Comparing a brain-inspired robot action selection mechanism with 'winner-takes-all'

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

* AnimatLab-LIP6
8, rue du capitaine Scott
75015 Paris, France
{benoit.girard,vincent.cuzin,agnes.guillot}@lip6.fr

** Department of Psychology
University of Sheffield
Western Bank, Sheffield S10 2TP, UK
{k.gurney,t.j.prescott}@sheffield.ac.uk

Abstract

We present a new robotic implementation of a brain-inspired model of action selection described by Gurney et al. (Gurney et al., 2001a, Gurney et al., 2001b) based on neural circuits located in the basal ganglia and thalamus of the vertebrate brain. Compared to an earlier robot implementation (Montes-Gonzalez et al., 2000), the new model demonstrates the capacity of the selection system to produce efficient 'energy' consumption/conversion in a 'feeding/resting' task whilst maintaining essential state variables within a 'zone of viability'. Generating appropriate action selection in this new setting entailed using biologically plausible Sigma-Pi units that can exploit correlated and anti-correlated dependencies between input signals when computing the 'salience' (urgency) of competing actions. A comparison between this brain-inspired selection mechanism and classical 'winner-takes-all' showed that the former can provide better behavioral persistence leading to more efficient energy intake.

1 Introduction

If the behavior of an animal or a robot is viewed as a discrete sequence of actions, then an understanding is needed of the mechanisms underlying the switching of behavior from one action to the next. In ethology several speculative hypotheses have been proposed concerning the action selection mechanisms underlying animal behavior switching. These hypotheses generally suppose that the motivational systems associated with a given act could win because they directly or indirectly activate, inhibit, or disinhibit their competitors. In the 1970's and 1980's, the mechanisms proposed for such interactions tried to explain the transitions between various behaviors in fishes, birds and rodents (Baerends et al., 1970, Ludlow, 1976, McFarland, 1977, Slater, 1978, Houston and Sumida, 1985). Eventually, ethologists lost interest in these models as they were unable to find a relationship between these speculative mechanisms and plausible biological equivalents.

Since the 1990's, with the rise of the animat approach, these models have been rediscovered and, with the improvement of computer methods, more precisely investigated (see Prescott et al. 1999, Guillot and Meyer, 2000, for reviews). However, many of the issues deriving from the earlier animal studies remain to be resolved (see Snaith and Holland 1991, Tyrrell, 1993, for reviews).

In recent years, a growing number of neurobiologists have become interested in a group of centrally-located brain structures known as basal ganglia as a possible neural substrate for action selection (for reviews see Redgrave et al., 1999, Prescott et al. 1999). According to Redgrave et al. (1999) centralised action selection could be important for large brains in order to achieve effective conflict resolution between competing sensorimotor systems whilst maintaining a cap on the connectivity and energy costs of the arbitration mechanisms. Several computational models of these neural structures have been investigated in a variety of simulation tasks (see Houk et al., 1995 for a representative selection). However, only that due to Gurney, Prescott and Redgrave (2001a,b) (henceforth the GPR model) has demonstrated the capacity of the basal ganglia to provide effective action selection in a real robot (Montes-Gonzalez et al., 2000). Based on the connectivity of the rat's basal ganglia, the GPR model (more precisely described below) is composed of two main circuits, one that computes the selection of the action *per se* and another that modulates the function of the first and which controls how this selection is done. The inputs to the model are variables called 'saliences', that are weighted functions computed from sensory, proprioceptive and contextual information, denoting the urgency associated with each act. The outputs of the model are inhibitions assigned to each potential action. At each time-step, the act which is least inhibited is performed. A third circuit provides, via the thalamus and cortex, a

feedback loop whereby the output of the basal ganglia can influence its own future input, and in particular, enhance the salience signals of currently selected actions (Humphries and Gurney, 2001).

As noted by the authors, the GPR model exhibits three properties that are important for such mechanisms (Snaith and Holland, 1991, Prescott et al., 1999). The first is *clean switching* between actions: a competitor with a slight edge over its rivals should see the competition resolved rapidly in its favor. The second is *lack of distortion*: the presence of other candidates for the control of an effector should not interfere with the performance of the winning sub-system, once the competition has been resolved. The third is *persistence*: a winning act should remain active with lower input levels than were initially required for it to overcome the competition.

When embedded in a complete 'creature', in this case a Khepera robot, the GPR model displayed effective transitions between five actions (Montes-Gonzalez et al., 2000). The task of this robot was to mimic some of the behaviors of a hungry rat placed in a novel environment. Specifically, the robot was required to avoid open-spaces by moving towards wall and corners when the level of simulated fear was high at the start of the experiment, and to forage (by collecting wooden cylinders) when simulated hunger was relatively high (and fear relatively low) later in the experiment. This work also focussed on the effects of simulated dopamine modulation on the behavioral display. Dopamine is a neuromodulator known to have a critical effect on the function of the basal ganglia and behavioral switching more generally (see Redgrave et al., 1999).

In the current paper, we describe a second robot implementation of this model using a different robot platform, the Lego Mindstorms robot, and a task more typical of the type used in earlier action selection studies. Here the robot is required to select efficiently between four actions –wandering, avoiding obstacles, 'feeding' and 'resting'– in order to 'survive' in an environment where it can find 'food places' and 'rest places'. Its control architecture should be sufficiently adaptive to generate sequences of actions allowing it to remain as long as possible in its so-called *viability zone* (Ashby, 1952). This requires maintaining two essential state variables above minimal levels: *Potential energy* (obtained via 'feeding') and *Energy* (converted from Potential energy via resting). Spier and McFarland (1996) note that a 'two resource' problem of this type is a minimal scenario for evaluating an action selection or decision-making mechanism.

A further objective of this work is to investigate if and how the GPR model implements more than a simple 'winner-takes-all' (WTA) mechanism; a classical selection mechanism proposed long ago by engineers and ethologists (Atkinson and Birch, 1970). 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. Whilst the GPR model has a superficially similar property of selecting (albeit by disinhibition) the most highly motivated action, this is modulated by the effects of the control and feedback circuits, potentially resulting in different pattern of behavior switching compared to simple WTA. For instance, according to Prescott et al. (1999), although a WTA can display both clean switching and lack of distortion, the lack of a mechanism to support appropriate persistence could lead it to generate unadaptive 'dithering' between actions, an issue in action selection previously noted by ethologists (Atkinson and Birch, 1970, Houston and Sumida, 1985). A comparison of the two control architectures, embedded in the same robot in the same environment, should therefore demonstrate precisely what benefits the GPR control circuits can bring to the action selection process.

Following a summary of the GPR model in section 2, we will describe, in section 3, how this model was reimplemented within the control architecture of a Lego Mindstorms robot. In section 4, the results obtained with the model will be presented and compared with those of a WTA, and these will be discussed, in section 5, from the perspective of biological plausibility.

2 The GPR model

The details of the computational model and its correspondence with the neural anatomy are fully described in Gurney et al.(2001a,b). We will only summarize here the main features of the model as shown in Fig. 1.

The terminology used for component structures is based on those comprising the basal ganglia: the striatum, the globus pallidus (with subcomponents GPe and GPi), the sub-thalamic nucleus (STN), and the substantia nigra (SNr). The selection and control sub-circuits of the Basal Ganglia-based model are designated here for conciseness by *BGI* and *BGII* respectively.

In each compoment structure, each action is associated with a discrete channel, which is represented by a single artificial neuron. Each artificial neuron consists of a leaky integrator whose activation is driven by a weighted sum of inputs (in the work presented here, this is modified to include nonlinear contributions). Each neuron is supposed to represent a biological neural population so that the activity in the model of each unit represents the mean activity of the population as a whole. While these model neurons are not as physiologically realistic as those that use conductance based methods with multiple membrane compartments, they are configured in circuits that are anatomically realistic and afford a useful tool for investigating models at the systems level of

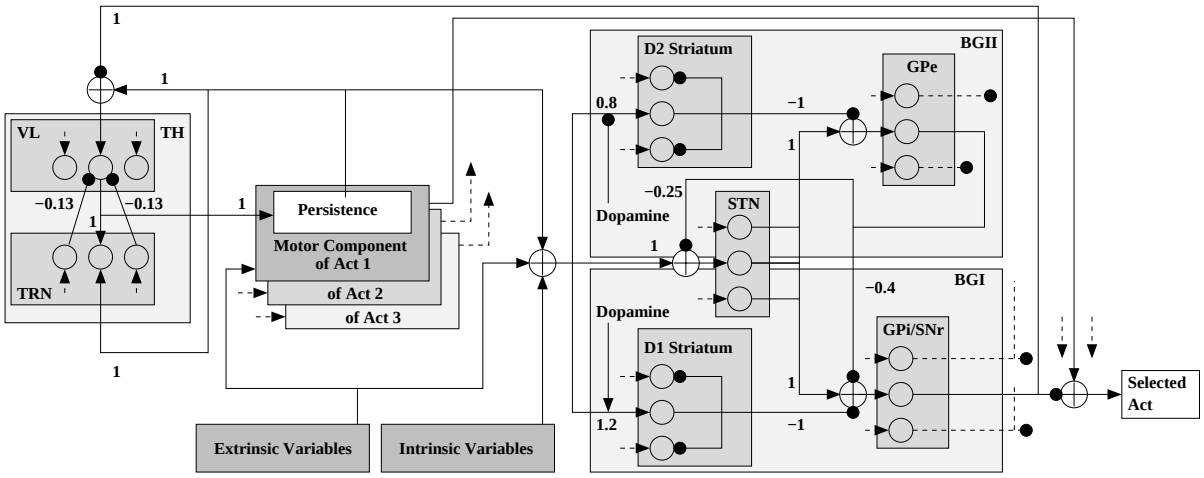

Figure 1: The GPR model. Arrows represent excitatory connections, blobs inhibitory connections. Weights are shown next to their respective pathways. See text for details.

description.

In *BGI*, selection is mediated by two separate mechanisms. First, there are local recurrent inhibitory circuits within the input component *D1 striatum*. [1] The second selection mechanism is comprised of an off-centre on-surround, feedforward network in which the 'on-surround' is supplied by excitation from STN and the 'off-centre' via inhibition from D1 striatum.

A similar arrangement prevails in *BGII*, except the 'output' of this structure (provided by the GPe) sends signals to *BGI*. In particular, it may be shown that the inhibition supplied to STN –the source of excitation for the feedforward selection network– is just sufficient to automatically scale this excitation with the number of channels n in the model, in such a way as to ensure appropriate selection. If this were not the case, the magnitude of the weights from STN and striatum would have to be crafted to be in an approximate ratio of $1 : n$. In the model, these weights have approximately the same magnitude and the scaling is performed by the automatic 'gain control' supplied by outputs from *BGII*.

Humphries and Gurney (2001) embedded the two circuits *BGI, BGII* into a wider anatomical context that included the thalamo-cortical excitatory recurrent loop. The thalamus was decomposed into two constituent structures: the thalamic reticular nucleus (TRN) and the ventro-lateral thalamus (VL). Both thalamic structures have the same segregated channels as *BGI* and *BGII*. This entire circuit is designated by *TH* in Fig. 1. The *TH* circuit not only improves the *clean switching* and *lack of distortion* mechanisms of the basic model, but also reinforces the salience of selected actions thereby fostering persistence of their state of being selected.

[1] the labels D1, D2 refer to types of dopamine synaptic receptor.

3 Implementation

3.1 The robot and its environment

The environment is a 2m x 1.60m flat surface surrounded by walls. It is covered by 40cm x 40cm tiles of three different kinds: 16 uniformly gray tiles (this neutral-gray represents 'barren' locations), 2 tiles with a gray to black gradient ('food' locations), and 2 tiles with gray to white gradient ('nest' locations) (Fig. 2). The robot is equipped with two frontal light sensors pointed to the ground –one behind the other– and with two bumpers, on the front-right and front-left sides (Fig. 2). These sensors provide the four *extrinsic* variables used in the salience calculations (see 3.4 below). Each light sensor produces a raw value corresponding to the color of the ground. The mean of these two values is filtered using a median filter with a 10 time-step window and then used to compute two variables, *Brightness* and *Darkness*, designated L_B, L_D respectively. L_B (resp. L_D) is equal to 0 for all grays darker (resp. brighter) than the neutral-gray, and increases linearly with brighter (resp. darker) grays, reaching 1 for the central white (resp. black) spots. Each of the two bumpers produces a binary value, B_L, B_R set to 1 when the robot hits an obstacle on the left and right respectively.

The 'metabolism' of the robot is based on two *intrinsic* variables: *Potential Energy,* E_{Pot} and *Energy,* E, that initially take on values between 0 and 255. Any action sub-system consumes *Energy* at a rate of 0.5 units per second (except for the variable rate of the resting behavior, see below). Then, these variables are normalised to lie between 0 and 1 for the salience computation.

When E reaches zero, the robot 'dies'. The procedure to reload *Energy* is:

1. to 'eat' on a black place, in order to get *Potential*

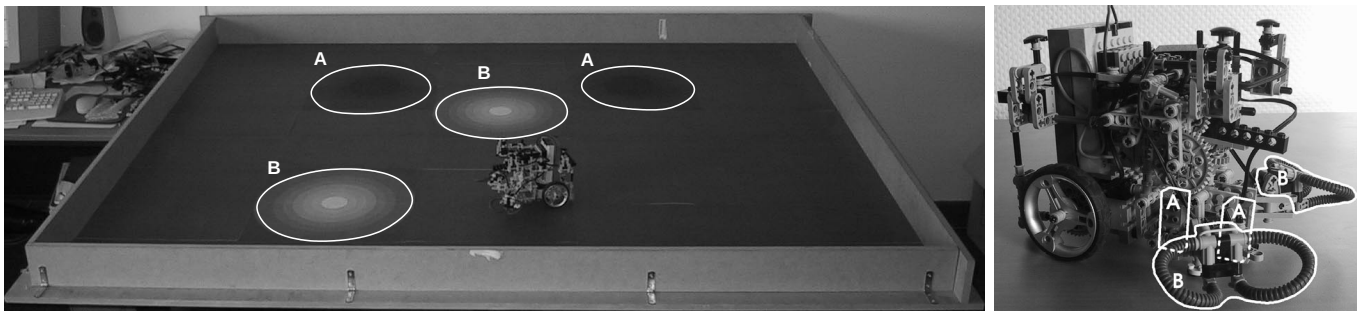

Figure 2: Left: The environment showing 'food' (A) and 'nest' (B) locations. Right: the Lego Mindstorm robot. (A): the light sensors; (B): the bumpers. See text for further details.

Energy, E_{Pot}. The gain $\Delta_E Pot$ in E_{Pot} during this time is proportional to the duration T_{eat} (in seconds) of the eating behavior and to the *Darkness*:

$$\Delta E_{Pot} = 7 T_{eat} L_D$$

2. to 'rest' on a white place, in order to 'assimilate' *Potential Energy* and convert it into *Energy*. When there is no *Potential Energy* to assimilate, *Energy* is decreased with the standard 0.5 units/sec rate, otherwise the changes in *Energy* and *Potential Energy* are proportional to the resting duration T_{rest}

$$\Delta E = T_{rest}(7 L_B - 0.5)$$

$$\Delta E_{Pot} = -7 T_{rest} L_B$$

These relations imply that, when the robot activates these action sub-systems at an inappropriate location (eating on a neutral-gray or bright place or resting on a neutral-gray or dark place), it consumes *Energy* without any benefit.

3.2 Robot: hardware details

The controller (the RCX) for the Lego Mindstorms robot has only 32 KB of memory, some of which is used by the operating system (LegOS). This limited the computation available on-board the robot to the sensory, metabolism and action sub-systems. A Linux-based PC performed all the GPR model-specific computations, calculating and returning inhibitory output signals based on the sensory inputs received from the RCX.

The RCX-PC communication occurred through the Lego MindStorms standard IR transceivers at roughly 10 Hz. This low communication rate required that the GPR model be allowed to compute up to four cycles with the same sensory data in order to have the GPR model working at equilibrium.

3.3 The action sub-systems

In all experiments, the robot has to select efficiently between four action sub-systems. Note that each of these sub-systems corresponds to one channel in the GPR model. When activated, each action sub-system generates a predefined, but interruptible, sequence of elementary acts chosen among the following four available commands for the wheel actuators: *move forward, move backward, turn on the spot, stop.*

The action sub-systems are:

1. *Wander*: a random walk in the environment, programmed as a succession of forward and turning acts of random duration. This action provides the only means for the robot to move around and find the black or white areas; it should, for instance, be activated when the robot is on neutral-gray places, when the current level of either *Energy* or *Potential Energy* is low.

2. *AvoidObstacles*: a short backward movement followed by a rotation triggered when one or both bumpers are activated. Note that there is no movement if the behavior is selected while no bumper is active, therefore it should only be activated when the robot detects it hit an obstacle.

3. *ReloadOnDark*: the robot stops, and, as previously stated, it 'eats' on a dark place, that is, it reloads the *Potential Energy*. This action should therefore only be activated when the robot is on a dark place while *Potential Energy* is low.

4. *ReloadOnBright*: the robot stops and 'rests', that is, it reloads *Energy* and consumes *Potential Energy* when activated on a white place. This action should therefore be activated only when the robot is on a white place while *Energy* is low and *Potential Energy* is high enough for assimilation to be productive.

3.4 The GPR model implementation

The configuration and parameters of the GPR model used in these experiments are the same as in the 'full' embodied model (with normal dopamine modulation) described in Montes-Gonzalez (2001) (see Fig. 1), but

there are also several key differences. These are concerned with modifications to processing of inputs and basal ganglia outputs which have been modified to take into account our different embodiment, environment, and tasks.

One important difference concerns the calculation of input saliences. In Montes-Gonzalez (2001), these were always computed as a linear, weighted sum of sensory, proprioceptive, and contextual variables. However, using any simple weighted sum does not allow salience to depend on a *coupling* of two variables. For instance, in our setting, the activation of *ReloadOnDark* should be correlated to the extrinsic variable *Darkness* and anti-correlated to the intrinsic variable *Potential Energy* (i.e. activated when the one is high and the other low). Activating it on a neutral-gray (or bright) place or while there is no need for *Potential Energy* 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 behavior from taking control of the robot. A similar problem also arises with *ReloadOnBright*. We therefore modified the salience computation to use Sigma-Pi units. These are artificial neurons that allow non-linear (multiplicative) combinations of inputs that can convey interdependencies between variables (Feldman and Ballard, 1982).

For the GPR and the WTA architectures, the weights of salience calculations were 'hand-crafted' over a series of pilot experiments in an attempt to find setting that were close to optimal. The following equations[2] show how the salience for each sub-system was computed as a function of the extrinsic sensory variables (*Brightness* L_B, *Darkness* L_D, *Bump left* B_L, *Bump right* B_L), the intrinsic sensory variables (*Potential Energy* E_{Pot}, *Energy* E) and the *Persistence* signal P for the given channel.

GPR salience calculations:

- *Wander*:
  $-B_L - B_R + 0.8(1 - E_{Pot}) + 0.9(1 - E)$
- *AvoidObstacles*:
  $3B_L + 3B_R$
- *ReloadOnDark*:
  $-2L_B - B_L - B_R + 3L_D(1 - E_{Pot})$
- *ReloadOnBright*:
  $-2L_D - B_L - B_R + 3L_B(1 - E)[1 - (1 - E_{Pot})^2]^{\frac{1}{2}}$

A second difference is in our use of the inhibitory output signal of the GPR model. A characteristic of the GPR model is that, in some cases where there is more than one channel with high salience, there can be partial disinhibition of the motor output of more than one channel. In the earlier robot implementation (Montes-Gonzalez, 2001) the motor outputs of all action sub-systems were therefore combined by weighting each one according to its degree of disinhibition, and Gurney et al (2001a) use the term 'soft switching' to describe an action selection mechanism that can generate a mixed/combined motor output of this kind. Clearly, when conflicting action sub-systems are involved, a merging of motor signals may result in distortion of the selected action(s). On the other hand, however, there are circumstances in which 'soft switching' may be desirable, for instance, where the outputs of two action sub-systems are fully compatible. For the current experiments, we were interested in making comparisons with the WTA mechanism which allows for only one winner (all losers are fully inhibited), a situation that can be termed 'hard switching'. In order to make comparisons between the two models the 'soft switching' characteristic of the GPR model was therefore disabled, in other words, the motor output of the most fully disinhibited action system was always enacted, and that of any partially disinhibited competitors ignored.

A final difference concerns the use in that model of an additional intrinsic variable termed the 'busy signal' whereby an active action sub-system could provide an additional signal to the selection mechanism that would give a temporary and short-term boost to its own salience. In the current robot task setting, the required behavior switching has so far been effectively implemented without including this feature of the original model.

Both architectures –GPR and WTA– were tested with the same robot, the same task, and in the same environment. As shown before, the saliences of the WTA and GPR were computed alike with the exception of the persistence signal P, which is included only in the GPR model. In the GPR architecture, the action sub-system with the least inhibition at each time-step is selected; in the WTA architecture, the action sub-system with the highest salience at each time-step is selected. In either architecture, where there were multiple winning outputs, the sub-system previously selected remained active.

WTA salience calculations:

- *Wander*:
  $-B_L - B_R + 0.5(1 - E_{Pot}) + 0.7(1 - E)$
- *AvoidObstacles*:
  $3B_L + 3B_R$
- *ReloadOnDark*:
  $-2L_B - B_L - B_R + 3L_D(1 - E_{Pot})$
- *ReloadOnBright*:
  $-2L_D - B_L - B_R + 3L_B(1 - E)[1 - (1 - E_{Pot})^2]^{\frac{1}{2}}$

[2] The term containing E_{Pot} in the *ReloadOnBright* salience is not a simple product. However, it may be reduced to such a form if we assume an intermediate variable $[1 - (1 - E_{Pot})^2]^{\frac{1}{2}}$ has been pre-computed first.

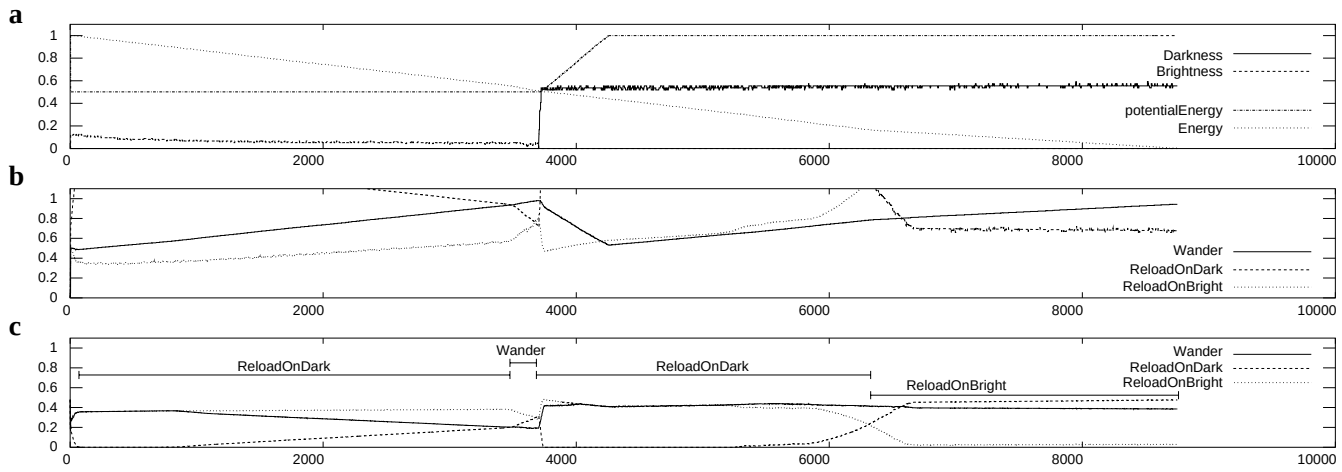

Figure 3: A typical GPR run without use of Sigma-Pi units, a) intrinsic and extrinsic sensory variables, b) corresponding saliences c) output GPi/SNr signals and behavioral sequence. The abscissa on all plots is the number of cycles where 1450 cycles correspond to 100 sec.

4 Results

4.1 Salience computation

Initial experiments with both architectures used simple linear weightings to compute action saliences. However, during the total 12hrs of experiments with the best hand-crafted weightings obtainable, the lifespan of the GPR and WTA robots never exceeded 1.5 time its minimum (8 minutes). Such a situation is depicted in Fig. 3. Here, the simple sum of *Potential Energy* and *Darkness* (3a, b) leads, with the same set of weights, to two inappropriate selections of *ReloadOnDark* (each for different reasons), and a final, and fatal inappropriate selection of *ReloadOnBright*. The first bout of *ReloadOnDark* occurs away from a dark square because the robot lacks *Potential Energy*; the second bout, on the other hand, is the result of a very dark sensor-reading, even though *Potential Energy* is no longer needed. The final ineffective *ReloadOnBright* occurs away from a bright tile because of a profound lack of *Energy*, and this act seals the fate of the animat.

Barring technical problems (such as communication glitches), the use of Sigma-Pi units enormously enhanced the life expectancy of both robots architectures (GPR and WTA), the longest uninterrupted experiment lasting 4 hrs and 20 minutes. Note that the robots can however still die, due to the intrinsic randomness in the *Wander* behavior. In the remainder of the paper we are exclusively concerned with experiments using the Sigma-Pi salience calculations.

4.2 GPR/WTA comparison

During experiments totalling more than 10hrs duration, we did not find any substantial difference between the GPR and WTA architectures with respect to life expectancy, simply because both robot architectures can outlive the time available for a single experiment. This first result led us to further analyze the structure of the behavior generated in the two conditions. In Fig. 4, graphs (a-c) shows the saliences, outputs, and behavior sequences of a typical run with the GPR model, while graphs (d) and (e) show the salience and behavior of the WTA architecture. Note first the substantial difference between the input saliences in the two runs Fig. 4 (a) and (d) which are primarily due to effects of persistence (positive feedback) in the GPR model. The output signals in Fig. 4 (b) show that the control circuit (*BGII*) and feedback loop (*TH*) have also increased the contrast between the action saliences (recall that, with GPR, the action sub-system with lowest inhibition is selected). Finally, in both behavioral sequences, we can observe that similar clean switching is displayed.

	activations per hour		avg. duration	
	GPR	WTA	GPR	WTA
W	302.3	488.0	4.0	3.8
ROD	41.7	62.5	16.0	8.8
ROB	65.1	81.6	15.1	8.0
AO	137.8	363.7	3.3	1.6

Table 1: Activation of each action sub-system showing average bouts per hour and average bout duration in seconds (W : Wander ; ROD : ReloadOnDark ; ROB : ReloadOnBright ; AO : Avoid Obstacle).

Table 1 shows that, with the exception of *Wander*, bouts of individual acts generally last longer with the GPR architecture than with WTA. This can be explained by effects of the persistence mechanism: posi-

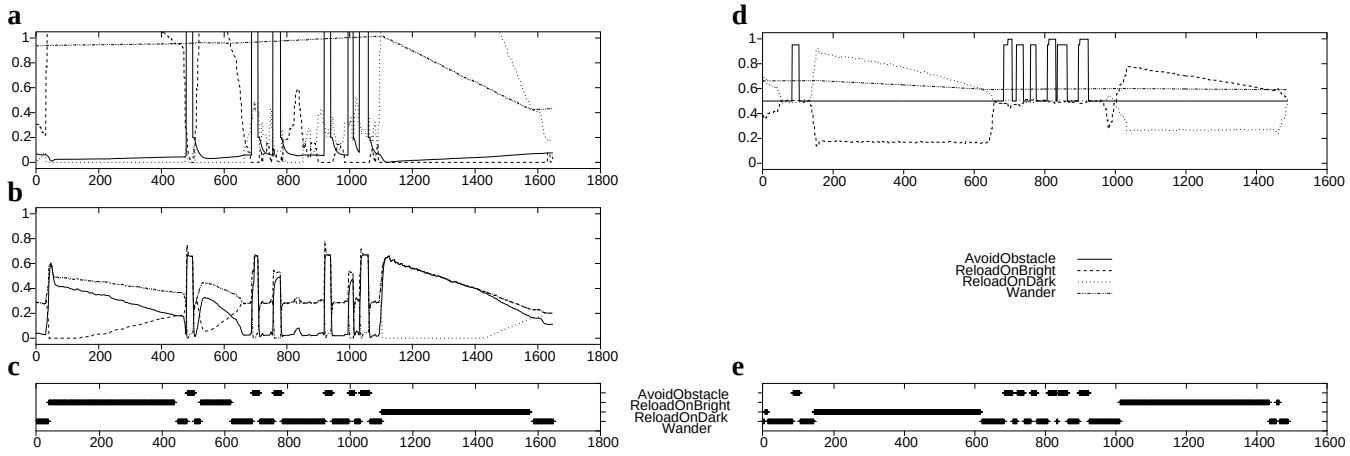

Figure 4: Left: a) Input saliences, b) output GPi/SNr signals and c) corresponding behavioral sequence generated by the GPR model. Right: d) Input and output signals (saliences) and e) the corresponding behavioral sequence generated by a WTA. The abscissa shows the number of cycles where 1450 cycles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

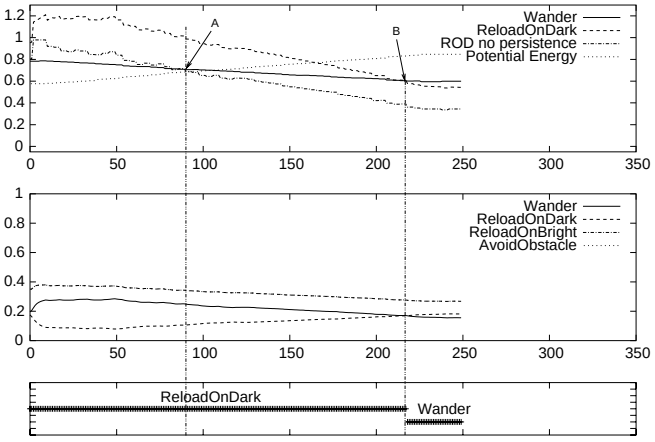

Figure 5: Effect of persistence in GPR. From top to bottom: 'raw' salience (i.e. without persistence) of ReloadOnDark ; output GPi/SNr signals ; the corresponding behavioral sequence generated by the GPR model where (A) points to the time where the switch would happen without persistence, and (B) points to where the switch actually takes place. The abscissa on all plots is the the number of computation cycles, where 1450 cycles correspond to 100 sec.

tive feedback allows a behavior to remain active for some time after its 'raw' salience has fallen below that of other behaviors (see Fig. 5 for an illustration). *Wander* is an exception to this pattern because there is zero weighting on the persistence input.

Although bouts of 'feeding' and 'resting' behavior are shorter in the WTA condition, their frequencies are correspondingly higher. This serves to substantially compensate for their shorter durations, to the point that the average *Potential Energy* and *Energy* end up having similar values ($E = 0.748, E_{Pot} = 0.711$ for WTA and $E = 0.76, E_{Pot} = 0.76$ for GPR). We suspect that this was helped by the relatively short distance between the *Energy* and *Potential Energy* sources in our environment. One may then ask whether the behavioral differences exhibited in Table 1 are reflected in the way the energies are collected.

As can be seen in Fig. 6, there is indeed a major difference between the temporal distribution of E_{Pot} in WTA and GPR. Specifically, the GPR manages to maintain its *Potential Energy* at over 95% of the maximum charge for 25% of the time, while the WTA does so for less than 10%. Unsurprisingly, since the energy sources are inexhaustible, the fact that a reloading action is allowed to last longer allows it to eventually reach the maximum charge most of the time.

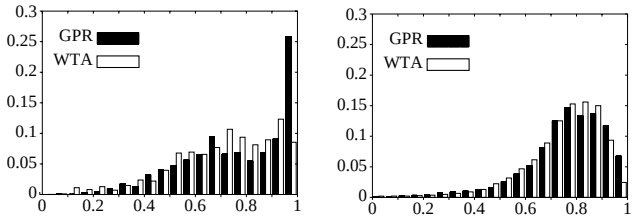

Figure 6: Histograms of the percentages of overall time during which Potential Energy (left) and Energy (right) are reloaded at the values shown on the abscissa.

The same occurs, though to a lesser extent, with the *Energy* (6.8% of maximal charge with GPR, 2.4% with WTA). The effect is less pronounced, because while *Potential Energy* only diminishes when assimilated by *ReloadOnBright*, all actions consume *Energy* therefore the constant decay levels the difference. Whilst persistence can be increased still further, a point is soon reached where the robot continues to recharge beyond

the point where further energy can be usefully consumed.

The preceding results showed that both models can display clean and efficient switching between actions but, due to the effects of *Persistence*, the GPR robot performs fewer transitions than the WTA robot, making it possible for it to load more energy.

5 Discussion

In the following, we discuss the biological plausibility of the modifications we have made to the computation of input and output signals and consider the role of persistence in our model in relation to the notion of positive feedback in animal behavior switching.

5.1 Merging sensory information

In our implementation, the action sub-systems are assumed to depend on saliences, which correspond to the causal motivational factors depicted by ethologists. The general issue of how the display of an action is related to internal and external stimuli is not yet resolved. It seems to depend highly on environmental context and on the animal's previous experiences. In our work, salience is calculated using nonlinear relationships processed by Sigma-Pi units. Such units have been already used for solving similar problems before, while dealing with learning in neural networks (Rumelhart and McClelland, 1986, Gurney, 1992) or context processing in animats (Balkenius and Moren, 2000), but the question arises as to whether such a computation is anything more than an engineering solution. Mel (1993) 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 saliences.

5.2 Merging motor signals

In the earlier robot implementation of GPR (Montes-Gonzalez et al., 2000), the motor components of all action sub-systems that were not fully inhibited could influence the displayed output behavior. In the current work, in order to facilitate comparison with a widely-used engineering solution to the action selection problem, we have not merged the output motor vectors. The literature on animal behavioral switching seems to indicate a wide-range of possible outcomes in situations where there is more than one highly salient action. Possibilities include the merging of the motor outputs (with potential positive or negative consequences), rapid switching between alternative actions (dithering), or the substitution of the salient actions by a third, non-salient 'displacement activity' that is unrelated to the current context (for example eating or grooming in a situation where both fight/flight are similarly primed (Hinde, 1970). The neurobiological substrates that support these various alternatives remain to be understood. However, it is worth noting that the behavior of animals in these (generally) unusual situations may reveal some of the processing characteristics and limitations of the underlying neural mechanisms.

The merging of multiple motor commands is also an issue with respect to the problem of generating appropriate behavior in animats with multiple actuators. In this case, action sub-systems with non-conflicting requirement to use different actuators (like walking and chewing-gum) can be selected at the same time. Neurobiological evidence of a somatotopic organization in the basal ganglia (Redgrave et al., 1999) suggests that there may, indeed, be distinct selection circuits subserving conflict resolution in relation to different limbs or body parts. Some preliminary work, derived from this idea, has been performed by replicating the GPR model as many times as there are actuators, with each GPR copy granting access to only one actuator.

5.3 Persistence and positive feedback in animats and animals

The main quality of the GPR model demonstrated in this study is that it provides a mechanism for ensuring appropriate persistence of a selected action. Though it is possible to add persistence to a WTA via a simple feedback loop, a control circuit (like the BGII in GPR) is then mandatory to avoid overload. The choice of such a circuit would divert the WTA from the zeroth-level action-selection mechanism we need to compare our system with.

Persistence has real adaptive effects. As stated before, it can maintain the animat's internal variables more effectively within their limits, helping it to survive any temporary upset in the availability of resources. It also serves to avoid dithering, which may be particular deleterious where there are significant costs associated with unnecessary switching between one action and another.

Another less intuitive effect of positive feedback is that it can 'prime' the animat to anticipate forthcoming opportunities for action. For instance, in our experiments we noticed that, due to the low communication frequency between RCX and PC, the WTA-robot often stops only after it has driven past the central brightest (or darkest) patch on the gradient tiles whereas the GPR version generally manages to stop closer to the center-most patch. What appears to be happening here is that the corresponding salience increases slightly as the robot enters the brighter (or darker) area. Although this is not, in itself, enough 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 center, it is able to select the appropriate action more rapidly. This increased responsiveness is possible because the lightness gradient serves to prime the appropriate behavior.

The importance of persistence as an adaptive process for animals has already been pointed out by ethologists (e.g. McFarland, 1971). They wished to explain how an activity could continue, in spite of a rapid decrease in its drive. To do so, they supposed that mechanisms of positive feedback or hysteresis are initiated at the start of a bout, enabling the animal to maintain its activity until sufficiently satisfied (Wiepkema, 1971). For instance, in the model of Houston and Sumida (1985), persistence was induced in a competition between two independent motivational systems by a positive feedback pathway similar to a simplified version of the current model. The current experiments provide a useful embodied demonstration of this principle, and of the hypothesis of Redgrave et al. (1999) that the basal ganglia thalamo-cortical loop may serve as the neural substrate that carries this feedback path.

Ethologists have also noted that persistence is more than simply the consequence of closed-loop positive feedback, and can emerge in a variety of different ways. For instance, in the model of Ludlow (1976), hysteresis emerges as the consequence of reciprocal inhibition between multiple motivational systems. Such a configuration can provide for a form of persistence in the selected action. However, the advantages and disadvantages of this solution compared with explicit positive feedback control remain to be fully explored. There is also a sense in which an action may show a 'hidden' persistence, even after its execution has been interrupted. For example, in the 'time-sharing' model of McFarland and Lloyd (1973), a 'dominant' act may be temporarily suspended to allow an alternative behavior to be expressed, only later resuming its performance. In this case, the 'salience' of the dominant act persists even though the behavior itself is deselected. The neural substrate that might underlie a time-sharing mechanism in the vertebrate brain has yet to be investigated. Finally, the duration of any observed behavioral persistence varies according to contextual factors. For example, McFarland (1971) pointed out that the duration of feeding bouts in rats could be diversely triggered by the stimulation of oral and of gut receptors. In Le Magnen (1985) and Guillot (1988), the persistence effect on feeding and drinking bouts in rats and mice was also shown to depend on learning, diurnal and nocturnal conditions. In the current model, the duration of behavioral persistence will also be sensitive to contextual variables since salience is a function of many factors of which positive feedback is only one. The weight on the persistence pathway could also, itself, be subjected to contextual modulation.

These considerations confirm the importance for biology of investigating biomimetic models of action selection such as the GPR model. Compared to the earlier ethological hypotheses, this model is fully computationally specified, is identified with specific neural circuits, and has now been tested in two different embodied implementations. This model is also significantly more complex than earlier proposals and further work is needed to determine both the consequences of this additional complexity for observable behavioral switching, and to consider what potential advantages these may bring to the animal.

6 Conclusion and Perspective

Building on the work of Gurney et al. (2001a,b) and Montes-Gonzalez et al. (2000), our objective was to demonstrate the robustness of the brain-inspired GPR model of action selection. We have shown that the model is able to generate adaptive behavioral sequences when embedded in a different robot, performing different actions, and situated in a different environment. The new implementation also revealed that more flexible mechanisms (Sigma-Pi) can be useful for salience computations than a simple weighted sum. Finally, the comparison with WTA served to highlight several adaptive properties specific to the GPR model, and in particular, its capacity to generate appropriate behavioral persistence.

Further research is planned in three principle directions. First, we will investigate the 'soft switching' capability of the GPR model (merging of output signals coming from multiple sub-systems), in order to explore the capacity of the model to generate compromise behaviors, and also to replicate some of the consequences of mixed motor output observed in animals. Second, we will submit the salience and persistence parameters of the model to learning processes, which will automate the process of tuning the system to new tasks and may also enhance switching efficiency. Third, we will utilise this model within an ongoing, multi-partner project which aims to synthesizing an 'artificial rat' in which biomimetic mechanisms for action selection are combined with a biomimetic mechanism for navigation, both inspired by existing structures in the rat brain.

Acknowledgements

This work was supported by Robea, an interdisciplinary program of the French Centre National de la Recherche Scientifique.

References

Ashby, W. (1952). *Design for a brain.* Chapman and Hall.

Atkinson, J. and Birch, D. (1970). *The dynamics of action*. John Wiley & Sons.

Baerends, G., Drent, R., Glas, P., and Groenewold, H. (1970). An ethological analysis of incubation behaviour in the herring gull. *Behaviour (Supplement)*, 17:135–235.

Balkenius, C. and Moren, J. (2000). A computational model of context processing. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H., and Wilson, S., (Eds.), *From animals to animats 6*, pages 256–265. Cambridge, MA: The MIT Press.

Feldman, J. and Ballard, D. (1982). Connectionist models and their properties. *Cognitive Science*, 6:205–254.

Guillot, A. (1988). *Contribution à l'étude des séquences comportementales de la souris: approches descriptive, causale et fonctionnelle*. PhD thesis, University of Paris 7.

Guillot, A. and Meyer, J.-A. (2000). From sab94 to sab2000: What's new animat? In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H., and Wilson, S., (Eds.), *From animals to animats 6*, pages 3–12. Cambridge, MA: The MIT Press.

Gurney, K. (1992). Training nets of hardware realizable sigma-pi units. *Neural Networks*, 5(2):289–303.

Gurney, K. N., Prescott, T. J., and Redgrave, P. (2001a). A computational model of action selection in the basal ganglia i. a new functional anatomy. *Biological Cybernetics*, 84:401–410.

Gurney, K. N., Prescott, T. J., and Redgrave, P. (2001b). A computational model of action selection in the basal ganglia ii. analysis and simulation of behaviour. *Biological Cybernetics*, 84:411–423.

Hinde, R. (1970). *Animal behaviour: a synthesis of ethology and comparative psychology*. Mc Graw Hill.

Houk, J., Davis, J., and Beiser, D., (Eds.) (1995). *Models of information processing in the basal ganglia*. Cambridge, MA: The MIT Press.

Houston, A. and Sumida, B. (1985). A positive feedback model for switching between two activities. *Animal Behaviour*, 33:315–325.

Humphries, M. D. and Gurney, K. N. (2001). The role of intra-thalamic and thalamocortical circuits in action selection. Submitted to : Network: Computation in Neural Systems.

LeMagnen, J. (1985). *Hunger*. Cambridge, UK: Cambridge University Press.

Ludlow, A. (1976). The behaviour of a model animal. *Behaviour*, 58:131–172.

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

McFarland, D. (1977). Decision making in animals. *Nature*, 269:15–21.

McFarland, D. and Lloyd, I. (1973). Time-shared feeding and drinking. *Quaterly Journal of Experimental Psychology*, 25:48–61.

Mel, B. W. (1993). Synaptic integration in an excitable dendritic tree. *Journal of Neurophysiology*, 70(3):1086–1101.

Montes-Gonzalez, F. (2001). *A robot model of action selection in the vertebrate brain*. PhD thesis, University of Sheffield, UK.

Montes-Gonzalez, F., Prescott, T. J., Gurney, K. N., Humphries, M., and Redgrave, P. (2000). An embodied model of action selection mechanisms in the vertebrate brain. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H., and Wilson, S. W., (Eds.), *From animals to animats 6*, volume 1, pages 157–166. Cambridge, MA: The MIT Press.

Prescott, T. J., Redgrave, P., and Gurney, K. N. (1999). Layered control architectures in robots and vertebrates. *Adaptive Behavior*, 7(1):99–127.

Redgrave, P., Prescott, T. J., and Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89:1009–1023.

Rumelhart, D. and McClelland, J. (1986). *Parallel Distributed Processing*, volume 1. Cambridge, MA: The MIT Press.

Slater, P. (1978). A simple model for competition between behaviour patterns. *Behaviour*, 57(3):236–257.

Snaith, S. and Holland, O. (1991). An investigation of two mediation strategies suitable for behavioural control in animals and animats. In Meyer, J.-A. and Wilson, S. W., (Eds.), *From animals to animats 1*, pages 255–262. Cambridge, MA: The MIT Press.

Spier, E. and McFarland, D. (1996). A fine-grained motivational model of behaviour sequencing. In Maes, P., Mataric, M. J., Meyer, J.-A., Pollack, J., and Wilson, S. W., (Eds.), *From Animals to Animats 4*, pages 255–263. Cambridge, MA: The MIT Press.

Tyrrell, T. (1993). The use of hierarchies for action selection. *Adaptive Behavior*, 1(4):387–420.

Wiepkema, P. (1971). Positive feedback at work during feeding. *Behaviour*, 39:266–273.