

Research Article

Action Selection and Operant Conditioning: A Neurobotic Implementation

André Cyr and Frédéric Thériault

Département d'Informatique, Université du Québec à Montréal (UQAM), Succursale Centre-Ville, Case Postale 8888, Montreal, QC, Canada H3C 3P8

Correspondence should be addressed to André Cyr; andre.cyr1@videotron.ca

Received 23 January 2015; Revised 7 April 2015; Accepted 14 May 2015

Academic Editor: Maki K. Habib

Copyright © 2015 A. Cyr and F. Thériault. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Action selection (AS) is thought to represent the mechanism involved by natural agents when deciding what should be the next move or action. Is there a functional elementary core sustaining this cognitive process? Could we reproduce the mechanism with an artificial agent and more specifically in a neurobotic paradigm? Unsupervised autonomous robots may require a decision-making skill to evolve in the real world and the bioinspired approach is the avenue explored through this paper. We propose simulating an AS process by using a small spiking neural network (SNN) as the lower neural organisms, in order to control virtual and physical robots. We base our AS process on a simple central pattern generator (CPG), decision neurons, sensory neurons, and motor neurons as the main circuit components. As novelty, this study targets a specific operant conditioning (OC) context which is relevant in an AS process; choices do influence future sensory feedback. Using a simple adaptive scenario, we show the complementarity interaction of both phenomena. We also suggest that this AS kernel could be a fast track model to efficiently design complex SNN which include a growing number of input stimuli and motor outputs. Our results demonstrate that merging AS and OC brings flexibility to the behavior in generic dynamical situations.

1. Introduction

The vast topic of action selection (AS), including decision-making, behavioral choice, and behavior-switch as nomenclatures, is thoroughly explored from different perspectives of comprehension, levels of resolution, and scientific communities [1]. The AS biological phenomenon results from a neural process that leads to the observation of an agent doing one action over several others. The precise neural substrate underpinning this mechanism is not yet discovered [2, 3]. Even though many insights [4, 5] point toward how to simulate the AS natural process in artificial agents, there is no consensus to approach this cognitive phenomenon. In this view, the neurobotic domain aims to study AS from bioinspirations but applied for artificial intelligence (AI) and robotics purposes [6]. As a premise, building controllers for unsupervised autonomous robots necessarily requires a dedicated mechanism to operate behavioral transitions. Moreover, in real world, these actions should be adaptive instead of being ruled by fixed-patterns. Thus, a simulated

AS process should be flexible enough to cope with changing environments. These adaptive behaviors could come from learning functions which act as modulators of the AS process.

Artificial spiking neural networks (SNN) [7] have been successfully used as brain-controllers for robots, and several researches have proposed different computational models implementing AS through this specific experimental paradigm [8, 9]. A major aspect of SNN is to understand the information process at the level of a single spike [10]. Therefore, timing of spikes can be used for temporal event correlations and associative learning. As such, it could be interesting to study an AS mechanism in combination with an operant conditioning (OC) process, since we anticipate that these processes add more flexibility to switch behavior from their interactions, sharing both the ability of specifying actions.

The function of an AS process is to decide between different actions depending on the context. As a matter of fact, invertebrate neural organisms like *C. elegans* [11, 12], cnidarians [13], and fruit flies [14] do well in choosing among several

actions with only small circuits of neurons. Those include command-center neurons and central pattern generators (CPG) [15–18], which are well recognized for their intrinsic oscillation property. A deduction that could be drawn is that modeling an AS process does not necessarily require complexity as in high brain structures. Thus, the working hypothesis for the emulation of an AS process states that a simple mechanism should then be derived. In this paper, we use a basic CPG neural structure which helps in simulating an AS process containing sensory inputs, motor outputs, and decision neurons [19].

We propose to study the AS process within a SNN framework, targeting bioinspired robots controllers. Our first motivation is to combine AS and OC processes in a single neuro-robotic model. The main goal is to build a simple yet adaptive AS mechanism merged with the plasticity feature of an OC learning rule, while both occur under a dynamical scenario. A second objective is to develop a fast track method for implementing general AS processes into SNN. This research was driven by the fact that it is still a challenge to create a robot controller with the ability to learn from multiple sensory cues and actions in a SNN paradigm.

Theoretical Background. In neurosciences, the drive to accomplish a behavior emerges from a real-time dynamic of external sensory cues and internal values, where the different competitive neural signals ultimately orient the agent toward one preferred action. In a psychological view of the AS problem, serial processes occur from sensors to motors ending in a behavioral choice. According to the literature in computational cognitive science, the affordance competition hypothesis [20] argues that such a process is parallel and implies a prior specification of possible actions from ongoing sensory inputs. Specifically, when dynamical processes include several feedback loops in high neural structures and an attentional mechanism, the brain focuses on a specific winning action while continuously searching for other actions to do, depending on the context [21].

In the robotic domain, computational models of the AS process have been proposed (stochastic accumulator, linear-ballistic accumulator, and integrated accumulator models) [22, 23] as well as CPG in conjunction with SNN [24]. Since only a few studies in that field have investigated the AS process using SNN as bioinspired brain-controllers, our study takes another step in this direction. Therefore, our focus is on the close interaction between AS and the OC learning function, which we propose as a novelty in the domain.

Empirically, the angle of the AS problem was often to reach an optimal solution with a statistical approach [5, 25] or reproducing biological data. In our research, we wanted to consider the modulation factors that may influence the dynamic of an AS model from its interaction with a learning rule. In this perspective, a learning skill may improve a robot’s choice of actions to determine the future solutions. OC consists in one of these primary learning functions allowing cognitive agents to associate a feedback from their own actions. The natural OC process is well understood at the level of invertebrates [26]. Therefore, among others [27, 28], learning with OC represents one potential modifier of the AS

mechanism, perhaps allowing more flexibility from synaptic plasticity features in their adaptive behaviors. From its own past actions, which gave rewarding or punishing feedbacks, a robot may eventually pick up a different action, accelerating or decelerating the bias toward an oriented alternative.

We address these questions of the AS process combined with OC by evaluating a simple scenario in virtual and physical robots. This current work does not focus on extensive tests nor evaluates the overall computational impact of the parameters involved in the AS-OC models. It was also beyond the scope of this paper to challenge other AS approaches. Despite these limitations, we show a biologically plausible core base of these mechanisms in a neuro-robotic implementation. A benefit for robots to include the AS and OC critical processes is undeniable, since most physical robots are now able to perform a rich selection of actions that may be organized in hierarchical priorities, sequential fixed-patterns, competitive actions, and conflicting parallel behaviors.

In resume, we show an AS mechanism based on a CPG structure and few elementary neural units. This AS process was subject to modulation when merged with an OC learning rule. Together, these processes offer more flexibility to choose the best action under dynamical and variable contexts. Further demonstrations in more complex scenarios remain to be studied.

2. Methodology

2.1. The Spiking Neural Model. We propose a simple scenario to explore the AS and OC interrelation, explaining both processes in a neuro-robotic paradigm. The robot’s controller consists in artificial neural units connected by synapses. Our SNN model [29], similar to standard leaky integrate-and-fire neuron models, is based on a membrane potential variation, integrating nonlinearly, and temporally ongoing inputs through the SNN (1). In these neurons, when the membrane potential reaches a specific threshold, an all-or-none action potential is triggered. To start the CPG dynamic at the beginning of a simulation, a realistic neural property of endogenous pacemaker is implemented from adding a stronger leak (see (1) and starter neuron in the SNN). Consecutive to a spike emission, an electrical flux is sent, transformed at the synapse into a local excitatory or inhibitory synaptic postpotential current. This is then received at the targeted elements (2). The synapse is computationally modeled as a dynamical weight and is subject to be modulated from learning functions. The learning rule we used in this SNN is an adapted spike-timing dependent plasticity (STDP) [30–32]. The result of a STDP function is to increase a synaptic weight if the preneuron spikes before the postneuron unit, in a defined short time window. If the prespike arises after the postspike, then the inverse correlation leads to decreasing the synaptic weight (3).

Equation 1: Discrete-Time Neural Input Integration Function. Consider the following:

$$v_m(k) = f(v_m(k-1) + v_t(k)) + l. \quad (1)$$

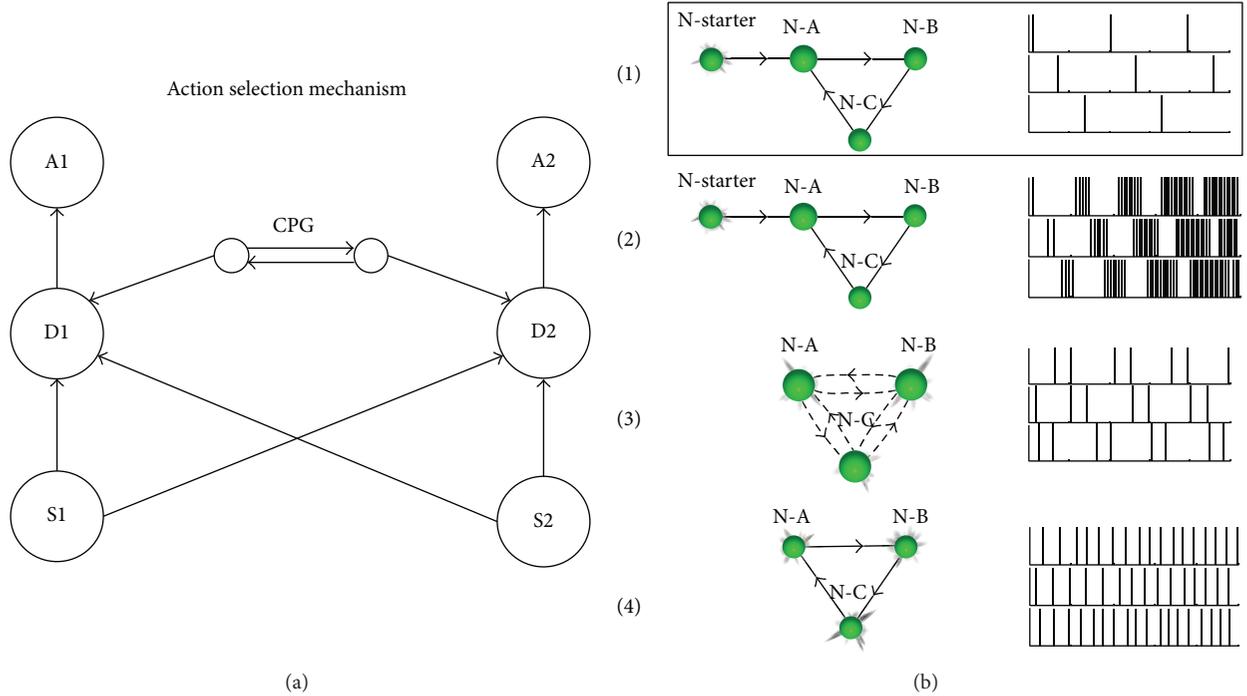


FIGURE 1: Dynamic of the AS mechanism and CPG topologies. (a) First, a stimulus is perceived from a sensory neuron. Then, the spikes emission from the sensory neuron targets all decision neurons, conserving a neutral aspect in the choice of possible actions to come. Finally, when the sensory input coincides with the CPG input at the decision neuron, the linked action will be performed (alternatively switching between actions). (b) In (1), the CPG represents the serial circular excitatory topology we used in this work. In (2), we show one option among several possibilities of tuning differently the internal neural parameters such as the postpotential state value, the refractory period, and the length of the neural curves. In (3), we show an all-to-all reciprocal inhibitory CPG circuit built with endogenous pacemaker neurons. In (4), we still keep the pacemaker neurons in the CPG but it is now linked with excitatory synapses. Overall, (b) (comparative spike bars in the same time scale on the right side) demonstrates the diversity of the output patterns reflected in different CPG, derived from small structural and functional changes.

$v_m(k)$ = membrane potential at cycle k , $v_t(k)$ = sum of the synaptic input as calculated in (2), f = ascending exponential function set between 0 and threshold (set as 65), and l = leak current for pacemaker property (set as 1).

Equation 2: General Alpha Function Representing the Postsynaptic Potential Curve. Consider

$$f(t) = ge^{-t/\pi}. \quad (2)$$

g = amplitude (set as 20), π = tau (set as 7), and t = time since spike (in cycle).

Equation 3: STDP Function Used. Consider the following:

$$\Delta w = \alpha_{t_{\text{post}} - t_{\text{pre}}} e^{(t_{\text{post}} - t_{\text{pre}})/\pi}. \quad (3)$$

Δw = synaptic weight change, $\alpha_{t_{\text{post}} - t_{\text{pre}}} = 1$ or -1 , depending on the sign of $t_{\text{post}} - t_{\text{pre}}$, and π = time constant.

STDP coefficient for Δw is as follows.

Maximum variation period = 3000 cycles.

Maximum synaptic change = 35%.

Maximum STDP time window = 25 cycles.

2.2. The AS Process. The elements in the AS mechanism consist of four basic groups of cells. The first group represents the decisions or command-neurons pointing to the action neurons which activates actuators (second group). The third group contains the sensory neurons providing contextual inputs linked to the decision neurons. Finally, the last group of cells contains the CPG neurons weakly connected to the decision neurons. The main function of the CPG is to provide a regular oscillation output pattern to bias one preferred decision neuron over the others. Notice that, a CPG neuron output is never allowed to trigger its linked decision neuron since the EPSP is too weak to reach the spike threshold. However, when pairing sensory and CPG inputs, only then can it reach its threshold and spike (see Figure 2). Therefore, the tuning of the parameters must overlap in relation of the CPG period and the sensory duration. In our experiments, a full CPG loop takes 90 cycles; hence, one CPG neuron spikes every 30 cycles. The sensory input duration last approximately 110 cycles. A second effect of the CPG is to disambiguate equal sensory inputs, a known conflicting problem difficult to resolve in the AI domain. Finally, CPG could also be understood as rhythmic internal values, feeding input in the AS process.

To graphically represent the AS process (Figure 1, left side), we show it in a complete generic scenario of two sensors

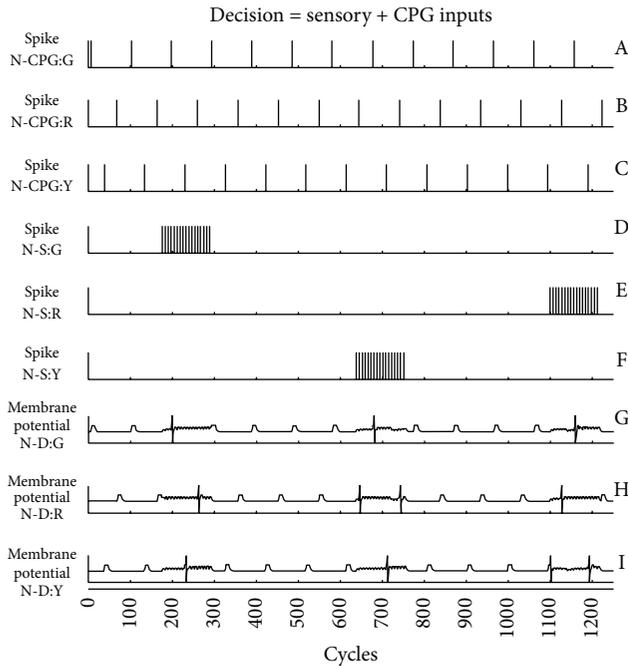


FIGURE 2: The figure shows the regular output spikes pattern from the serial circular excitatory CPG neurons (graphics A to C), which is the exact topology we used in our experiments. The mathematical neural dynamic behind these neurons is the same as the other spiking cells. When the CPG spike coincides with the sensory spikes (graphics D to F), the membrane potential of the decision neurons (graphics G to I) reaches the threshold for a spike emission (see the AS process in a complete SNN architecture 4).

and two actions. The SNN architecture is divided into three distinct layers: the sensory inputs, the internal integrative states, and the external action outputs. In Figure 3, the AS components are also clustered in a singular module within the generic but detailed SNN.

For the CPG’s kernel, we chose to embed the most regular and minimalist structure (see option 1 in the highlighted right side of Figure 1). The synaptic weights were all set to 100%, in order to have a continuous spike loop. To start the CPG, we used a biologically plausible endogenous pacemaker that shuts down just after initiating the dynamic. This starter option could be understood in terms of an internal value (i.e., low batteries, attentional process, and sensory-motor input) or could also be any other kind of triggers. As a result, the three neurons of the CPG are stimulated one after the other because of the circular serial excitatory connections 2.

2.3. The OC Learning Procedure. The cellular components included in the OC process consist in sensor neurons that provide the contextual inputs for Decision-to-Action neurons to generate the behaviors. Also, external reinforcer pointing to predictor neurons are also connected to Decision-to-Action neurons. Since sensor neurons are weakly linked to the predictor neurons but contain an STDP rule, the repetitive coincidence of the reinforcer (following the desire action) and the sensory input at the predictor neurons will increase

the synaptic weight. Therefore, sensory inputs will eventually trigger actions without any further needs of reinforcers [29].

2.4. The SNN Architecture. Specifically used for our results in a three-sensor and three-action context (Figure 4), the sensory neurons are composed of three color sensors (green, yellow, and red) in addition to one light sensor to perceive the rewarding light. The motor output neurons are represented by three LEDs (green, yellow, and red). Our AS process includes, as a modulating element, CPG placed at the intermediary neural layer. It contains three neurons paired with the same number of the possible actions. The proposed CPG kernel consists in excitatory neurons organized in a serial circular topology.

Each CPG neuron in the CPG network is connected to its own decision neuron with a small synaptic weight. One target of a decision neuron is its connected action neuron with a strong synaptic weight between these units; when a decision neuron spikes, the linked action neuron spikes as well. Each decision neuron is also weakly connected to its own predictor neuron for the learning context interrelation. In this experiment, the predictor neurons target their output to all other decision neurons with inhibitory strong synaptic links. Therefore, when the sequence sensor-action-reward is learned from a precise predictor neuron, it will shut down all other possible actions. This arbitrator mechanism could be understood as a type of neural competition. Initial synaptic values used in our SNN were manually tuned and can be retrieved in Table 1.

2.5. The Task and the Actions. In the virtual experiment (Figure 5), the SNN is implemented in a static robot. The robot’s task consists in learning to match colors between its three possible actions of LED emission (green, yellow, and red) and the color blocks perception. Our 3D simulation software environment (SIMCOG-NeuroSim, AI-Future) allows three different color blocks (green, yellow, and red) to move continuously, at a constant speed, in a clockwise circular trajectory, passing one at a time just in front of the robot. The time frame perceptible contact enables the robot to, at least, produce one different action for each block over 1000 cycles. In the first part of the experiment (0–10000 cycles), a rewarding light (not shown) triggers only when the color LED action matches the block of the same color. At cycle 10500, the robot was moved temporarily for 3000 cycles (cycle 10500–13500) to a location where no sensory input is received, allowing a forgetting factor to operate and reset the synaptic weights. Then, the robot was replaced to its initial position for another round. However, in this part, the rewarding light follows only when the LED emission is on the next color block. The purpose of this part is to show how efficient the AS and OC dynamics can modify the behavior, since these novel learning associations are achieved in a single trial.

As a proof of concept and endpoint in the robotic domain, we reproduced the virtual setup in a physical experiment (Figure 6). The SNN is totally identical and we simply transferred it into the physical robot without any further adjustments. For simplicity, we chose the EV3 Lego Mindstorm (Lego Inc.) as physical platform. The main processor is an

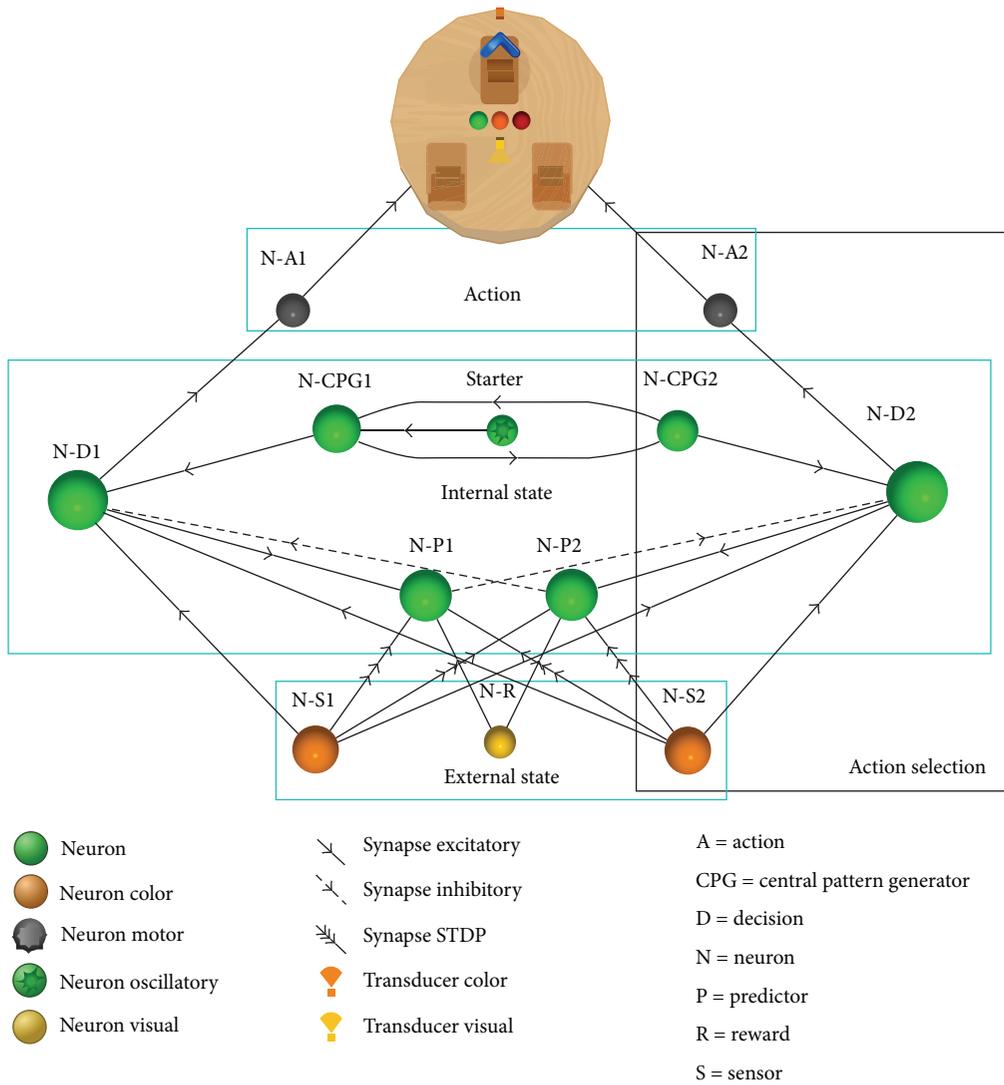


FIGURE 3: Generic SNN of AS and OC in a two-sensor and two-motor configuration.

ARM9 core clocking at 300 MHz and it contains 64 MB RAM. The LEDs are similar to the virtual scenario except that there are only two colors available, the green and the red. When the two of them are opened at the same time, the resulting color is orange, hence having our third color for the experimentation. A light sensor is also used to read the external rewarding light, which was synchronized and delivered from a Raspberry Pi board, just after a desired action is done by the robot. A NXT Lego Mindstorm controller (Lego Inc.), mounted on a shaft, controls the rotation of three color bricks (green, orange, and red) using one attached motor. A slow stepwise speed was set with no possibility of modulation from the robot. In this configuration, the bricks pass just in front of the color sensor. When a sensor catches a color block, the numerical value is converted in an artificial electrical current with an adapted scaling factor for the SNN. Only the first learning part was done for the demonstration. Supplementary material is available at <https://www.youtube.com/watch?v=8MXA4wxJSpE> and consists of a video of the experiment.

3. Results

The results from the virtual experiment were obtained in a single trial. The following graphic data will refer to Figure 4 for the SNN architecture and Table 1 for its associated synaptic weight matrix. In Figure 7, we can observe at the beginning of the simulation that when the Green Sensor neuron (N-S:G) spikes (black bars in graphic A), the robot tries alternative actions of lighting up each LED (graphic B: green, graphic G: yellow, and graphic L: red). Since there was no reward for any actions triggered prior to cycle 300, no learning from the STDP rule was observed at the synapse going from the sensor to the predictor neuron (D, I, and N). At around cycle 500, a first yellow block is perceived from the sensor yellow (F), while the CPG continuously provides alternative actions of LED emissions. Specifically, with the lighting up of the yellow LED (G), and with the following light reward (not shown), the associated predictor neuron spikes (H). Consequently, a positive association between the Yellow

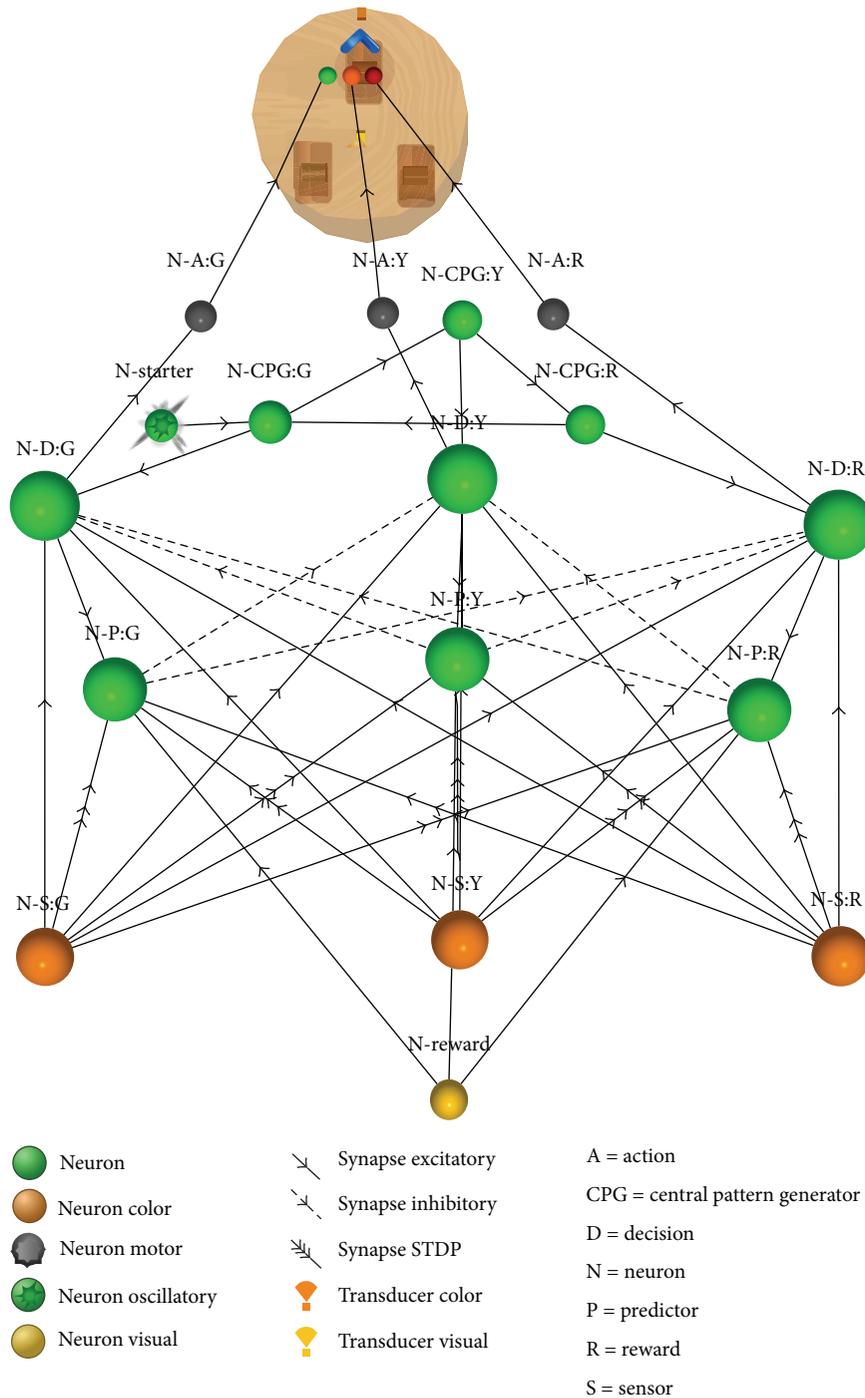


FIGURE 4: SNN architecture used for our experiments, demonstrating the interrelation of the AS process and the OC learning rule in a three-sensor and three-motor configuration. See also Figure 3: A = action, CPG = central pattern generator, D = decision, G = green, N = neuron, P = predictor, S = synapse, R = red, Y = yellow, and STDP = three consecutive arrows on a synapse.

Sensor neuron and this predictor neuron starts to increase the STDP coefficient (I). This affects the synaptic weight to a bound limit when several associations occurred, stabilizing at around cycle 4000. The role of the predictor neuron (H) in this SNN is to inhibit the other decision neurons and their connected action neurons (B, L). At around cycle 8500, one

can see that the robot has fully learned the three sensory-motor contexts by pairing the good LED action with the good perceived color block. Since the period of the CPG neurons and the rotation of the color blocks did not fit perfectly, the learning time frame for each sensory-motor pair is not identical.

TABLE 1: Synaptic weights associated with the SNN of Figure 4. These are represented by numerical percentage values. Delays in cycle as well as the type of the synapse are also mentioned.

Synaptic weight table			
Name	Strength	Delay	Type
S-A:G/E:G	5	0	Excitatory
S-A:R/E:R	5	0	Excitatory
S-A:Y/E:Y	5	0	Excitatory
S-CPG:G/CPG:Y	100	30	Excitatory
S-CPG:G/D:G	30	0	Excitatory
S-CPG:R/CPG:G	100	30	Excitatory
S-CPG:R/D:R	30	0	Excitatory
S-CPG:Y/CPG:R	100	30	Excitatory
S-CPG:Y/D:Y	30	0	Excitatory
S-D:G/A:G	100	0	Excitatory
S-D:G/P:G	15	0	Excitatory
S-D:R/A:R	100	0	Excitatory
S-D:R/P:R	15	0	Excitatory
S-D:Y/A:Y	100	0	Excitatory
S-D:Y/P:Y	15	0	Excitatory
S-Reward/P:G	8	0	Excitatory
S-Reward/P:R	8	0	Excitatory
S-Reward/P:Y	8	0	Excitatory
S-S:G/D:G	15	0	Excitatory
S-S:G/D:R	15	0	Excitatory
S-S:G/D:Y	15	0	Excitatory
S-S:G/P:G	15	0	Excitatory
S-S:G/P:R	15	0	Excitatory
S-S:G/P:Y	15	0	Excitatory
S-S:R/D:G	15	0	Excitatory
S-S:R/D:R	15	0	Excitatory
S-S:R/D:Y	15	0	Excitatory
S-S:R/P:G	15	0	Excitatory
S-S:R/P:R	15	0	Excitatory
S-S:R/P:Y	15	0	Excitatory
S-S:Y/D:G	15	0	Excitatory
S-S:Y/D:R	15	0	Excitatory
S-S:Y/D:Y	15	0	Excitatory
S-S:Y/P:G	15	0	Excitatory
S-S:Y/P:R	15	0	Excitatory
S-S:Y/P:Y	15	0	Excitatory
S-Starter/CPG:G	50	0	Excitatory
S-P:G/D:R	10	0	Inhibitory
S-P:G/D:Y	10	0	Inhibitory
S-P:R/D:G	10	0	Inhibitory
S-P:R/D:Y	10	0	Inhibitory
S-P:Y/D:G	10	0	Inhibitory
S-P:Y/D:R	10	0	Inhibitory

Action selection and operant conditioning

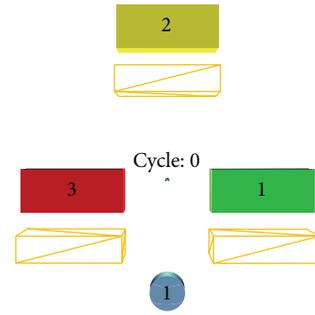


FIGURE 5: Snapshot of the 3D environment including the robot and three color blocks rotating constantly clockwise and passing just in front of the color sensor of the robot.

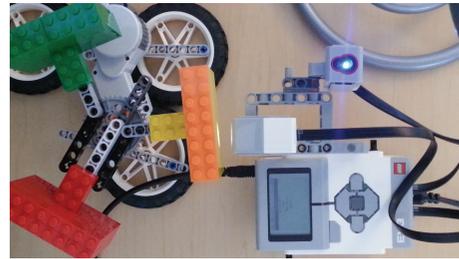


FIGURE 6: Picture of the EV3 brick Lego Mindstorm and the rotative color blocks setup. The color sensor points to the blocks and the light sensor points up.

Between cycle 10500 and 13500, we changed the robot's location to avoid perception of the color blocks. This was done in order to allow the SNN to reset the synaptic weights to their initial values (using a forgetting parameter present in the STDP rule). This was optional, and learning forever would be the result scenario if the feature was not active. Unlearning could also be obtained from inverting the temporal sequence of the sensor, action, and reward. If there is no correlation anymore, the STDP rule will progressively decrease the synaptic weight. In another simulation setup, a punishment (inhibition) could also serve as a fast negative modulation factor of the synaptic weights.

The last part (>13500 cycle) of Figure 7 demonstrates the online adaptive behavior aspect of the SNN embedding OC and AS. One can observe that the robot must choose a different action in order to receive the reward. In this case, lighting up a green LED on a yellow block, a yellow LED on a red block, and a red LED on a green block triggers the reward. The corresponding STDP factors (E, J, and N) match these three learning sets.

As for our physical simulation, the architecture was not modified in any way, except for the binding of logical sensors and motors to the robot. Figure 8 shows the results of the simulation, which were obtained in a single trial. They show approximately the same data, with more or less precision and small artifacts. This is due to the fact that it is indeed much easier to configure variables and the context of virtual environments than it is in the real world.

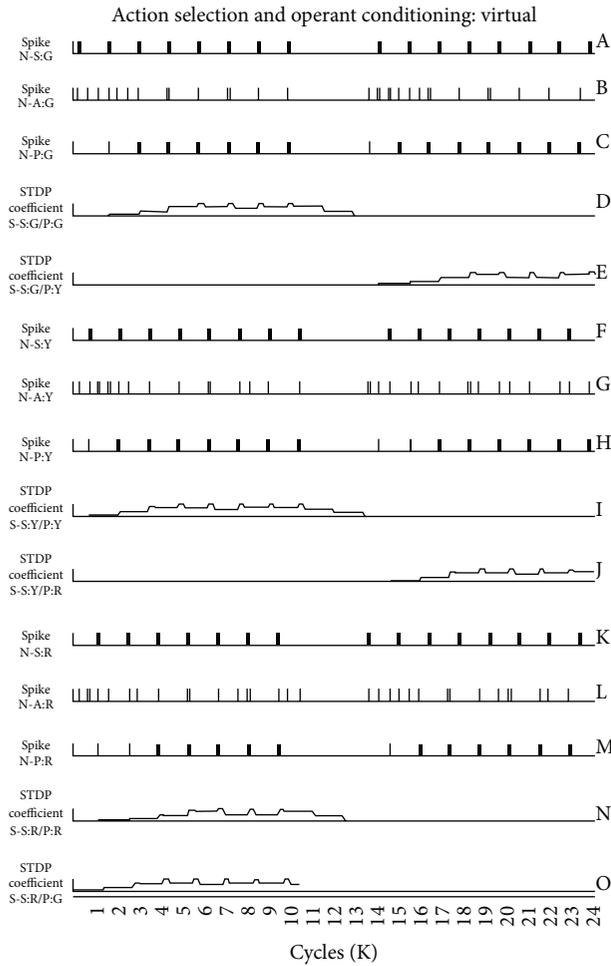


FIGURE 7: Graphics of the neural spikes and the STDP rule factors occurring over 24000 cycles. The first five (graphics A to E) concern the green block and green LED. The middle five capture the logic of the yellow block and the yellow LED. The last group of five is associated with the red color. Around cycle 8500, the robot has learned to perform the right action with the appropriate cue. This means that, upon triggering an action, if the predictor spikes from a following reward, the STDP coefficient will increase, boosting the synaptic weight between the sensor and the predictor. Once this synaptic weight reaches a critical threshold, the sensor input will trigger a spike to the reinforced predictor neuron, without the need of the reward.

4. Discussion

In this paper, we explored the AS process through a neuro-robotic perspective. Since this general mechanism directly involves actions, we demonstrated the phenomenon in the context of OC procedures which also imply a selection of actions from reinforcer. Our main objective was to study the benefit effects of merging this learning rule with an AS process. A second concern was to provide a fast track solution to efficiently design more complex SNN used as brain's controller for virtual and physical robots that include several motor outputs. We propose a basic CPG motif as one

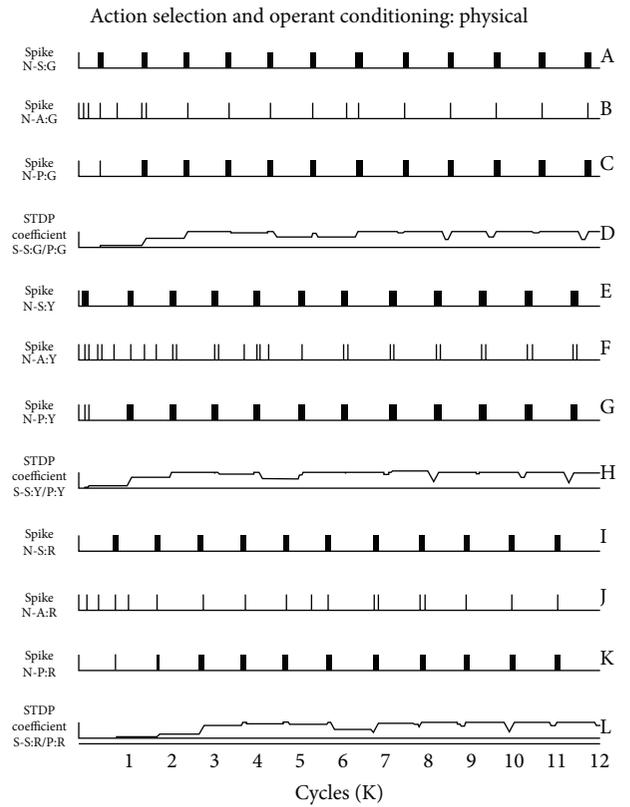


FIGURE 8: Graphics of the neural spikes and the STDP rule factors occurring over 12000 cycles.

key component of an AS process, in order to neutrally switch between its available actions. With the CPG structure used in relation with a sensory input context, a decision neuron gets all the information needed to bias toward one preferred action. We also showed that the OC learning function influences the AS process, conferring supplementary adaptive behaviors from synaptic plasticity.

We chose a simple CPG topology as one component of the AS mechanism. Other CPG configurations are possible [33], including those built with reciprocal inhibitory synaptic links and endogenous pacemaker neurons, though the analytical issues are more complex to track and predict. Tuning the parameters (i.e., postpotential spike value, threshold) of individual neuron differently could also influence the rhythm, affecting the CPG network by increasing or decreasing their output periods. After several options, we found that a serial excitatory circular CPG motif is a good trade-off between simplicity and benefits. In our AS model, without any other synaptic feed, this CPG configuration will spike one unit after the other, indefinitely and at a constant rate. We showed the AS model in a generic two-two example and in a specific three-three sensory input and motor output configurations. Adding more sensors and actions will necessary require other neurons in the CPG network, though their numbers are linearly related to their attached decision and action neurons, acting as a premotor structures. In this case, hierarchical

groups of CPG/actions could also replace the serial circular topology, possibly avoiding useless spikes or triggering other networks. Allowing different combinations and compositions of CPG units also dramatically increase the behavior possibilities, without considering a one-to-one CPG-action, though it was not explored in this paper.

Adding the decision neurons (equivalent to command-neuron in invertebrates) into the AS model allowed flexibility, regarding several contextual input sources. The CPG units bring the decision neurons membrane potential to a subthreshold firing level. Since the CPG period is fixed, the speed-accuracy trade-off (SAT) of the decision-making [34] result is fast and accurate. Unfortunately, but no adjustment is possible, a major point to consider when modeling an AS process. Sometimes, cognitive agents must take decisions quickly while in other conditions, it is necessary to take the time to compute the best decision. According to a recent hypothesis [35], flexibility of the SAT's response variables depends on adjusting the baseline firing rate, the sensory gain and noise inputs, the firing threshold, and related bound parameters of the receiving neurons [25, 36].

Having those various AS modulator factors in mind, designing complex SNN with several populations of neurons including heterogeneous individual neural parameters values is possible, but highly complex to tune properly. In this perspective, progressively integrating stronger/lower and faster/slower CPG inputs could add discriminative and flexible response advantages as well as offering more realistic behavioral features of the AS mechanism. A computational challenge in an AS model within the neurobotic field is to allow the SNN to dynamically change all these initial fixed parameters values, conferring considerable adaptive properties at the level of the cognitive agent. In this vein, perhaps a question remains about the SAT: what are the variables biases in the AS process when there is no emergency to choose one action?

In our experiments, reversing the rotation order for the block to counter clockwise, accelerating or decelerating the speed of rotation as well as mixing the color order do not change the qualitative aspect of the learning curve. However, the temporal relation between the perception time contact of the sensor inputs and the timing of the CPG influences the number of occurrences of these associations and, thus, the length of time needed to learn. In any case, the EPSP timing between the CPG and the sensory neurons is of major importance and needs a full coherence into the whole dynamic system. In this perspective, the physical experiment shed some light on the temporal robustness of our AS and OC models, justifying its inclusion in the study. Without changing any parameters in the SNN, the EV3 robot was able to learn very well how to receive the reward when achieving the good action, even if the rotation of the blocks was irregular due to the imprecision of the material.

We explained how the OC learning rule modulates the AS process in a SNN paradigm. The sensory-motor context does influence the decision to do one action over others. These decisions were not just built-in reflexes. Moreover, the behavioral plasticity was observed even if the CPG dynamic was fixed. At this point, some interesting variances could be to add

other learning rules that are not limited by an OC procedure. As such, integrating nonassociative (habituation) and other associative learning (classical conditioning) functions could complete the design of an AS model, but it was beyond the scope of this paper.

One concern we avoid in this paper is the attentional problem. We chose to ignore this major cognitive component, mostly because of the current lack of neural mechanism and theory when applied in lower neural system organisms. We understand that basal ganglion or any subcortical or cortical implications in the AS process seem to be relevant structures in higher biological neural systems such as in humans or primates. However, our present perspective on the AS problem lies in the AI neurobotic domain, which is still beyond the reach of lower cognitive natural species. Our strategy aims to engineer a bioinspired minimalist solution from emulating simple neural organisms such as in *C. elegans*, which selects its actions without involving huge structures. Instead, command-neurons and CPG neurons are common cellular elements found in the primitive invertebrate neural circuits. No doubt, complex neural layers may extend and add profitable values in the simulation of an AS process but should not be a necessary requirement to achieve a basic one. These evolutionary concerns may eventually find an echo in a multiplicity and hierarchical AS mechanisms.

The AS scope obtained from these results is theoretically not limited to only a few simple actions or unimodal single sensory stimuli. The generic aspect of the AS process comes from the parsimonious components and parameters inside the kernel. The simplicity of this AS module already allows to be adaptive from a 2×2 to a 3×3 sensors-actions scenario without much changes in the SNN architecture. In these two scenarios, as long as there are the same numbers of possible ending actions and sensors as input, the AS core process will operate and tune the same. Therefore, building more complex SNN including several actions should be anticipated as faster and easier, though it remains to be proven in other situations. In that sense, we currently work on a shaping behavior learning technique based on the AS process, while simulating an indoor dynamical navigation task with several possible behaviors. This is an example that can demonstrate how this bioinspired AS process could help in concrete application in the robotic field.

5. Conclusion

This paper showed an AS process made from simple cellular elements. It is based on CPG and sensory neurons which influence decision neurons in their choice to generate a behavior from the action neurons. We demonstrated this basic AS mechanism in an OC learning context that allows behavioral flexibility from their mutual influences. The experiments were conducted under a biologically inspired paradigm, specifically with a SNN acting as brain-controller for virtual and physical robots. In addition, the simplicity and the generic aspect of our AS model may provide a fast track solution to build more complex SNN, including multiple actions in different dynamic scenarios.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] A. K. Seth, T. J. Prescott, and J. J. Bryson, *Modelling Natural Action Selection*, Cambridge University Press, Cambridge, UK, 2011.
- [2] M. D. Humphries, K. Gurney, and T. J. Prescott, "Is there a brainstem substrate for action selection?" *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1485, pp. 1627–1639, 2007.
- [3] J. I. Gold and M. N. Shadlen, "The neural basis of decision making," *Annual Review of Neuroscience*, vol. 30, pp. 535–574, 2007.
- [4] 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," *Adaptive Behavior*, vol. 13, no. 2, pp. 131–148, 2005.
- [5] N. F. Lepora, C. W. Fox, M. H. Evans, M. E. Diamond, K. Gurney, and T. J. Prescott, "Optimal decision-making in mammals: insights from a robot study of rodent texture discrimination," *Journal of the Royal Society Interface*, vol. 9, no. 72, pp. 1517–1528, 2012.
- [6] T. J. Prescott, F. M. M. González, K. Gurney, M. D. Humphries, and P. Redgrave, "A robot model of the basal ganglia: behavior and intrinsic processing," *Neural Networks*, vol. 19, no. 1, pp. 31–61, 2006.
- [7] W. Maass, "Networks of spiking neurons: the third generation of neural network models," *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, 1997.
- [8] T. C. Stewart, T. Bekolay, and C. Eliasmith, "Learning to select actions with spiking neurons in the basal ganglia," *Frontiers in Neuroscience*, vol. 6, article 2, 2012.
- [9] S. Skorheim, P. Lonjers, and M. Bazhenov, "A spiking network model of decision making employing rewarded STDP," *PLoS ONE*, vol. 9, no. 3, Article ID e90821, 2014.
- [10] E. M. Izhikevich, "Simple model of spiking neurons," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1569–1572, 2003.
- [11] S. Faumont, T. H. Lindsay, and S. R. Lockery, "Neuronal microcircuits for decision making in *C. elegans*," *Current Opinion in Neurobiology*, vol. 22, no. 4, pp. 580–591, 2012.
- [12] M. Zhang, W. R. Schafer, and R. Breitling, "A circuit model of the temporal pattern generator of *Caenorhabditis* egg-laying behavior," *BMC Systems Biology*, vol. 4, no. 1, article 81, 2010.
- [13] A. L. Stöckl, R. Petie, and D.-E. Nilsson, "Setting the pace: new insights into central pattern generator interactions in box jellyfish swimming," *PLoS ONE*, vol. 6, no. 11, Article ID e27201, 2011.
- [14] C. R. von Reyn, P. Breads, M. Y. Peek et al., "A spike-timing mechanism for action selection," *Nature Neuroscience*, 2014.
- [15] S. Grillner, "Neurobiological bases of rhythmic motor acts in vertebrates," *Science*, vol. 228, no. 4696, pp. 143–149, 1985.
- [16] K. Matsuoka, "Mechanisms of frequency and pattern control in the neural rhythm generators," *Biological Cybernetics*, vol. 56, no. 5–6, pp. 345–353, 1987.
- [17] A. J. Ijspeert, "Central pattern generators for locomotion control in animals and robots: a review," *Neural Networks*, vol. 21, no. 4, pp. 642–653, 2008.
- [18] M. Heisenberg, "Action selection," in *Invertebrate Learning and Memory*, R. Menzel and P. Benjamin, Eds., vol. 22, Academic Press, 2013.
- [19] P. Redgrave, T. J. Prescott, and K. Gurney, "The basal ganglia: a vertebrate solution to the selection problem?" *Neuroscience*, vol. 89, no. 4, pp. 1009–1023, 1999.
- [20] P. Cisek, "Making decisions through a distributed consensus," *Current opinion in neurobiology*, vol. 22, no. 6, pp. 927–936, 2012.
- [21] P. Cisek, G. A. Puskas, and S. El-Murr, "Decisions in changing conditions: the urgency-gating model," *The Journal of Neuroscience*, vol. 29, no. 37, pp. 11560–11571, 2009.
- [22] S. D. Brown and A. Heathcote, "The simplest complete model of choice response time: linear ballistic accumulation," *Cognitive Psychology*, vol. 57, no. 3, pp. 153–178, 2008.
- [23] M. Usher and J. L. McClelland, "The time course of perceptual choice: the leaky, competing accumulator model," *Psychological Review*, vol. 108, no. 3, pp. 550–592, 2001.
- [24] M. Kassim, N. Zainal, and M. Arshad, "Central pattern generator in bio-inspired simulation using matlab," in *Proceedings of the MEDINFO*, vol. 98, 1998.
- [25] D. Standage, D. Wang, and G. Blohm, "Neural dynamics implement a flexible decision bound with a fixed firing rate for choice: a model-based hypothesis," *Frontiers in Neuroscience*, vol. 8, article 318, 2014.
- [26] B. Brembs, F. D. Lorenzetti, F. D. Reyes, D. A. Baxter, and J. H. Byrne, "Operant reward learning in *Aplysia*: neuronal correlates and mechanisms," *Science*, vol. 296, no. 5573, pp. 1706–1709, 2002.
- [27] K. M. Crisp and K. A. Mesce, "Beyond the central pattern generator: amine modulation of decision-making neural pathways descending from the brain of the medicinal leech," *Journal of Experimental Biology*, vol. 209, no. 9, pp. 1746–1756, 2006.
- [28] T. M. Wright Jr. and R. L. Calabrese, "Patterns of presynaptic activity and synaptic strength interact to produce motor output," *The Journal of Neuroscience*, vol. 31, no. 48, pp. 17555–17571, 2011.
- [29] A. Cyr, M. Boukadoum, and F. Thériault, "Operant conditioning: a minimal components requirement in artificial spiking neurons designed for bio-inspired robot's controller," *Frontiers in Neuroinformatics*, vol. 8, article 21, 2014.
- [30] N. Caporale and Y. Dan, "Spike timing-dependent plasticity: a Hebbian learning rule," *Annual Review of Neuroscience*, vol. 31, pp. 25–46, 2008.
- [31] A. Cyr and M. Boukadoum, "Classical conditioning in different temporal constraints: an STDP learning rule for robots controlled by spiking neural networks," *Adaptive Behavior*, vol. 20, no. 4, pp. 257–272, 2012.
- [32] H. Markram, W. Gerstner, and P. J. Sjöström, "A history of spike-timing-dependent plasticity," *Frontiers in Synaptic Neuroscience*, vol. 3, article 4, 2011.
- [33] A. I. Selverston, "Invertebrate central pattern generator circuits," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365, no. 1551, pp. 2329–2345, 2010.
- [34] P. Cassey, A. Heathcote, S. D. Brown, and O. Sporns, "Brain and behavior in decision-making," *PLoS Computational Biology*, vol. 10, no. 7, Article ID e1003700, 2014.
- [35] R. Bogacz, E.-J. Wagenmakers, B. U. Forstmann, and S. Nieuwenhuis, "The neural basis of the speed-accuracy tradeoff," *Trends in Neurosciences*, vol. 33, no. 1, pp. 10–16, 2010.
- [36] R. P. Heitz and J. D. Schall, "Neural mechanisms of speed-accuracy tradeoff," *Neuron*, vol. 76, no. 3, pp. 616–628, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

