Cognitive Architecture and Software Environment for the Design and Experimentation of Survival Behaviors in Artificial Agents

Bhargav Teja Nallapu[1,2,3] and Frédéric Alexandre[1,2,3]
[1]*INRIA Bordeaux Sud-Ouest, 200 Avenue de la Vieille Tour, 33405 Talence, France*
[2]*LaBRI, Université de Bordeaux, Bordeaux INP, CNRS, UMR 5800, Talence, France*
[3]*IMN, Université de Bordeaux, CNRS, UMR 5293, Bordeaux, France*

Keywords: Cognitive Architecture, Cerebral Systems, Survival.

Abstract: We discuss here the characteristics of a software environment appropriate for the development of a bio-inspired cognitive architecture, which can emulate the behavior of autonomous intelligent agents. First, it is reminded that, while the focus is often set on the more abstract aspects of cognitive abilities, studying the fundamental bases of intelligence that allow for autonomy is a prerequisite for well defined intelligent systems. Secondly, we highlight functional loops associating cerebral structures including the basal ganglia in the brain of most species along the evolution. They are dedicated to the organization of behavior under the constraint of reinforcement, corresponding in their simplest expression to the selection of action for survival. Lastly, concerning the simulation of such models, we describe a software environment to study such relations in a more controlled way than hardware implementations, by adapting a platform built on the top of a video game for the development of classical artificial intelligence models. We explain here how our neuronal model exhibits bodily and internal characteristics necessary for survival tasks and how these characteristics are plugged in the simulation platform. Some scenarios of survival are reported as an illustration of this environment.

1 INTRODUCTION

Autonomous behavior and ability to survive forms a key prerequisite for any cognitive agent. High level cognitive agents, on the other hand, are often characterized to solve specific complex problems while their capacities to survive and evolve autonomously are often disregarded. We describe a bio-inspired approach that (i) forms the basis of understanding these capacities and (ii) defines the characteristics of a software environment in which the problem of survival is demonstrated using a virtual agent.

From a simple organism such as C-elegans with 302 neurons to a human brain with billions of neurons, the changes that led to more complex behavior are quite challenging to understand. So is extracting the invariants that underlie the survival capacities of these species. Within this scope, at all stages of phylogeny, two processes can be mentioned: signaling and regulation. The first indicates that, by various means, the organism is informed about the state of the environment and its own (bodily and mental) state. The second emphasizes that the organism has different ways (motor, chemical, decisional) of responding to adapt (its body and mental analysis) to the situation. These processes are central in the concept of autopoiesis introduced by (Maturana and Varela, 1991), as the property of a system to produce itself, permanently and in interaction with its environment, and thus to maintain its organization despite the changes. First introduced to characterize living cells, it was then extended to organisms and became equally important in cognitive science (Varela et al., 1992).

In vertebrates, the existence of central and peripheral nervous systems makes it possible to better emphasize this distinction between the environment, body and brain and to introduce two types of perception, exteroception for sensations coming from the environment (e.g., visual or auditory) and interoception for those coming from the body (including pleasure, pain and needs), as well as different types of responses aimed at controlling these two classes of signals (Craig, 2003).

The basal ganglia (BG) are brain structures that play a central role in the selection of responses adapted to internal and external states (Redgrave et al., 1999). These structures are present in all vertebrates and an homologous structure has even been

Nallapu, B. and Alexandre, F.
Cognitive Architecture and Software Environment for the Design and Experimentation of Survival Behaviors in Artificial Agents.
DOI: 10.5220/0006920301520159
In *Proceedings of the 10th International Joint Conference on Computational Intelligence (IJCCI 2018)*, pages 152-159
ISBN: 978-989-758-327-8

found in arthropods (Strausfeld and Hirth, 2013), with similar neuronal activities and connectivity allowing behavioral regulation. This structural consistency across species makes them core structures within survival neural loops and we will therefore present these structures in more detail in the next section.

Developing models of bio-inspired neuronal architectures for survival functions is challenging at multiple levels because the focus is usually set in neuronal models on local perceptual analyses or cognitive functions associated with direct performance measures. It is sometimes difficult to distinguish the general principles that aim at survival of autonomous systems. The same is true for the software and hardware systems used to implement and evaluate these architectures. They are generally oriented towards solving specific complex and sometimes abstract problems, particularly in the context of Artificial Intelligence (Stoeter and Papanikolopoulos, 2005).

In the rest of this paper, we present a bio-inspired neuronal model of loops involving the BG and its implementation and experimentation using a software platform dedicated to video games. The key point, here, is to demonstrate to which extent this arrangement is relevant to account for some fundamental mechanisms of agent behavior related to survival and to observe and manipulate them experimentally. The joint use of these systems showcases several survival scenarios and provides a road-map for future works.

2 A FUNCTIONAL DESCRIPTION OF THE BASAL GANGLIA LOOPS

The basal ganglia (BG) are a set of interconnected sub-cortical nuclei, organized in loops with many other brain structures (Parent and Hazrati, 1995), as elaborated here. These loops are described as parallel and segregated (Alexander et al., 1986) because they correspond to distinct and related territories of the structures involved, and because they are structurally similar (in terms of involved neural populations and connectivity), suggesting that the same kind of processing is applied generically to different information.

The association of (interoceptive or exteroceptive) sensations with (internal or external) responses is sometimes straightforward and a simple sensorimotor structure is enough to trigger the response. But the involvement of BG is essential when the selection of the response (e.g., goal or action) is based on ambivalent or uncertain criteria (Floresco, 2015).

In the experiments reported in this paper, we consider four loops as identified in primates in (Alexander et al., 1986), involving different regions of the Striatum, the largest nucleus of the BG. The first two loops are called limbic and are based on interoceptive information. They are organized around the selection of the goal of the behavior, according to its motivational value, in response to perceived needs or according to its hedonic value. The other two are called sensori-motor and they process exteroceptive information. They are organized around the motor behavior allowing to reach the goal, according to its spatial position (orientation) or according to the physical characteristics involved (handling). We refer to these four loops by the question each loop attempts to answer, detailed as follows.

1. The *Why* loop selects the current motivation (satisfying hunger or thirst in our task) from the interoception of needs and possibly the costs of actions. The motivation is expressed in the anterior cingulate cortex (ACC) and the loop also associates the ventral striatum (the core of the nucleus accumbens), lateral hypothalamus and insula for interoception.
2. The *What* loop selects the goal according to the preferences (e.g., gustative preferences, quantity), innate or acquired and represented in the amygdala. Preferences are expressed in the orbitofrontal cortex (OFC) and the loop also combines the ventral striatum (the shell of the nucleus accumbens), amygdala and insula for gustative interoception. The goal object can be consumed if it is directly available, otherwise it will become the goal for the spatial and temporal organization of the behavior.
3. The *Where* loop considers the spatial location of the goal and selects the orientation behavior relevant to face it, which can concern eye movement as well as body orientation, as also observed in the superior colliculus. The orientation strategy is expressed in the Frontal Eye Field (FEF) in the frontal cortex and the loop also combines the dorsolateral striatum, the parietal cortex and the superior colliculus.
4. The *How* loop supports the latest postural adjustments when the goal is attainable, by simply reducing the distance or possibly manipulating the object before consuming it. This concerns the motor areas, the parietal cortex and the dorsolateral striatum.

This functional description highlights that the generic processing of response selection by the BG is ascribed in a generic loop, also associating the frontal

cortex, sub-cortical and cortical sensory structures as shown in figure 1.

We now describe the implementation of the loops in a bio-inspired neuronal model and use them to emulate a survival behavior in a simulation platform. From a technical point of view, the description of the tasks accomplished by each loop reveals certain characteristics that are not conventionally considered in neural models (e.g. notions of motivation or goal). Furthermore, implementation in a video game simulator requires dedicated specifications (e.g. bodily characteristics, bodily needs and biological constraints). We address these requirements in the design and implementation of the system.

In addition, we have so far discussed each behavioral loop individually, whereas describing the possible association and interactions of these loops is a major topic in neuroscience (Haber et al., 2000). In the case of survival tasks like the one we demonstrate, we can particularly wonder how to model the functional interaction between these loops and if different forms of survival strategies (e. g. goal-driven or stimulus-driven) can be performed on this basis. The latter question forms one of the open problems in computational neuroscience (Daw et al., 2005), thus motivating our digital experiments.

3 IMPLEMENTING A MODEL OF BG LOOPS

As highlighted in the previous section, although each loop addresses a different issue, the principle behind the operations of these loops appears quite generic in terms of their physical connectivity and computational dynamics. In this perspective, several generic computational neuronal models of BG loops have been proposed (Gurney et al., 2001; Guthrie et al., 2013; Hazy et al., 2006). These neuronal models exploit several pathways observed between the nuclei of the BG to implement the decisional process, with a globally excitatory (direct) pathway for selecting the best response and other inhibitory pathways (hyper-direct or indirect) that will penalize inappropriate responses. The BG implementation in our model description is directly inspired from the 'Go-No Go' process implemented in (Hazy et al., 2006) for the decision process. Above mentioned models also agree on the critical role of neuromodulators, in particular dopamine, in updating contextual associations according to the prediction errors. We do not consider this aspect in the current implementation and will address it in further work.

Based on this computational formalism, we implemented the four loops discussed in section above, each in the form of a generic loop. A loop is formed between 3 components (see Figure 1), with the input information coming to the sensory module of the loop (blue component). This sensory information will elicit possible actions in the frontal module (green component). These actions will compete until one is selected and triggered. Basically, this selection is made by BG (red component), also informed by the sensory context. We will see below that the selection process might also involve the influence of other loops. When an action is triggered, it remains active with a sustained activity until some sensory information is received, informing about the end of the action.

From a more practical point of view, in each loop, the processing happens in the following stages, in a given small time interval - *information acquisition*, *action evaluation and selection* and *sustained activation by feedback control*. For each loop *(i)* acquire sensory information through exteroception and interoception, *(ii)* evaluate alternative responses, select the most appropriate one and set the corresponding goal, and finally *(iii)* sustain the activation of the response by a constant feedback until the goal has been achieved.

This generic mechanism of response maintenance and goal monitoring is an important aspect of our computational model, implemented in each loop. Selecting a response to be executed means defining a sensory state that must be achieved (the goal). As a consequence, the rule of response execution is implemented as a sustained activation of the response which terminates, thanks to a feedback mechanism, when the goal is met. As it is elaborated in the section *Scenarios*, the goal is not always reached simply by activating the response, but sometimes requires other responses and secondary goals to be defined, still within the same generic mechanism. To implement this, we define a *desired* state of activation for goals that asks for additional responses until it becomes *actual*.

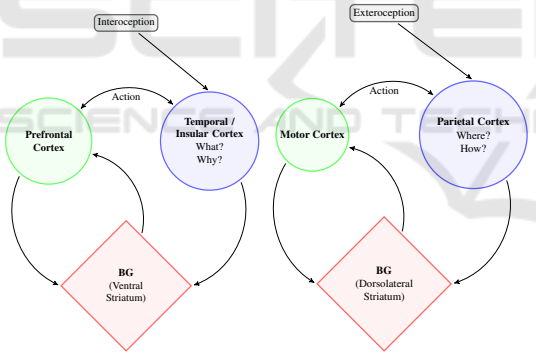
To give a more concrete understanding of the information processed in each loop, we stick to the very simple example that will be developed below, describing the connection of the model to an experimentation platform and the unfolding of survival scenarios. In this example, the agent is given two needs (representing thirst and hunger), each as a variable to maintain between bounds to survive. It can also detect and possibly reach objects (representing food and drinks).

In that context, the two limbic loops can be defined as follows. The *Why* loop is responsible for the selection of the need. It receives sensory information about the levels of need through interoception

and about the kinds of objects perceived by exteroception. Responses it can trigger correspond to the decision to go for food or drink, until the need is satisfied (by consuming upon reaching). The *What* loop is responsible for the selection of an object. It receives sensory information about the levels of preference through interoception and about the identity of the objects perceived by exteroception. Responses it can trigger correspond to the decision to select one object until the object is reached.

Similarly, the two sensori-motor loops can be defined, still using the same framework, as follows. The *Where* loop is responsible for the orientation of the agent in space. It receives the azimuth of each object perceived by exteroception and when one is selected, triggers a movement of orientation which stops when the agent is facing the object. The *How* loop is responsible for the reaching of an object. It receives the distance to each object it is facing by exteroception and when one is selected, it moves forward until the object is reached.

Before describing how these loops can be associated in organized behavior, we present the Malmo platform used for the experimentations and describe the connection of the platform to the cerebral model of loops.



(a) limbic (b) sensori-motor

Figure 1: Implementation of two classes of generic loops in the model. In both the classes, a blue component represents a region from Sensory Cortex, green represents a region from Frontal cortex and the red BG represents the corresponding sub-cortical BG nuclei involved in the loop. A black component, Information (through *Interoception* or *Exteroception*) feeds the blue component of the sensory cortex which propagates it to both the Frontal Cortical regions and the BG. (a) generic limbic loop - based on which the *What?* and *Why?* loops are implemented. (b) generic sensori-motor loop - based on which, the *Where?* and *How?* loops are implemented.

4 THE MALMO PLATFORM

Minecraft is a well-known video game, with a block based 3D world, allowing virtual exploration, resource gathering, and including survival task scenarios, this for a single or multiple players. It has been adapted for a systemic neuroscience simulation platform called 'Virtual Enaction'. Later, an experimentation platform called Malmo was built on the top of Minecraft (Johnson et al., 2016), and is dedicated to support research in various AI related areas. Malmo allows to incorporate various models of reinforcement learning, planning and related problems into the Minecraft game environment, ranging from a basic Q-learning algorithms on a single agent to more collaborative and competitive strategies among multiple agents.

The advancements in the research on intelligent systems and their behavior requires to be able to test, study and visualize the models in a more elaborative setting, as opposed to the traditional symbolic representations and numerical experimentations. Given the complexity of the survival task that we target to demonstrate, it is considerably difficult to choose the right kinds and number of attributes to be encoded in the model. Malmo, exploiting the power of Minecraft environment, precisely provides such great convenience to study our model of generic loops. Here, since our goal is to explain the dynamics of emergence of the behavior using and organizing the loops defined in the model presented above, we use only a specific set of Malmo features that are adapted according to the task and the model. We have designed on the top of Malmo a minimal software layer to accommodate our adaptations concerning the interoceptive and exteroceptive attributes relevant to the survival task.

Particularly, Malmo provides a set of attributes representing the vital characteristics of the agent. We have built on them more precise variables relevant to the task together with their functional dynamics, as a part of the software layer. Similarly, the adaptations related to the agent's vision and the sensori-motor responses can be conveniently implemented using the related features of Malmo. These attributes in conjunction with our adaptations concisely explain the generic dynamics in the loops as a result of which several behaviors emerge. Furthermore, it is interesting to observe that Malmo invariantly supports demonstrating these different behaviors with no or minimal changes to itself but only from the changes in the state of the agent or that of the environment.

In the rest of this section, we explain the attributes of Malmo that we have used in the context of the

survival task. In the subsequent section, we describe the additional software layer with the adaptations that also demonstrates the embodiment of the agent.

World. This is the environment in which an agent is free to move around and explore, besides other objects (*items*) present in it. It is a simplified environment of the Minecraft world, designed to have a complete control on the external objects, and simple enough to understand the causal relationships with respect to the simulated agent behavior. It is 3 dimensional, allowing the items to be at a height above the ground and allowing the agent(s) to jump if necessary. The ground (*floor*) is defined in terms of blocks which have properties like texture, type and color. Such block properties like the color play the role of the environment *context*. In behavioral scenarios like fear learning or fear extinction, the *context* is a useful attribute because it adds an extra dimension to processing the stimulus information and attributes a preferential relevance to it (either from previous learning or from memory).

Agent. Malmo allows multiple agents to interact simultaneously in a given environment. An agent, at any given point of time, has access to its vital variables like *life*, its current position and its current orientation with respect to the *World*. Like in the case of an animal, the variables are affected by the external world - for e.g, components like *fire* or an *attack* could reduce *life*. In the context of our task, we consider only a single agent.

Items Malmo provides a list of *items* that can be procedurally placed in the environment. When the *items* are in the configured vicinity of the agent, the positions and the orientations of the *items* are available for the agent. Each item can be configured with a certain *reward* value at the beginning of the task. As a part of a task, the *reward* can be awarded to the agent, either for collecting the *item* or *discarding* it. There are several such items, from which we use *apple*, *cake*, *water bucket* and *stew*. The distance within which the agent can *collect* the item can also be configured.

Actions. Suitable to the 3D world, the agent is capable of doing actions like moving, turning and jumping. In order for the agent to reach an object or a position, it uses, from the in-built set of actions, predominantly the *turn* action (to orient towards a stimulus) and *move* (to approach a stimulus or keep exploring). Any of the actions can be stopped when required.

State. The state of the world, at any given instant, is constituted by the attributes of both the agent and the objects present in the *vicinity* of the agent. At any instant, the agent has information about its current levels of vital variables and how far they are from critical or fatal limits. It also has information about its own position and orientation with respect to the environment. Information about the item like its name, position and the reward it carries is also accessible the agent. As explained earlier, *context* also is a part of the state, describing the type of the *floor* for a requested subset of blocks.

5 EMBODIMENT OF THE AGENT

We describe here, the adaptations that we made to Malmo, in order to connect our model of cerebral loops to the world simulated in Malmo, including the characteristics of the environment and the agent. In addition to the technical considerations, these adaptations allow us to distinguish several actors in our tasks, namely the brain, the body and the environment, which have been often reported to constitute embodied cognition (Varela et al., 1992). Hence, attributing bodily features to the agent and associating them to its motivational and emotional characteristics form a key aspect of our model. These characteristics have been respectively implemented in terms of needs and preferences. Also, from a functional point of view, we adapted few aspects like *visibility* of the agent and the information about the *positions* (of items as well as the agent itself). These adaptations were important to add certain biologically plausible restrictions to the task.

Needs and Preferences. The agent has two vital variables - *hunger* and *thirst* - which increase with time as well as with its efforts (meaning a *move* or *turn* action). Instead of a one dimensional *reward*, each *item* carries a value that is relevant to the *hunger* or the *thirst* level it would satisfy, and a value indicating the level of *preference* of the agent for this item.

Visibility. Malmo provides information about the items all around the agent's vicinity of chosen range. However, we restrict its 'Field Of Vision' to a biologically plausible value (in this case, 120°), which is further divided into 3 different zones viz., *Appear*, *See* and *Reach*, depending on the distance from the agent. When the agent is moving and some *items* are present in the *Appear* zone, the agent has no precise information about the stimuli (the *items* that are perceived by the agent) such as the precise location of each, or their preference appetitive values. Rather, the agent has minimal information about the presence or absence of some *items* in some direction. When the stimuli are within the *See* zone, all the information about the stimuli is provided as inputs to the model. In the *Reach* zone, an additional information is provided, that the stimuli are accessible for the agent to

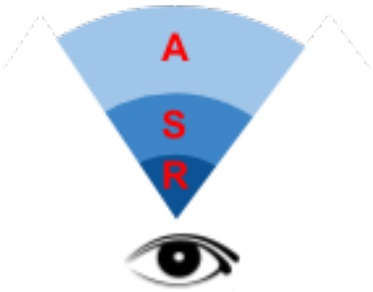


Figure 2: Zones of visibility in the field of agent's vision. Zone marked '**R**' is *Reach*, '**S**' is *See* and '**A**' is *Appear*.

consume.

Positions In regard to the positions of the agent and the *items* in the environment, Malmo provides their exact coordinates, the absolute yaw details with respect to the environment. But to demonstrate a very important feature within each loop of the model, we avoid using these exact position details. Instead, we take the agent as the origin, convert the relative distance and orientations of the items into signals that regulate the activity within the loops of the model. It is usually these feedback signals relative to the desired state and the current state of the agent that sustain the execution of a selected goal.

6 SCENARIOS

We intentionally define a world with simple and few characteristics, in order to study precisely, not only the functioning of each loop, but also the way they interact with one another to emulate specific kinds of behavior. We present here for illustration, some behaviors that could be demonstrated with the model to emulate and others that are part of our ongoing work.

6.1 Exploration Behavior

Actions for spatial exploration can be triggered for several reasons. They can be triggered if, as in goal-directed behaviors described below, the agent must explore the environment to find desired stimuli. They can be also triggered, as in the stimulus-driven behavior described below, if the agent has no current need, to give it the opportunity to discover new options. For any kind of behavior, the agent can also apply an exploration/exploitation strategy (Humphries et al., 2012) and at any moment interrupt the current behavior to explore. The exploration behavior is also particularly important at the beginning of the task, when the agent rotates until it can perceives some stimuli. If nothing is perceived, the agent selects a random direction, moves by a random distance and rotates again. When some stimuli are perceived, if they are in the Appear zone, the agent moves in that direction until they are in the See zone and can be discriminated. Then depending on the current behavior, several actions can be triggered as described below.

This basic behavior also forms an interesting basis to learn or update the contingencies in the environment or between characteristics of the environment and those of the agent. Particularly, in the limbic loops, this can contribute to set the values of the preferences and help connect some items to the needs they can satisfy. In the sensori-motor loops, this can help calibrate the movements of the agent and learn the consequences (in terms of modification in the perception) of their activation. However, since the work we report here concerns the dynamics of the loops, we provide the agent with this initial learning of fundamental contingencies as a pre-existing set of values, thereby enabling the agent to exploit them in its behavior.

6.2 Goal Directed Behavior (GD)

The system has been initially designed for a simple survival task, corresponding to activate the loops in a hierarchical way. First the *Why* limbic loop monitors the levels of the needs and when one of them passes above a critical threshold, satisfying this need becomes the primary goal of the agent. In the What limbic loop, the objects associated to the satisfaction of this need, are set as potential secondary goals and are activated as desired. If none of them is presently perceived, an exploration behavior is triggered in the *Where* sensori-motor loop, making the agent rotate until it perceives some stimuli. If they are too far (in the Appear zone), the agent approaches for them to be in the See zone and gathers their characteristics including the preferences. If several stimuli are perceived, the one with the highest preference is selected as the secondary goal of the behavior. The agent approaches the stimulus until it reaches and consumes it, thus satisfying the goals of the behavior. An implementation of this behavior is described in the section *Illustrations*.

6.3 Stimulus Driven Behavior (SD)

Without a specific goal or motivation, the agent can wander in the environment and discover by chance one or several items. In this case, the *What* limbic loop (estimating agent's *preference* from external information) is triggered and can select the most preferred item. Although the Why limbic loop (defin-

ing the levels of *need*) has not triggered the decision making process beforehand, it can be activated by this *preference* and, depending on the corresponding level of *need*, it can decide to activate the sensori-motor loops to execute an action in order to reach and consume the selected item.

6.4 Opportunistic Behavior

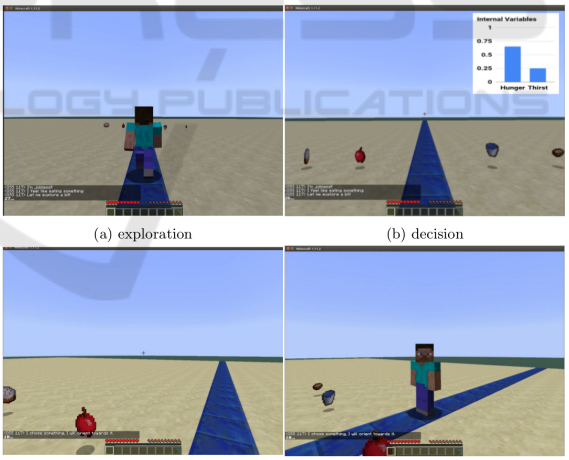
As a part of our ongoing work, we would like to be able to interrupt a behavior in an opportunistic way. This is specifically the case when the agent is engaged in a goal-driven behavior and suddenly perceives a stimulus corresponding to the non-selected need but with a strong preference. In this case, it is conceivable to reason that, in some condition, it is preferable to choose a stimulus with a strong impact on a minor need as compared to another stimulus with a minor impact on the current need. This could be particularly the case if the stimulus with the strong preference is rare or if stimuli detected to satisfy the major need have very low levels of preference (or both). Better understanding the mutual influences between the limbic loops (Haber et al., 2000) is one of the major challenges in our ongoing work.

7 ILLUSTRATIONS

Figure 3 shows a sequence of snapshots from a goal-directed behavior, as implemented in Malmo. In this basic scenario, the agent has already selected its most urgent need, (*hunger* in figure 3(b) inset). The satisfaction of this need now becomes a *desired* state as a primary goal, which remains active in the *Why* loop until the need is satisfied. From previous experience to address the current need, the *Why* loop triggers the *desired* state of the stimuli known to satisfy the need, in the *What* loop. Figure 3(a) illustrates this exploration behavior, where the agent starts to move with a *desired* activation for some items (*apple* and *cake* in this case). If, as it is the case here, perceived stimuli are too far (in the Appear zone), the agent will have to move until they are in the See zone and can be discriminated.. When the detected stimuli are in the See zone, the simplest of the cases is when only one of the *items* is desired and can be directly selected. However, when encountered with multiple desired *items*, the agent has to *decide* the suitable choice among the *items*, suitable meaning the one with the highest preference (as illustrated in figure 3(b)).

Once the decision has been made, as a secondary goal, the selected stimulus becomes *desired* in the *What* loop and remains active until it is reached. The execution of the behavior involves two steps. *(i)* To evoke the necessary sequence of actions to reach the goal and *(ii)* to sustain the selected goal until the agent actually reaches the selected stimulus. In the case considered here, once an *item* is chosen, the other secondary goals are to orient towards the selected item and the reach it. The agent starts *turning* towards the chosen *item*. Here *turning* doesn't stop by using the target yaw provided by Malmo. Instead, we derive a feedback signal to the *Where* loop to sustain the act of *turning* until the agent is oriented towards the *item*, as illustrated in figure 3(c)&(d).

Then, the agent can move towards the target to reach it. With exactly the same mechanisms as described above, but here applied in the *How* loop, the goal of reaching is maintained until the item is reached. And once reached, this goal is considered achieved. In our current implementation, the internal action of consumption is automatically triggered when an item is reached. In this case, the goal in the *What* loop (sustained from the initial selection of the goal) is considered achieved. This consumption will also modify the level of need and similarly, the goal in the *Why* loop is also considered satisfied. This terminates the behavior as this was the primary goal of the scenario illustrated.



(a) exploration (b) decision

(c) orienting (d) (before) approaching

Figure 3: Snapshots at different stages in the task. The figure shows several steps involved in a goal-directed behavior of the agent. (a) *exploration* until the agent finds some stimuli in the *Appear* zone. (b) *decision* among the stimuli in the *See* zone, corresponding to the current need (inset:hunger). (c) *orienting* towards the selected stimulus until it is in the line of sight. (d) ready to *approach* towards the oriented stimulus.

8 DISCUSSION

The work that we have presented here can be considered under three points of view.

Firstly, this work is original by its deep anchoring in biological inspiration. The behavior of our autonomous agent is elaborated thanks to a model built on four loops described as playing an important role in the brain of most animals (Alexander et al., 1986). This inspiration is anatomical, considering the nature of information flows brought by several sensory and motor regions. This inspiration is also functional, particularly considering mechanisms to select actions and to sustain goals until they are achieved. A major characteristic of this work is to consider similar architectural and functional properties to build four loops and to build all the considered behaviors only by emergence, on the basis of the loops and their interactions. This biological inspiration is also very precious because most of the questions and orientations for future works we have evoked in the paper will be addressed by going deeper into biological details.

Secondly, we have argued that, even if these loops are mostly studied for the understanding of higher cognitive functions like reward-based decision making, considering them to implement survival scenarios is very important to design autonomous systems. Particularly, considering such basic scenarios is very convenient to study all the loops together, which is hardly addressed in the modeling literature. It is also interesting to understand how the two basic processes of signaling and regulation have evolved to allow for more abstract behaviors, which can still be described as interactions between limbic and sensorimotor loops, originally built for survival.

Thirdly, another major innovation of this paper is to propose that Malmo, a platform originally designed for experimentation in Artificial Intelligence, is a very powerful tool to build basic autonomous systems performing survival tasks. Not only it offers many interesting characteristics for the simulation and the visualization of a survival task, but it also eases the design of the most critical part, corresponding to the interface between the computational parts of the model and the internal and bodily aspects of the agent. In addition, this platform has another usefulness. To address the critical questions we have evoked in this work, we are currently seeking insights from biology, to improve and augment our model. Some of these questions are clearly unanswered by the current state of the art and must be investigated jointly with neuroscientists. In this perspective, Malmo offers a striking advantage to describe our model and its behavior to them, who are more accustomed to biological observations than algorithms and equations.

REFERENCES

Alexander, G., DeLong, M., and Strick, P. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Ann. Rev. Neurosci.*, 9:357–381.

Craig, A. D. (2003). Interoception: the sense of the physiological condition of the body. *Current Opinion in Neurobiology*, 13(4):500–505.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, 8(12):1704.

Floresco, S. B. (2015). The nucleus accumbens: an interface between cognition, emotion, and action. *Annual review of psychology*, 66:25–52.

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

Guthrie, M., Leblois, A., Garenne, A., and Boraud, T. (2013). Interaction between cognitive and motor cortico-basal ganglia loops during decision making: A computational study. *Journal of Neurophysiology*.

Haber, S., Fudge, J., and McFarland, N. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J Neurosci*, 20(6):2369–2382.

Hazy, T. E., Frank, M. J., and O'Reilly, R. C. (2006). Banishing the homunculus: making working memory work. *Neuroscience*, 139(1):105–118.

Humphries, M., Khamassi, M., and Gurney, K. (2012). Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. *Frontiers in Neuroscience*, 6:9.

Johnson, M., Hofmann, K., Hutton, T., and Bignell, D. (2016). The malmo platform for artificial intelligence experimentation. In *IJCAI*, pages 4246–4247.

Maturana, H. R. and Varela, F. J. (1991). *Autopoiesis and Cognition: The Realization of the Living (Boston Studies in the Philosophy of Science, Vol. 42)*. D. Reidel Publishing Company, 1st edition.

Parent, A. and Hazrati, L. N. (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Res Brain Res Rev*, 20(1):91–127.

Redgrave, P., Prescott, T. J., and Gurney, K. (1999). The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience*, 89(4):1009–1023.

Stoeter, S. A. and Papanikolopoulos, N. (2005). Autonomous stair-climbing with miniature jumping robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 35(2):313–325.

Strausfeld, N. J. and Hirth, F. (2013). Deep Homology of Arthropod Central Complex and Vertebrate Basal Ganglia. *Science*, 340(6129):157–161.

Varela, F. J., Thompson, E. T., and Rosch, E. (1992). *The Embodied Mind: Cognitive Science and Human Experience*. The MIT Press.