A modular neural-network model of the basal ganglia’s role in learning and selecting motor behaviours

Action editors: Wayne Gray and Christian Schunn

Gianluca Baldassarre*

Via di Pietralata 430 G18, 00158 Roma, Italy

Received 1 March 2001; accepted 1 September 2001

Abstract

This work presents a modular neural-network model (based on reinforcement-learning actor–critic methods) that tries to capture some of the most relevant known aspects of the role that basal ganglia play in learning and selecting motor behavior related to different goals. The model uses a mixture of experts network for the critic and a hierarchical network with two levels for the actor. Some simulations with the model show that basal ganglia select ‘chunks’ of behavior whose ‘details’ are specified by direct sensory-motor pathways, and how emergent modularity can help to deal with tasks with asynchronous multiple goals. A ‘top-down’ approach is adopted that first analyses some adaptive non-trivial interaction of a whole (simulated) organism with the environment, and its capacity to learn, and then attempts to implement these functions with neural architectures and mechanisms that have an empirical neuroanatomical and neurophysiological foundation. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Basal ganglia; Double-inhibition mechanism; Dopamine; Modularity; Neural networks; Mixture of experts networks; Reinforcement learning; Actor–critic methods; Multi-goal tasks

1. Introduction and methodology

What is the role that basal ganglia play in mammals’ sensory-motor behavior? When organisms have different needs and goals, sometimes they have to associate slightly different behaviours with the same perception patterns, some other times they have to associate completely different behaviours with them. This work presents some simulations that suggest that in the former case the differences are dealt with by the same sensory-motor pathway (implemented by a neural module), while in the latter cases different sensory-motor pathways are selected. In fact if the behavioral response to associate with a given perception were different with different needs and goals, using the same neural synapses and pathways would only cause interference. In this context the basal ganglia could play a role in selecting different sensory-motor pathways when necessary.

This work follows a ‘top-down’ approach, where the starting point of analysis is organisms’ behavior and learning processes (cf. Meyer & Guillot, 1990). With this purpose it presents a simulation of an...
organism that has different needs (signals coming from the body and indicating a physiological unbalance (cf. Rolls, 1999)) or, alternatively, different goals (desired states of the body and the world) associated with different positions in the environment (for example we can assume that these different positions are occupied by resources that satisfy different needs). The organism learns through classical and instrumental learning ((Lieberman, 1993); in Baldassarre & Parisi, 2000), these two learning mechanisms are integrated in a comprehensive actor–critic model. Cf. (Barto, 1995; Sutton & Barto, 1998), for this model) to navigate in the environment in order to reach those positions. Given this behavior, the work attempts to produce it by building a neural-network controller that satisfies (some of) the constraints coming from the known empirical evidence about basal ganglia. Since the starting point of this approach is to simulate sophisticated organisms’ behaviors, sometimes there is no empirical data suggesting which mechanisms underlie them. In these cases some computational solutions are adopted that do not have a known empirical correspondent (they will be referred to as ‘arbitrary’ in the rest of the paper). These solutions should be considered as a useful theoretical exercise, eventually suggesting interesting ideas for empirical investigation, and should not be judged too severely on the basis of the current neural evidence.

2. The basal ganglia and the empirical evidence addressed

This section first presents a brief review of the basic anatomy of basal ganglia, and then illustrates the empirical evidence specifically addressed by this research. Basal ganglia (Houk, 1995; Houk, Adams & Barto, 1995) are a set of nuclei situated in the deep part of the two cerebral hemispheres, inside the white substance under the cortex. Basal ganglia receive their input from the whole cerebral cortex and send their output to the frontal areas of cortex, to some non-cortical motor systems (superior colliculus and substantia nigra pars reticulata) and to dopaminergic systems (such as the substantia nigra pars compacta) in the midbrain. Fig. 1 shows the main nuclei that make up the basal ganglia, and the other nuclei and areas to whom they are connected. The striatum (mainly made up by the putamen and nucleus caudatus) constitutes the input component of the basal ganglia, while the globus pallidus (internal and external) constitutes their output component. The substantia nigra pars reticulata in the midbrain plays an output role for the basal ganglia functionally similar to the role of the internal globus pallidus. The connections between the striatum and these two nuclei form the ‘direct pathway’. This pathway is particularly important for the control of movement. In fact, the substantia nigra pars reticulata contri-
butes to control the eyes’ movement via the superior colliculus, while the internal globus pallidus contributes to control the skeletal movements via the thalamus and the frontal areas (prefrontal, premotor and motor areas). The striatum also projects to the external globus pallidus that, in turn, projects, via the subthalamus, to the internal globus pallidus and substantia nigra pars reticulata. This is called the ‘indirect pathway’. The striatum also projects, directly and indirectly, to the substantia nigra pars compacta. This is an important pathway since this nucleus contains dopaminergic neurons that project to the striatum itself and to the frontal areas and play an important role in learning (see below).

The anatomical and physiological evidence specifically addressed in this work is now illustrated. Chevalier and Deniau (1990) propose that a double-inhibition mechanism is the basic process of the basal ganglia’s functioning (Fig. 1). They report that in some experiments where monkeys have to carry out a delayed saccade to a remembered target, some striatal cells (usually mute) are induced to fire with local injection of glutamate. The striatal discharge inhibits (via GABAergic connections) a group of cells in the substantia nigra pars reticulata (usually tonically active) that release from (GABAergic) inhibition a subset of cells of the superior colliculus responsible for the saccade. In the case of skeletal movements the striatum–globus pallidus–thalamus pathway implements the double inhibition releasing mechanism. The authors report that while in rodents this mechanism is sufficient to trigger movements, in the reported experiments the execution of a saccade requires temporal coincidence of basal ganglia disinhibition with command signals from other sources. The model presented here reproduces this aspect: basal ganglia select a particular sensory-motor pathway that then yields the detailed behavioral output.

Graybiel (1998), addressing the role that the basal ganglia’s neural modules play in human slow habit learning and animal stimulus response association, draws an abstract parallel between the striatum’s anatomical organization in partly interconnected zones, called ‘matrisomes’, and the modular architecture of the neural networks proposed by Jacobs, Jordan, Nowlan & Hinton (1991). As we shall see, the computational model presented here proposes a possible way to specify such a parallel.

Houk, Adams & Barto (1995) suggest a possible correspondence between the actor–critic models’ architecture and functioning (Barto, 1995; Sutton & Barto, 1998) and the architecture of the basal ganglia. In particular they propose that the circumscribed regions called ‘striosomes’ (differently from matrisomes, they are identifiable for their chemical make-up and output connectivity, cf. Fig. 1) may implement the function of the critic: predicting future rewards and yielding a step-by-step reward signal in cases of delayed rewards. The surrounding ‘matrix’ regions may implement the function of the actor: selecting actions or, as in the model presented here, sensory-motor pathways. As we shall see, the actor–critic model is at the base of the model presented here.

Lots of other aspects of these contributions have been incorporated into the model, and will be presented in detail in the next section. The numerous brain-imaging studies of basal ganglia’s role in sequence learning are not directly addressed in this paper (see (Graybiel, 1998), for some references). Similarly the important role that basal ganglia play in working memory is not addressed in this paper (see Goldman-Rakic, 1995; Frank, Loughry & O’Reilly, 2000).

3. Scenario and model of basal ganglia

The environment used in the simulations is a square arena with sides measuring 1 unit (Fig. 2). The organism cannot see the boundaries of the arena and cannot exit it. Inside the arena there are 5 circular landmarks/obstacles that the organism can see with a one-dimension horizontal retina covering 360 degrees with 50 contiguous sensor units. Each unit gets an activation of 1 if a landmark is in its scope, 0 otherwise, and is affected by noise (0.01 probability of flipping). The signals coming from the retina are aligned with the magnetic north through a simulated compass whose reading is affected by Gaussian noise (0 mean, 1 degree variance). Before being sent to the controller, these signals are re-mapped into 100 binary units representing the image ‘contrasts’ (these units implement edge detection). Two contiguous retinal units activate one contrast unit if they are respectively on and off, another
Fig. 2. (Top) The scenario of the simulations containing three goals (marked with ×), five landmarks (black circles), the scope of the organism’s 50 visual sensors (delimited by the rays), and the organism (white circle at origin of rays). (Bottom) The activation of the visual sensors, its re-mapping into contrasts, and the bottom left goal (contrast pattern).

Fig. 3. The components of the organism’s controller. Labels in italics indicate the possible brain areas and nuclei corresponding to the model’s components. Thin arcs indicate one-to-one connections with weight +1 when not differently indicated. Dashed thin arrows indicate unit-to-unit or unit-to-area inhibitory connections (strong enough to make the target units and areas silent). Bold arrows indicate connections updated on the basis of the dopaminergic signal. Dashed bold arrows indicate the dopaminergic signal.
and the current input contrasts. A goal is the contrasts’ pattern corresponding to the goal position. When these patterns have at least 94% of bits with same value, the matcher returns 1 otherwise it returns 0. The threshold 94% has been chosen because it produces a satisfactory small size of the area recognized as goal. It is assumed that some memory process, not simulated in the model, evokes the goal patterns. When a goal is reached, another goal is evoked that is randomly chosen between the three goals. In real brains, goal patterns may be generated within frontal areas, for example by the frontal eye fields in the case of saccades, and recognition could take place here or in the sensory areas themselves (see Fig. 3; cf. (Kosslyn, 1999), on this issue).

There is an alternative way to view this part of the model. Animals are endowed with innate neural structures that take input from the environment and map it into a ‘reward’ or ‘punishment’ internal signal. This usually happens when some states of the environment are achieved that are relevant for adaptation, for example some food is ingested or the body is hurt (primary reinforcements). Notice that these signals are produced only if a correspondent appetitive need (e.g. hunger) is present (Rolls, 1999). In the model the presence of a certain need could be thought of as corresponding to an arbitrary pattern (the goal pattern) coming from the body, while the signal relevant for adaptation is the signal coming from the sensors, for example from the sensors in the mouth that detect the ingestion of food. In this case the matcher would yield a rewarding signal when a need and the corresponding satisfying input pattern are present together, and it would correspond to limbic structures (cf. Rolls, 1999). In both cases the matcher’s signal arrives to the substantia nigra pars compacta, and this generates a dopaminergic signal that controls learning.

The actor, with the 6 ‘expert’ networks (6 different input areas–thalamus–frontal areas pathways), implements the organism’s ‘action–selection policy’. Each expert is a two-layer feed-forward neural network that gets the goal and the visual contrasts as input, and has 8 sigmoidal output units that locally encode the actions. The experts may correspond to neural assemblies of the thalamus or frontal areas: here the details of the model are quite arbitrary. To select one action, the activation $m_k$ (interpretable as ‘action merit’) of the output units is sent to the frontal areas where a stochastic winner-takes-all competition takes place (cf. (Hanes & Schall, 1996), regarding this possibility). The execution of one action has to be thought of involving the activation of a particular muscle template. The probability $P$ that a given action $a_k$ becomes the winning action $a_w$ to execute is given by

$$P[a_k = a_w] = m_k / \sum_j m_j.$$  

The role of the basal ganglia is to select an expert which, in its turn, has to select the actions to be executed through the mechanism of double inhibition illustrated previously involving the matrix of the striatum and the globus pallidus. This is done with another winner-takes-all competition analogous to the previous one, but this time involving the experts instead of the actions (could this mechanism correspond to the bistable behavior of the striatum spiny cells?). Notice that the basal ganglia can only prevent the proper expert from being inhibited, but cannot trigger an action directly.

The critic is a ‘mixture of experts network’ (Jacobs, Jordan, Nowlan & Hinton, 1991) based on 6 ‘expert’ networks. Each expert is a two-layer feed-forward neural network that gets the goal and the visual contrasts as input and has one linear output unit. The critic learns to produce the estimation $V'[s]$ of the ‘evaluation’ $V[s]$ of the current contrast pattern $s$. $V'[s]$ is defined as the expected discounted sum of all future reinforcements $r$, given the current action-selection policy $\pi$ expressed by the actor

$$V'[s] = E[\gamma^0 r_{t+1} + \gamma^1 r_{t+2} + \gamma^2 r_{t+3} + \cdots],$$

where $\gamma \in [0, 1]$ is the discount factor, set to 0.95 in the simulations, and $E$ is the mean operator. In order to compute $V'[s]$ the output $v_k$ of the experts is weighted and summed,

$$V'[s] = \sum_k [v_k g_k].$$

The weight $g_k$ is computed as the softmax activation function of the output units $o_k$ of the gating network,
\[ g_k = \exp[\alpha_k] / \sum_j[\exp[\alpha_j]]. \]

This part of the model is arbitrary, but the modularity of the striosomes confers some plausibility: the model is an implementation of what is suggested by Houk, Adams & Barto (1995) according to whom different striosomes may be specialized in dealing with different behavioral tasks. As we shall see, this is an emergent feature of the model presented here. The last component of the critic (subthalamic loop and substantia nigra pars compacta) is a neural implementation of the computation of the ‘temporal-difference error’ \( e \) defined as (Houk, Adams & Barto, 1995)

\[ e_i = (r_i + \gamma V'_{s_i} - V_{s_i}). \]

Each critic’s expert has a specific error defined as

\[ e_{ki} = (r_{i+1} + \gamma V'_{s_{i+1}} - v_i[s_i]). \]

These error signals correspond to the dopaminergic signals and are at the base of the learning processes of the actor and critic.

Each critic’s expert is trained on the basis of the expert’s dopaminergic error signal that assumes the role of error in the estimation of \( V'_{s_i} \) in a supervised learning algorithm. The weights of the experts are updated so that their estimation \( v_i[s_i] \) tends to be closer to the target value \( (r_i + \gamma V_{s_{i+1}}) \). This target is a more precise evaluation of \( s_i \) because it is expressed at time \( t + 1 \) on the basis of the observed \( r_i \) and the new estimation \( V_{s_{i+1}} \). The formula (a modified Widrow–Hoff rule, cf. (Widrow & Hoff, 1960)) to update the weights of each expert is

\[ \Delta w_{ki} = \eta e_{ki} y_i h_k, \]

where \( w_{ki} \) is a weight of the expert \( k \), \( \eta \) is a learning rate (set to 0.01 in the simulations) and \( y_i \) is the activation of the goal and contrast units. \( h_k \) (absent in the Widrow–Hoff rule) is the (updated) contribution of the expert \( k \) to the global answer \( V'_{s_i} \), and is defined as

\[ h_k = g_k c_k / \sum_j[g_j c_j], \]

where \( c_k \) is a measure of the ‘correctness’ of the expert \( k \) defined as

\[ c_k = \exp[-0.5 e_k^2]. \]

The gating network weights \( z_{ki} \) are updated to increase the contribution in yielding \( V'_{s_i} \) of the experts who had low errors,

\[ \Delta z_{ki} = \xi (h_k - g_k) y_i, \]

where \( \xi \) is a learning rate set to 0.1 in the simulations. This algorithm leads the experts to specialize in the different regions of the input-goal space. Notice that \( \xi = \eta = 0.01 \) the experts did not specialize and interference between different goals prevented learning.

The actor is trained according to the dopaminergic signal \( e_i \). In this case this signal is interpreted as the actor’s capacity to select actions that lead the organism to new states with an evaluation higher than the average evaluation experienced previously departing from that same state. The updating of the action merits of the selected expert (and only this) is done by updating the weights of the neural unit corresponding to the selected action \( a_w \) (and only this) as follows:

\[ \Delta w_{wi} = \zeta e_i (4 \eta (1 - m_w) v_i), \]

where \( \zeta \) is a learning rate (0.01) and \( m_w (1 - m_w) \) is the derivative of the sigmoid function multiplied by 4 to homogenize the size of the learning rates of the actor and the linear critic. The model’s dopaminergic signal affecting the sensory-motor pathways may correspond to the brain dopaminergic signal targeting the frontal areas downstream the thalamus. For simplicity in the model these dopamine-sensitive areas have been designed upstream the thalamus. The weights of the winning gating network’s unit are updated in the same way used for the experts’ merits (learning rate 0.01).

The learning mechanism of the critic and the actor differ because in the case of the actor it is not possible to have a teaching pattern to implement a supervised learning algorithm. The stochastic nature of the actor is necessary in order to produce new behaviors that are then strengthened or weakened.
according to their outcome in terms of rewards. At the beginning of the simulations the weights of the critic and actor (only those affected by the dopamine) are randomized in the interval \([-0.001, +0.001]\). This implies that the evaluations expressed by the linear critic are around 0, and the merits (probabilities) expressed by the ‘sigmoidal’ actor (stochastic selector) are around 0.5 (0.125). This implies that initially, the organism’s behavior is a random walk. Then the critic and the actor are trained simultaneously (policy iteration): the evaluator learns to determine the evaluation is over 0.99. The first column of Fig. 5 shows the resulting gradient field of the evaluations for the three goals). This probably means that the positions in the arena need to receive a different evaluation for the three different goals, so that using the same weights (same expert) would only cause negative interference. This also means that the connections from the (contrast) input pattern to the critic’s gating network are redundant. The fact that different parts of the striosomes specialize for different goals, as in the model, is an interesting hypothesis that has not yet been verified empirically. Notice that the controller is capable of not using some of the resources available (expert 1, 3, 4).

4. Simulations, results, interpretations

As mentioned, the task of the organism is to reach one of the three goal positions shown in Fig. 2. When a goal is reached a new one (randomly chosen between the three goals) is assigned to the organism which then has to reach it from its current position. Fig. 4 shows the organism’s learning curve in terms of number of steps taken to reach a goal (mobile average for 100 successes, average for 10 random seeds). The performance improves from about 1000 to about 30 steps.

Fig. 5 presents some data about how the neural-network controller of one of the 10 simulations has self-organized during learning. The other random seeds have produced results with analogous quality. Concerning the critic, we see that each goal is dealt with by a different expert (in each possible position of the arena the contribution of this expert in determining the evaluation is over 0.99. The first column of Fig. 5 shows the resulting gradient field of the evaluations for the three goals). This probably means that the positions in the arena need to receive a different evaluation for the three different goals, so that using the same weights (same expert) would only cause negative interference. This also means that the connections from the (contrast) input pattern to the critic’s gating network are redundant. The fact that different parts of the striosomes specialize for different goals, as in the model, is an interesting hypothesis that has not yet been verified empirically. Notice that the controller is capable of not using some of the resources available (expert 1, 3, 4). These resources could be used for other goals.

With regard to the actor, Fig. 5 shows that the specialization of the experts is much less pronounced. In particular the graphs of the second and third column of the Fig. 5 show that, while pursuing a goal, the actor uses different experts in different position in the arena. The histograms show the frequency of use of the different experts for the different goals. Clearly the controller tends to use different experts when dealing with different goals, but now (differently from what is observed in the critic) the visual input plays an important role. An interesting fact coming out from the second and third column of Fig. 5 is that the same experts are being used for different goals (e.g. expert 1 for goal 1 and 3). Further investigation should show if this different use of experts in the critic and in the actor are due to the differences in the role they play or if it is due to the difference between the algorithms employed (supervised learning and stochastic unsupervised learning; cf. Calabretta, Nolfi, Parisi & Wagner, 1998), on the evolutionary emergence of modular networks’ function through genetic algorithms).

Notice that in the actor, as in the case of the critic, there is a partial use of the resources available (marginal role of expert 3, 4 and 5).

The exploration of some parameters and simulation conditions has shown some limits of the controller. Too high learning rates (especially for the critic) produce instability, while too low rates produce slow
Fig. 5. Data about the self-organization of the controller during learning (1 out of 10 random seeds). The three rows of graphs are relative to the three different goals. The first column of graphs shows the gradient field of evaluations $V^c[s]$ produced by the critic in 400 different positions (corresponding to the $20 \times 20$ cells of the grid). The area of the white (positive evaluations) and black (negative evaluations) cells is proportional to the evaluation produced. The patches where no evaluation is present correspond to the obstacles. The number in each graph indicates the expert that is used by the critic for the particular goal (only 1 expert per goal). The second column of graphs shows the order number of the actor’s expert with highest probability of being selected (for the same 400 positions of the previous column). The last column of graphs shows the histograms that summarize the frequencies of the experts illustrated in the previous column.

learning. The system is also quite sensitive to the ‘aliasing’ problem (this is the problem that occurs when there are states of the world that appear to be the same or very similar, cf. (Whitehead & Ballard, 1991)). In particular if there are positions that are similar to the goal positions, the organism tends to waste time searching around them. This happens because they will tend to have a high evaluation. With more goals some problems also occur: with some random seeds the same critic’s expert is used for more than one goal. This produces a gradient field with more than one peak. This causes the organism to pursue the positions corresponding to these peaks at the same time so that the behavior results to be dithering.

5. Conclusion

This work has presented a computational model that attempts to summarize in a coherent picture some of the most relevant properties of basal ganglia
Regarding motor behavior. An attempt has been made to design a model that on one side is capable of controlling an organism in a non-trivial behavioral task, and on the other side is based on architectures and mechanisms possibly grounded on empirical evidence about the anatomy and physiology of basal ganglia. The model has shown that the role of the striosomes in the striatum might be that of producing an evaluation of the expected future rewards, and to build a dopaminergic signal corresponding to previously neutral input patterns on the basis of some primary reinforcers. The dopaminergic signal is used to learn to express the evaluations themselves on the basis of a supervised learning algorithm. The simulations have shown that the modularity of the striosomes is used to deal with different behavioral tasks the organism meets during its life. The model has also shown that the role of the matrix in the striatum might be that of learning to generate stochastic variants of behavior, eventually consolidated on the basis of the dopaminergic signal. Here the role of the basal ganglia’s double-inhibition mechanism is not that of directly triggering particular patterns of behavior, but that of releasing from inhibition sensory-motor pathways that then yield a particular behavior suitably related to the current goals and precepts.

Acknowledgements

The Department of Computer Science, University of Essex, funded the author’s research. Special thanks are expressed to Prof. Jim Doran (University of Essex) and Prof. Domenico Parisi (Italian National Research Council) for their valuable contribution of ideas, and James Adam for his precious help in the preparation of the article.

References


