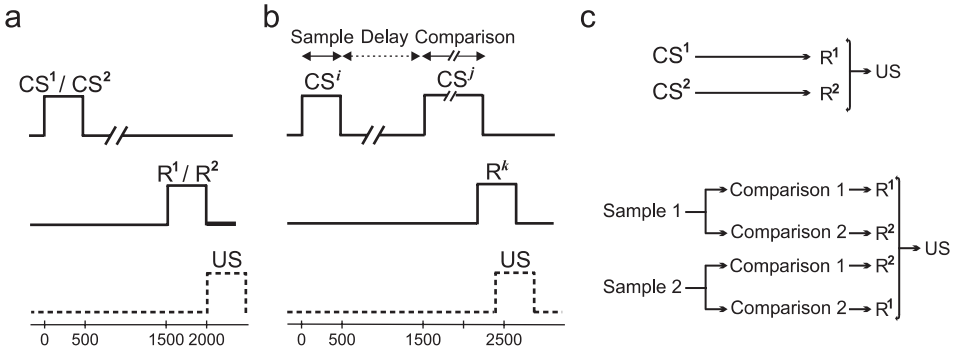

Fig. 1. Schematic diagram of the behavioral tasks. In both cases the delay period can be set to zero. (a) VD task. R^1 and R^2 represent saccadic movements in opposite directions. (b) DMTS task. R^1 and R^2 represent the Go and No–Go responses, respectively. The proper response to be executed depends on which stimuli were presented as sample and comparison. (c) Reinforcement contingencies for the VD task (top) and the DMTS task (bottom). For the non-matching rule, the response index must be exchanged.

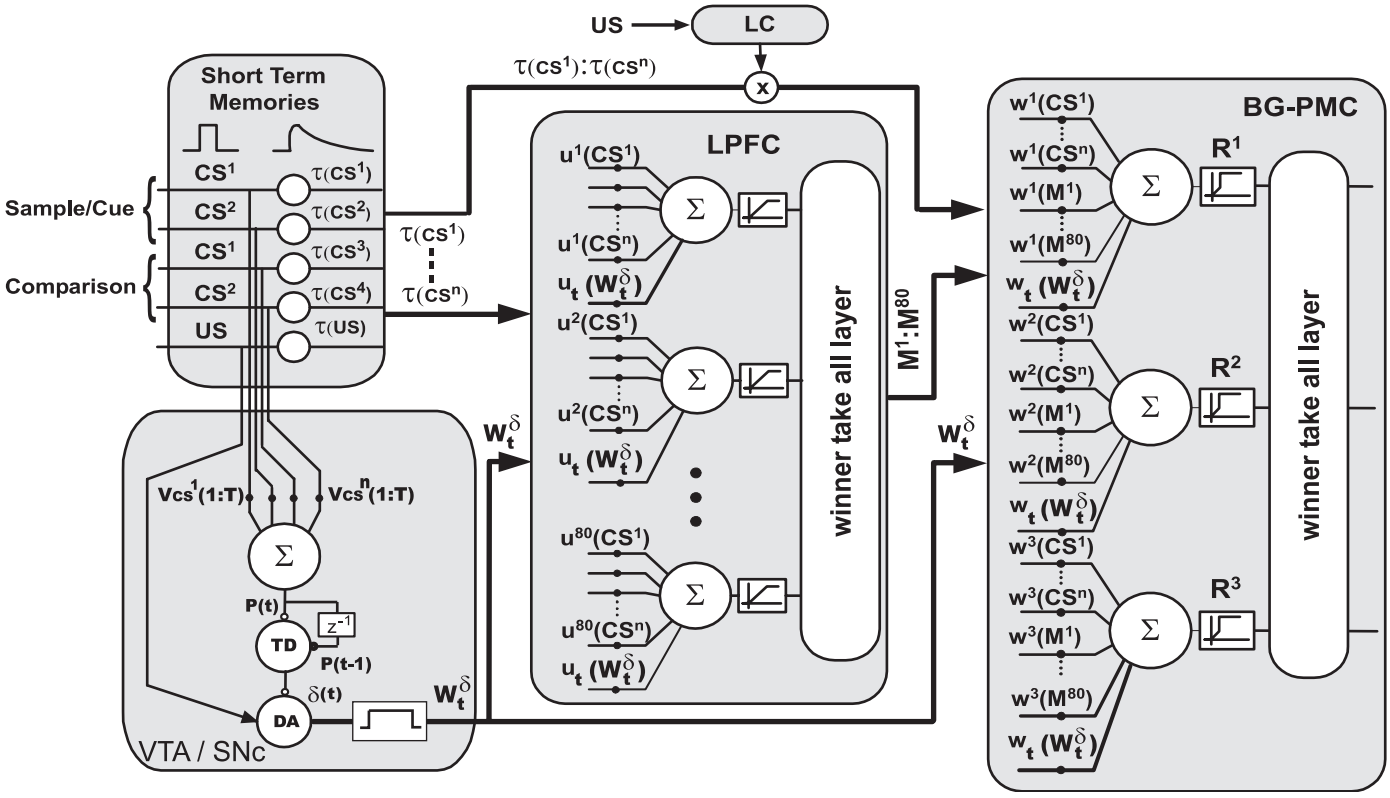

Fig. 2. Scheme of the neural network model. The first layer generates short-term memories of the stimuli as a result of the interaction between different structures such as ventromedial PFC, inferotemporal ctx., posterior parietal ctx., hippocampus and amygdala. We used 80 neurons in the lPFC, three in the BG–PMC and a TD(λ_{TD}) model in the VTA/SNc. The LC represents a modulation exerted by the locus coeruleus. In Section 4 there is a qualitative description on how the model works. The parameters used during the simulations are: $\alpha = 0.45$; $\beta = 0.02$; $\lambda_{TD} = 0.02$; $\lambda_{TD} = 0.9$; $\theta_{hebb} = 0.1$; $\theta_{anti\text{-}hebb} = -0.05$; $u_t(P) = w_t(P) = -0.05$; $\theta_{BG\text{–}PMC} = 0.4$; $basal_{PFC} = basal_{BG\text{–}PMC} = 0.2$; $\alpha_{lc} = 0.003$; $\mu_{PFC} = 0.99875$; $v_{PFC} = 0.0025$; $\mu_{BG\text{–}PMC} = 0.99975$; $v_{BG\text{–}PMC} = 0.00125$; $B_{winner} = 0.14$.

where S^i represents the *CS* or *US* stimuli, t is a time step and τ_t represent the stimuli traces.

Considering that activity changes in the BG and the PMC occur at the same time-course during VD learning [5], we model the BG and the PMC as a single layer of neurons.

If throughout a trial the activity of the BG–PMC neurons does not exceed the activation threshold, a random response is executed with probability 1/3. Random responses are executed 400 or 600 ms after the end of the delay period for the VD and DMTS tasks, respectively. When a response is executed, the activity of its associated neuron is set to 1 along a period of 5 time steps, while the others are forced to 0 along the same time period.

In a VD task these responses represent saccadic movements to the right (R1), left (R2) and to any other non-rewarded direction (R3). In the DMTS task these responses represent a Go (R1), a No–Go (R2) and any other response non-related with the task (R3). All of them are codified at the motor-related structures layer.

Along different phylogenetic species such as rats and primates DA neurons have been shown to respond to unpredicted rewards [32,42]. Moreover, after repeated paired presentation of CS–US, neurons in midbrain dopaminergic structures as the ventro tegmental area (VTA) and the sustantia nigra pars compacta (SNc), change their firing pattern codifying the prediction error of being rewarded. Time difference models (TD) [28,32,44] predict the firing of DA neurons for different paradigms (classical and operant) employing one or multiple conditioned stimuli.

In Fig. 2, the VTA/SNc block is a TD model whose inputs are the CS's and the US. The reward predicted by each conditioned stimulus at time step t is calculated as follows:

$$P_{CS^i}(t) = V_{CS^i} \bullet \; x_{CS^i}(t), \tag{2}$$

where $x_{CS^i}(t)$ is a vector whose t component (at time t) is set to 1 if the CS^i onset takes place at time t or before, and 0 elsewhere. Its dimension is equal to the number of time steps in a trial. Each $x_{CS^i}(t)$ has an associated weight vector V_{CS^i}. Based on the reward and its predictions for each CS^i, $\delta(t)$, the prediction error at time step t is computed as

$$\delta(t) = r(t) + \sum_{\forall CS} (P_{CS^i}(t+1) - P_{CS^i}(t)), \tag{3}$$

where $r(t) = US_t$ at the onset of the reward and zero elsewhere. Because the low basal firing rate of the DA neurons, positive prediction errors produce an increment in activity of about 270% over basal levels, while negative

ones only decrease the firing rate 55% below baseline [10]. Because of this, in the model, negative prediction errors are scaled by a factor of 1/6. At each time step, the synaptic weight vector V_{CS^i} is updated according to the values of $\delta(t)$ and $e_{CS^i}(t)$, the eligibility trace for the conditioned stimulus CS^i

$$\Delta V_{CS^i} = \alpha_{\mathrm{TD}}\delta(t)e_{CS^i}(t), \tag{4}$$

where α_{TD} is the learning rate parameter and the vector $e_{CS^i}(t)$ is computed as follows:

$$e_{CS^i}(t) = \lambda_{\mathrm{TD}}e_{CS^i}(t-1) + x_{CS^i}(t). \tag{5}$$

The TD(λ_{TD}) used here account for the experimental results shown in [32], where a sequence of two conditioned stimuli precede the reward delivery.

Although the TD model reproduces the phasic firing pattern of dopaminergic neurons, postsynaptic effects of DA on the PFC and motor-related structures persist for longer periods of time [13,15]. In our model, we introduce the variable W_t^δ as a gating window for the DA postsynaptics effects. When DA bursts occur, if $\delta(t) > \theta_{\mathrm{hebb}}$, $W_t^\delta = 1$ for the following T steps. The duration of this window (T) depends on the amplitude of $\delta(t)$, in accordance with the experimental results obtained in [13,15]. We propose to choose it according to

$$T = \min(50\delta(t) + 10, 35). \tag{6}$$

When the predicted reward is omitted DA firing goes below baseline. If $\delta(t) < \theta_{\text{anti-hebb}}$, $W_t^\delta = 0$ for the following 15 time steps. When the DA firing is close to baseline, i.e. $\theta_{\text{anti-hebb}} < \delta(t) < \theta_{\mathrm{hebb}}$, $W_t^\delta = 0.5$.

VTA stimulation decreases the spontaneous firing of PFC pyramidal neurons, mainly by exciting interneurons [19]. In our model, such inhibition is represented by clamped negative synaptic weight $u_t(W_t^\delta)$ from the VTA to the PFC. However, due to the synergism between NMDA and D1 receptors [48], we postulate that initially inhibited PFC pyramidal neurons will fire strongly when afferent inputs release large amounts of glutamate. This activated cluster will then inhibit the other clusters [9]. To model this effect, we apply a winner-take-all mechanism at PFC output [22]. The following equations show the calculation of the PFC neurons outputs:

$$O_t^k = \sum_{\forall CS^i} u_t^k(CS^i)\tau_t(CS^i) + u_t(W_t^\delta)W_t^\delta + B_{\mathrm{winner}}W_t^\delta + basal_{PFC} \quad \text{if } O_t^k > 0; \text{ else } O_t^k = 0, \tag{7}$$

$$M_t^k = \begin{cases} O_t^k & \text{if } k = k^*, \\ 0 & \text{otherwise}, \end{cases} \tag{8}$$

where $k^* = \arg\max_k O_t^k$ represents the index of the winner neuron, B_{winner} stands for the synergism between D1 DA receptors and NMDA receptors, and $basal_{PFC}$ is the baseline firing rate of PFC neurons.

It has been hypothesized that DA modulates the excitability of striatal neurons allowing the BG to inhibit competent programmes and to release the correct one [27]. As in the PFC, in our model the released DA inhibits the motor area through clamped negative synaptic weight $w_t(W_t^\delta)$, and, in contrast to this general inhibition, the winner neuron is excited proportionally to the released DA. The effect of this mechanism is to apply a "brake" over all possible motor programmes and to release the one whose activity surpasses a fixed threshold. The output of the response neurons is computed as

$$R_t^j = \sum_{\forall CS^i} w_t^j(CS^i)\tau_t(CS^i)lc_t + \sum_{\forall M^k} w_t^j(M^k)M_t^k + w_t(W_t^\delta)W_t^\delta + B_{\mathrm{winner}}W_t^\delta + basal_{BG-PMC}, \tag{9}$$

where $basal_{PFC}$ is the baseline firing rate of BG–PMC neurons and lc_t represents a modulation exerted by noradrenergic neurons of the locus coeruleus (LC) over visual and somatosensory cortical neurons. Effects of NE on the modulation of glutamate-evoked responses have been proved to have an inverted U shape [4], that is, low and high doses of NE produce a decrease on neuron excitability, while medium doses increase their excitability. In behaving monkeys, tonic firing of LC neurons shows a defined correlation with performance [46]. Tonic frequencies of 2–3 Hz are associated with good performance periods while frequencies >3 Hz are related to periods where erratic performance and distractibility are observed. This gives a hint of the function of the noradrenergic system in the regulation of exploratory behavior [2]. We model the tonic firing of LC neurons as a function of the received reward in a time window that includes many trials:

$$\tau_t^{\mathrm{long}}(US) = (1-\alpha_{lc})\tau_{t-1}^{\mathrm{long}}(US) + \alpha_{lc}US_t, \quad lc_t = 1 - 5\tau_t^{\mathrm{long}}(US). \tag{10}$$

Short-term memories for the response neurons are computed according to

$$\tau_t(R^j) = (1-\alpha)\tau_{t-1}(R^j) + \alpha R_t^j \tag{11}$$

and as in (8) for the PFC area, a winner-take-all rule is applied.

In addition to excitability, DA effects on PFC pyramidal neurons are also related to modifications of synaptic efficacy via LTP and LTD [30,38]. For this reason, previous models have used the DA signal in the modulation of synaptic weights modifications [24,29]. When $W_t^\delta = 1$, Hebbian learning is applied to both PFC and BG–PMC neurons. The opposite occurs when $W_t^\delta = 0$. No modifications in synaptic weights take place when VTA/SNc firing is close to baseline. Thus, when $W_t^\delta = 1$ or 0, the PFC winner neuron k^* and BG–PMC neurons update their synaptic weights, which belong to (0,1), according to

$$u_t^k(CS^i) = \mu_{PFC}u_{t-1}^k(CS^i) - (-1)^{W_t^\delta}\nu_{PFC}\tau_t(CS^i)O_t^k \quad \text{if } k = k^*, \tag{12}$$

$$w_t^j(CS^i) = \mu_{BG-PMC} w_{t-1}^j(CS^i) - (-1)^{W_t^\delta} v_{BG-PMC} \tau_t(CS^i) \tau_t(R^j) lc_t,$$
$$w_t^j(M^k) = \mu_{BG-PMC} w_{t-1}^j(M^k) - (-1)^{W_t^\delta} v_{BG-PMC} M_t^k \tau_t(R^j). \quad (13)$$

In the previous equations, μ_{PFC} and μ_{BG-PMC} are first-order momentum constants while v_{PFC} and v_{GB-PMC} are learning rates for the PFC and BG–PMC, respectively.

4. How the model works

The input to the model is composed by different stimuli from the environment (CSs and US). The first layer of the model generates short-term traces for those stimuli, computed as shown in (1). These short-term memories maintain information about stimuli during the delay period, decaying exponentially after their offset. Biologically, this is not the result of the information processing by a single structure. Actually, it is the result of the interaction between sensory cortices, associative cortices (mPFC, ITC) and subcortical structures (amygdala, hippocampus).

The second layer also involves the interaction of many structures. The lPFC conveys information from mPFC and also from ITC, cingulate and parietal cortex. It allows further filtering of task relevant stimuli (by leading to highly selective lPFC clusters that encode the S–R mappings).

However, it is a key component for learning the DMTS task, as joint representations of the sample and comparison will be generated in this layer. These clusters will then be associated to the proper response (in the third layer) according to the contingencies of the task. Thus, although mPFC and lPFC are engaged in working memory processes, they represent different kinds of information [29]. Moreover, we share the approach of Atallah et al. [3] as the second layer selects the initially most activated clusters in order to establish the mappings, while the remaining clusters are inhibited (winner-take-all rule).

Responses executed first randomly, simulating a motivational state generated by deprivation, proceed after acquisition of S–R mappings represented in the synaptic weights of the model. Reinforcement information reaches the VTA/SNc and LC, mainly from frontal structures related to its processing, as the mPFC. However, both the VTA/SNc and the LC process such information in a different way in order to produce different patterns of neural responses and/or learning mechanisms. Once information about reinforcement reaches the VTA/SNc, DA is released, and those conditioned stimuli traces active at that time can be associated with reinforcement. As learning proceeds, the probability of being reinforced increases and the synaptic weights $V_{CS}{}^i$ in VTA/SNc will represent the association between conditioned stimuli and reinforcement. Consequently, every time that a conditioned stimulus is presented to the model, the VTA/SNc will fire strongly, releasing DA over the lPFC and BG–PMC structures. Effects of DA on these structures are represented by W_t^δ. Hebbian learning in the lPFC and BG–PMC neurons takes place whenever W_t^δ is 1, increasing the associative strength between stimuli/lPFC and stimuli/BG–PMC neurons. The opposite is applied if W_t^δ is 0 (anti-Hebbian learning). This mechanism allows the model to learn paradigms based on reinforcement and to build S–R mappings that, unless the task contingencies change, remain over time.

Related to the LC processing of reinforcement, we propose it as a mechanism that allows the model to react to environment or schedule changes in order to find best strategies related with plausible behavioral responses (exploration strategies). We propose that tonic levels of NE are a function of a long-term processing of reinforcement. Due to the effect of NE over the excitability of visual and dopaminergic neurons, the action of LC activation modulates the energy flow between layer 1 to the response layer.

Despite the fact that in three-term contingencies, as in the VD task case, direct S–R mappings between input and output neurons are enough to learn the paradigms, high-order contingencies need additional resources to be learned. A previous version of this model that does not include the lPFC structure (layer 2 in the model) can explain experiments such as the matching law, partial reinforcement extinction, blocking, learned helplessness, response selection, successive negative contrast effect, modulation of the avoidance response, transfer of control between conditioned stimuli and spontaneous recovery [17,18,24,41]. However, when tested in DMTS paradigm, its performance is close to chance level. Contrary to the VD task and those previously mentioned experiments, input patterns in the DMTS paradigm can be proved to be linearly non-separable [39], that is, there is not a plane in the input space that separates those input patterns associated to the Go response from those associated to the No–Go response. Under this condition, the DMTS paradigm cannot be learned using a unique layer of linear threshold neurons. Although only four lPFC neurons are sufficient to map the rules involved in a DMTS task, more neurons give more variability to the network and consequently rules are learned faster. We used 80 neurons in the lPFC structure, a number that allows to learn the task at least in 80% of the simulations, reaching an averaged performance of 90% in approximately 300 trials.

5. Model parameters

Although the model works properly over a wide subspace of the parameter space, the parameters used in the simulations were chosen in order to explain behavioral and physiological results from classical and operant paradigms. Simple operant learning allows a wide combination of parameter values with model performance over criterion, while in DMTS paradigm, the parameters range

must be restricted to a subset of the ones for the VD task in order to predict behavioural and physiological predictions.

We adjusted the learning rate of the TD model to predict future rewards in a range of conditioning trials varying from 30 to 100 and to exhibit DA bursting at the comparison presentation time even at trial 400 [32].

Once these parameters were adjusted, learning rates of neurons in the lPFC (μ_{PFC} and v_{PFC}) and in the BG–PMC (μ_{BG-PMC} and v_{BG-PMC}) structures were found in agreement with behavioral results of monkeys performing operant conditioning in response selection, VD and in delayed matching to sample experiments [33,43] reaching criterion in approximately 50, 100–200 and 150–250 trials, respectively.

The dynamics of the LC tonic firing were chosen to show high values when performance is erratic and low values when the correct rule is exploited.

Additional parameters as the dopaminergic inhibition over pyramidal neurons ($u_t(W_t^\delta)$ and $w_t(W_t^\delta)$), the synergetic effect between DA and glutamate (B_{winner}) and the baseline firing rates ($basal_{PFC}$ and $basal_{BG-PMC}$) were explored to predict results related to neuron selectivity and their correlation to behavioral performance in a VD task [23,33].

6. Simulations and results

The duration of each trial is 50 time steps (5 s). Each experiment is conducted during 1000 trials. All the simulations are the result of the ensemble average of 100 independent experiments. The estimated performance (which will be correlated with the time to half-maximum selectivity (HTMS)) at each correct trial was estimated by averaging the results of previous responses within a sliding window of 16 trials length.

6.1. Visual discrimination and DMTS tasks

In the VD task the subject must learn the association between the cue stimuli CS^1 and CS^2 with the responses R^1 and R^2, respectively [1]. A 100% of the simulations learned with more than 90% performance, reaching an averaged performance of 100% in 300 trials on average. Fig. 3 shows the average performance, NE level and reaction times. As expected, the NE level and reaction times are large when the performance is near chance, and they decrease as the performance is improved.

In the DMTS paradigm, the subject must learn to match (or non-match) a comparison stimulus with a previously presented sample stimulus. A certain stimulus is codified by different neurons at the input layer depending on its presentation at the sample or comparison periods (see Fig. 2). From the three possible responses, two responses are related to reward, i.e. to release (R^1) or to hold (R^2) the lever. Therefore, a total of four rules are learned using stimuli CS^1 and CS^2 (see Fig. 1c).

Eighty percent of the simulations learned with more than 90% performance, reaching an averaged performance

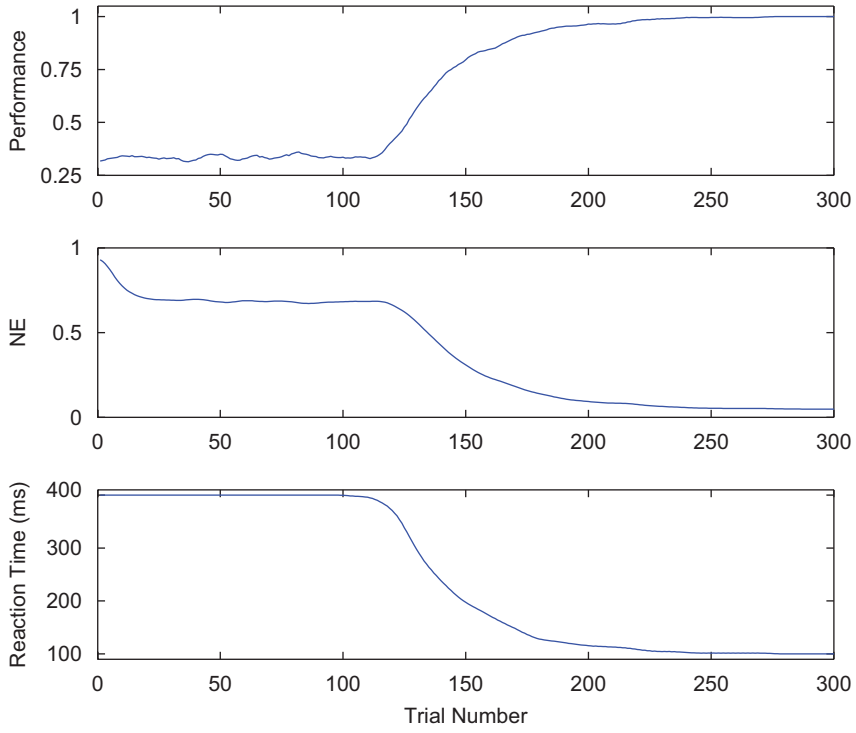

Fig. 3. For the VD task, averaged values from 100 out of 100 (100%) simulations that reach a performance of 90% were computed. Performance acquisition, NE (lc_t is calculated using (10)) level and reaction times (number of steps to the response execution from the end of the delay period, multiplied by 100) are shown. The curves are smoothed using a Gaussian window with a span of 10 samples.

of 90% in approximately 300 trials on average. Fig. 4 shows the average performance, NE level and reaction times for the DMTS task. As in the VD task, the NE level and reaction times are negatively correlated with the performance.

In Fig. 5, the DA firing for both tasks is shown for a single run of the model. Three rewarded trials were chosen for each task in three different stages of the learning process. At the beginning of training, non-predicted rewards elicit a phasic DA response at the time when

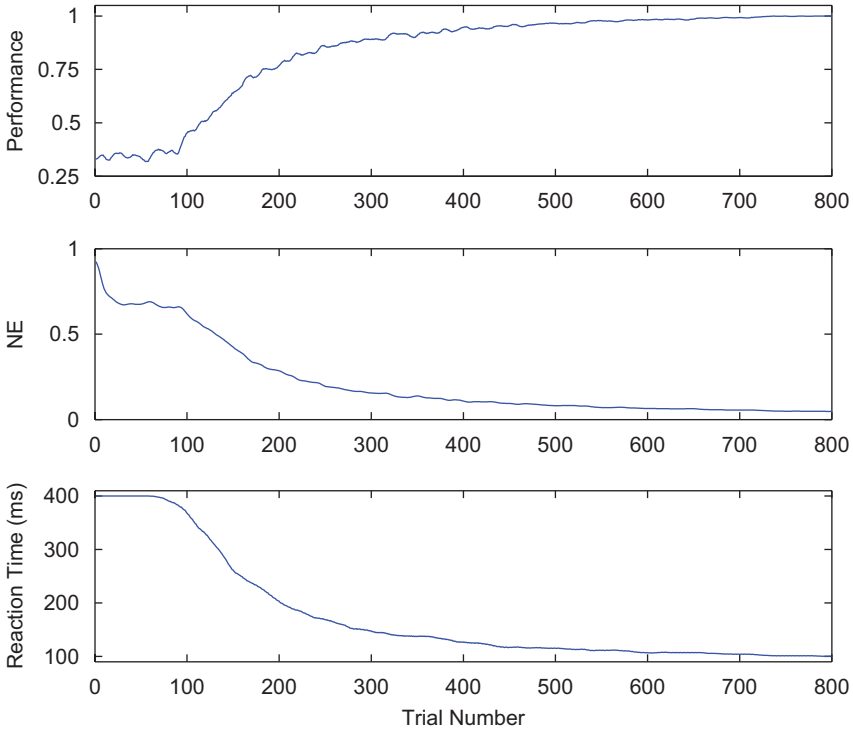

Fig. 4. Averaged values of performance acquisition, NE (lc_t) level and reaction times from 100 out of 125 (80%) DMTS experiments that reach a performance of 90% are shown. The curves are computed as in Fig. 3.

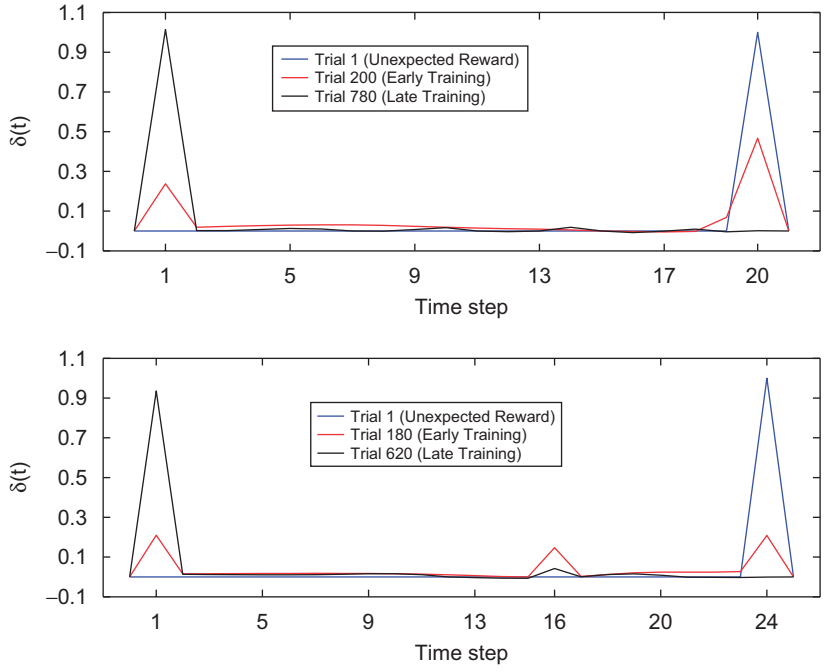

Fig. 5. Prediction error ($\delta(t)$) in three different rewarded trials for both tasks (VD task, top panel; DMTS task, bottom panel). The curves are taken from a single experiment (that reach criterion) of each task. The three trials are chosen at different stages of the learning process, i.e. unpredicted reward, early training and late training. During the remaining of each of the trials shown, the prediction error is 0 (not shown).

reward is delivered. Early training DA responses show phasic activity at both cues and reward onset. In late training stages, DA neurons show activation at the CS's onset, while no activation is observed when the predicted reward is delivered.

6.2. Analysis of neuron selectivity index (SI) for VD and DMTS tasks

Motivated by the results shown in [33] we analyze the changes in the SI of PFC and BG–PMC neurons for both tasks.

For the VD task, in each experiment, the most selective PFC neuron for stimulus CS^1 was found. As in [33], only correct trials were included in the analysis. This means that the correct response for CS^1 is R^1. Then the SI across correct trials for the most selective PFC neuron and for the one in the BG–PMC (associated with R^1) were computed as

$$SI_t(PFC) = \frac{O_t^k(CS^1) - O_t^k(CS^2)}{O_t^k(CS^1)}, \tag{14}$$

$$SI_t(BG-PMC) = \frac{R_t^j(CS^1) - R_t^j(CS^2)}{R_t^j(CS^1)}, \tag{15}$$

where $O_t^k(CS^1)$ and $R_t^j(CS^1)$ are the activities of neurons in the lPFC and BG–PMC structures, respectively, when stimulus CS^1 is presented. If the SI was negative on a certain step, it was set to 0 (a negative SI represents that at that time the neuron was more selective to the other stimulus/response). The average values of PFC and BG–PMC neuron SI are shown in Fig. 6. At the beginning of training, neurons in both, the PFC and BG–PMC structures show no preference or selectivity for any particular stimulus. As learning progresses, the selectivity increases earlier in the trial and neurons in the PFC and the BG–PMC respond strongly for those stimuli they are selective. Even though the selectivity behaves in a similar way in both, the PFC and the BG–PMC, faster changes can be observed in the BG–PMC (Fig. 6b). Around trial 20, the SI surpass the half of its maximum value near-time step 2 (200 ms), whereas the same behavior can be observed around trial 30 in the PFC (Fig. 6a).

For the averaged results, the THMS was computed for each correct trail and then correlated with the estimated mean performance (Fig. 7). The resulting values were −0.95 and −0.78 for PFC and BG—PMC, respectively, which are in agreement with the experimental results for the same task [33].

We repeated the same procedure for the DMTS task. In this case we chose one of the four possible rules and calculate the SI across correct trials as

$$SI_t(PFC) = \frac{3O_t^k(rule^1) - \sum_{i=2}^{4} O_t^k(rule^i)}{3O_t^k(rule^1)}, \tag{16}$$

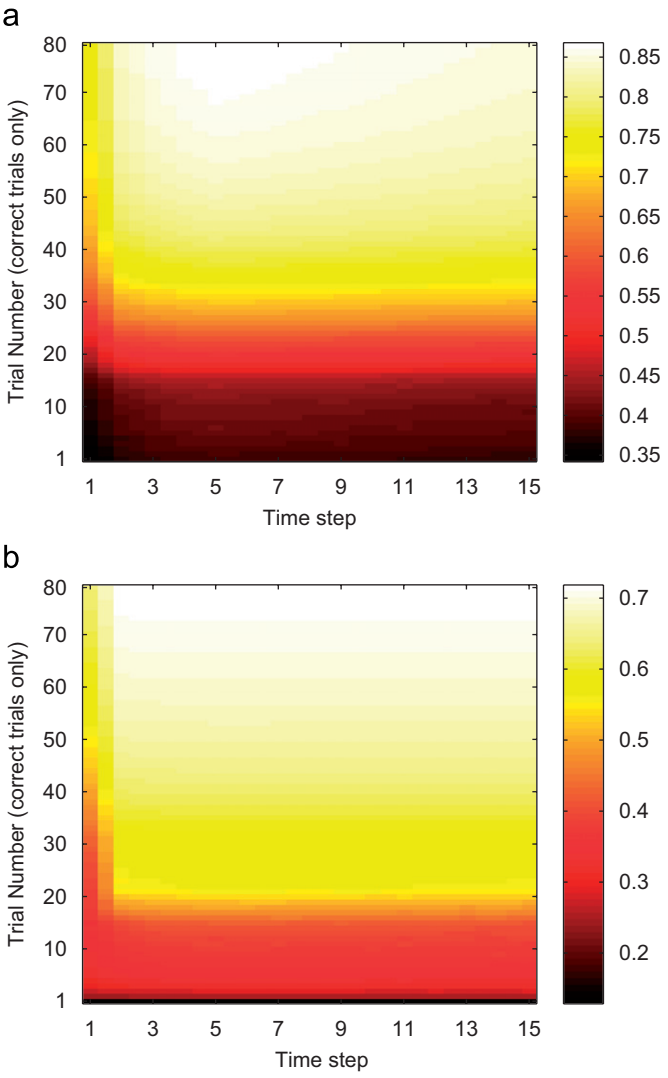

Fig. 6. VD task. (a) Averaged SI for the most selective PFC neuron to stimulus CS^1 (computed using (14)). (b) Averaged SI for the BG–PMC neuron associated with R^1 (computed using (15)). The X-axis is showing the time step during each trial, whereas the Y-axis shows the correct trial number during the learning process for the rule under analysis. The color scale represents the SI computed at each time step. The scale is generated using the minimum and maximum values of the plot. Each trial was oversampled by a scale factor of 2.

$$SI_t(BG-PMC) = \frac{3R_t^j(rule^1) - \sum_{i=2}^{4} R_t^j(rule^i)}{3R_t^j(rule^1)}. \tag{17}$$

Although Eqs. (16) and (17) look different to Eqs. (14) and (15), as the SI index has never been experimentally computed for the DMTS task, we propose these expressions in order to make the comparison with the VD task as fair as possible. The average values of PFC and BG–PMC neuron SI for the DMTS task are shown in Fig. 8.

When the THMSs were computed, contrary to the results shown in the VD task, we found that the correlation with the estimated performance was −0.91 and −0.9 for the PFC and BG–PMC, respectively (Fig. 9). It should also be noted in Fig. 8b that the SI for BG–PMC neurons is

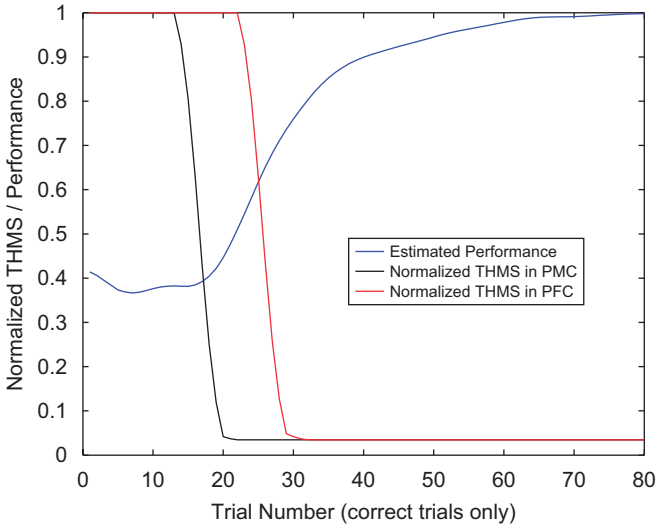

Fig. 7. VD task. Normalized averaged THMS, for the PFC and the BG–PMC, and estimated performance as a function of correct trial number. On each plot, the THMS is calculated as the time step where the SI surpasses the half value (maximum–minimum)/2. Then, it is normalized by the maximum THMS. The performance computed here is different to the one shown in Fig. 3. Here, it is computed only at correct trials for the rule under analysis. On each one of them, the performance arise averaging the correct responses over the previous 16 trials (regardless of the rule presented on each one of them). All the curves were smoothed using a gaussian window with a span of five samples.

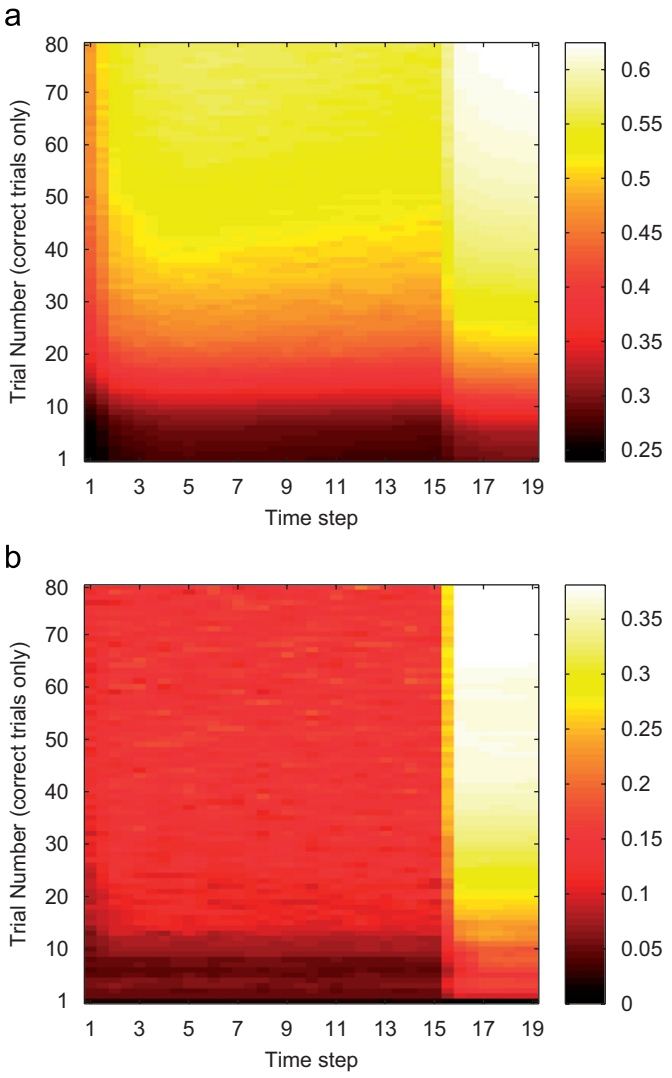

Fig. 8. DMTS task. (a) Averaged SI for the most selective PFC neuron to rule 1 (matching using the CS^1 as sample and comparison), computed using (16). (b) Averaged SI for the BG–PMC neuron associated with R^1 (Go), computed using (17). Other details are the same as in Fig. 6.

very low during the sample and delay periods due to the fact that, at this time, it is not possible to predict if the correct response is Go or No–Go. Actually, the neuron under analysis in the PFC is activated by two out of the four possible during the sample and delay periods. On the other hand, during the same period of time, the neuron under analysis in the BG–PMC is activated with probability $\frac{1}{2}$ by the four possible rules, leading the numerator of (17) to a very low average value.

As can be seen from Fig. 9 and from the correlation between THMSs and performance, in the DMTS task no lead is observed in PFC and BG–PMC circuits. This result can be interpreted as that learning occurs simultaneously in the PFC and BG–PMC during the execution of complex tasks, whereas in simple ones a direct pathway from input to motor structures is enough to enable accurate performance.

7. Discussion

We have introduced a biologically plausible neural network model that explains changes in activity in PFC neurons and motor structures during learning of simple and complex contingencies.

Although the neural network model has several parameters, and might seem complex at a first glance, the key assumptions are simple. The model relies on a layer with STM traces, a reward prediction neural cluster, the PFC layer, a winner-take-all mechanism, the layer with possible responses and a mechanism to balance the energy flow between input–output layers governed by NE release. All these assumptions are widely accepted in the field. The fact that we observed different predictions in the role of BG–PMC/PFC depending on the task is due to the way the model can balance its different structures to accomplish the tasks. The pathway from layer 1 to the responses layer allows the model to select those responses associated with reward with a consequent fast increment in the selectivity of BG–PMC neurons. In the absence of a PFC like structure, a simple task can be solved mainly by this direct pathway, while a more complex one requires both the direct and the PFC-(BG–PMC) circuitry. This makes sense from an evolutionary perspective, where animals with simpler structures than the PFC are unable to perform complex tasks as DMTS, while animals with a more developed PFC can do it. This is mainly because the

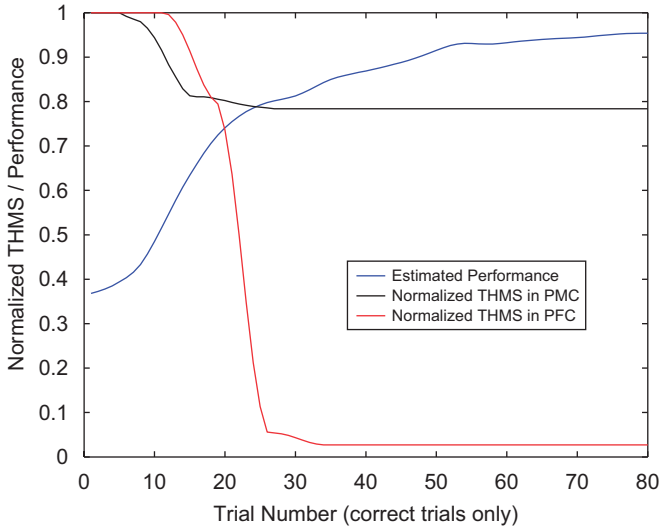

Fig. 9. DMTS task. Normalized averaged THMS, for the PFC and the BG–PMC, and estimated performance as a function of correct trial number. The details are the same as in Fig. 7.

addition of the PFC allows the animals to solve XOR type of problems, which cannot be solved with a single layer (the DMTS task is a linearly non-separable problem). Nonetheless, our previous models [17,18,24,41] lacking the PFC layer proved to be able to explain many of the key operant conditioning paradigms. We believe that the overall structure of our neural network model, and not its particular implementation, is responsible for the findings in this work.

Tested on a simple discrimination paradigm, the model fits behavioral and physiological results. As suggested in [33], the fact that the THMS of the BG–PMC is faster than the one of the PFC during the VD task, can be interpreted as a leadership of the former structures over the latter one. Here, we reproduced these results without a direct path from BG–PMC to cortical areas. Although it might seem natural to require a direct path for supporting the "training" hypothesis introduced in [33], this interpretation cannot be ruled out as the control can be done indirectly. The activity in the BG–PMC, and the resulting responses, will change the reward delivery probability. This will affect the DA firing, which in turn can modulate synaptic changes in PFC according to the mechanisms proposed in this work. Actually, other models of BG include feedback connections via the thalamus or the superior colliculus, but in both cases the purpose of this pathway is to inhibit cortical motor neurons after the selected response is executed [20,25]. However the "training" hypothesis has been criticized using different arguments [35].

On the other hand, from the results on the DMTS task, it is not possible to conclude that there is a lead in activity changes from neither PFC nor BG–PMC. This suggests that the hypothesis proposed in [33] would be paradigm-dependent. In this task, the input–output path cannot provide the necessary information to solve the task correctly. Each CS presented contributes to R^1 and R^2 in the same way, and, therefore, it does not allow an early increase in BG–PMC selectivity. It would be interesting to test this prediction experimentally. We postulate that when the task requires learning high-order contingencies, such as in the DMTS case, motor structures quickly select the subset of responses allowing the subject to obtain reward, but learning in the cortico-basal ganglia loop progresses in a concurrent way in order to maximize reward.

References

[1] W.F. Asaad, G. Rainer, E.K. Miller, Neural activity in the primate prefrontal cortex during associative learning, Neuron 21 (6) (1998) 1399–1407.
[2] G. Aston-Jones, J.D. Cohen, An integrative theory of locus coeruleus–norepinephrine function: adaptive gain and optimal performance, Annu. Rev. Neurosci. 28 (2005) 403–450.
[3] H.E. Atallah, M.J. Frank, R.C. O'Reilly, Hippocampus, cortex, and basal ganglia: insights from computational models of complementary learning systems, Neurobiol. Learn. Mem. 82 (3) (2004) 253–267.
[4] C.W. Berridge, B.D. Waterhouse, The locus coeruleus-noradrenergic system: modulation of behavioral state and state-dependent cognitive processes, Brain Res. Rev. 42 (1) (2003) 33–84.
[5] P.J. Brasted, S.P. Wise, Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum, Eur. J. Neurosci. 19 (3) (2004) 721–740.
[6] C. Cavada, T. Company, J. Tejedor, R.J. Cruz-Rizzolo, F. Reinoso-Suarez, The anatomical connections of the macaque monkey orbitofrontal cortex. A review, Cereb. Cortex 10 (3) (2000) 220–242.
[7] A. Compte, N. Brunel, P.S. Goldman-Rakic, X.J. Wang, Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model, Cereb. Cortex 10 (9) (2000) 910–923.
[8] G. Deco, E.T. Rolls, Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex, Eur. J. Neurosci. 18 (8) (2003) 2374–2390.
[9] D. Durstewitz, J.K. Seamans, T.J. Sejnowski, Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex, J. Neurophysiol. 83 (3) (2000) 1733–1750.
[10] C. Fiorillo, P. Tobler, W. Schultz, Discrete coding of reward probability and uncertainty by dopamine neurons, Science 299 (5614) (2003) 1898–1902.
[11] J.M. Fuster, J.R. Jervey, Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task, J. Neurosci. 2 (3) (1982) 361–375.
[12] J.M. Fuster, The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe, Raven Press, New York, NY, 1989.
[13] P.A. Garris, R.M. Wightman, Different kinetics govern dopaminergic transmission in the amygdala, prefrontal cortex, and striatum: an *in vivo* voltammetric study, J. Neurosci. 14 (1) (1994) 442–450.
[14] T. Gisiger, M. Kerszberg, J.P. Changeux, Acquisition and performance of delayed-response tasks: a neural network model, Cereb. Cortex 15 (5) (2005) 489–506.
[15] F. Gonon, Prolonged and extrasynaptic excitatory action of dopamine mediated by D1 receptors in the rat striatum *in vivo*, J. Neurosci. 17 (15) (1997) 5972–5978.
[17] D.A. Gutnisky, B.S. Zanutto, Cooperation in the iterated prisoner's dilemma is learned by operant conditioning mechanisms, Artif. Life 10 (4) (2004) 433–461.

[18] D.A. Gutnisky, B.S. Zanutto, Learning obstacle avoidance with an operant behavior model, Artif. Life 10 (1) (2004) 65–81.
[19] A. Lavin, L. Nogueira, C.C. Lapish, R.M. Wightman, P.E. Phillips, J.K. Seamans, Mesocortical dopamine neurons operate in distinct temporal domains using multimodal signalling, J. Neurosci. 25 (20) (2005) 5013–5023.
[20] A. Leblois, T. Boraud, W. Meissner, H. Bergman, D. Hansel, Competition between feedback loops underlies normal and pathological dynamics in the basal ganglia, J. Neurosci. 26 (13) (2006) 3567–3583.
[21] H. Lee, G.V. Simpson, N.K. Logothetis, G. Rainer, Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex, Neuron 45 (1) (2005) 147–156.
[22] S.E. Lew, H.G. Rey, S.B. Zanutto, A Theory of Perceptual Categorization in Primates. Program No. 538.7, Society for Neuroscience, Washington, DC, 2005.
[23] S.L. Lew, H.G. Rey, B.S. Zanutto, Task Complexity and Learning Dynamics in Prefrontal and Motor Structures: A Neural Network Model. Program No. 65.15, Society for Neuroscience, Atlanta, GA, 2006.
[24] S.E. Lew, C. Wedemeyer, B.S. Zanutto, Role of unconditioned stimulus prediction in the operant learning: a neural network model, in: Proceedings of IJCNN '01, Washington, DC, USA, pp. 331–6.
[25] C.C. Lo, X.J. Wang, Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks, Nat. Neurosci. 9 (7) (2006) 956–963.
[26] E.K. Miller, J.D. Cohen, An integrative theory of prefrontal cortex function, Annu. Rev. Neurosci. (24) (2001) 167–202 Review.
[27] J.W. Mink, The basal ganglia: focused selection and inhibition of competing motor programs, Prog. Neurobiol. 50 (4) (1996) 381–425.
[28] P.R. Montague, S.E. Hyman, J.D. Cohen, Computational roles for dopamine in behavioural control, Nature 431 (7010) (2004) 760–767.
[29] R.C. O'Reilly, D.C. Noelle, T.S. Braver, J.D. Cohen, Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control, Cereb. Cortex 12 (3) (2002) 246–257.
[30] S. Otani, H. Daniel, M.P. Roisin, F. Crepel, Dopaminergic modulation of long-term synaptic plasticity in rat prefrontal neurons, Cereb. Cortex 13 (11) (2003) 1251–1256.
[32] W.X. Pan, R. Schmidt, J.R. Wickens, B.I. Hyland, Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network, J. Neurosci. 25 (26) (2005) 6235–6242.
[33] A. Pasupathy, E.K. Miller, Different time courses of learning-related activity in the prefrontal cortex and striatum, Nature 433 (7028) (2005) 873–876.
[34] M. Petrides, Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory, J. Neurosci. 20 (19) (2000) 7496–7503.
[35] S. Ravel, B.J. Richmond, Where did the time go?, Nat. Neurosci. 8 (6) (2005) 705–707.
[36] G. Rainer, S.C. Rao, E.K. Miller, Prospective coding for objects in primate prefrontal cortex, J. Neurosci. 19 (13) (1999) 5493–5505.
[38] J.N.J. Reynolds, J.R. Wickens, Dopamine-dependent plasticity of corticostriatal synapses, Neural Networks 15 (4–6) (2002) 507–521.
[39] V.P. Roychowdhury, K.-Y. Siu, T. Kailath, Classification of linearly nonseparable patterns by linear threshold elements, IEEE Trans. Neural Networks 6 (2) (1995) 318–331.
[40] K. Sakai, J.B. Rowe, R.E. Passingham, Active maintenance in prefrontal area 46 creates distractor-resistant memory, Nat. Neurosci. 5 (5) (2002) 479–484.
[41] N.A. Schmajuk, B.S. Zanutto, Escape, avoidance, and imitation: a neural network approach, Adapt. Behav. 6 (1) (1997) 63–129.
[42] W. Schultz, Getting formal with dopamine and reward, Neuron 36 (2) (2002) 241–263.
[43] M. Sidman, R. Rauzin, R. Lazar, S. Cunningham, W. Tailby, P. Carrigan, A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children, J. Exp. Anal. Behav. 37 (1) (1982) 23–44.
[44] R.S. Sutton, Learning to predict by the method of temporal difference, Mach. Learn. (3) (1988) 9–44.
[45] H. Tomita, M. Ohbayashi, K. Nakahara, I. Hasegawa, Y. Miyashita, Top-down signal from prefrontal cortex in executive control of memory retrieval, Nature 401 (6754) (1999) 699–703.
[46] M. Usher, J.D. Cohen, D. Servan-Schreiber, J. Rajkowski, G. Aston-Jones, The role of locus coeruleus in the regulation of cognitive performance, Science 283 (5401) (1999) 549–554.
[47] H.B. Uylings, H.J. Groenewegen, B. Kolb, Do rats have a prefrontal cortex?, Behav. Brain Res. 146 (1–2) (2003) 3–17.
[48] J. Wang, P. O'Donnell, D(1) dopamine receptors potentiate NMDA-mediated excitability increase in layer V prefrontal cortical pyramidal neurons, Cereb. Cortex 11 (5) (2001) 452–462.

Glossary

Stimuli and short-term memory traces
CS^i: conditioned stimulus i
US: unconditioned stimulus (reward)
$\tau_t(CS^i)$: short-term memory (STM) of CS^i at time step t
$\tau_t(R_j)$: short-term memory (STM) of R^j at time step t
$\tau_t^{long}(US)$: long-term memory of the US

Structures and neuron outputs
VTA: ventro tegmental area
SNc: sustantia nigra pars compacta
LC: locus coeruleus
PFC: prefrontal cortex
$BG–PMC$: basal ganglia and premotor cortex
$P_{CS}{}^i(t)$: prediction of the US due the presence of CS^i at time step t
$\delta(t)$: prediction error (phasic dopaminergic activity) at time step t
$W^{\delta}{}_t$: gating window for the DA postsynaptic effects
$\mathbf{e}_{CS}{}^i(t)$: eligibility trace for the conditioned stimulus $CS^i R_t^j$ output of the BG–PMC neuron j at time step t
O_t^k: output of the PFC neuron k at time step t
M_t^k: output of the PFC neuron k after the winner-take-all at time step t
lc_t: tonic activity of the locus coeruleus at time step t

Excitability and plasticity
$Basal_{PFC,BG–PMC}$: basal activity of PFC and BG–PMC neurons.
$\mathbf{V}_{CS}{}^i$: association of $\tau_t(CS^i)$ with the US
α_{TD}: learning constant for the TD model
$U_t^k(CS^i)$: synaptic weight from CS^i to the kth neuron of the PFC
$w_t^j(CS^i)$: synaptic weight from CS^i to the jth neuron of the BG–PMC
$w_t^j(M^k)$: synaptic weight that connect the kth neuron of the PFC with the jth neuron of the BG–PMC
θ_{hebb}, $\theta_{anti\text{-}hebb}$: thresholds for Hebbian or anti-Hebbian learning
B_{winner}: synergism D1-NMDA
v_{PFC}, $v_{BG–PMC}$: learning constant for synaptic weights of PFC and BG–PMC neurons
μ_{PFC}, $\mu_{BG–PMC}$: first-order momentum for synaptic weights of PFC and BG–PMC neurons
SI: selectivity index

Sergio E. Lew was born in Chaco, Argentina, in 1966. He received the B.Eng. degree in electronic engineering from the University of Buenos Aires (UBA) in 1992 and the Ph.D. degree in Engineering from UBA in 2007. Since 1994, he has been a Research Assistant with the Department of Electronics at the University of Buenos Aires. In 1999, he joined the Institute of Biomedical Engineering at the University of Buenos Aires. His research interests include complex systems, neural networks and computational neuroscience.

Hernan G. Rey was born in Buenos Aires, Argentina, in 1978. He received the B.Eng. degree in electronic engineering from the University of Buenos Aires, Buenos Aires, Argentina, in 2002. Since 2002, he has been a Research Assistant with the Department of Electronics at the University of Buenos Aires. In 2004, he joined the Institute of Biomedical Engineering at the University of Buenos Aires, where he is currently pursuing the Ph.D. degree. His research interests include adaptive filter theory, neural networks and computational neuroscience.

Diego A. Gutnisky was born in Buenos Aires, Argentina, in 1978. He received the B.Eng. degree in electronic engineering from the University of Buenos Aires, Buenos Aires, Argentina, in 2002. From 2001 to 2004, he has been a Research Assistant in the Institute of Biomedical Engineering at the University of Buenos Aires. In 2004 he joined the laboratory of Dr. Dragoi at University of Texas-Health Science Center at Houston, where he is currently pursuing a Ph.D. degree in Neuroscience. His research interests include theoretical neuroscience, vision, machine learning and cooperative behavior.

B. Silvano Zanutto received the Electronic Engineer degree from the University of Buenos Aires (UBA) in 1980 and the Ph.D. degree in Biology from UBA in 1993. He was Research Associate at the Department of Psychology and Neuroscience of Duke University from 1993 to 1997. He was Assistant Professor of Electronic Department (UBA) from 1998 to 2002 and has been Professor and Director of the Instituto de Ingeniería Biomédica (UBA) since 2002. He has been Researcher of IBYME-CONICET since 1998. He is interested in theoretical neuroscience, neural circuits, experimental psychology and machine learning. Current topics are computational models of operant learning and categorization based on psychological, neurobiological and ethological data.