2016

MECANISMO DE TOMA DE DECISIONES EMOCIONAL BIOINSPIRADO APLICADO COMO CONTROLADOR DE UN AGENTE AUTÓNOMO

NETTLE VACHER, CRISTÓBAL JESÚS

Universidad Técnica Federico Santa María

Improving surviving: tonic dopamine control improves bio-inspired robot controller behaviour in a simple survival task

UNIVERSIDAD TECNICA FEDERICO SANTA MARIA

Cristóbal Nettle[1,3], María-Jose Escobar[1] and Arthur Leblois[2].

[1] Department of Electronics Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile; [2] CNRS, Laboratoire de neurophysique et physiologie, UMR 8119, Université Paris Descartes, Paris, France; [3] Innovación y Robótica Estudiantil, UTFSM, Chile.

DEPARTAMENTO DE ELECTRONICA

UNIVERSITÉ PARIS DESCARTES

Introduction

- Several models of the human's decision making system has been proposed. Multiple loop models explain separable selections from electrophysiologic data.
- Dopamine has been related to the exploration-exploitation trade-off.
- Striatal dopamine receptors are reduced in patients with anxiety disorders.

For artificial intelligent agents

- Exploration-exploitation trade-off is fundamental.
- The integration of emotions might be a great benefit in terms of:
  - learning processes.
  - interaction with humans and the environment.

Proposal

- Integration of tonic DA effects in a multiple loop selection model.
- Integrate the model in a robot controller to:
  - control its exploration-exploitation trade-off.
  - indirectly integrate an emotional related effect.

Cortico-Basal Ganglia (CBG) model

From Guthrie et al. (2013)

- Considerates two competing pathways:
  - Direct pathway: focussed positive feedback

    Cortex → Striatum → Globus pallidus (internal) → Thalamus ↔ Cortex
  - Hyper-direct pathway: spread negative feedback

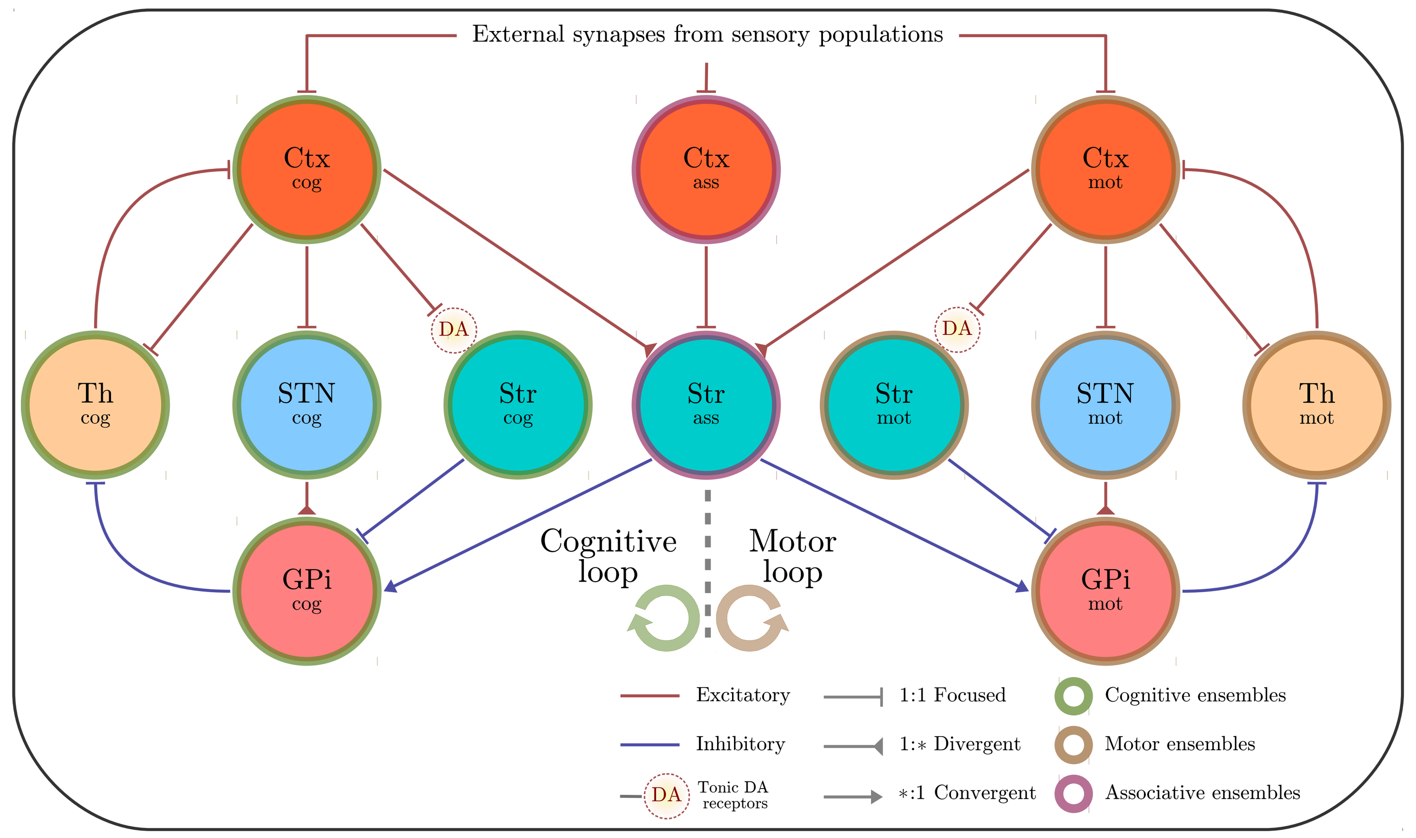
    Cortex → Subthalamic nucleus → Globus pallidus (internal) → Thalamus ↔ Cortex
- Multiple parallel loops implementation: a $Cognitive$ and a $Motor$ loop
  - Symmetrical.
  - Two distinguishable selections.
  - Crosstalk between loops in the striatal *associative* populations.
- Populations are simulated using a simple neural rate model, considerating:

$$\tau\frac{dm(t)}{dt} = -m(t) + \mu(t) \qquad \mu(t) = S(I^T(t) - T)$$

τ: decay time constant. $\mu_i(t)$: instantaneous activity. $I^T(t)$: synaptic input. $m_i(t)$: synaptic output activity. $S(\cdot)$: transfer function. T: threshold.

  - Threshold linear transfer functions for almost all populations.
  - Sigmoidal transfer function for striatal populations.
  - Gaussian noise is added to synaptical inputs, proportional to the inputs amplitude.
- Dopaminergic learning modifies corticostriatal synapses.
  - Learning based in reward.
  - Rewarded selections strengthen its corticostriatal connections.
  - Non-rewarded selections attenuate them instead.

  ⇒ Learning leads to the selection of the better option.



Tonic dopamine (DA) integration

Tonic dopamine type D1 has a strong effect in terms of the control of the exploration-exploitation trade off (Humphries et al., 2012).

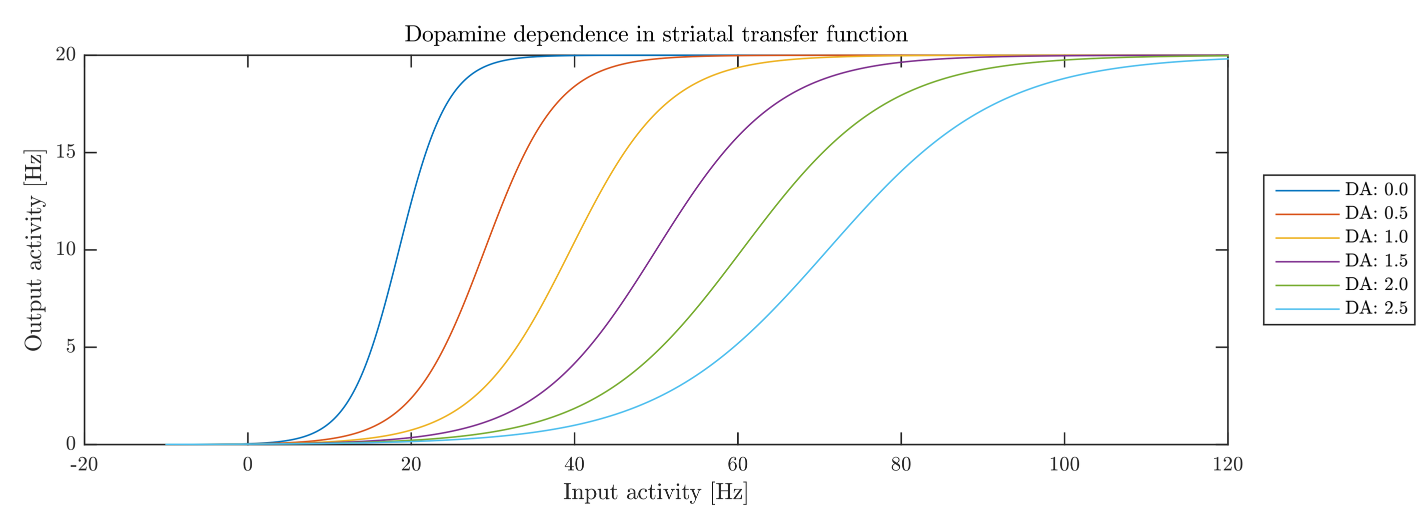
- Affects cognitive and motor corticostriatal synapses.
- Simulates D1-type as a multiplicative factor:

$$I^T(t) = (1 + DA)\ I(t)$$

- Tonic DA also modifies the threshold V_h and slope V_c of the striatum:

$$V_h = V_{h_{DA}}\ DA + 18.5 \qquad V_c = 3.0\ (1 + DA)$$

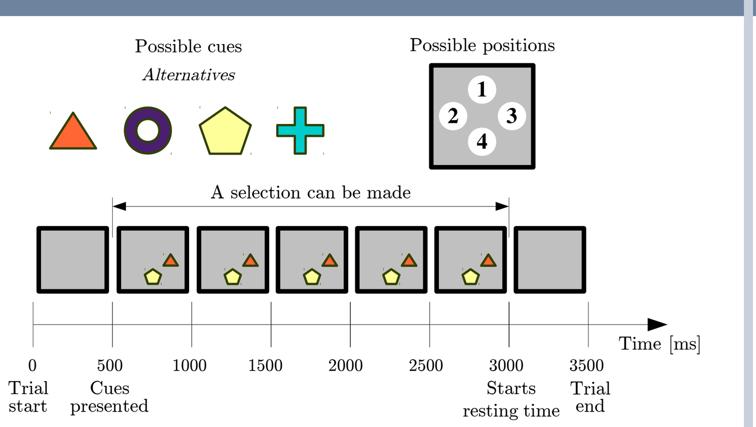
- Cognitive and motor inputs noise is modified as:
  - Has constant variance.
  - Presents temporal correlation.



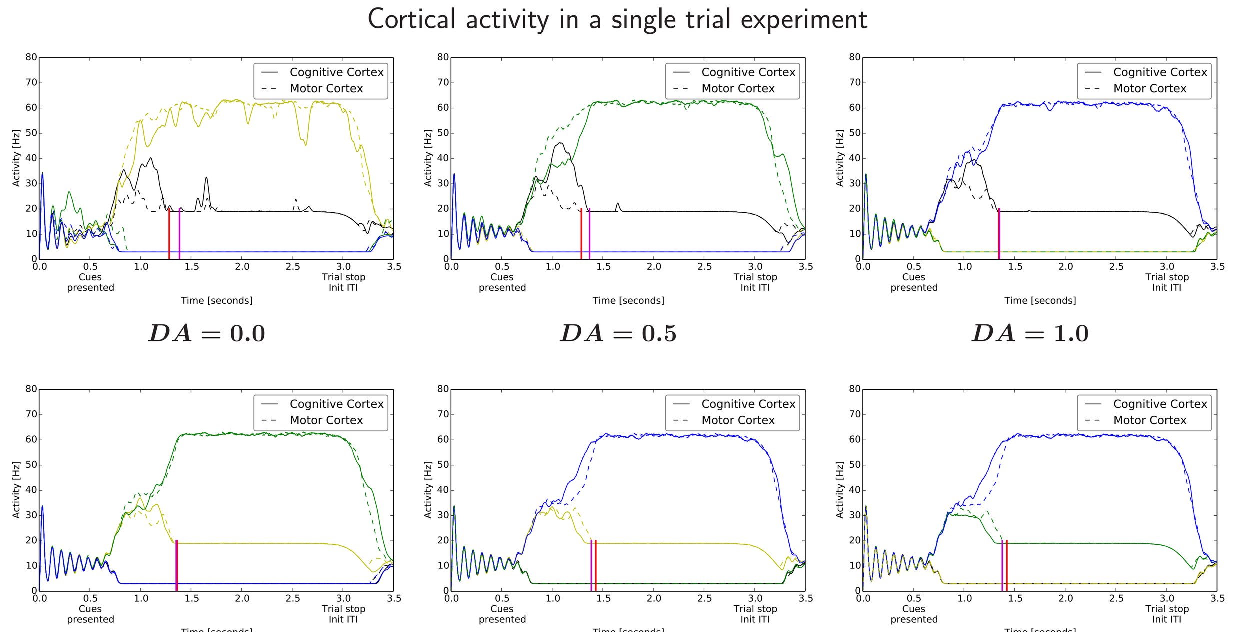
Experimental evaluation - Performing a two-choice forced selection task

Trials

- Selection between four different shapes.
  - Reward probabilities: 0/3 1/3 2/3 3/3
- Two different shapes presented simultaneously.
- Four possible positions.
- Selection has to be made in 2.5 [s].

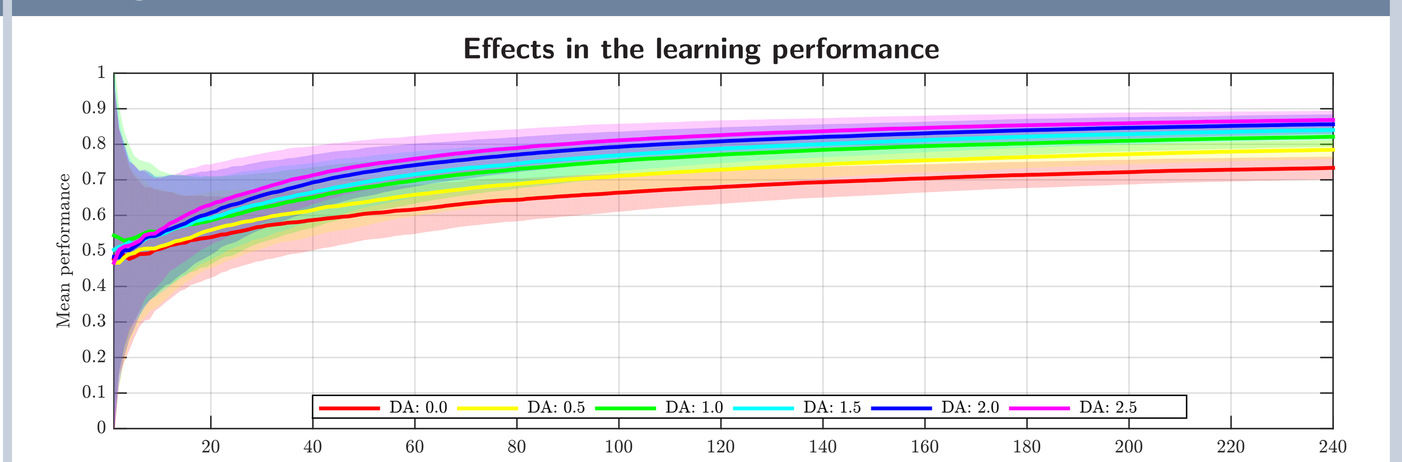


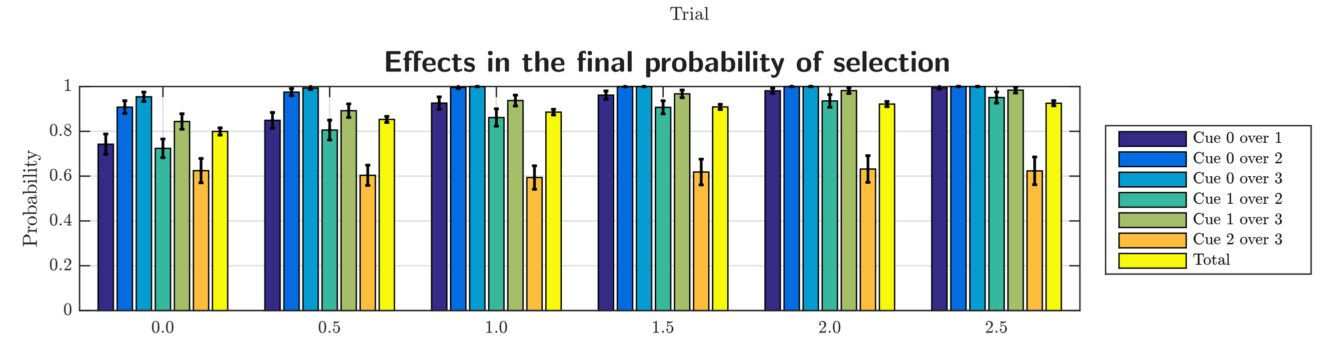
Cortical activity in a single trial experiment

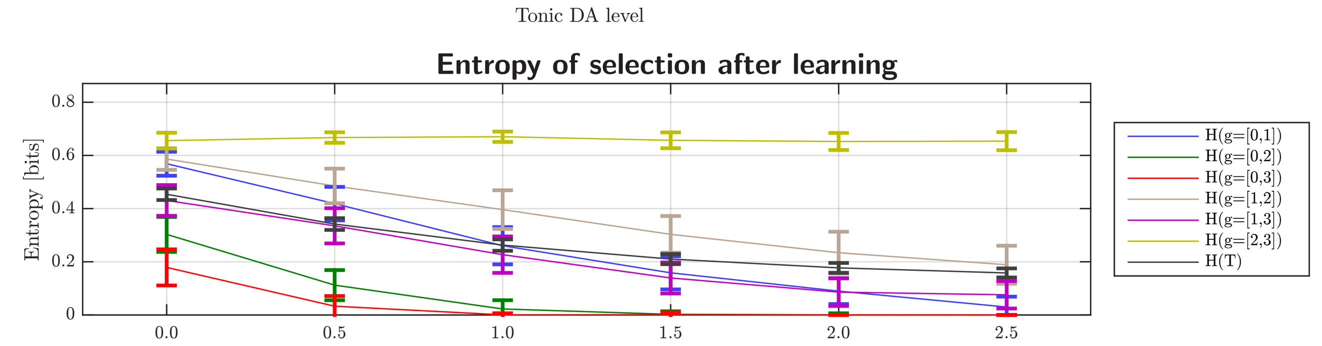


$DA = 0.0$ $DA = 0.5$ $DA = 1.0$ $DA = 1.5$ $DA = 2.0$ $DA = 2.5$

Learning results







Implementation of the CBG model as a robot controller

Using the CBG model as a decision making mechanism

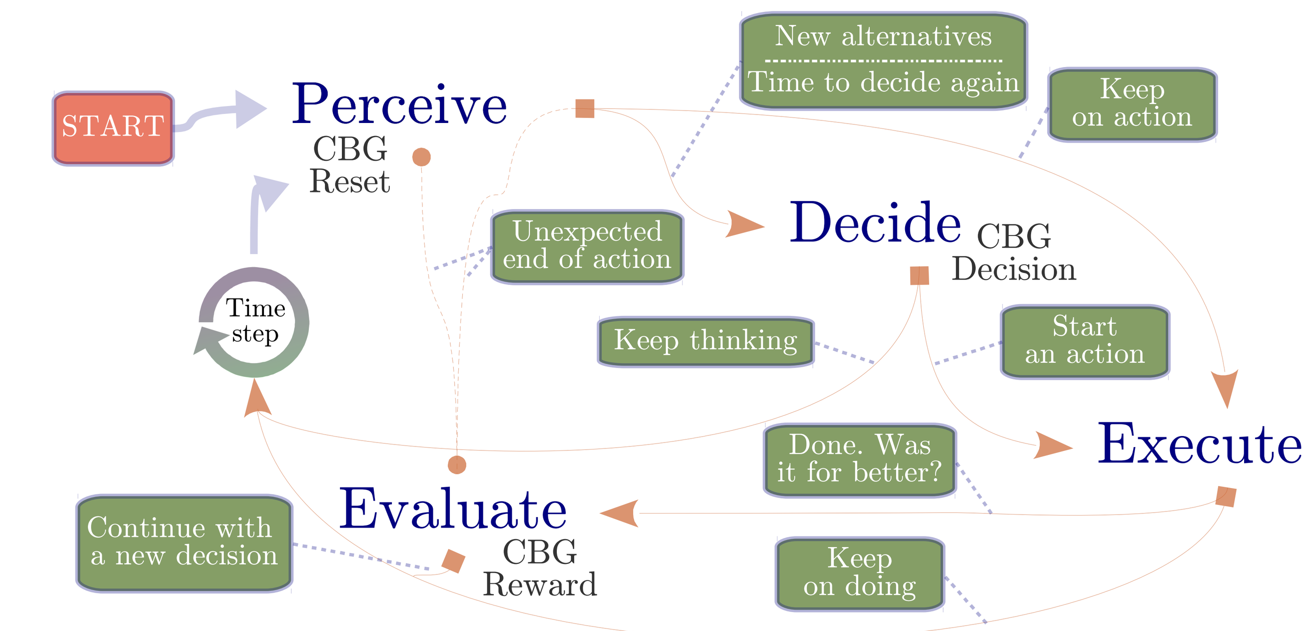
- The model is integrated in a robot controller.
- The robot has to learn on-line which option is better for its current state.

The robot controller

- Implemented as a finite state machine.
  1. Perceive: detects the environment and then, the robot's possible actions.
  2. Decide: performs a selection using the CBG model.
  3. Execute: controls the robot movements.
  4. Evaluate: performs the rewarding evaluation and learns.

About the implementation

- The system is tested using the Virtual Robot Experimentation Platform.
  - Performs all the physics related calculations.
  - Can be controlled by an external software using a ready to use API (Application Programming Interface).

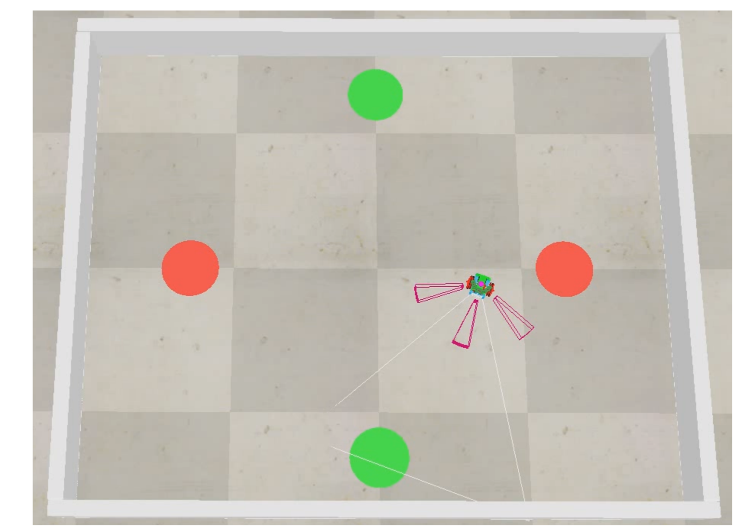


Two-source survival task

- Minimal scenario for evaluating decision making mechanisms.
- The agent has two intrinsic energy levels:
  - Potential energy PE: a food like energy level.
  - Vital energy VE: decreases in time leading to death. Its acquisition requires potential energy.
- To reload its energy levels, the robot has to be placed in an energy source.
  - Two energy sources are considered for each energy type.

$$\Delta VE = \begin{cases} \alpha_{PE} & \text{if } Reload_{VE} \\ 0.5 \cdot \alpha_{VE} & \text{if } Rest \\ \alpha_{VE} & \text{Otherwise} \end{cases}$$

$$\Delta PE = \begin{cases} -\alpha_{PE} & \text{if } Reload_{VE} \\ \alpha_{PE} & \text{if } Reload_{PE} \\ 0 & \text{Otherwise} \end{cases}$$

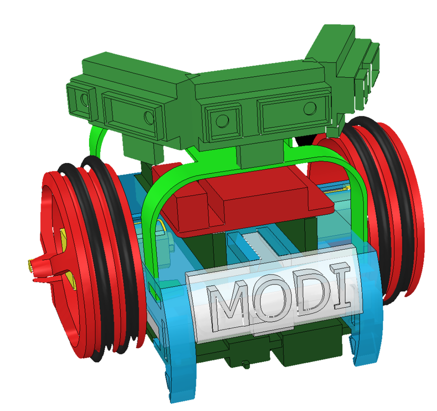


Virtual scenario

Agent: the MODI robot

- MODI (MODular Intelligent) is a compact open-hardware sensorless robot.
- Made with wireless capabilities for swarm robotics applications.
- Proximity sensors where attached.
- Virtual sensing of the energy sources is considered.
  - The robot instantaneously knows the position of any energy source inside a range of vision.
- Reward conditions prioritized from top to bottom

Seeking behaviour	Activation	Reward condition
PE seeker	PE $\leq$ 0.2	Is the robot closer to VE_s?
VE seeker	VE $\leq$ 0.5	Is the robot closer to PE_s?
Both	Otherwise	Is the robot closer to either VE_s or PE_s?

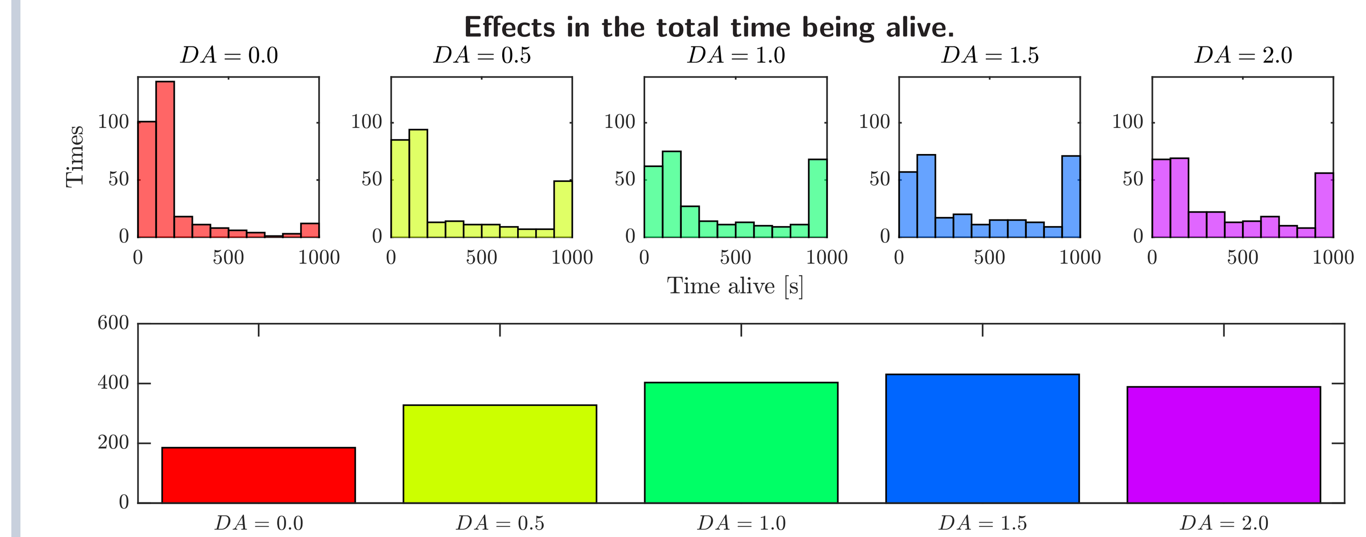


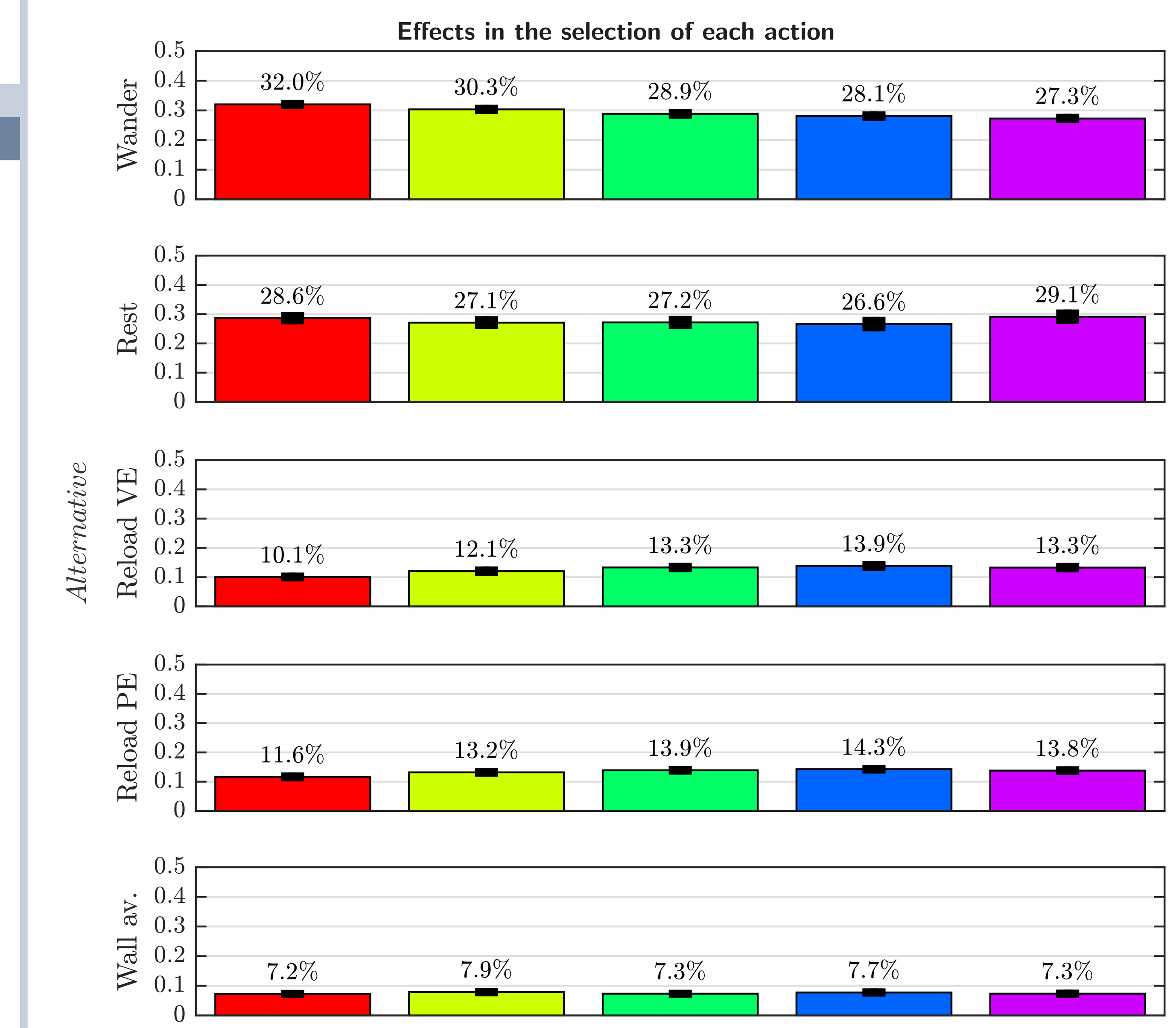
- Agent capabilities:

Motor actions	
$forward$	Move forward.
$turn\_left$	Turn left maintaining its position.
$turn\_right$	Turn right maintaining its position.
$rest$	Stop its movements in order to reduce to the half its vital energy consumption ratio.
$reload$	Without physical response, reload energy while being above a source.

Cognitive alternatives	
$Wander$	The agent randomly selects and executes a movement between $forward$, $turn\_left$ and $turn\_right$.
$Rest$	Reduce vital energy consumption executing the $rest$ motor movement.
$Wall_{av}$	Avoid collisions with walls. Depending on where is the wall placed, the agent turns or moves away.
$Reload_{VE}$	Has the goal to increase its VE level. The agent moves closer to a VE_s turning or moving forward, or, if the agent is close enough, reload.
$Reload_{PE}$	Similarly as $Reload_{VE}$, the agent acts in order to increase its PE level.

Tonic DA level effects in the robot's performance





Conclusions

About the tonic DA effects in the CBG loop

- The presented model is able to correctly perform a selection and, among trials, learn.
- About learning:
  - Lower tonic DA produces a lower signal to noise ratio.
  - For lower signal to noise ratios, noise leads the selections.
  - There is a tonic DA range with direct relation with performance.

About the robot controller implementation

- The controller implementation shows the factibility of using the proposed CBG model as a decision making mechanism in an artificial intelligent agent.
- The system is able to learn on-line its best option, given its current energy levels.
- Tonic DA controls the agent's exploration-exploitation trade-off. Direct related to its tonic DA level, the agent
  - modifies its probability of selection of its better option.
  - modifies its surviving skills.

Main references: Leblois, Arthur, et al. "Competition between feedback loops underlies normal and pathological dynamics in the basal ganglia." The Journal of Neuroscience 26.13 (2006). Humphries, et al. "Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia." Frontiers in neuroscience 6.9 (2012).

cristobal.nettle@gmail.com