ADAPTIVE MOTIVATION IN A BIOMIMETIC ACTION SELECTION MECHANISM

A. Coninx,[1,2*] A. Guillot,[2] B. Girard[1]
1 – Laboratoire de Physiologie de la Perception et de l'Action
CNRS - Collège de France
11 place Marcelin Berthelot,
75231 Paris Cedex 05
2 – Institut des Systèmes Intelligents et de Robotique
CNRS - UPMC
4 place Jussieu,
75252 Paris Cedex 05

ABSTRACT
In this paper, we extend a basal ganglia action selection model with an adaptive motivational system. We evaluate the resulting model in a minimal survival task, where the adaptive motivational system is used to take the resource density in the environment into account. We show that the model exhibits an increased decisional autonomy, which allows the robot to behave more economically when possible, without jeopardizing its survival chances. We discuss the effect of motivational adaptation on the robot's behaviour in various initial conditions, and find that prior adaptation strongly affects selection between two unequally abundant resources. Finally, we propose hypotheses on the neurobiological mechanisms of motivational adaptation.

KEY WORDS
Motivation, action selection, basal ganglia, adaptation

1 Introduction

Action selection can be defined as the resolution of conflicts between functional units competing to access motor resources. It is a classical problem for both computational neuroscience and autonomous robotics, and includes an evaluation component (how to learn the actions' relative priorities) and a selection component (how to solve the conflict between the concurrent actions). Neuroscientists have shown natural action selection to be strongly related to the basal ganglia (BG) [1], and many neurocomputational models deal with either the evaluation aspect of action selection, using reinforcement learning (RL) algorithms [2], or the selection aspect [3].

A way to evaluate the relevance of those models is to enclose them into a sensorimotor loop by a simulated or robotic embodiment, and to measure their efficiency while solving a given task. In this study, we address the question of the adaptation of the motivational processes during selection, according to the environmental conditions. Our purpose is to allow the agent to efficiently adapt its behaviour in situations where the task is the same (and the efficient action sequences are hence identical) but the environment is different, e.g., various resource densities.

We use the basal ganglia action selection model proposed in [4] (CBG model), and extend it by adding motivational adaptation capabilities. A recent work from Konidaris *et al.* [5] contains an interesting proposal for an adaptive motivational system, but does not use such a neuromimetic action selection model. It learns a policy by a RL system whose reward value is generated from the satiation levels of resources using Hullian drives. A priority parameter is computed so as to reflect the agent's long-term beliefs about his environment, and biases each drive's reward generation process so that a resource perceived as "rare" will also be perceived as more rewarding. The task used to test this model differs from the one used to evaluate the CBG model. We thus adapt this proposal to the CBG model: we use a similar priority parameter to bias the selection process, without modifying the underlying behavioural rules.

This paper first presents the motivational adaptation model, studies its performance and benefits, and discusses its possible neurobiological substrates.

2 Material and methods

2.1 Action selection model

The CBG model, proposed in [4], is a neurocomputational basal ganglia action selection model. It closely follows the recent neuroscientific research on the anatomy and connectivity of the dorsal circuits of the basal ganglia, and has been shown to perform efficient action selection. Like the older GPR model [6, 7], on which it is based, the CBG model selects an action within a repertoire by disinhibiting it (see Fig. 1). It is made of a number of channels in competition, each of them being associated to an action. The inputs of these channels are numerical values called *saliences*, which model a cortical input representing the agent's propensity to execute the associated action. The channels' inhibitory output, which represent projections to the brainstem motor centers, is tonically active except in the selected action's channel : the selected action is therefore disinhibited.

*Corresponding author, email: alexandre.coninx@college-de-france.fr. A.C. and A.G. acknowledge the support of the European Project ICEA (Integrating Cognition, Emotion and Autonomy), FP6-IST-027819

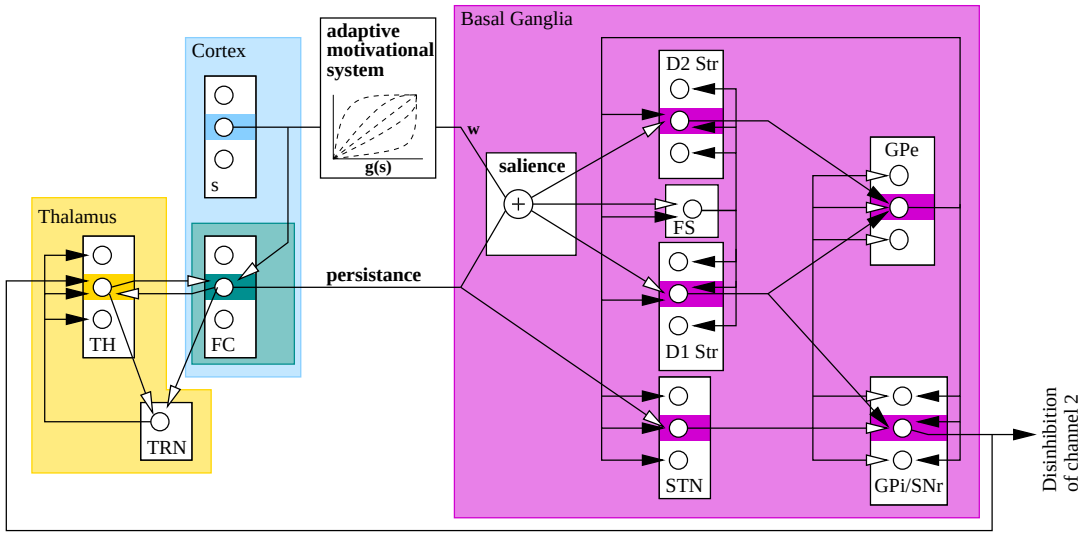


Figure 1: The CBG model, modified to take adaptive motivational modulation into account. Three competing channels are shown. For better readability, only the projections from the second channel are represented, but similar projections exist for all channels. Black arrows are inhibitory projections, white arrows are excitatory projections. TH : thalamus. TRN : thalamic reticular nucleus. FC : frontal cortex. Str : striatum. D1 Str, D2 Str and FS (fast-spiking) : various cell populations within the striatum. STN : subthalamic nucleus. GPe : external globus pallidus. GPi : internal globus pallidus. SNr : substantia nigra pars reticulata. Motivational modulation is applied to the cortical input s (which represents sensory and inner state information) in the cortico-striatal synapses. See [4] for a comprehensive description of the CBG model.

The input saliences are computed in the following way :

$$\text{salience}_i = w_i \times f(s_i) + p_i \quad (1)$$

The parameters are defined as follows :

- w_i is a **weight**, which defines action's i priority with regard to the others. For example, the saliences for consumptive actions must be greater than those for appetitive actions in order to take precedence over them. These weights' values were hand-tuned in the present work, but previous research has shown that they could be efficiently learnt by a RL system [8].
- s_i is a product of the **sensory and inner state variables** (such as satiation levels) relevant for action i. For example, an energy-seeking behaviour must not be activated if the energy need is fully satiated.
- f is a sigmoid transfer function.
- p_i denotes a **persistance** term, a positive feedback provided by the CBG, which can be used to give a bonus to the salience of the currently selected action.

Konidaris [5] uses a family of nonlinear transfer functions g parametrized by a ρ value as shown in Fig. 2. These functions replace the f transfer function in our salience computation and allow to influence each drive's motivational weight by varying the ρ parameter. By having the robot learn ρ_i values adapted to its environment, we can then make its motivational system adaptive.

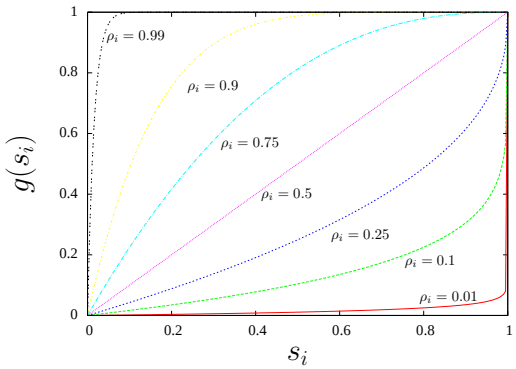


Figure 2: Transfer functions proposed by Konidaris, adapted to the CBG model. $g(s_i, \rho_i) = 1 - (1 - s_i)^{tan\frac{\rho_i \pi}{2}}$, where s_i is the sensory and inner state input value related to action i. For $\rho_i = 0.5$, the function is linear (no bias). For $\rho_i > 0.5$, the motivational drive is overstressed (the related sensory and inner state input is given more importance). For $\rho_i < 0.5$, it is understressed.

2.2 Survival task

The CBG model, as well as the older GPR model [6, 7], have already been evaluated in a simulated robotic task, which has been designed as a minimal survival task to evaluate action selection mechanisms. We used the same task to evaluate our model. In this task, a simulated robot has an artificial metabolism based on two inner variables, Energy (E) and Potential Energy (E_p), taking values between 0 and 1. In order to succeed, the robot has to regularly reload

E_p and then transform it into usable E. (See Fig. 3)

- E decreases at a rate of 0.007 unit per second. If it reaches 0, the trial is stopped. In order to prevent that, the robot must activate the *ReloadE* behaviour on an Energy resource. This transforms E_p into E at a rate of 0.2 unit per second.
- E_p only decreases during the above-mentioned transformation process, and can be reloaded at a rate of 0.2 unit per second by activating the *ReloadEp* behaviour on a Potential Energy resource.

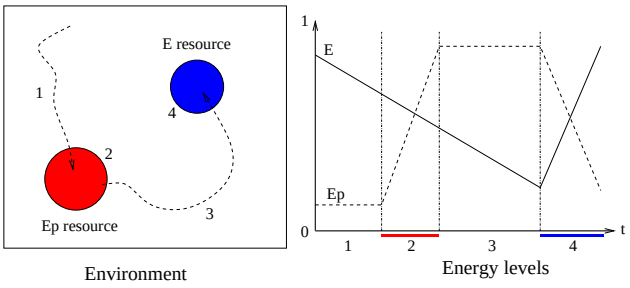


Figure 3: Structure of the survival task. In order to keep its E level above 0 and to survive, the robot must 1) reach a E_p resource ; 2) reload E_p ; 3) reach a E resource ; and 4) transform E_p into E.

The robot must alternate between locations containing the two different types of resources. Its action selection system uses a repertoire of 8 atomic actions:

- *WanderEp* and *ApproachEp* are appetitive actions used to reach E_p resources. *WanderEp* is a random exploration behaviour used to discover E_p resources, and *ApproachEp* is a visual approach behaviour activated when an E_p resource is in sight.
- *WanderE* and *ApproachE* are similar actions used to reach E resources. Note that *WanderEp* and *WanderE* activate the same random exploration behaviour but are different actions for the action selection system since they are not directed to the same goal.
- *ReloadEp* and *ReloadE* are the above-described consumptive actions.
- *AvoidObstacle* activates a simple obstacle avoidance behaviour.
- *Sleep* halves the energy consumption rate (0.0035 unit per second), but stops the agent, which prevents it from gathering resources. It must therefore be activated only when both E and E_p are satiated.

The experiments were conducted using the Player/Stage simulation software [9]. The environment was a $15 \times 15\,m$ square arena containing one or four resources of each type depending on the tested condition (Fig. 4). Those resources are neither destroyed

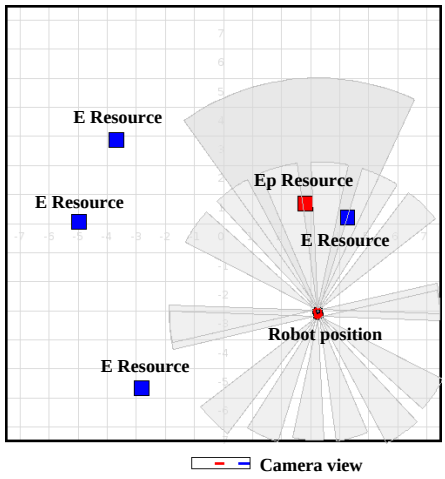


Figure 4: Simulated environment in the (4 E, 1 E_p) condition. (See 2.2 for a detailed explanation.)

nor depleted by the consumptive actions. The simulated robot and its sensorimotor abilities are the same as in [4].

The basic principle of our adaptive motivational system is to modify the sensibility of salience computations with regard to the agent's belief about the resource density in the environment. This belief is encoded in the rho_i parameters, whose modification affects the shape of the g functions. In his experiment [5], Konidaris uses the two resources problem [10] : the agent needs two independant resources, which are available as randomly placed lumps that must be gathered and consumed. In this task, he computes the ratio of the number of consumptive acts of the two resources to learn the ρ_i parameters. Our task is significantly different, first because the two resources are coupled by the simulated metabolism, and second because they are available in unlimited supply at fixed locations. A robot surviving in this task necessarily alternates between the two resources, and thus has a ratio of the two resources' number of consumptive acts close to 1. Consequently, this ratio is not a relevant indicator in our task.

We propose using visual input to assess each resource's rarity. On each period of 25 seconds, we count the number of times each type of resource appears in its field of view. A moving average of the last counts is then used as a measure of the availability of each resource a_{E_p} and a_E.

We use as a reference condition environments with one of each resource (1 E, 1 E_p). In this condition, we tune the initial value of the ρ_i parameters so as to ensure good survival performances and we associate them to the measured a_{E_p} and a_E values. The adaptive motivation mechanism then adjusts the ρ_i parameters depending on the variations of a_{E_p} and a_E:

- The appetitive actions' ρ parameters (ρ_{WanderE}, $\rho_{\text{ApproachE}}$, ρ_{WanderEp} and $\rho_{\text{ApproachEp}}$) linearly decrease with the availability of their related resource: these

behaviours must be overstressed when that resource is scarce and are less critical when it is abundant.

- The consumptive actions' ρ parameters (ρ_{ReloadE} and ρ_{ReloadEp}) do not change, because in this task the reloading actions operate relatively fast, so that there is no point in interrupting an ongoing reload.
- $\rho_{\text{AvoidObstacle}}$ does not change either since the *AvoidObstacle* behaviour has no link with resources gathering.
- ρ_{Sleep} linearly grows with both a_{E_p} and a_E : the *Sleep* behaviour can be activated longer without danger when the resources are abundant, but must be avoided otherwise.

The agent therefore continuously ajusts its beliefs about the resource density and consequently adapts its behaviour.

Three resources repartitions were tested : (1 E, 1 E_p), (1 E, 4 E_p), and (4 E, 1 E_p) (Fig. 4). For each condition, 50 different environments were generated (with random resources placement), and the robotic task was run in each environment twice, with and without the motivational adaptation. Each trial was stopped when the robot ran out of energy ($E = 0$), or after 30 minutes (and the survival trial was then considered as successful). The simulated robot starts each trial with an empty potential energy reserve ($E_p = 0$) and a full energy reserve ($E = 1$), which allows to survive for 2 minutes and 23 seconds if not reloaded.

Moreover, in order to evaluate the behavioural consequences of the motivational adaptation, we tested the robot's choices when faced to two equally distant resources (Fig. 5), after having been exposed to the various resource settings, and when varying the internal state.

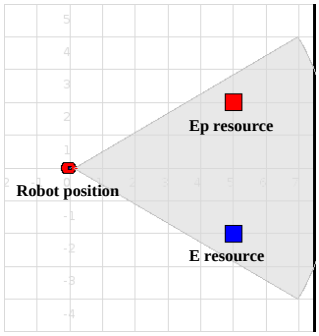


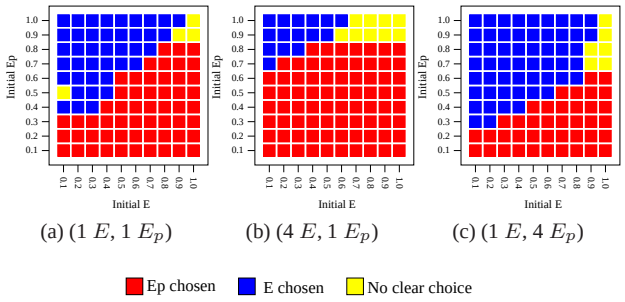
Figure 5: Simulated environment for the behavioural choice tests. The robot can see both resources and reach any of them. Trials were run for various initial E and E_p levels (see Fig. 6 for results).

3 Results

We found out that motivational adaptation does not modify the agent's survival performances, but changes its behaviour and makes it more economical. The first general result is that, in all resource settings, the number of successful runs is quite similar when comparing the system with and without adaptation (Table. 1). Moreover, the distribution of survival durations for the unsuccessful runs were compared with and without adaptation, for each resource settings, using the Kolmogorov-Smirnov (KS) test. They are not significantly different.

The stability of the adaptation algorithm is assessed by the facts that in the reference (1 E, 1 E_p) condition, a_{E_p} and a_E remain very close to their initial values, the behavioural allocation is similar to the one without adaptation and finally, the distributions of the survival times and energy consumption rates (see Table. 1) are not significantly different (KS test, $D_{KS} = 0.11, p = 0.97$ and $D_{KS} = 0.12, p = 0.84$ respectively).

In an environment with an increased E resources density, the *Sleep* action is used much more often with the adaptive system than without, as the increase in global resource density is taken into account. For example in the (4 E, 1 E_p) setting, *Sleep* is activated 50.0% of the time with adaptation vs. 17.4% without. This results in a significantly lower energy consumption (see Table. 1): the two-tailed KS shows that energy consumptions with and without motivational adaptation are drawn from significantly different distributions ($D_{KS} = 0.86, p < 0.001$). The behavioural test shows that after adaptation to this environment, the robot preferentially approaches the less abundant E_p (compare Fig. 6a and Fig. 6b).



(a) (1 E, 1 E_p) (b) (4 E, 1 E_p) (c) (1 E, 4 E_p)

Figure 6: Resource chosen in the behavioural choice task, depending on the initial E (horizontal) and E_p (vertical) levels. Note the shift of the choice boundary towards the less abundant resource.

Surprisingly, in the high E_p density environments, *Sleep* activation is not significantly modified by the adaptive system: 31.5% of the time vs. 28.7% without adaptation. Energy consumption (see Table. 1) is thus not affected; the KS test shows no significant difference ($D_{KS} = 0.2, p = 0.24$), despite a behavioural modification driving the robot towards E resources in the choice test as shown by Fig. 6c. This is caused by the specific chaining of the two resources in the artificial metabolism of this task. The robot can afford sleeping only when both E and E_p are high, and this can only happens when it just finished reloading E_p (after reloading E, the E_p level is always low, as E_p has been consumed in the process). In the high E_p density

environments, the ρ_{Sleep} increase, caused by the global resource density increase, is not sufficient to generate longer *Sleep* bouts, because *Sleep* is then in competition with *WanderE* or *ApproachE*, whose ρ is also high. Nevertheless, the behaviour of the robot is affected by the resource imbalance, as it preferentially approaches the E resource (compare Fig. 6a and Fig. 6c).

Resources	Model	$< T_{\text{survival}} >$	$< C_{E_p} >$	n_{survival}
(1 E, 1 E_p)	NA	926.5	0.0062	11/50
	A	1092.0	0.0062	18/50
(4 E, 1 E_p)	NA	1591.4	0.0060	37/50
	A	1504.4	0.0050	33/50
(1 E, 4 E_p)	NA	1559.5	0.0057	34/50
	A	1386.4	0.0056	29/50

Table 1: Survival time $< T_{\text{survival}} >$ (in seconds); average energy consumption $< C_{E_p} >$ (in E_p/s) and number of successful trials (where the simulated robot survived for 30 minutes) in each resources conditions. NA: no adaptation; A: adaptation.

4 Discussion & Conclusion

We presented an adaptive motivational mecanism added to an existing neuromimetic action selection model, so as to take into account the variations of the availability of resources in the environment. It affects the behavioural selection as expected and allows economical use of energy when possible, which would be of great interest in an environment with limited resources.

This adaptive mecanism is a purely algorithmic and has not been matched with the operation of any brain region. We suggest that the involvement of the ventral basal ganglia circuits, especially the one including the nucleus acumbens shell [11], in the incentive salience processes [12], makes it a possible substrate of this adaptive process. This circuit is in position to modulate the selection in the dorsal BG circuits – those which are simulated in the CBG model – especially through its dopaminergic projections, which are precisely supposed to affect motor and arousal processes [13]. In a future work, we thus plan to improve the neuroinsipration of our system by using a shell circuit model developed within the ICEA consortium, projecting to our dorsal basal ganglia model.

References

[1] J. W. Mink. The basal ganglia: Focused selection and inhibition of competing motor programs. *Prog Neurobiol*, 50(4):381–425, 1996.

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

[3] K. Gurney, T.J. Prescott, J. Wickens, and P. Redgrave. Computational models of the basal ganglia: from membranes to robots. *Trends Neurosci*, 27:453–459, 2004.

[4] B. Girard, N. Tabareau, Q.C. Pham, A. Berthoz, and J.-J. Slotine. Where neuroscience and dynamic system theory meet autonomous robotics: a contracting basal ganglia model for action selection. *Neural Netw*, 2008.

[5] G.D. Konidaris and A.G. Barto. An adaptive robot motivational system. In S. Nolfi, G. Baldassarre, R. Calabretta, J.C.T. Hallam, D. Marocco, J.-A. Meyer, O. Miglino, and D. Parisi, editors, *SAB06*, volume 4095 of *LNAI*, pages 346–356, Berlin, Germany, 2006. Springer.

[6] K. Gurney, T. J. Prescott, and P. Redgrave. A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological Cybernetics*, 84:401–410, 2001.

[7] K. Gurney, T. J. Prescott, and P. Redgrave. A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. *Biological Cybernetics*, 84:411–423, 2001.

[8] M. Khamassi, L. Lachèze, B. Girard, A. Berthoz, and A. Guillot. Actor-critic models of reinforcement learning in the basal ganglia: From natural to artificial rats. *Adapt Behav*, 13(2):131–148, 2005.

[9] B.P. Gerkey, R.T. Vaughan, and A. Howard. The Player/Stage project: Tools for multi-robot and distributed sensor systems. In *ICAR 2003*, pages 317–323, Coimbra, Portugal, 2003.

[10] E. Spier and D. McFarland. Possibly optimal decision-making under self-sufficiency and autonomy. *J Theor Biol*, 189(3):317–331, 1997.

[11] A. E. Kelley. Neural integrative activities of nucleus accumbens subregions in relation to learning and motivation. *Psychobiol*, 27:198–213, 1999.

[12] K. C. Berridge and T. E. Robinson. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience. *Brain Res Rev*, 28:309–369, 1998.

[13] A.E. Kelley, B.A. Baldo, W.E. Pratt, and M.J. Will. Corticostriatal-hypothalamic circuitry and food motivation: integration of energy, action and reward. *Physiol Behav*, 86(5):773–795, Dec 2005.