White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/107036/

Version: Accepted Version

Article:
Girard, B., Cuzin, V., Guillot, A. et al. (2003) A basal ganglia inspired model of action selection evaluated in a robotic survival task. Journal of Integrative Neuroscience, 2 (2). pp. 179-200. ISSN: 0219-6352

https://doi.org/10.1142/S0219635203000299

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

A basal ganglia inspired model of action selection evaluated in a robotic survival task

Benoît Girard*, Vincent Cuzin*, Agnès Guillot*, Kevin N. Gurney** and Tony J. Prescott**

*AnimatLab/LIP6, Paris, France

** Department of Psychology, University of Sheffield, Sheffield, U.K.

Basal ganglia model embedded in a robot
Benoît Girard, AnimatLab/LIP6, 8 rue du capitaine Scott, 75015 Paris, France
Tel/fax : +33 1 44 27 88 09 – email : benoit.girard@lip6.fr

Abstract :

The basal ganglia system has been proposed as a possible neural substrate for action selection in the vertebrate brain. We describe a robotic implementation of a model of the basal ganglia and demonstrate the capacity of this system to generate adaptive switching between several acts when embedded in a robot that has to 'survive' in a laboratory environment. A comparison between this brain-inspired selection mechanism and classical winner-takes-all selection highlights some adaptive properties specific to the model, such as avoidance of dithering and energy-saving. These properties derive, in part, from the capacity of simulated basal ganglia-thalamo-cortical loops to generate appropriate behavioural persistence.

Key words: Biorobotics; basal ganglia; action selection; computational model; survival task.

1.Introduction

Computer simulation is one approach to gain insight into the details of the understanding of biological mechanisms. It can be useful for predicting the activations of cells or biological circuits but, as far as behaviour is concerned, it cannot replace real world experiments in evaluating sensorimotor systems. Biorobotics is a recent field at the intersection of biology and robotics that designs the architectures of robots as models of specific biological mechanisms [38]. The realization of these artificial systems can be used to evaluate and compare biological hypotheses, as well as to estimate the efficiency of biological mechanisms within a robot control setting.

The objective of the current paper is to evaluate, on a robotic platform, the hypothesis that the vertebrate basal ganglia provide a possible neural substrate for action selection [10],[29]. Action selection concerns the issue of solving conflicts between multiple sensorimotor systems so as to display relevant behavioural sequences. Several computational models of decision-making involving these neural structures have been investigated in a variety of simulation tasks, like trajectories control, multi-armed bandit or Wisconsin card sorting tasks (for reviews see [19],[7], [33], [34], [14]), but few have been faced with the reality of a robotic device. An exception is the computational model of Gurney, Prescott and Redgrave [17], [18] which has been embedded in a Khepera robot (©Kteam) and used to simulate the behavioural sequences of a hungry rat placed in a novel environment [30]. In the current paper, our objective is to test the same model with a more classical action selection task—a 'two resources' survival problem, which is well-known as the minimal scenario for evaluating this kind of mechanism [36]. This investigation, which uses alternative robot platform (©Lego Mindstorms), requires the embedded basal ganglia model to

select efficiently between several actions to allow the robot to 'survive' in an environment where it can find `ingesting places' and `digesting places'. A key requirement is that the control architecture should be sufficiently adaptive to generate a chaining of actions allowing it to remain as long as possible in its so-called *viability zone* [4]. This entails, at each time step, maintaining its essential state variables above minimal thresholds.

--Insert Fig.1 about there --

As illustrated on Fig.(1), the basal ganglia (BG) is a group of interconnected sub-cortical nuclei. In the rat brain, the principle basal ganglia structures are the striatum, the globus pallidus (GP), the entopeduncular nucleus (EP), the subthalamic nucleus (STN), and the substantia nigra pars reticulata (SNr). The striatum receives somatotopic cortical input from the sensory, motor, and association areas. The major output structures are EP and SNr. They maintain a tonic inhibition on thalamic nuclei that project on the frontal cortex, in particular on motor areas. The internal connectivity of the BG has long been interpreted as a dual pathway [2], a *direct pathway,* consisting of inhibitory striatal efferents which projects to the EP/SNr, and a parallel excitatory *indirect pathway,* projecting to EP/SNr by way of GP and STN. This interpretation has been shown to have several shortcomings, in particular, it fails to account for several anatomically important pathways within the BG and to accommodate recent clinical data.

The model proposed by Gurney, Prescott and Redgrave (henceforth the GPR model, [17],[18]) reinterprets the basal ganglia anatomy as a set of neural mechanisms for selection in a new, dual-pathway functional architecture. A *selection* pathway, including D1 striatum (i.e., striatal neurons with D1 dopamine synaptic receptors) and STN, operates through disinhibition of the output nuclei (EP/SNr). A *control* pathway, involving D2 striatum (i.e., striatal neurons with D2 dopamine synaptic receptors) and STN, modulates the selection process in the first pathway via

innervation from GP. Moreover, Humphries and Gurney [21] embedded the two circuits into a wider anatomical context that included a thalamo-cortical excitatory recurrent loop whereby the output of the basal ganglia can influence its own future input.

The current work, which is an extension of previous experiments [15], will specifically investigate whether the GPR model implements more than a simple `winner-takes-all' (WTA) mechanism, a classical selection mechanism proposed long ago by engineers and ethologists [5]. The WTA is based on selecting for execution the action that corresponds to the highest `motivation' (integration of internal and external factors), whilst inhibiting all competitors. Although the GPR model has a superficially similar property of selecting (albeit by disinhibition) the most highly motivated action, it is modulated by the effects of the control and feedback circuits, potentially resulting in different patterns of behavioural switching, compared to simple WTA. In particular, the GPR feedback loops can induce 'behavioural persistence'. The importance of persistence as an adaptive process for animals has long been recognised by ethologists (e.g. [25], [39]) due to its functional role in allowing an activity to endure in spite of a rapid decrease in its drive. Some authors have also speculated about the possible mechanisms that could generate behavioural persistence such as positive feedback [20] or reciprocal inhibition between multiple motivational systems [24], [8]. In the current we paper we investigate the possibility, proposed by Redgrave, Prescott and Gurney [34], that basal ganglia-thalamocortical control loops may serve as a neural substrate for a behavioural persistence effect. A comparison of the two control architectures GPR and WTA, embedded in the same robot in the same environment, should therefore demonstrate precisely if and how the GPR control circuits can bring benefits to the action selection process.

Following a description of the GPR computational model, we will detail how this model was implemented within the control architecture of the Lego Mindstorms robot. The results obtained with the model will be presented and compared with those of a WTA, and these will be discussed from the perspective of biological plausibility.

2.Method

2.1. The GPR computational model

The details of the computational model and its correspondence with the neural anatomy have been fully described in [17], [18], [21], therefore we will introduce here only its main characteristics. As shown in Fig.(2), the *selection* and *control* pathways will be designated for conciseness by *BGI* and *BGII* respectively.

--Insert Fig.2 about there –

The inputs for the GPR model are speculative variables called '*saliences*' that represent the commitment toward displaying a given action. The saliences allow the representation of actions at the initial stage of the model in terms of a 'common currency' [27]. Each action is associated with a given salience, which is supposed to be computed in the sensorimotor cortex as a weighted function of all the external information and internal needs relevant for this action. The values are provided as inputs to STN, D1 and D2 striatum through segregated parallel channels each one simulated by a leaky-integrator artificial neuron, which is the simplest neural model incorporating the notion of a dynamic membrane potential [40], [3].

The *BGI selection* effect is mediated by two mechanisms. The first one concerns local recurrent inhibitory circuits within D1 striatum, which select a single winner salience – the highest one - and generate an output value proportional to that winner. This value is given as an inhibitory input to the corresponding channel into EP/SNr. The EP/SNr channels are tonically active and direct a continuous flow of inhibition at neural centres that generate the actions. When the signal emanating from D1 striatum inhibits a particular inhibitory channel in EP/SNr, it thereby removes the inhibition from the corresponding action.

EP/SNr affords a second selection mechanism with an ‘off-centre on-surround’ feedforward network, in which the `on-surround' is supplied by a diffuse excitation of all channels from STN, and the `off-centre' by the inhibitory signal from D1 striatum. This mechanism serves to reinforce the discrepancies between the winning EP/SNr channel value and all the others.

The properties of this selection pathway are modulated by *control* signals computed by the *BGII* circuit, in which a arrangement similar to *BGI* prevails: local recurrent inhibitory circuits within D2 striatum and an ‘off-centre on-surround’ feedforward network, in which the `on-surround' is supplied by a diffuse excitation of all the GP channels from STN, and the `off-centre' by the inhibitory signal from D2 striatum. Outputs from GP inhibit the EP/SNr and STN channels. The role of this control pathway appears to be three-fold.

First, the output GP signal directed to EP/SNr enhances the selectivity under inter-channel competition. Second, the inhibition from GP to STN serves to automatically regulate STN activity allowing for effective selection irrespective of the number channels in the model. Without this mechanism, as the number of active channels is increased, the STN diffuse excitation will also increase, eventually shutting down all the output channels of the model. The negative feedback provided by GP is therefore just sufficient to automatically scale this excitation in such a way as to ensure appropriate selection. The third role of *BGII* concerns dopaminergic

modulation via the inhibitory D2 receptors, which proceeds synergetically - with excitatory D1 striatal receptors - to increase the inter-channel output activation interval.

In the thalamic excitatory recurrent loop, the thalamus (*TH* on Fig.(2) is decomposed into the thalamic reticular nucleus (TRN) and the ventro-lateral thalamus (VL) [21]. Both structures have the same segregated channels as *BGI* and *BGII*. This positive feedback loop reinforces the salience of selected actions thereby fostering *persistence* of their state of being selected.

The outputs of the whole model, provided by the EP/SNr channels, are disinhibitions assigned to each action. At this stage, there can be partial disinhibitions of the motor components associated to more than one channel. As a consequence, the behaviour eventually displayed can be a combination of several actions, by weighting each one according to its degree of disinhibition. This kind of selection is called ‘soft switching’, in contrast to ‘hard switching’ denoting the selection of a single action [17].

2.2. The robot and its environment

The environment is a 2m x 1.60m flat surface surrounded by walls (Fig.(3), left). It is covered by 40cm x 40cm tiles of three different kinds: 16 uniformly gray tiles (neutral-gray representing `barren' locations), 2 tiles with a circular gray to black gradient (‘ingesting' locations, with inexhaustible resource), and 2 tiles with a circular gray to white gradient (‘digesting' locations).

--Insert Fig.3 about there –

Depending on the experimental settings, the robot has to select efficiently among the following actions:

- *ReloadOnDark (ROD)*: the robot stops on a dark place in order to 'ingest' a virtual food (i.e., it reloads a 'potential energy' that it cannot consume).
- *ReloadOnBright (ROB)*: the robot stops on a white place in order to 'digest' the eaten food (i.e., it transforms its amount of potential energy into an amount of actual energy that it can consume).
- *Wander(W)*: The robot wheels randomly in the environment (i.e., forward and turning acts of random duration).
- *AvoidObstacles (AO)* : the robot achieves a backward movement followed by a 45° rotation when one bumper is activated or a 180° rotation when both are.
- *Rest (R)* : the robot stops for 'resting'.
- *Grooming (G)* : the robot stops and displays a 'virtual grooming' (without moving any effector).

External and internal variables contribute to the calculation of the *saliences* associated to these actions.

External variables are provided by four sensors. The robot (Fig.(3), right) has two frontal light sensors, one behind the other, pointed to the ground. The mean of both values produced by these sensors is used to compute two variables, *Brightness* and *Darkness* (resp. L_B and L_D). The variable L_B (resp. L_D) is null for all colours darker (resp. brighter) than the neutral-gray, and increases linearly with brighter (resp. darker) colours, reaching 1 for the central white (resp. black) spots. Two bumpers, situated on the front-right and the front-left, produce each a binary value (resp. B_L and B_R) set to 1 when the robot hits an obstacle on the left, or right respectively.

The robot has a 'virtual metabolism' based on the two internal variables: *Potential Energy* (E_{Pot}) and *Energy* (E). The procedure to reload *Potential Energy* is to display *ReloadOnDark*. The gain

Delta E_{Pot} in E_{Pot} is proportional to the duration T_{Ingest} (in seconds) of the ingesting behaviour *and* to the Darkness of the ground:

$$\Delta E_{Pot} = 7^* T_{Ingest} * L_D \quad (1)$$

Then, the variable *Potential Energy* increases when the robot is on a dark place, but it decreases when it is converted into *Energy*.

The procedure to transform *Potential Energy* into *Energy* is to display *ReloadOnBright*. The changes in *Energy* and *Potential Energy* are proportional to the assimilation duration T_{Digest} and to the *Brightness* of the ground:

$$\Delta E = T_{Digest} \ (7^* L_B - 0.5) \quad (2)$$

$$\Delta E_{Pot} = - 7^* T_{Digest}^* L_B \quad (3)$$

When the variable *Potential Energy* is not null, the display of *ReloadOnBright* produces *Energy*. Otherwise, it consumes it at a rate of 0.5 units/s. The other actions consume *Energy* at a rate of 0.5 units/s, with the exception of *Rest*, which consumes half less (0.25 units/s).

Initially, *Potential Energy* and *Energy* take on values between 0 and 255. Then, these variables are normalized to lie between 0 and 1 before being used for salience computation (note that the *Energy* consumption rate is 2e-3 per second and –1e-3 per second for *Rest*– after normalization). When *E* reaches zero, the robot `dies'.

A third internal variable, *Dirtiness*, is used in only one experiment and has no effect on the metabolism. It increases at a rate of 1 units/s and is lowered by the activation of the action *Grooming* at a rate of 4 units/s.

2.3.Details of the GPR model implementation

As mentioned before, the level of commitment toward displaying a given action is expressed by a specific salience value, which will be given as an input for a specific channel in *BGI* and *BGII*. For each action, the salience is computed as a function of the involved external variables (L_B, L_D, B_L or B_R) and internal variables (E_{Pot}, E or *Dirtiness*), with the addition of the *Persistence* signal (P), coming from the positive feedback of the *TH* circuit, for the corresponding channel. Transfer functions are applied to the sensory inputs, in order to allow, for instance, salience dependencies on lack on energies (Table I).

Because some salience values depend on a coupling between two variables, Sigma-Pi units - allowing non-linear combinations of inputs conveying interdependencies between variables [12] - were also added. For instance, in our setting, *ReloadOnDark* should be activated when *Darkness* and *Potential Energy* are low. Activating it on non-dark places or if *Potential Energy* is high just wastes *Energy* without any benefit. This situation can eventually lead to 'death', because the salience corresponding to this channel is reinforced by its feedback persistence and prevents other behaviour from taking control of the robot. The computation by Sigma-Pi units is more than an engineering solution, as Mel [28] argues that the dendritic trees of neocortical pyramidal cells can compute complex functions of this type. Thus it is at least plausible to assume that second-order functions of the relevant contextual variables could be extracted by the neurons in either the cortex or the striatum that compute action salience.

Concerning the computation of all the channel values within *BGI*, *BGII* and *TH*, we use the same parameters - registered on Fig.(2) and Appendix 1- as in [30], except that the dopamine modulation was kept constant. To stabilise the system, the basal ganglia module was allowed to

compute up to four cycles with the same sensory data before waiting for new ones. For the output computations, the modality proposed in the original GPR model was modified, for the sake of the comparison with a WTA mechanism. Initially, the outputs of all actions were combined, leading to a ‘soft switching’. However, a WTA mechanism allows for only one winner, a situation that can be termed ‘hard switching'. In our experiments, the ‘soft switching' of the GPR output was disabled, in order to compare the selection of the winning act, and not the way this winning act is displayed *after* this selection. Accordingly, the motor output of the most fully disinhibited action was always enacted, and that of any partially disinhibited competitors always ignored.

2.4. The experiments

The GPR and the WTA architectures are embedded in the same robot. Both ‘GPR robot’ and ‘WTA robot’ have to achieve the same task independently, in the same environment. The input saliences are computed alike for both conditions, with the exception of the persistence signal P which is only included in the GPR saliences (Table I). For all experiments, the parameters were `hand-crafted' in an attempt to find settings that were close to optimal.

As an output, the GPR robot will display at each time-step the least inhibited action, and the WTA will display the action associated with the highest salience. If, in either architecture, there are multiple winning outputs, the action previously selected remains active.

--Insert Table I about there –

Experiment 1 compares both architectures with the smallest set of actions enabling survival (i.e., *Wander, ReloadOnDark, ReloadOnBright and AvoidObstacle*), in order to

determine their efficiency in selecting a chaining of actions for this minimal two resources task. Two further experiments were also performed to explore the potential of the GPR architecture.

Experiment 2 compares the capacity of both architectures to avoid the so-called 'dithering effect', a classical issue in action selection corresponding to a rapid oscillation between two acts [37]. For this purpose, a competing action, *Grooming*, is added to the behavioural repertoire used for Experiment 1. This action does not influence the robot's metabolism, but its salience is weighted in order to enhance the competition between this action and a current winning act.

Experiment 3 compares the capacity of both architectures to save energy, by having the opportunity to display a low-energy cost act, i.e., *Rest*. This action, added to the behavioural repertoire used in Experiment 1, can only be exhibited when the robot does not need either to reload *Energy* or *Potential Energy*, or to *Wander*.

All these experiments are composed of a set of runs for each architecture. At the beginning of a run, *Energy* is set to 1 and *Potential Energy* to 0.5, which allows less than 9 minutes of survival without appropriate reloading behaviours. The robot operates as long as its *Energy* is above 0 and its real batteries are loaded (up to 5 hours of continuous functioning). An action selection mechanism will be considered as successful if it is able to ensure at least 1 hour of survival per run.

Data describing internal variables and active behaviours is recorded approximately 15 times per second. For each run, this data is used to compute the following value sets:

- Medians of *Energy*, *Potential Energy* and acts duration
- Frequencies of activation of each act

- Average amount of *Potential Energy* extracted per second from the inexhaustible resources. This value is computed by adding up the variations of *Potential Energy* during ROD activations and by dividing the result by the total duration of the run.

For each experiment, all the value sets per run obtained for the GPR and WTA will be compared using the non-parametric Mann-Whitney U test.

3. Results

3.1. Experiment 1 (GPR: 9 runs; WTA: 10 runs)

With both GPR and WTA architectures, the robot achieved efficient action selection. All runs were successful, as all of them lasted more than one hour. Both architectures clearly succeeded in keeping the robot's essential variables within the *viability zone*. This first result prompted us to further analyze the structure of the behavioural sequences generated in the two conditions.

On Fig.(4), graphs (a) (b) and (c) show the input saliences, the inhibitory outputs (EP/SNr) of the GPR model and the corresponding behavioural sequence displayed by the GPR robot, while graphs (d) and (e) show the input salience and corresponding behavioural sequence displayed by the WTA robot. Note the substantial difference between the input saliences in the two runs (a) and (d), which are primarily due to effects of persistence in the GPR model. The output signals in (b) show that the control circuit (*BGII*) and the positive feedback loop (TH) have increased the contrast between the saliences.

--Insert Fig.4 about there –

As shown in Table II, ROB, ROD and AO bouts generally last longer with the GPR architecture than with the WTA. This can be explained by effects of persistence, allowing an action to remain active for some time after its 'raw' salience has fallen below that of other actions. Although bouts of `ingesting' and `digesting' are shorter in the WTA condition, their frequencies are correspondingly higher. One may then ask whether these behavioural differences are reflected in the way the energies are collected.

--Insert Fig.5 about there –

--Insert Tables II & III about there –

Table III indicates that the higher frequencies of acts of the WTA robot serve to substantially compensate for their shorter durations, to the point that the medians of *Potential Energy* and *Energy* between both architectures end up having similar values. The histogram of Fig.(6), which depicts the percentages of overall time (on y-axis) during which *Potential Energy* is gained at the levels shown on x-axis, also reveals a similarity in reloading, except for the last class: the GPR robot maintains E_{Pot} at over 95% of the maximum charge during 25% of time, compared to less than 13% for the WTA robot. This observation suggests that the GPR robot could reach this 'state of comfort' for a longer time than the WTA. However, it does not seem to take full advantage of this efficient reloading strategy, as both E_{Pot} medians are similar (Fig. (5) and Table III). Since the transformation from E_{Pot} to E is dissipative (see Eq. (2) and (3)), the E_{Pot} extracted from the environment is just larger (2.2e-3) than the rate of *Energy* consumption. In this experiment, this value is similar for both systems, because all the available behaviours consume *Energy* at the same rates.

--Insert Fig.6 about there –

Experiment 1 has shown that both models can display relevant switching between actions, but with different survival strategies. The purpose of the following experiments is to explore some consequences of such behavioural discrepancies.

3.2. Experiment 2: analysis of persistence.

The main differences between the behaviour of the GPR and WTA robots derive from the duration of their activity bouts. Specifically, in the GPR robot, the bout duration of a given action - due to the *BGI* selection - is extended by both the positive feedback of *TH* and the control of *BGII*. The persistence value computed by *TH* increases the winning salience, favouring the selection of the corresponding act for the forthcoming time steps. In parallel, the inhibitory signals coming from *BGII* to STN and EP/SNr decrease the global activation of the model. Without this control, a winning act could reinforce itself endlessly preventing itself from being deselected.

Figure (7) illustrates the persistence effect on an example involving *ReloadOnDark*. The salience of *ReloadOnDark* is proportional to the lack of *Potential Energy*, and while the robot reloads, its salience decreases. With the WTA architecture, another salience can easily interrupt this action before the *Potential Energy* has been fully reloaded. In contrast, if correctly tuned, the GPR robot is able to completely reload, because the salience for *ReloadOnDark* is reinforced by the persistence signal which therefore increases the duration of the bout.

--Insert Fig.7 about there –

One of the advantages of having a persistence effect is to prevent 'dithering', that is, switching frequently between different actions. As mentioned before, avoiding this oscillation is an issue for most of the engineering-designed architectures for action selection**.** Dithering may be particularly deleterious where there are significant costs associated with unnecessary switching between one action and another. However, in other situations, frequent interruptions of a selected action may actually be appropriate. For example, if an animal eats in a dangerous area, it should break away from its meal on a regular basis to check for predators. As a consequence, an efficient action selection system should be able to control the level of behavioural persistence according to circumstances. The following will show that it is the case for the GPR architecture, not for the WTA one.

In order to demonstrate the importance of appropriate persistence, the *Grooming* act was added to the previous behavioural repertoire of the robot. Note that this act has no effect on the robot's metabolism, but that it can compete with other actions. For example, a robot that is situated on a black tile (potential energy source), is currently starved of E_{Pot}, and has a high *Dirtiness* value, may 'hesitate' to display either an ingesting or a grooming act, as shown on Fig.(8). By appropriate tuning of the *Persistence* weights the GPR robot can overcome this difficulty for both actions. Lacking this possibility, the WTA robot necessarily oscillates between these behaviours. As a consequence, the GPR robot can show persistence or rapid switching as required by circumstances.

--Insert Fig.8 about there –

3.3. Experiment 3 (GPR: 5 runs; WTA: 6 runs)

Despite its lack of persistence, the WTA robot was able to survive during Experiment 1 by increasing the frequency of its reloading actions. However, the short duration of these bouts and the low percentage of time during which it is completely reloaded suggest that it spends less time than the GPR robot in a 'comfort zone' well away from the boundary of viability. The conditions of Experiment 1 did not allow the GPR robot to make use of this benefit. Indeed, when the GPR robot had both high *Energy* and *Potential Energy* levels, none of its four actions were relevant to its situation, however, it was nevertheless required to choose one. The addition of *Rest* – a low energy cost act – to the behavioural repertoire of the robots in Experiment 1 will test the capability of both selection architectures to save energy when other actions are not particularly relevant, that is, when their saliences are close to zero.

--Insert Fig.9 about there –

--Insert Tables IV & V about there –

In this experiment, all runs for both robots were successful and lasted more than one hour. As in Experiment 1, both architectures clearly succeeded in keeping the robot's essential variables within the *viability zone*.

The results of Table IV show that both robots activate *Rest* with a similar frequency per hour. However, as expected, the duration of a *Rest* bout is significantly longer for the GPR robot. As a consequence, it consumes less *Energy* than the WTA and needs to perform fewer reloading

actions (Table IV) than in Experiment 1, in which *Rest* was not available. In this situation, it can extract from the environment significantly less E_{Pot} than the WTA (Table V).

The more frequent reloading actions displayed by the WTA robot allow it to maintain a higher level of *Energy,* but the conjunction of more frequent conversions of E_{Pot} into E and incomplete reloads prevent it from reaching a higher level of *Potential Energy* (Fig. (9) and Table V). In contrast, the GPR robot can maintain a high level because, on the one hand, it can take advantage of its ability to save *Energy* and, on the second hand, it can display more efficient reloading actions. Indeed, despite extracting less E_{Pot} over the course of the experiment the GPR achieves a higher median level of E_{Pot} overall (see Fig.(10)), reaching over 95% of the maximum charge during more than 45% of the time.

--Insert Fig.10 about there –

4. Discussion

Building on the work of Gurney, Prescott and Redgrave [17], [18], our objective was to evaluate the vertebrate basal ganglia as a possible neural substrate for action selection. We have shown that the model is able to generate adaptive switching between several acts when embedded in a robot that has to 'survive' in a real environment. The comparison with WTA served to highlight some adaptive properties specific to the GPR model, such as the avoidance of dithering and energy-saving that derives from its capacity to generate appropriate behavioural persistence. Though it is possible to add persistence to the WTA architecture via a simple feedback loop - like *TH* loop in GPR -, a control mechanism - like *BGII* in GPR - is then mandatory to avoid overload.

One adaptive effect of persistence is that it can maintain the robot internal variables at more comfortable levels, helping it to survive any temporary upset in the availability of resources (Experiments 1 and 3). Another effect is that it serves to avoid dithering, the main issue of most engineering architectures of action selection (Experiment 2). It also allows the GPR robot to save energy with a longer display of a low-cost act when other actions are not contextually relevant (Experiment 3). A final less intuitive adaptive effect is that persistence can `prime' the robot to anticipate forthcoming opportunities for action. For instance, due to the low communication frequency between the robot and the PC, we noticed that the WTA robot often stops *after* it has driven past the central brightest (or darkest) patch on the gradient tiles, whereas the GPR robot generally manages to stop closer to the patch centre. What happens is that the corresponding salience increases slightly as the GPR robot enters the brighter (or darker) area. Although this is not enough, in itself, to prompt a change in the selected action, the positive feedback begins to build up the salience so that, when the robot eventually reaches the centre, it is able to select the appropriate action more rapidly. This increased responsiveness is possible because the brightness (or darkness) gradient serves to prime the appropriate action.

This work lends additional weight to the proposal that the basal ganglia control loops implemented by the GPR model may serve as a neural substrate for the adaptive benefits of persistence. It also suggests some improvements to the model. In particular, as stated by physiologists and ethologists, we know that persistence varies accordingly to various contextual factors. For example, McFarland [25] pointed out that the duration of feeding bouts in rats could be diversely triggered by the stimulation of oral or of gut receptors. The persistence effect on feeding and drinking bouts in rats was also shown to depend on learning, diurnal and nocturnal conditions [23]. According to McFarland and Lloyd [26] and the 'time-sharing' hypothesis, an action may also show a 'hidden' persistence, even after its execution has been interrupted. For

example, a 'dominant' act may be temporarily suspended to allow an alternative behaviour to be expressed, only later resuming its performance. In this case, the 'salience' of the dominant act persists even though the behaviour itself is deselected for a short while.

In the GPR model, the duration of behavioural persistence could also be sensitive to contextual variables, since salience is a function of many factors of which positive feedback is only one. But the persistence weights - together with the salience weights – are still tuned 'by hand' to suit different environmental situations. One way to improve the model is to consider biological hypotheses on learning in the basal ganglia that have already been implemented in various computational models (see [22] for a review). The activation patterns of dopamine neurons within the striatum - e.g., shifting back from responding to a primary reward to a reward-predicting stimulus [35] - has been shown to be very similar to that generated by machine learning algorithms, in particular Temporal Difference models [6]. These models distinguish an 'actor', which learns to display actions so as to maximize the weighted sum of future rewards, and a 'critic', which computes this sum on line. The GPR model can be assimilated to an 'actor' – with details usually neglected by the preceding models – to which a critic should be added to exhibit efficient learning capabilities. Although, according to [31], [22], TD learning inspired models of the basal ganglia are built on suppositions that are incompatible with observed features in the basal ganglia anatomy and physiology, some of them have succeeded in closely simulating the observed activations of striatal neurons in conditioning responses [11]. Accordingly, these models - or their alternative [31] - can be a future support for inclusion of learning processes in the GPR control architecture.

This robotic embodiment is part of an ongoing, multi-partner project which aims to synthesize *Psikharpax*, an `artificial rat', in which such biomimetic mechanisms for action selection will be combined with biomimetic mechanisms for navigation [13], both inspired by existing structures

in the rat brain. For that purpose, the current model will be further refined to match the particular characteristics of the nucleus accumbens (ventral striatum) in order to investigate its role in integrating spatial, sensorimotor and motivational information [16], [1], [9].

Acknowledgements

This research has been granted by the Project *Robotics and Artificial Entities* (ROBEA) of the Centre National de la Recherche Scientifique, France. The authors thank Sébastien Laithier for his useful contribution.

Abbreviations

AO *AvoidObstacle* action

BG basal ganglia

BGI 'selection part' of the GPR

BGII 'control part' of the GPR

B_L left bumper sensor value

B_R right bumper sensor value

D1 striatal neurons containing D1 dopamine receptors

D2 striatal neurons containing D2 dopamine receptors

E energy

E_{Pot} potential energy

EP entopeduncular nucleus

G *Grooming* action

GP globus pallidus

GPR Gurney, Prescott and Redgrave model of basal ganglia

L_B brightness sensor value

L_D darkness sensor value

P persistence

R *Rest* action

ROB *ReloadOnBright* action

ROD *ReloadOnDark* action

SNr substantia nigra pars reticulata

STN sub-thalamic nucleus

T_{Digest} duration of ROB action

TH 'thalamic part' of the GPR

T_{Ingest} duration of ROD action

TRN thalamic reticular nucleus

VL ventro-lateral thalamus

W *Wander* action

WTA winner-takes-all action selection mechanism

5. References

[1] Albertin, S., Mulder, A., Tabuchi, E., Zugaro, M., and Wiener, S. Lesion of the medial shell of the nucleus accumbens impair rats in finding larger rewards, but spare reward-seeking behavior, *Behav. Brain Res.* 117 (2000) pp. 173-183.

[2] Albin, R. L., Young, A. B., and Penney, J. B. The functional anatomy of basal ganglia disorders, *Trends Neurosci*. 12 (1989) pp. 366-375.

[3] Arbib, M. Introducing the neuron. In *The handbook of brain theory and neural networks*, ed. By Arbib, M. (The MIT Press, Cambridge MA, 1995).

[4] Ashby, W.R. *Design of a Brain* (New York: John Wiley & Sons, 1952)

[5] Atkinson, J. and Birch, D. *The dynamics of action* (New York: John Wiley & Sons, 1970)

[6] Barto, A. G. Adaptive critics and the basal ganglia, In *Models of Information Processing in the Basal Ganglia* ed. by Houk, J. C., Davis, J. and Beiser D. (The MIT Press, Cambridge, MA, 1995).

[7] Beiser, D. G., Hua, S. E., and Houk, J. C. Network models of the basal ganglia, *Current Opinion Neurobiol.* 7 (1997) pp. 185-190.

[8] Blumberg, B. Action-Selection in Hamsterdam: Lessons from Ethology. In *From Animals to Animats 3. Proceedings of the Third International Conference on Simulation of Adaptive*

Behavior, ed. by Cliff, D., Husbands, P., Meyer, J-A., & Wilson, S.W. (The MIT Press, Cambridge, MA, 1994).

[9] Cardinal, R.N., Parkinson, J.A., Hall, J. and Everitt, B.J. Emotion and motivation: the role of amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev*. 26 (2002) pp.321-352.

[10] Chevalier, G. and Deniau, M. Disinhibition as a basic process of striatal functions, *Trends Neurosci.* 13 (1990) pp. 277-280.

[11] Contreras-Vidal, J. L. and Schultz, W. A predictive reinforcement model of dopamine neurons for learning approach behavior, *J. Comp. Neurosci.* 6 (1999) pp. 191-214.

[12] Feldman, J. and Ballard, D. Connectionist models and their properties, *Cog. Sci.* 6 (1982) pp.205-254.

[13] Filliat, D. and Meyer, J.-A. Global localization and topological map learning for robot navigation. In *From Animals to Animats 7. Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, ed. by Hallam, B., Floreano, D., Hallam, J., Hayes, G., and Meyer, J.-A. (The MIT Press, Cambridge, MA, 2002).

[14] Gillies, A. and Arbruthnott, G. Computational models of the basal ganglia. *Movement Disorders*, 15 (2000) pp. 762-770.

[15] Girard, B., Cuzin, V., Guillot, A., Gurney, K. N., and Prescott, T. J. Comparing a bioinspired robot action selection mechanism with winner-takes-all. In *From Animals to Animats 7. Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, ed. by Hallam, B., Floreano, D., Hallam, J., Hayes, G., and Meyer, J.-A. (The MIT Press, Cambridge, MA, 2002).

[16] Groenewegen, H. J., Mulder, A. B., Beijer, A., Wright, C. I., Silva, F. L. D., and Pennartz, C. M. A. Hippocampal and amygdaloid interactions in the nucleus accumbens, *Psychobiol.* 27 (1999), pp. 149-164.

[17] Gurney, K., Prescott, T. J., and Redgrave, P. A computational model of action selection in the basal ganglia. I. A new functional anatomy, *Biol. Cyber.* 84 (2001a) pp. 401-410.

[18] Gurney, K., Prescott, T. J., and Redgrave, P. A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour, *Biological Cybernetics*, 84 (2001b) pp. 411-423.

[19] Houk, J. C., Davis, J. and Beiser D. (Eds.) *Models of Information Processing in the Basal Ganglia* (Cambridge, MA: MIT Press, 1995)

[20] Houston, A. and Sumida, B. A positive feedback model for switching between two activities. *Anim. Behav.* 33 (1985) pp. 315-325.

[21] Humphries, M. D. and Gurney, K. N. The role of intra-thalamic and thalamocortical circuits in action selection. *Network.* 13 (2002) pp. 131-156.

[22] Joel, D., Niv, Y., and Ruppin, E. Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Networks*, 15 (2002). pp.4-6.

[23] Le Magnen, J. *Hunger*. (Cambridge University Press, Cambridge, UK, 1985).

[24] Ludlow, A. The behaviour of a model animal. *Behav.* 58 (1976) pp.31-172.

[25] McFarland, D. *Feedback mechanisms in animal behaviour* (Academic Press, London, 1971).

[26] McFarland, D. and Lloyd, D Time-shared feeding and drinking. *Quart. J. Exp. Psychol.* 25 (1973) pp. 48-61.

[27] McFarland, D. and Sibly, R. The behavioural final common path. *Phil. Trans. Royal Soc. London*, 270 (1975) pp.265-293.

[28] Mel, B. W. Synaptic integration in an excitable dendritic tree. *J. Neurophysiol*, 70 (1993) pp.1086-1101.

[29] Mink, J. W. The basal ganglia: Focused selection and inhibition or competing motor programs, *Prog. Neurobiol.* 50 (1996) pp.381-425.

[30] Montes Gonzalez, F. Prescott, T.J., Gurney, K. Humphries, M., and Redgrave, P. An embodied model of action selection mechanisms in the vertebrate brain. In *From Animals to Animats 6: Proceedings of the 6th International Conference on Simulation of Adaptive Behaviour*, ed. by Meyer, J-A. , Berthoz, A., Floreano, D. , Roitblat, H. and Wilson, S.W. (The MIT Press: Cambridge MA, 2000).

[31] Pennartz, C. M. A. Reinforcement learning by hebbian synapses with adaptive threshold. *Neurosci.* 81 (1997) pp.303-319.

[32] Pennartz, C., M. A., McNaughton, B. L., and Mulder, A. B. The glutamate hypothesis of reinforcement. *Prog. Brain Res*. 126 (2000) pp.231-253.

[33] Prescott, T. J., Redgrave, P., and Gurney, K. N. Layered control architectures in robot and vertebrates. *Adap. Behav*. 7 (1999) pp.99-127.

[34] Redgrave, P., Prescott, T. J., and Gurney, K. The basal ganglia: a vertebrate solution to the selection problem? *Neurosci*. 89 (1999) pp.1009-1023.

[35] Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80 (1998) pp.1-127.

[36] Spier, E. and McFarland, D. A fine-grained motivational model of behaviour sequencing. In *From Animals to Animats 4. Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, ed. by In Maes, P., Mataric, M. J., Meyer, J.-A., Pollack, J., and Wilson, S. W. (The MIT Press: Cambridge, MA, 1996).

[37] Tyrrell, T. The use of hierarchies for action selection. *Adap. Behav*. 1 (1993) pp.387-420.

[38] Webb, B. and Consi, T.R. (Eds) *Biorobotics: Methods and Applications.* (The MIT Press: Cambridge, MA, 2001).

[39] Wiepkema, P. Positive feedback at work during feeding. *Behav.* 39 (1971) pp.266-273.

[40] Yamada, W., Kock, C., and Adams, P. Multiple channels and calcium dynamics. In *Methods in neuronal modellings*, ed. by Koch, C., Segev, I. (The MIT Press, Cambridge, MA, 1989)

Appendix 1. Weight parameters for the GPR model

Module	Threshold	Slope
D1 striatum	0.2	0.35
D2 striatum	0.2	0.35
STN	-0.25	0.35
GP	-0.2	1
EP/SNr	-0.2	1
Persistence	0	1
TRN	0	0.5
VL	-0.8	0.62

Neurons parameters

Time constant K= 0.25

Lambda of the Dopamine levels = 0.2

Table I. Salience calculations. The transfert functions Rev(x) and Circ(x) stand respectively for (1-x) and the square root of (1-x²).

Actions	Saliences calculations	
Wander	WTA	$-B_L - B_R + 0.5 \times \mathrm{Rev}(E_{Pot}) + 0.7 \times \mathrm{Rev}(E)$
	GPR	$-B_L - B_R + 0.8 \times \mathrm{Rev}(E_{Pot}) + 0.9 \times \mathrm{Rev}(E)$
AvoidObstacle	WTA	$3\ B_L + 3\ B_R$
	GPR	$2\ B_L + 2\ B_R + 0.5\ P_{AO}$
ReloadOnDark	WTA	$-2\ L_B - B_L - B_R + 3\ L_D \times \mathrm{Rev}(E_{Pot})$
	GPR	$-2\ L_B - B_L - B_R + 3\ L_D \times \mathrm{Rev}(E_{Pot}) + 0.4\ P_{ROD}$
ReloadOnBright	WTA	$-2\ L_D - B_L - B_R + 3\ L_B \times \mathrm{Circ}(\mathrm{Rev}(E_{Pot})) \times \mathrm{Rev}(E)$
	GPR	$-2\ L_D - B_L - B_R + 3\ L_B \times \mathrm{Circ}(\mathrm{Rev}(E_{Pot})) \times \mathrm{Rev}(E) + 0.5\ P_{ROB}$
Rest	WTA	$-B_L - B_R + 0.1$
	GPR	$-B_L - B_R + 0.6\ P_R$

Table II. Experiment 1. Comparison (U Mann-Whitney test) between the GPR and WTA robots of all runs cumulated. Top: Act durations. Bottom: Frequency of activations of each act per hour.

Duration	W	ROD	ROB	AO
GPR				
M	50	253	212	34
range	46 : 52	133.5 : 356	145 : 268	31 : 38
WTA				
M	46	141	139	20
range	40 : 48	98 : 246	112 : 152	20 : 20
U =	24	8	3	3
	p> 0.05	**p< 0.01**	**p< 0.01**	**p< 0.01**

Frequency	W	ROD	ROB	AO
GPR				
M	272.52	28.79	49.98	233.05
range	259.12 : 294.12	26.57 : 45.58	40.78 : 59.08	220.13 : 257.68
WTA				
M	433.79	40.58	51.96	331.42
range	393.37 : 470.97	24.71 : 49.17	48.29 : 61.42	295.20 : 403.69
U =	0	18	20	0
	p<0.01	**p<0.05**	**p<0.05**	**p<0.01**

Table III. Experiment 1. Comparison (U Mann-Whitney test) between the GPR and WTA robots of the medians of *Energy* (*E*), of *Potential Energy* (E_{Pot}), and of *Potential Energy extracted* (E_{Pot} *extr*.) per run.

		E	E_{Pot}	E_{Pot} *extr.* (10^{-3})
GPR	M	0.78	0.75	2.3
	range	0.68 : 0.81	0.65 : 0.86	2.4 : 2.2
WTA	M	0.77	0.77	2.2
	range	0.72 : 0.81	0.71 : 0.83	2.0 : 2.6
	U =	26	43	27
		p> 0.05	p> 0.05	p> 0.05

Table IV. Experiment 3. Comparison (U Mann-Whitney test) between the GPR and WTA robots of all runs cumulated. Top: Act durations. Bottom: Frequency of activations of each act per hour.

Duration		W	ROD	ROB	AO	R
GPR	M	52.5	302	294	34	1728
	range	49 : 56	233.5 : 348	233 : 308	33 : 35	1445 : 2116.5
WTA	M	48	161	150	20	485
	range	48 : 49	132 : 192	120 : 168	20 : 24	340 : 568
	U =	1	0	0	2	0
		p< 0.01	**p< 0.01**	**p< 0.01**	**p< 0.05**	**p< 0.01**

Frequency		W	ROD	ROB	AO	R
GPR	M	195.35	27.20	44.99	143.64	10.06
	range	182.42 : 239.65	16.01 : 32.30	29.61 : 56.94	137.52 : 175.58	9.29 : 12.74
WTA	M	427.22	52.12	74.22	306.40	8.75
	range	408.67 : 436.53	45.05 : 61.47	57.19 : 81.97	301.41 : 313.96	4.71 : 9.62
	U =	0	0	0	0	13
		p< 0.01	**p< 0.01**	**p< 0.01**	**p< 0.01**	p> 0.05

Table V. Experiment 3. Comparison (U Mann-Whitney test) between the GPR and WTA robots of the medians of *Energy* (*E*), of *Potential Energy* (E_{Pot}), and of *Potential Energy extracted* (E_{Pot} *extr*.) per run.

		E	E_{Pot}	E_{Pot} *extr.* (10^{-3})
GPR	M	0.78	0.937	1.8
	range	0.757 : 0.792	0.875 : 0.973	1.7 : 1.9
WTA	M	0.8	0.81	2.2
	range	0.788:0.816	0.796 : 0.855	2.1 : 2.2
	U =	1	0	0
		p< 0.01	**p< 0.01**	**p< 0.01**

Figure captions

Figure 1. Basal ganglia in the rat brain (gray areas).

Figure 2. The GPR model (see text for explanations). Arrows represent excitatory connections, blobs inhibitory connections. Weight parameters are shown next to their respective pathways.

Figure 3.
Left: The environment showing `food' (A) and `nest' (B) locations.
Right: the Lego Mindstorm robot. (A): the light sensors; (B): the bumpers.

Figure 4.
Left: (a) Input saliences, (b) output EP/SNr signals and (c) the corresponding behavioural sequence generated by the GPR model.
Right:(d) Input and output signals (saliences) and (e) the corresponding behavioural sequence generated by a WTA. The abscissa shows the number of cycles where 1450 cycles correspond to 100 sec.

Figure 5. Scatter plots of the median values of Energy (top) / Potential Energy (bottom) in Experiment 1 for each run of the GPR and the WTA robots.

Figure 6. Percentages of overall time (on y-axis) during which Potential Energy is reloaded at the levels shown on x-axis (the maximum charge is 1). GPR: white, and WTA: black (all runs cumulated).

Figure 7. Effect of persistence in GPR.
From top to bottom:
`Raw' salience (i.e. without persistence) of ReloadOnDark; Output EP/SNr signals; the corresponding behavioural sequence generated by the GPR robot: (A) points where the switch would happen without persistence, (B) points where the switch actually takes place. On x-axis: number of computation cycles (250 cycles correspond to approx. 17 sec).

Figure 8. Control of dithering: with a GPR architecture (left) it is possible to control the oscillation length by adjusting the persistence parameters, while a WTA (right) necessarily dithers between acts.

Figure 9. Scatter plots of the median values of Energy (top) / Potential Energy (bottom) in Experiment 3 for each run of the GPR and WTA robots.

Figure 10. Percentages of overall time (on y-axis) during which Potential Energy is reloaded at the levels shown on x-axis (the maximum charge is 1). GPR: white and WTA: black (all runs cumulated).

Fig. 1

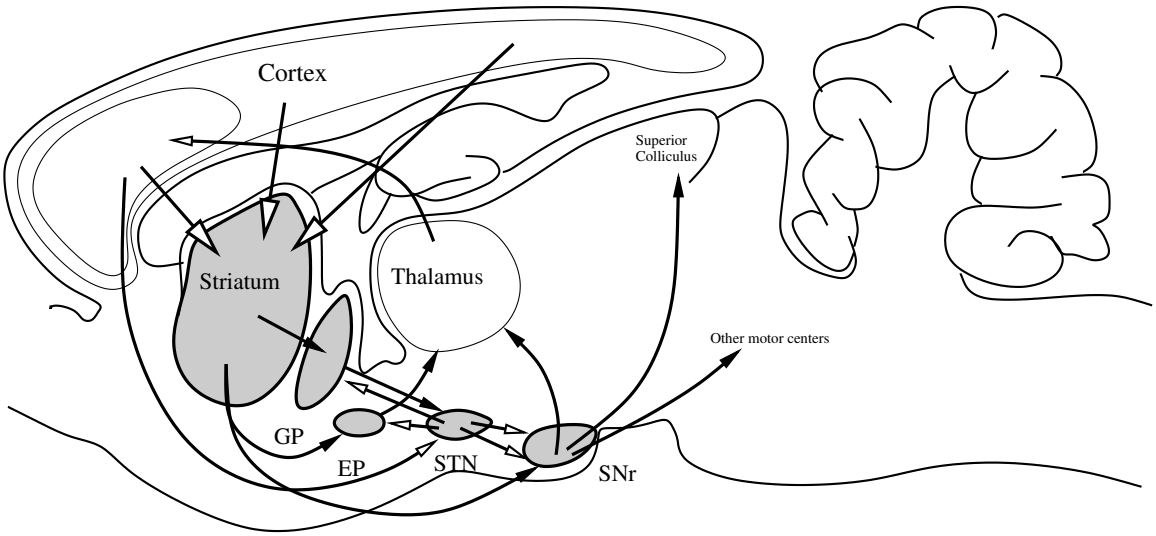

Cortex
Superior
Colliculus
Striatum
Thalamus
Other motor centers
GP
EP
STN
SNr

Fig.2

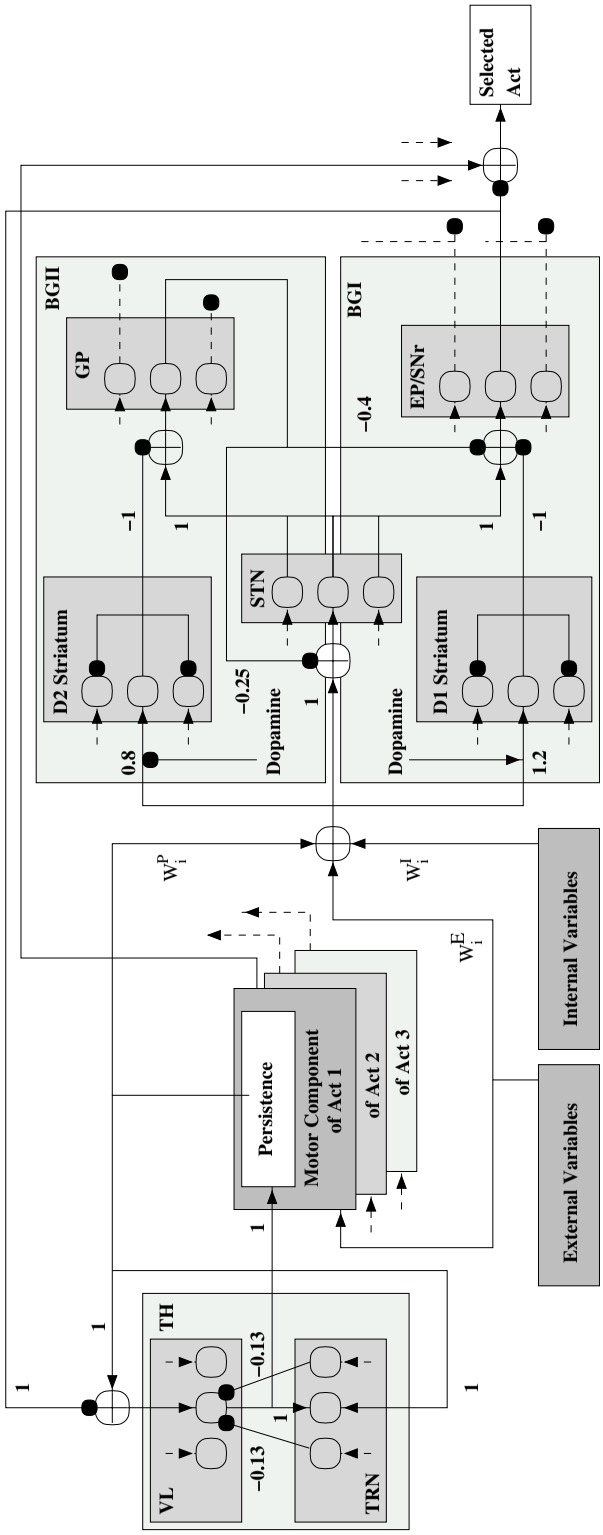
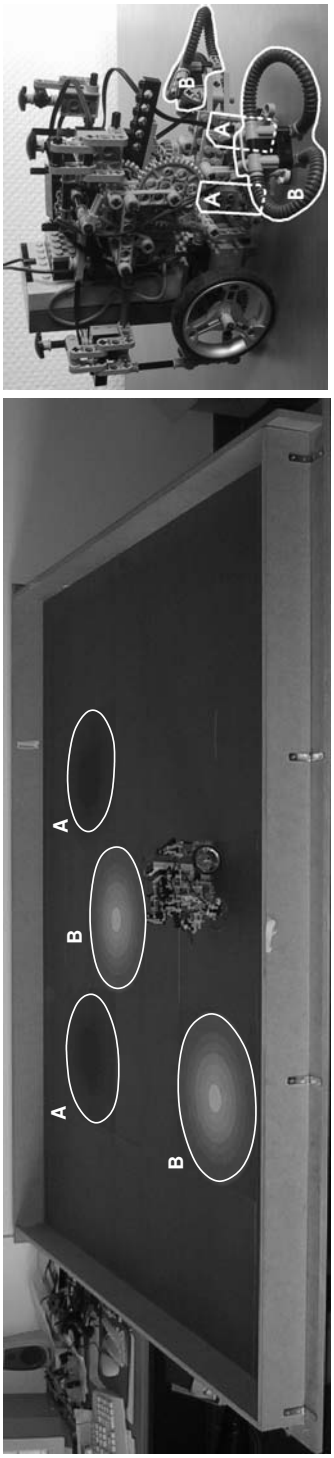

Selected Act
BGII
GP
BGI
EP/SNr
−0.4
−1
1
1
−1
STN
D2 Striatum
D1 Striatum
−0.25
Dopamine
1
Dopamine
0.8
1.2
W_i^P
W_i^I
W_i^E
Internal Variables
External Variables
Persistence
Motor Component of Act 1
of Act 2
of Act 3
1
TH
VL
TRN
−0.13
−0.13
1
1
1

Fig.3



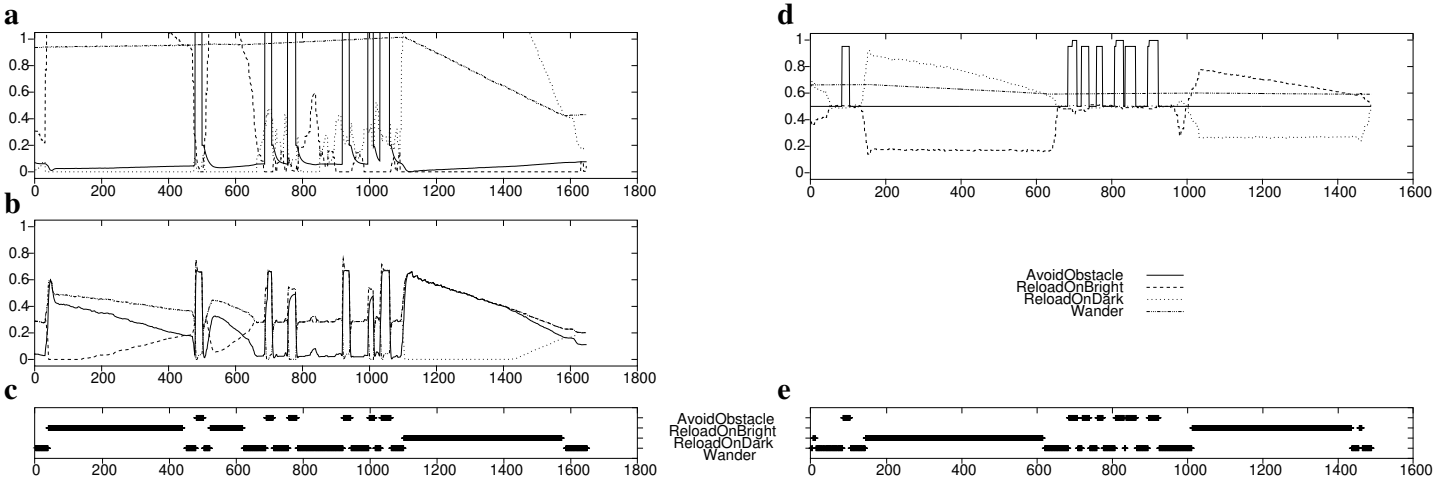

a
b
c
d
e
AvoidObstacle
ReloadOnBright
ReloadOnDark
Wander

Fig.4

Fig.5

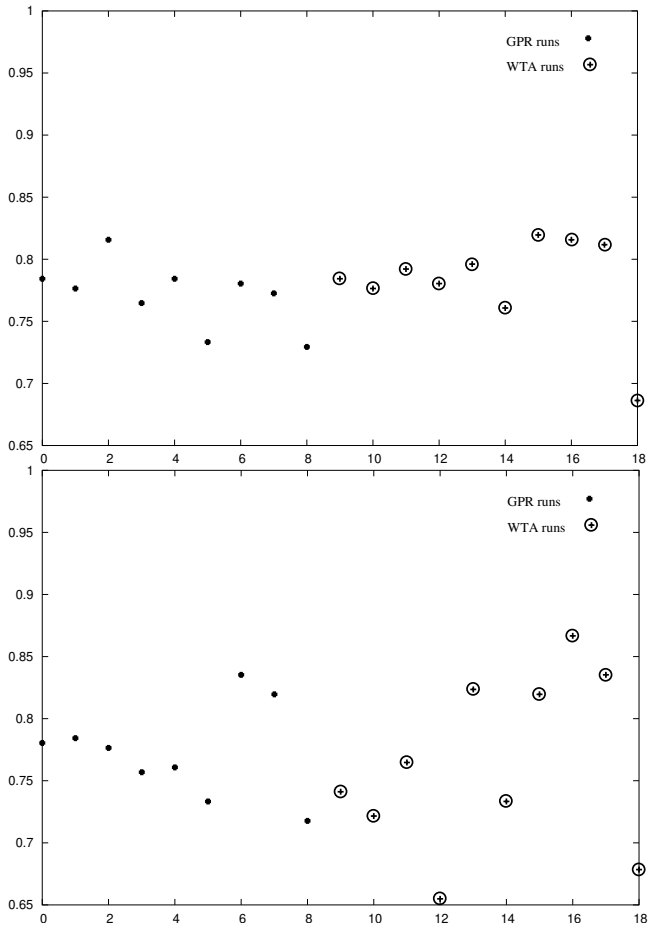

GPR runs
WTA runs
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0
2
4
6
8
10
12
14
16
18
GPR runs
WTA runs
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0
2
4
6
8
10
12
14
16
18

Fig.6

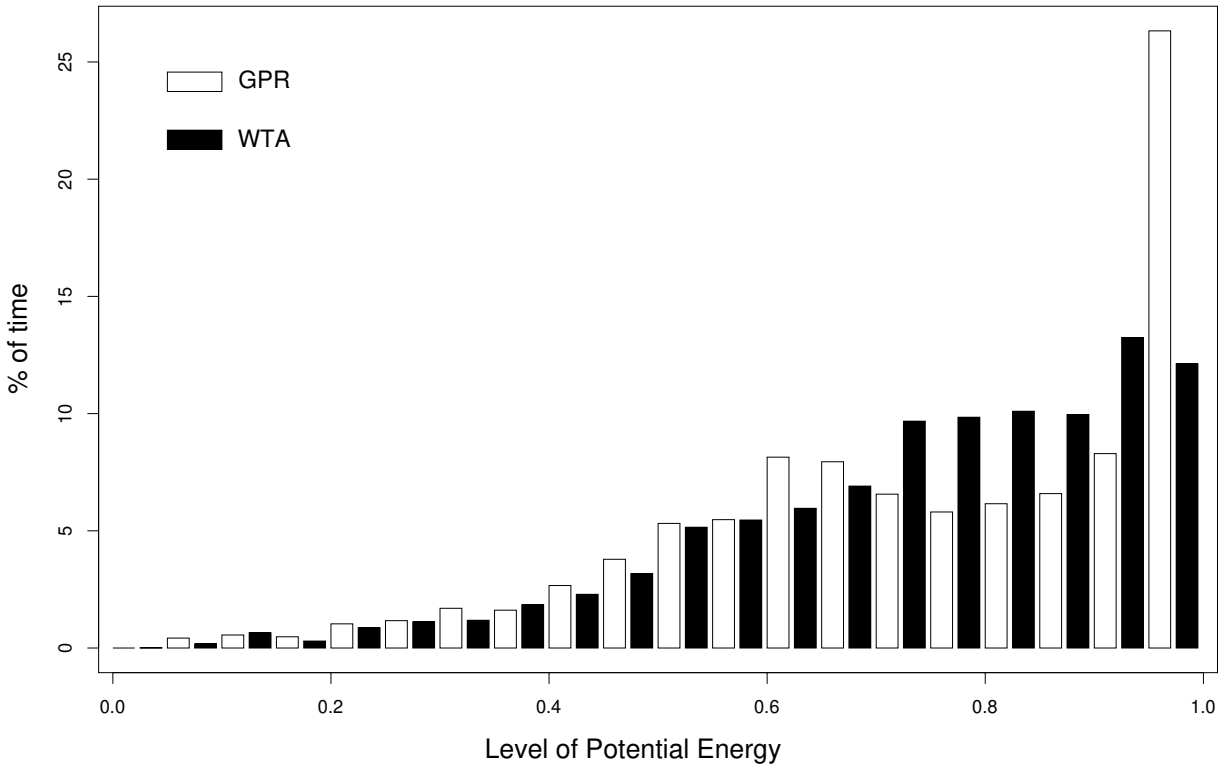

GPR
WTA
% of time
0
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0
Level of Potential Energy

Fig.7

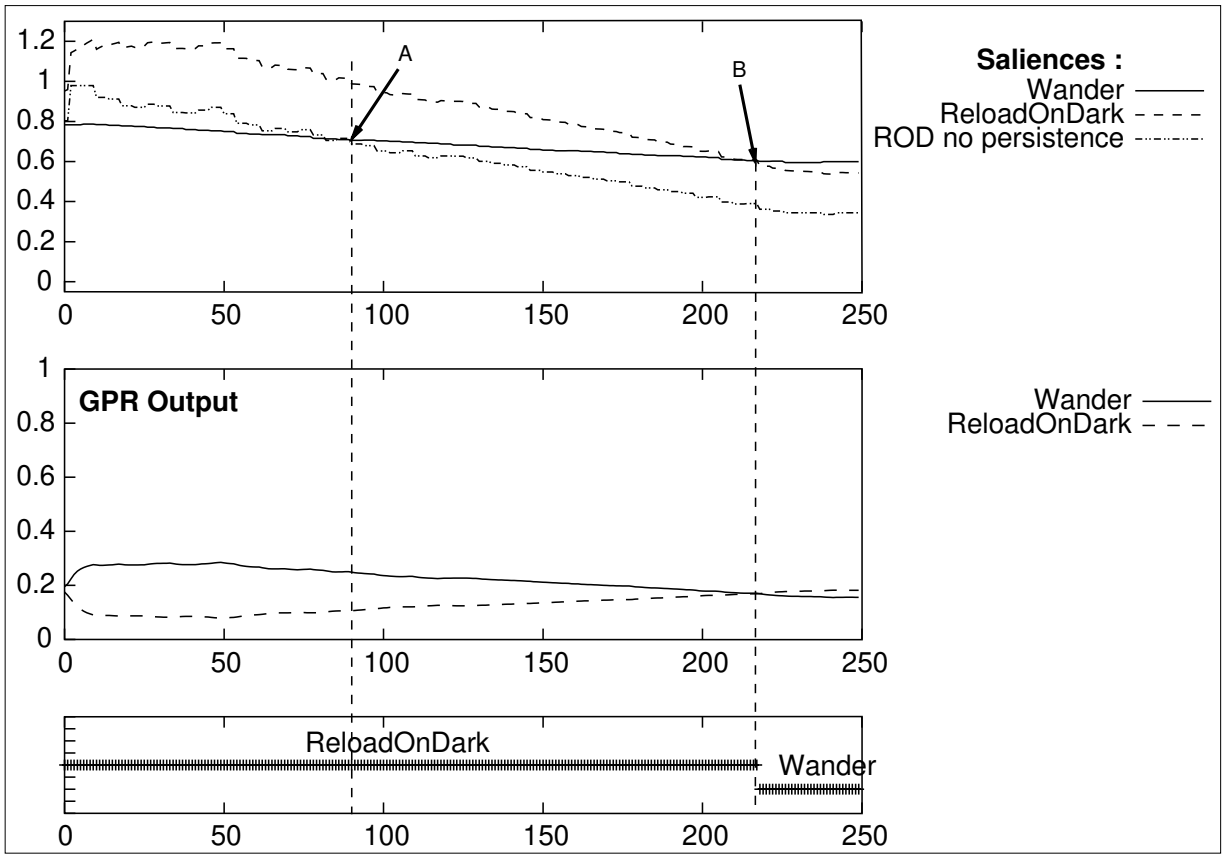

1.2
1
0.8
0.6
0.4
0.2
0
A
B
0
50
100
150
200
250
Saliences :
Wander
ReloadOnDark
ROD no persistence
1
GPR Output
0.8
0.6
0.4
0.2
0
Wander
ReloadOnDark
ReloadOnDark
Wander

Fig.8

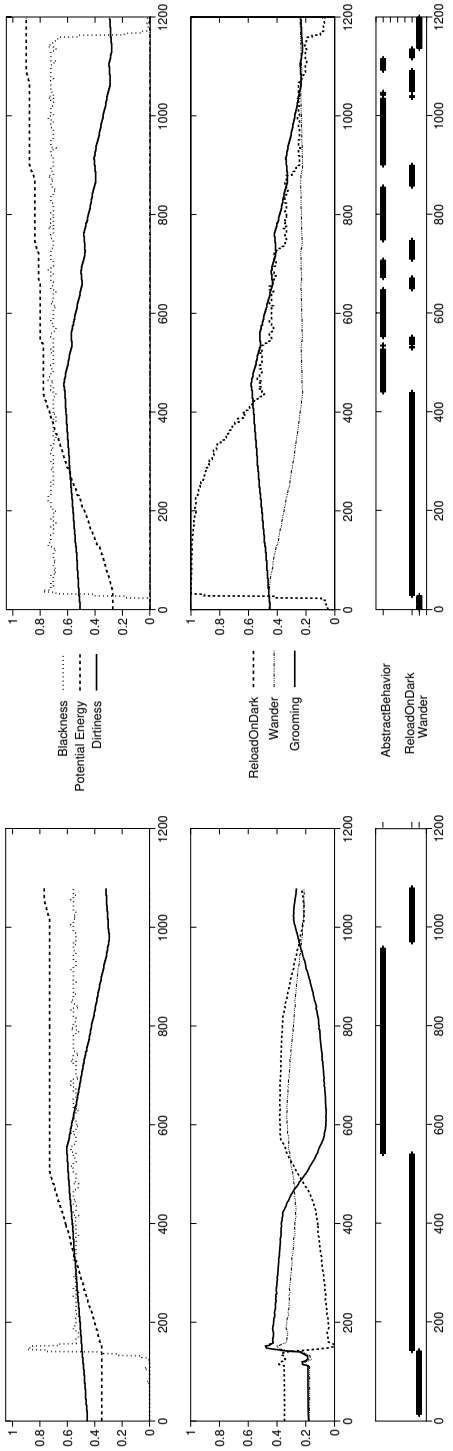

Blackness
Potential Energy
Dirtiness
ReloadOnDark
Wander
Grooming
AbstractBehavior
ReloadOnDark
Wander

Fig.9

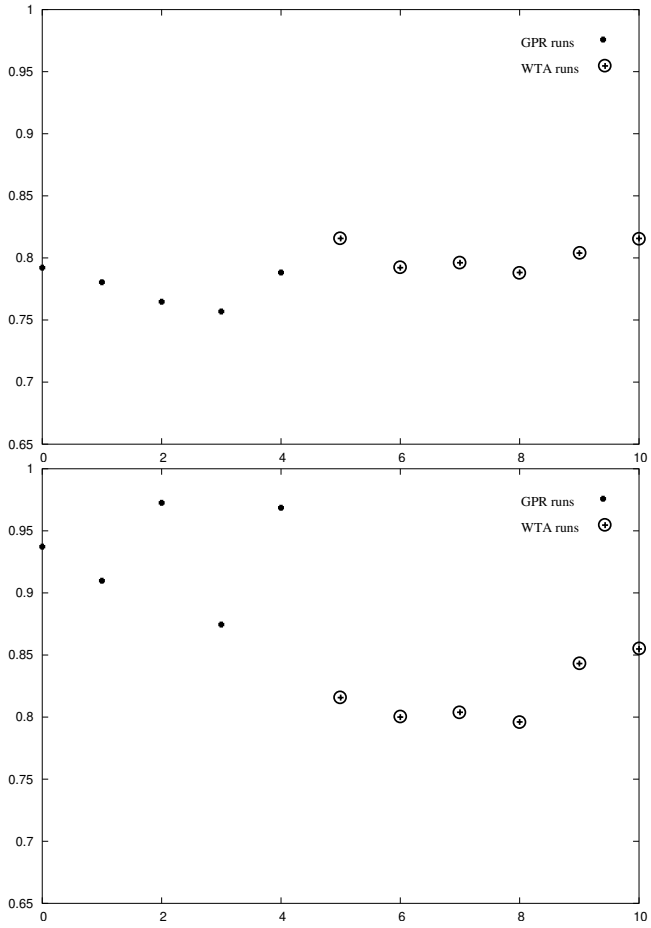
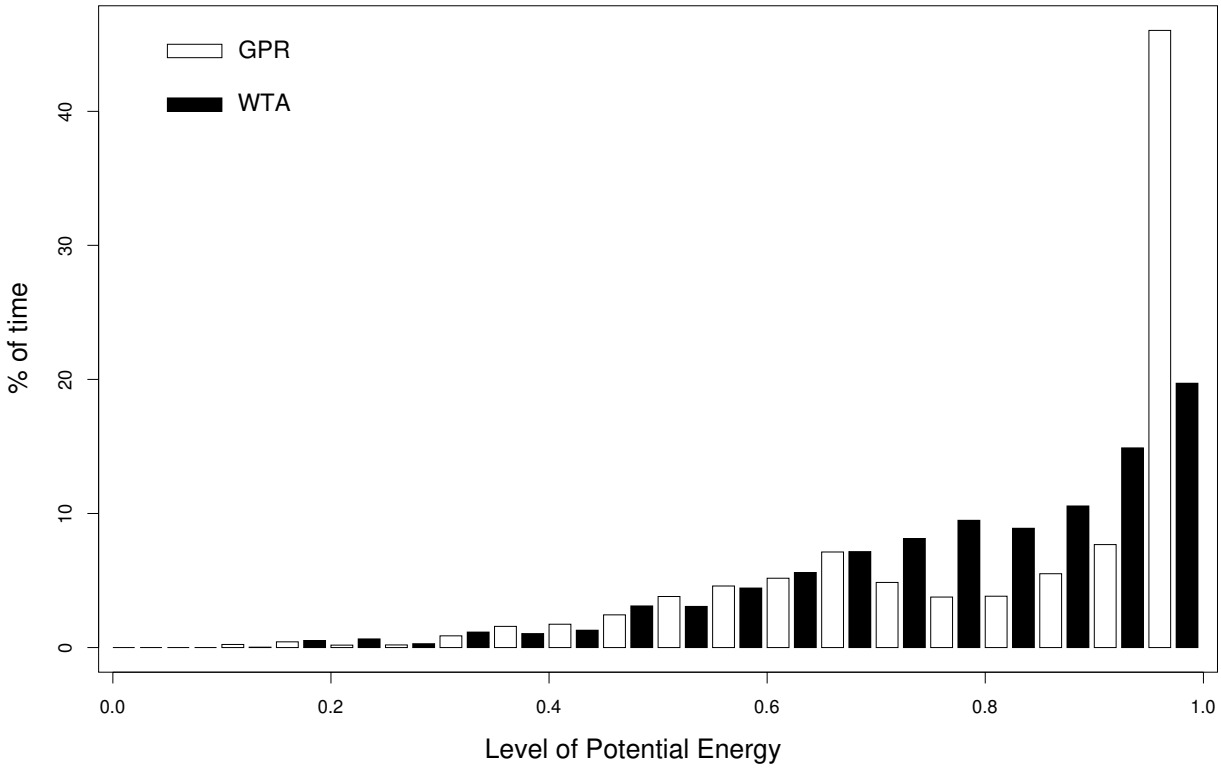

GPR runs
WTA runs
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0
2
4
6
8
10
GPR runs
WTA runs
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0
2
4
6
8
10

Fig.10


GPR
WTA
% of time
0
10
20
30
40
0.0
0.2
0.4
0.6
0.8
1.0
Level of Potential Energy