A survival task for the design and the assessment of an autonomous agent

Bhargav Teja Nallapu, Frédéric Alexandre

▶ **To cite this version:**

HAL Id: hal-01931573
https://inria.hal.science/hal-01931573

Submitted on 22 Nov 2018

A survival task for the design and the assessment of an autonomous agent

Bhargav Teja Nallapu[1] and Frédéric Alexandre[1]

INRIA Bordeaux Sud-Ouest, 200 Avenue de la Vieille Tour, 33405 Talence, France
LaBRI, Université de Bordeaux, Bordeaux INP, CNRS, UMR 5800, Talence, France
IMN, Université de Bordeaux, CNRS, UMR 5293, Bordeaux, France
`bhargav.teja-nallapu@inria.fr`

Abstract. Learning to survive in a complex environment is a more relevant task for the design of intelligent autonomous machines than complex problem solving. To establish this statement, we emphasize that autonomy requires emotional- and motivational-like characteristics that are much more straightforward to define in the context of survival tasks than in problem solving which is classically considered with intelligent agents. We also propose that using a simulation platform is a good preliminary step before the design of real machines because it allows to consider another fundamental challenge, related to the association of these emotional and motivational characteristics to higher cognitive functions. These considerations are illustrated by current simulations that we are carrying out with a bio-inspired neuronal model of the cerebral architecture of primates.

Keywords: computational neuroscience, autonomy, simulation platform

1 Introduction

This paper presents a neuromimetic approach to emulate intelligent behavior in autonomous machines. Lately, striking progresses have been reported in the field of Artificial Intelligence, on hierarchical neuronal models such as Deep Networks applied to specific tasks like playing Go or answering general knowledge questions, setting the focus on dedicated complex problem solving tasks. However, in reference to some characteristics of intelligence in animals, autonomy raises other challenges at least as important as abstract intelligence does. Firstly, in reference to the emotional domain, an autonomous agent should be able to set its own goals by anticipating outcomes and to acquire its preferences by learning. Similarly, motivation should also be considered, to energize potential behaviors according to a level of need felt or to the anticipated cost of an action [Dickinson and Balleine, 2002]. As these characteristics have been often associated to bodily aspects [Cardinal et al., 2002], these challenges will require further studies about the body and its physiology. Secondly, it has been shown that these embodied aspects play an important role in the most complex aspects of cognition, including decision-making, planning or social skills [Damasio et al., 1996], but the underlying neuronal mechanisms are to be understood in more details.

The above mentioned challenges correspond to two very ambitious research programs of different nature, that are difficult to carry out in tandem, although based on common ground. One issue is to modify the material characteristics of machines to allow them to reflect certain sensory and motor characteristics associated to emotions and motivations, as we observe in the case of animals. The other issue is to understand how these characteristics influence the most elaborate forms of reasoning, particularly allowing to associate subjective and incomparable quantities such as the level of need for a resource and its cost or its social impact.

We present here some elements of a modeling work that bridges the challenges between the above mentioned directions. This work considers the possibility to define an autonomous agent by adapting a well established neuronal model proposed in [Hazy et al., 2006], mimicking the function of major loops in the brain of primates and primarily involving two cerebral structures, the cortex and the basal ganglia. The so-called cortico-basal loops have been associated to the generic function of decision making and action selection and several such loops have been experimentally observed [Alexander et al., 1986]. Based on these observations, we have implemented four loops. Two loops are called sensorimotor and are fed with external spatial and visual information and select actions of orientation and object reaching. The other two loops are called limbic and are fed with internal information carrying emotional and motivational characteristics and make decisions about the need to satisfy and the object to select.

In addition to implementing each of the four loops and their interactions, a major question in our work was to assess the ability of the whole model to emulate an autonomous behavior, by defining what could stand for body and an environment to interact with this model of cerebral architecture and what could be the task to address, to quantify the behavior of the brain+body system, possibly under different dimensions.

In the next section, we propose that learning to survive in a complex environment is a more relevant task than complex problem solving. We define a survival task where the autonomous agent must monitor basic needs such as hunger or thirst to select the resource to be reached in its environment. In the subsequent section, we introduce a simulation platform that simulates the agent, the body as well as its environment, and allows us to control the complexity level of the environment and modify its properties. This also leads to the definition of the emotional and motivational characteristics of the agent and their encoding in the model. We propose that, even if this software platform brings a radical simplification compared to a hardware machine, it is an excellent first step towards building autonomous machines and addresses fundamental questions about embodiment. We then report for illustration the results of the first experiments that we have carried out and finally mention further work that could be considered.

2   A survival task

Let us first state that, if there is a precise task that characterizes an autonomous agent, it is a task of survival. An autonomous agent has to be able to survive in an unknown environment, being aware of its vital variables and being able to keep those variables in acceptable bounds at any given moment and for the forthcoming moments. Consequently, it has to minimize uncertainty and avoid *surprise* to ensure that it accomplishes the goal of keeping variables in bound. This has been formalized in, e.g., [Friston et al., 2016] and is revisited differently here. A step further, biological systems often fail, e.g. have limited generalization capability or learning performances, and the utopian vision that they are "very general autonomous systems" is questionable: They may more "adapt" than "learn", i.e. change some parameters of some versatile behaviors, instead of "inventing" new behaviors.

To demonstrate this principle, we consider an agent in an environment that has 'objects' that are primarily given an incentive value - their relevance to the amount of hunger they satisfy or the amount of thirst they quench if these needs are considered. The objects also carry a *preference* value with respect to the agent, which could be retrieved from the agent's pre-existing knowledge (or being learned, this very important aspect will not be considered in the present study but see [Carrere and Alexandre, 2015] for previous work we made on pavlovian learning, known to contribute to preference learning). The two limbic loops mentioned above that we have implemented are respectively in charge of selecting the need to be considered and the preferred object. We will mention below how they are associated, through the simulation platform, to the internal perception of needs (for motivational aspects) and to the representation of preference as an anticipation of a certain level of pleasure (for emotional aspects).

At any given instant, it is possible that there are no objects perceived by the agent in the vicinity. In this case, the other two loops, the sensorimotor loops, are responsible for the orientation of the agent in space until objects are perceived. If any reveal pertinent to the task, the loops are also responsible for navigation, reaching and consumption of this object. We mention below how the spatial and visual aspects to be processed in the sensorimotor loops are extracted from the simulation platform.

For the moment in our description, the loops are presented independent to each other and each of them respectively carries out a specific function. It is of course necessary to consider some coordination between the loops to emulate a really intelligent and autonomous behavior and we will evoke below several strategies of coordination and how they can be implemented, particularly to give rise to a possibility of different starting points of the survival task. Before all these developments, we describe, in the following section, the generic algorithm implemented for each loop in the model and the Malmo platform we used to create an environment for the task.

3   A model of cerebral loops

In the brief description above, each loop has been assigned a very different function, but it is also noticeable to mention that each loop is dynamically computed with the same algorithm, since they are all part of the same circuitry linking the basal ganglia and the sensory and frontal cortical areas.

Firstly, sensory cues, corresponding to actual or desired sensations activate candidate actions in the frontal area. A primitive strategy is to trigger the action most often associated to these sensations. This corresponds to *habits*. Else, a selection process takes place to make a decision based on a deeper contextual analysis. This is attributed as one of the major roles of the basal ganglia.

The basal ganglia is often presented as a competitive system (Go-No Go in [Hazy et al., 2006]), based on several excitatory and inhibitory pathways respectively adding more weight to the selection and the inhibition of each candidate action until one reaches a threshold and is triggered. This dual evaluation is based on the estimated values of the expected sensory consequences of the action and other modulatory contextual information.

There have been computational accounts of such action selection in the basal ganglia using a two-arm bandit task [Pasquereau et al., 2007], that go on to demonstrate how the outcomes of such actions are learned, thereby being able to choose the best rewarding one later [Redgrave et al., 1999,Topalidou and Rougier, 2015]. The conflict between two almost equivalent options is resolved by an external noise driving one of them towards selection. Even without a complicated task in context, the same selection between two equivalent stimuli can be demonstrated in a single channel of neural populations. This kind of selection is achieved by implementing the above mentioned complementary pathways in the basal ganglia as neural ensembles with excitatory and inhibitory connections [Leblois et al., 2006].

Each loop described in our model has a respective substrate of the basal ganglia involved for the local selection. Given the space constraints in the paper and the main emphasis being on the high level behaviour, we restrict the description of the dynamics within the basal ganglia in the equation 1 as in [Leblois et al., 2006]. Assuming each unit represents an ensemble tuned towards a particular stimulus : I_{ext} is the external input representing the salience of the stimulus, I_s is the input to the unit from its connections (synaptic input) and τ is the decay time constant of the synaptic input and m is the output of the unit.

$$\tau \frac{dm}{dt} = \text{-}m + I_s + I_{ext} \tag{1}$$

The synaptic input to a unit j, I_s^j, which is the input as a result of the connections from units of other structures (say i), depends on the connections weights (w_{ij}) between units i and j , as shown in the equation 2. We would like to emphasize that the dynamics given by the equations 1 and 2 are phenomenological and are not constrained to a specific neuronal architecture.

$$I_s^j = \Sigma_i w_{ij} * m_i \tag{2}$$

When an action is triggered, its expected sensory consequences are also activated to a specific desired level, representing the goal of the action. The action will be maintained until the expected sensory consequences (or other conditions for interruption, not developed here) are met.

In some cases, triggering the action is not sufficient to reach the goal (e.g. deciding to eat is not sufficient to get some food) and the desired activity can itself trigger new actions in other loops (e.g. finding some food). This process can recursively trigger other secondary goals, until some goal is immediately achieved, stopping the corresponding action.

This generic algorithm is applied in parallel in the four loops. Among the limbic loops, one is responsible for selecting the need, based on the interoception of internal levels. We refer to this loop as *Why* loop as it corresponds to the motivation behind selecting the need. The other is responsible for selecting perceived items known to satisfy some needs, based on their levels of preference. This loop is referred to as *What* loop as it signifies the qualitative nature of each stimulus owing to its preference level. Among the sensorimotor loops, one is responsible for orienting the agent towards the selected target, which we call *Where* loop. The other, referred to as *How* loop, is responsible for moving to reach the target. The *Why* loop also receives an input from the sensori-motor loops, representing the cost of action involved in reaching a certain stimulus, to be able to compare with the appetitive gain the stimulus might offer. This can be simply implemented using negative connection weights w_{ij} in equation 2.

In addition, this selection made locally in a specific loop is also modulated by the selection made in a different loop. For example, the activity strength of selection in *What* loop modulates the activities competing for selection in the *Why* loop. Similarly, the *Why* loop modulates the *Where* loop which in turn modulates the *How* loop. To keep the modulation simple and tractable, we implement a simple biasing factor b_{ij} from unit i to j, as a function of the source activity a_i as shown in equation 3.

$$b_{ij} = a_i^{\tan \sigma_{IJ} \frac{\pi}{4}} \tag{3}$$

In equation 3, σ_{IJ} is a bias strength parameter that is specific between two populations I and J. For instance, the bias that the preference-based choice might have on the need-based choice could be less than the bias that the overall limbic choice has on that of any of the sensorimotor loops. This kind of interactions between different cortico-basal ganglial loops in animals (including primates) is a question of wide interest in the field of neuroscience. Here in this work, we stick to a rather simple implementation for the ease of the demonstration of the principle. Therefore, it has to be noted that all the equations given above represent a generalized principle (lateral interactions, modulation etc.) between the neuronal structures and their intra-loop and inter loop dynamics.

4 The Malmo platform

Minecraft is a sandbox video game with a virtual 3D world that can be generated procedurally. With single or multiple agents in it, activities like exploration and resource gathering can be simulated. Malmo is an Artificial Intelligence experimentation platform on the top of Minecraft [Johnson et al., 2016]. Malmo allows to incorporate various models of reinforcement learning, planning and collaborative and competitive strategies into the Minecraft game environment among the agents. Malmo has been used for various kinds of specific experimentations like learning to navigate in the Minecraft world [Matiisen et al., 2017] and computational models of animals living in block world [Strannegård et al., 2017a]. In most of such cases, very specific feature of Malmo is used (either the 3D space in the environment or the block nature of the world or the agent to perform a task. We exploit, simultaneously, various features of Malmo like the agent's internal body attributes, external constraints like vision, and uncertainty that can be induced into the scenario owing to the vast environment and its contents. Furthermore, Malmo provides great convenience to connect our model of generic loops to the Minecraft world, since the model inevitably requires an elaborative environment to study and visualize the behavior in a survival task.

Malmo provides an agent, on which we have a full control on its actions, either through the tool or through our model. The agent by default has an attribute *life* that is affected by the external world (e.g, when in contact with fire or attacked by other agents). Malmo allows to construct a 3D world of the size of choice, in which the agent is present. We can procedurally place certain *items* in the world, in the vicinity of the agent or elsewhere. *Items* can be attributed respective reward values that the agent is able to gain when it collects them. The agent is allowed to execute finely controlled actions like *turn*, *move*, *jump*, etc. At any given time step, the framework provides the *state* of both the agent and the environment with respect to the agent. The *state* includes attributes like agent's life, its position and orientation in absolute coordinates with respect to the world, the items that are present in the chosen accessibility range around the agent and their positions and attributes.

For the survival task we have chosen, attributing bodily and other emotional and motivational characteristics to the agent forms a key aspect of our model. Besides the relevant features of Malmo we use, we augment these characteristics to the agent as well as to the *items* in the task. Agent has two vital variables - *hunger* and *thirst* - which increase with time as well as with the level of efforts (meaning an action *move* or *turn*). Instead of a one dimensional reward, each item carries a motivational index that is relevant to the *hunger* or the *thirst* level it would satisfy, and a level of *preference* with respect to the agent, expressing its emotional value. 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). For instance, as the model is desired to be bio-inspired, we consider few bodily, biological constraints to the agent and restrict the agent's field of vision to a plausible value (in our case, $120°$). The information about the objects in the environment is also restricted depending on the distance from

the agent within the field of vision. These adaptations were important to add certain biologically plausible restrictions to the task given that we would not model them in detail as they don't fall in the scope of our goals.

5 Illustrations

Figure 1 shows several moments from an episode in the task implemented in Malmo, primarily concerned with different questions each loop in the model addresses. In a basic scenario, the agent starts exploring the environment (figure 1.a) with a *desired* activation for a particular item that (known from previous experience) would satisfy the current major *need* (figure 1.a, inset). This is a result of the internal state processing in the *Why* loop. When the agent perceives multiple stimuli (figure 1.b), along with the appetitive relevance of each of the stimuli, the action costs, depending on their positions, are also provided (implemented as a negative signal from the sensory-motor loops). Furthermore, the choice on the pre-existing preferences towards the stimuli corresponding to the selected *need* is made in the *What* loop and it modulates the selection in the *Why* loop.

Once the decision has been made in the *Why* loop and the goal has been set, the execution of goal involves two steps in the two sensorimotor loops. Once a stimulus is chosen, the goal is to orient towards it. The *Where* loop is responsible for the agent to start *turning* towards it (see *apple* in figure ??.c) until the selected stimulus is in the sight of the agent. And finally, owing to the processing in the *How* loop, the agent moves to reach the stimulus that it has oriented towards and *consumes* it (an imaginary action which we equate to the agent reaching the *item* and updates the corresponding *hunger* and *thirst* values).

Sustaining the selection of goal until it is achieved is at the core of the processing in each loop. The *Where* loop, after choosing the orientation to turn, sustains the activity until the object is in sight. And the *How* loop sustains its activity from the point of orienting to the point where it has reached closer to the stimulus, to be able to *consume* it. This would now cascade back to the *limbic* loops which have been sustaining their selected responses. The *Why* loop, which has been active since selecting the current need, is sustained until the need levels are modified by the consumption of the stimulus. Similarly, the *What* loop, which has been sustaining activity since selecting a preferred stimulus, continues until verifying the consumed stimulus has the expected value.

6 Discussion

In the wake of the field of Artificial General Intelligence that advocates intelligent agents should be able to solve general problems, it is rather easy to classify *survival* of an autonomous agent as a general problem. We argue that *survival* of an autonomous agent, in fact, addresses a very specific question, that requires the agent to keep its vital variables within acceptable bounds in the current as well as a future, anticipated state. We emphasize the *body*, which completes

(a) Why?　　(b) What?　　(c) Where?　　(d) How?
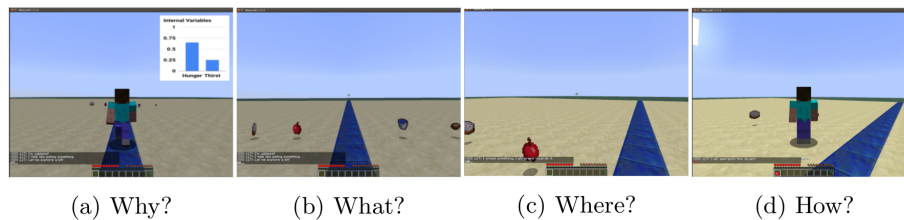
Fig. 1. Snapshots at different stages in the task. (a) Internal needs monitored in the *Why* loop (inset). (b) Processing information about the stimuli in the *What* loop. (c) Orienting towards the selected stimulus using the *Where* loop. (d) Reaching the selected stimulus using the *How* loop, once oriented. *Note:* The different third person and first person views in (a) and (b) are only chosen to show the change in proximity of stimuli to the agent, but these views have no effect on the task. They can be switched while watching the task. Similarly, for (c) and (d)

a closed-loop of interaction for the agent with the environment, as the key to the process of survival. Consequently, we derive our basis from neuroscience of animal cognition and behavior to build a bio-inspired model that helps an artificial agent survive. The components of our model, representing the cortical and basal ganglia structures in the brain, are inspired to be close to the biological plausibility in their working dynamics. As mentioned earlier, understanding the interactions between these cortico basal-ganglial loops is an active ongoing work in the field of neuroscience and we, through the description and further development of our model, would like to be in a position of useful contribution to that work.

In the field of robotics, there have been accounts of motivation driven robotic systems and a few design frameworks [Strannegård et al., 2017b,Konidaris and Barto, 2006]. Most of them address the task of making a choice based on motivation. We derive inspiration from such formalisms and begin to build much more comprehensive scenarios involving the behaviour of agent, in a bio-inspired manner, in which action selection forms a key component. We remain aware that building such scenarios in a bio-inspired manner requires further more understanding of higher level neural mechanisms involved in animal foraging behavior [Kolling et al., 2012,Constantino and Daw, 2015]. However, we would like to emphasize that the fundamental basis of demonstrating comprehensive animal behaviours in artifical agent lies in the description of the underlying generic characteristics of the survival task. Therefore, in this work, we contain ourselves to the description of the basic task and required attributes, keeping the developments and performance studies to a much finer study elsewhere. In fact, using the neural structures already involved in certain loops, we account for context-specific preferences of the agent (in the *What* loop) and the action costs involved in choosing a stimulus (in the *Why* loop). In the context of the task that we target to demonstrate, given its complexity, it is considerably difficult

to choose the right kind and number of attributes to be encoded in the model. Rightly so, we chose an experimentation platform that inspires our concrete choice of emotional and motivational characteristics of the model.

Different starting points characterize certain scenarios in which agent exhibits a distinct behavior, quite often, that belongs to a set of behaviors extensively studied in neuroscience related to primate cognition. The example scenario illustrated in the section *Illustrations* describes a very prominent animal behavior in neuroscience, called goal-directed behavior [Balleine et al., 2009]. As a part of our ongoing work, we plan to study other key animal behaviors like stimulus-driven behavior where, even in the absence of internal need, the external stimuli in the environment drive the motivation of the animal. We also are interested in extending the model to explain much more interesting, opportunistic behaviors where the agent's chosen current behavior can be interrupted to switch to another, anticipating a higher satisfaction for a future need as opposed to a lower satisfaction of a current need. It is the emergence of such versatile behaviors that helps the animal survive in a dynamic external world. The platform we chose, Malmo, invariantly supports demonstrating various behaviors with no or minimal changes to itself but only from the changes in the state of the agent or the objects in the environment. Unlike other works that used Malmo to interface with Minecraft gaming environment, we utilize its features to have a finer control on the environment and to demonstrate the modulation between the cognitive behaviour and action execution in the environment.

Interestingly enough, understanding the dynamics between such different behaviors is still an open problem in computational neuroscience. That encourages us to strengthen the basis of our model more from biology, exploit the convenient visualization platform like Malmo to demonstrate the model to neuroscientists and gain their insights. After all, we, the intelligent primates, are a perfect example of an autonomous agent trying to survive, only in a real world.

X

Acknowledgements

References

[Alexander et al., 1986] 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.

[Balleine et al., 2009] Balleine, B. W., Liljeholm, M., and Ostlund, S. B. (2009). The integrative function of the basal ganglia in instrumental conditioning. *Behav. Brain Res.*, 199(1):43–52.

[Cardinal et al., 2002] Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience & Biobehavioral Reviews*, 26(3):321–352.

[Carrere and Alexandre, 2015] Carrere, M. and Alexandre, F. (2015). A pavlovian model of the amygdala and its influence within the medial temporal lobe. *Frontiers in Systems Neuroscience*, 9(41).

[Constantino and Daw, 2015] Constantino, S. M. and Daw, N. D. (2015). Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective, & Behavioral Neuroscience*, 15(4):837–853.

[Damasio et al., 1996] Damasio, A. R., Everitt, B. J., and Bishop, D. (1996). The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1346):1413–1420.

[Dickinson and Balleine, 2002] Dickinson, A. and Balleine, B. W. (2002). The Role of Learning in the Operation of Motivational Systems. In Pashler, H. and Gallistel, R., editors, *Stevens' Handbook of Experimental Psychology*, volume 3: Learning, Motivation and Emotion, pages 497–533. John Wiley & Sons, New York, 3rd edition.

[Friston et al., 2016] Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., Pezzulo, G., et al. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, 68:862–879.

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

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

[Kolling et al., 2012] Kolling, N., Behrens, T. E., Mars, R. B., and Rushworth, M. F. (2012). Neural mechanisms of foraging. *Science*, 336(6077):95–98.

[Konidaris and Barto, 2006] Konidaris, G. and Barto, A. (2006). An adaptive robot motivational system. In *International Conference on Simulation of Adaptive Behavior*, pages 346–356. Springer.

[Leblois et al., 2006] Leblois, A., Boraud, T., Meissner, W., Bergman, H., and Hansel, D. (2006). Competition between feedback loops underlies normal and pathological dynamics in the basal ganglia. *Journal of Neuroscience*, 26(13):3567–3583.

[Matiisen et al., 2017] Matiisen, T., Oliver, A., Cohen, T., and Schulman, J. (2017). Teacher-student curriculum learning. *arXiv preprint arXiv:1707.00183*.

[Pasquereau et al., 2007] Pasquereau, B., Nadjar, A., Arkadir, D., Bezard, E., Goillandeau, M., Bioulac, B., Gross, C. E., and Boraud, T. (2007). Shaping of motor responses by incentive values through the basal ganglia. *Journal of Neuroscience*, 27(5):1176–1183.

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

[Strannegård et al., 2017a] Strannegård, C., Svangård, N., Bach, J., and Steunebrink, B. (2017a). Generic animats. In *International Conference on Artificial General Intelligence*, pages 23–32. Springer.

[Strannegård et al., 2017b] Strannegård, C., Svangård, N., Lindström, D., Bach, J., and Steunebrink, B. (2017b). The animat path to artificial general intelligence. In *Proceedings of IJCAI-17 Workshop on Architectures for Generality & Autonomy*.

[Topalidou and Rougier, 2015] Topalidou, M. and Rougier, N. P. (2015). [re] interaction between cognitive and motor cortico-basal ganglia loops during decision making: a computational study. *The ReScience journal*, 1(1).