Analyzing the Performance of “Winner-Take-All” and “Voting-Based” Action Selection Policies within the Two-Resource Problem

Orlando Avila-García, Lola Cañamero, René te Boekhorst

Adaptive Systems Research Group
Department of Computer Science, University of Hertfordshire
College Lane, Hatfield, Herts AL10 9AB, UK
O.Avila-Garcia@herts.ac.uk, L.Canamero@herts.ac.uk, R.TeBoekhorst@herts.ac.uk

Abstract. The problem of action selection for an autonomous creature implies resolving conflicts between competing behavioral alternatives. These conflicts can be resolved either via competition, following a “winner-take-all” approach, or via cooperation in a “voting-based” approach. In this paper we present two robotic architectures implementing these approaches, and report on experiments we have performed to compare their underlying optimization policies. We have framed this study within the context of the “two-resource problem,” as it provides a widely used standard that favors systematic experimentation, analysis, and comparison of results.

1 Introduction

The problem of behavior or action selection (AS) for an autonomous creature, animal or robot, consists in making a decision as to what behavior to execute next in order to satisfy internal needs and guarantee survival in a given environment and situation. This implies resolving conflicts between competing behavioral alternatives. Different behavior selection models have been proposed in the (robotics and ethological) literature (see [13] for a good overview), although not all of them address the same problems. Models can be classified according to various dimensions, such as their approach to behaviors (emergent functionality versus ethologically-inspired behavioral competencies), the use of external and internal factors to drive the selection (reactive versus motivated architectures), or the organization of the different architectural elements (in a “flat,” parallel network versus a hierarchy of behavioral subsystems). Another dimension often referred to when comparing architectures is their arbitration policy: Conflicts among different elements in the architecture can be resolved either via competition, following a “winner-take-all” (WTA) approach, or via cooperation in a “voting-based” (VB) approach. Competition and cooperation (and therefore WTA or VB policies) can also occur at several levels[1], giving rise to different behavior selection results.

Our analysis focuses on motivated behavior-based architectures with a combination of flat and hierarchical organization, following [4]. Within this common framework,

---

[1] Namely between motivational states, between behavioral subsystems, between simpler behaviors or actions, and between motor commands.

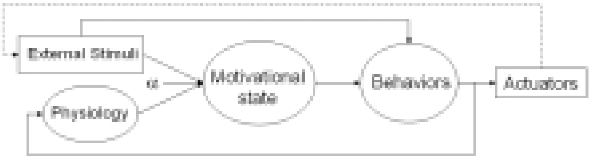

Fig. 1. Motivated behavior-based model underlying our architectures.

we have defined two architectures with different arbitration policies—WTA and VB. The work presented in this paper is part of our ongoing effort to investigate the adequacy of these AS policies for different problems and environments. In previous work [5, 2] we built and compared a WTA and various VB architectures in static and dynamic environments using different performance indicators drawn from the notion of viability [1]. Results obtained regarding the performance of WTA and VB architectures were complementary, and pointed to the possibility of different underlying optimization policies. To investigate this issue, we decided to place our study within the so-called "two-resource problem", in which a self-sufficient creature must continuously choose between consuming one of two resources available in the environment. This framework presents the advantages of providing a widely accepted standard that allows to compare results with those of other researchers, and of being simple enough to permit systematic experimentation and analysis. The remainder of the paper is as follows. Section 2 describes our architectures, which are framed within the two-resource problem in Section 3. Section 4 reports the robotic experiments we have performed to compare the optimization policies of our architectures. Finally, Section 5 draws some conclusions.

2 Behavior Selection Architectures

Our architectures are neither strictly flat (parallel) nor hierarchical (structured), but a combination of both. They consist of two layers—motivational and behavioral—linked through a synthetic physiology, leading to a two-step computation of intensity. This computation is parallel within each layer, but motivational intensity must be computed prior to the calculation of behavioral intensity, since the latter depends on the former. Both architectures (WTA and VB) have the same organization and elements, but they vary in the way in which these are combined (their arbitration mechanisms).

2.1 Architectural elements

The common architectural model (Figure 1) has three main elements: a synthetic physiology, motivational states, and behaviors.

The *physiology* consists of a number of survival-related, controlled homeostatic variables—abstractions representing the level of internal resources that the agent needs in order to survive. They must be kept within a range of values for the robot to stay alive, thus defining a physiological space [8] or viability zone [1] within which survival (continued existence) is guaranteed, whereas transgression of these boundaries leads to death. These variables vary as a function of internal body dynamics and of the interactions of the robot with its environment.

Motivational states are abstractions representing tendencies to behave in particular ways as a consequence of internal and external factors. The main elements defining them are bodily needs (traditionally known as "drives") and external (incentive) stimuli. Survival-related *bodily needs* set urges to action to maintain the controlled physiological variables within a certain range. A feedback detector generates an error signal—the drive—when the value of this variable departs from its ideal value (set point), and this triggers inhibitory and excitatory controlling elements (in this case, the execution of behaviors) to adjust the variable in the adequate direction. Each motivation receiving an error signal from its feedback detector receives an intensity (activation level) proportional to the magnitude of the error. The motivational state of the robot is also influenced by the presence of *external stimuli* or environmental cues that allow to execute (consummatory) behaviors and hence to satisfy bodily needs. In addition to bodily needs and the presence of environmental cues, *other factors* can influence motivational states, such as the quality of the stimulus (e.g., palatability of food), abnormal bodily states (e.g., "illness"), etc. To account for these factors we have introduced a parameter (α), as we will see in Section 2.2. As a consequence of these combined influences, several motivations can be active at the same time (and either "competing" or "cooperating" to be satisfied), with varying degrees of intensity.

Our *behaviors* are coarse-grained subsystems (embedding simpler actions) that implement different competencies, similarly to those proposed in [7, 4]. Following the classical distinction in ethology, motivated behaviors can be consummatory (goal-achieving and needing the presence of an incentive stimulus to be executed) or appetitive (goal-directed search for a particular external stimulus). In addition to modifying the external environment, the execution of a behavior has an impact on (increases or decreases) the level of specific physiological variables. Behaviors can be activated with different intensities that depend on the intensities of the motivations related to them. However, only one behavior can be executed at a time, following the assumption of the behavioral final common path [10].

2.2 Arbitration mechanisms

This is the point in which our architectures differ. With a WTA policy the robot will execute the behavior that best satisfies its most urgent motivation, i.e., only this motivation drives behavior selection. With a VB policy, all the motivations influence the final selection since the robot will execute the behavior that best satisfies a subset. Their respective behavior selection loops are:

Behavior selection loop in WTA. Repeat forever:

1. The winner motivation j_{winner} is calculated:
   (a) For each motivation j:
      i. Compute the intensity of the motivation's drive as proportional to the error[2] of its controlled variable v ($e_{v_j} \in [0, 1]$).

---

[2] The error is calculated as follows in both behavior selection loops:
A. Calculate the distance between the set point and the limit: $ld_v = abs(l_v - p_v)$.
B. Calculate the distance between the actual variable value and the set point: $vd_v = abs(v_v - p_v)$

      ii. Compute the effect of the presence of external stimuli k influencing the intensity of the motivation j (i.e., of incentive stimuli): $s_{k_j} \in [0, 1]$.
      iii. $m_j = e_{v_j} + (e_{v_j} \times \alpha s_{k_j})$ is the final intensity of j, where $\alpha \in [0, 1]$ is a weight controlling the influence of the incentive stimulus.
   (b) The motivation with highest intensity (j_{winner}) is selected.
2. The intensity of each behavior linked (through the physiology) with the winner motivation is computed as $b_i = m_{j_{winner}} \times f_{iv}$, where b_i, $m_{j_{winner}}$ are the intensities of behavior i and the winner motivation, respectively, and f_{iv} is the effect that the execution of behavior i has on v, which is the physiological variable controlled by j_{winner}.
3. The behavior with highest intensity is selected to be executed.

Behavior selection loop in VB. Repeat forever:

1. Calculate the intensity of each motivation j:
   (a) Compute the intensity of the motivation's drive as proportional to the error of its controlled variable v ($e_{v_j} \in [0, 1]$).
   (b) Compute the effect of the presence of external stimuli k influencing the intensity of the motivation j (i.e., of incentive stimuli): $s_{k_j} \in [0, 1]$.
   (c) $m_j = e_{v_j} + (e_{v_j} \times \alpha s_{k_j})$ is the final intensity of j, where $\alpha \in [0, 1]$ is a weight controlling the influence of the incentive stimulus.
2. The intensity of each behavior is computed as $b_i = \sum(m_j \times f_{iv})$, where b_i, m_j are the intensities of behavior i and motivation j, respectively, and f_{iv} is the effect that the execution of behavior i has on v, the physiological variable controlled by j.
3. The behavior with highest intensity is selected to be executed.

The way in which motivational and behavioral intensity are computed in the above algorithms deserves particular attention. When computing motivational intensity, the main problem is to define an appropriate combination rule between external (s_{k_j}) and internal (e_{v_j}) influences. This problem has been discussed extensively by ethologists [9]. Using an artificial creature, Spier and McFarland demonstrated how a simple multiplicative rule that they call *Cue* × *Deficit* model ($m_j = e_{v_j} \times s_{k_j}$) can show opportunism and persistence [11], and can generate optimal behavior equivalent to the quadratic utility function model [12]. One of the problems of this simple combination rule is the lack of motivational arousal (and hence, of behavioral tendency) when external incentive stimuli are absent ($s_{k_j} = 0$). This problem would be fatal in models in which the motivational state leads to the selection of the behavior to be executed and that also take into account appetitive behaviors (executed in the absence of external stimuli), as in our case. Therefore, following the formula proposed by [13] we have extended the previous model to *Deficit* + *Cue* × *Deficit*: $m_j = e_{v_j} + (e_{v_j} \times \alpha s_{k_j})$. Note that now there is motivational tendency to execute appetitive behaviors when there is no external stimulus, while we still get the benefits of the *Cue* × *Deficit* model. Another important enhancement of our model is the introduction of a weight (α) to control the

---

C. Calculate the normalized error:

$$e_v = \begin{cases} vd_v / ld_v & \text{if } ld_v > abs(l_v - v_v) \\ 0 & \text{otherwise} \end{cases}$$

influence of external stimuli. This weight can be taken to stand for other factors influencing motivational states[3], as explained in Section 2.1 and can have a big impact on the way optimization is achieved, as we will see in Section 4.3.

Computation of behavioral intensity is the point in which both arbitration mechanisms differ. Note that, when calculating behavioral intensity in the selection loop of WTA (step 2), only one motivation drives behavior selection; however, in VB (step 2), all the motivations influence (positively or negatively) the selection of a behavior.

3 The Two-Resource Problem

The framework we have used to compare our WTA and VB architectures and analyze their optimization policies is known as the two-resource problem. In this scenario, a self-sufficient (biological or artificial) creature must continuously decide which of its two survival-related needs to satisfy by choosing between two resources available in the environment. This framework presents the advantage of providing a standard that allows to compare results with those of other researchers, since it has been widely used to study action selection both in animals (e.g., to study feeding and drinking patterns in doves [8]) and in artificial creatures (see e.g., [3, 12, 6]). Also, its simplicity (although not devoid of problems) favors a systematic analysis of results, and it has actually been characterized as the minimal scenario to test action selection mechanisms [12].

Our particular implementation of the two-resource problem, detailed in Table 1, has been as follows. Like any implementation of this problem, we have used two environmental resources (external stimuli) and two physiological needs giving rise to two motivational states. However, departing from other implementations in which each motivational state can only be satisfied by one behavior, in our case two consummatory behaviors can satisfy each motivation[4] to varying degrees. The reason for this is that, with only one behavior to satisfy each motivation, a VB architecture would become a WTA one, and therefore we would not be able to compare these policies.

Consummatory behaviors can only be executed when the intensity of their related external stimulus is greater than a threshold (0.85); the robot then stops on top of the resource and “consumes” it, modifying its physiology accordingly. If a consummatory behavior cannot be executed, Search is triggered to reach the center of the resource (where $>$85% of stimulus is present) using both light sensors for orientation over the light gradient. Avoid is a reflex behavior that uses both bumpers to avoid the wall.

3.1 Analysis Criteria

Spier and McFarland [11] noted that, within the framework of the two-resource problem, a self-sufficient robot must perform a *basic cycle* of activities to maintain viability. The study of the physiological space should give us these cycles. Figure 2 (left) shows the basic cycle for our implementation. During phase *A*, our robot executes the Search

---

[3] In addition, this weight could be useful to introduce lifetime learning or even evolution to adapt the influence of different external stimuli to varying ecological circumstances [9].

[4] And adjust its controlled variable, also affecting the other physiological variable.

Table 1. Motivations (left) and behaviors (right) used. Controlled (physiological) variables have set points fixed at 100 and lethal boundaries at 0 ($Var_i \in [0, 100]$).

Motivation	**Physiological Drive**	**Ext. Stim.**
Cold	$\uparrow Temperature$	Heat
Fatigue	$\uparrow Energy$	Food

Behavior	**Type**	**Stimulus**	**Effects on physiology**
Avoid	*Reflex.*	Obstacle	$\downarrow 0.2 Temperature, \downarrow 0.2 Energy$
WarmUp1	*Consum.*	Heat	$\uparrow 0.9 Temperature, \downarrow 0.1 Energy$
WarmUp2	*Consum.*	Heat	$\uparrow 1.0 Temperature, \downarrow 0.3 Energy$
Feed1	*Consum.*	Food	$\downarrow 0.1 Temperature, \uparrow 0.9 Energy$
Feed2	*Consum.*	Food	$\downarrow 0.3 Temperature, \uparrow 1.0 Energy$
Search	*Appet.*	None	$\downarrow 0.2 Temperature, \downarrow 0.2 Energy$

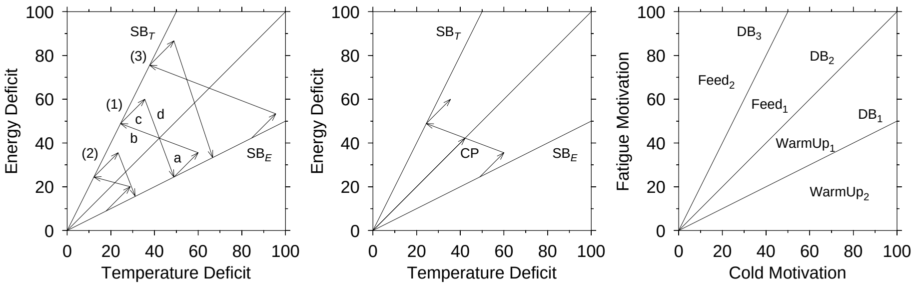

Fig. 2. Left: basic cycles in the physiological space; (1) is a stable cycle, (2) a cycle that climbs up the diagonal, (3) a cycle that descends it. Center: Optimization error as the magnitude of the vector OpE. Right: Motivational space with dominance boundaries for behavioral competition.

behavior to find the resource Heat. The phase labeled *B* corresponds to the execution of consummatory behaviors that increase Temperature (WarmUp1, WarmUp2), also decreasing Energy to a lesser extent. When this need is satisfied, another episode of Search is executed (label *C*), this time to obtain Food. This is followed by the execution of consummatory behaviors that help to correct the Energy deficit (Feed1, Feed2, that also decrement Temperature slightly) in phase *D*. Upon satiation, the same basic cycle is repeated again. Note that in our implementation the cycle is symmetric with respect to the diagonal of the physiological space because both aspects of our problem are symmetric in all respects (physiology, resources, motivations and behaviors). Figure 2 (left) shows how the shape of the basic cycle depends on its location on the diagonal with respect to the origin, and that this cycle becomes stable (i.e., its initial and final points coincide) only at a particular location.

Satiation boundaries. The point at which the satisfaction of a particular need is achieved can be termed *satiation boundary* (Figure 2, left). This limit defines the amplitude of the cycle, and can be obtained analytically. Since our two problems are symmetric, we can assert that the point at which a motivation that is being satisfied (m_{cold}) loses the competition against the other ($m_{fatigue}$) is defined by $m_{cold} = m_{fatigue}$. At this moment, the robot is consuming a resource leading to the satisfaction of m_{cold}, and hence the intensity of this external stimulus must be near the maximum ($s_{heat} \simeq 1$). In this situation, the incentive stimulus of the second motivation ($m_{fatigue}$) is disregarded ($s_{food} \simeq 0$). Therefore, using the *Deficit + Cue × Deficit* formula we have adopted to calculate motivational intensity, the final equation will be:

$$e_{energy} + (e_{energy} \times \alpha s_{food}) = e_{temper} + (e_{temper} \times \alpha s_{heat}), \qquad (1)$$

$$e_{energy} \simeq (1 + \alpha) e_{temper} \qquad (2)$$

The line determined by this equation is the satiation boundary of m_{cold} (Figure 2, left, SB_T). Note that the parameter α, determining the influence of the external stimulus, determines also the slope of that boundary, and therefore the amplitude and location of the stable cycle; this implies that α determines the distance of the stable cycle with respect to the origin of the physiological space. This parameter thus seems to play an important role in determining how much the cycles approach the origin or ideal physiological state (Figure 2, left), and hence, how the evolution of the cycles is optimized.

Optimization criteria. Since the cycles are going to shift along the axis determined by the diagonal of the physiological space, our optimization criterion is going to be the Euclidian distance from the origin to the point where the cycle crosses the diagonal (Figure 2, center). Note that the shorter the distance, the closer the cycle gets to the origin. We have used the mean and minimum optimization error (normalized between [0,1]) obtained during the robot lifetime as performance indicators.

Dominance boundaries. Motivational states define a motivational space (Figure 2, right). Decisions made to select behaviors under different motivational states are reflected in this space, and this defines dominance boundaries [8] that divide the motivational space into regions in which a particular behavior is “dominant,” i.e., is being executed. For WTA, only one dominance boundary is drawn (Dominance Boundary 2 in Figure 2, right) since in fact this policy only uses two of the four behaviors available—those that satisfy the motivations to a greater extent, Feed2 and WarmUp2. On the contrary, for VB behavioral competition defines a more complex set of dominance boundaries—three in our case, dividing the dominance regions of our four behaviors. This difference observed between WTA and VB is reflected in the different consummatory phases of basic cycles, as we will see in Section 4.3.

4 Experiments

4.1 The robot and its environment

Tauron, our robot, is built using Lego Mindstorms. It has a simple “roverbot-like” design to facilitate replication of our experiments, and it is equipped with (Figure 3, left and center): two bumpers for obstacle avoidance, two frontal light sensors to detect brightness and darkness, and two motors providing differential steering to the wheels. The environment (Figure 3, right) is a $1m \times 1m$ arena surrounded by a wall. The floor is made of 9 tiles[5] (of $33cm \times 33cm$ each) of three types: “empty space” (uniform gray), two “heat sources” (white gradient), and two “food sources” (black gradient).

---

[5] The use of tiles does not mean that our environment is a “grid world” since our tiles have continuous physics.

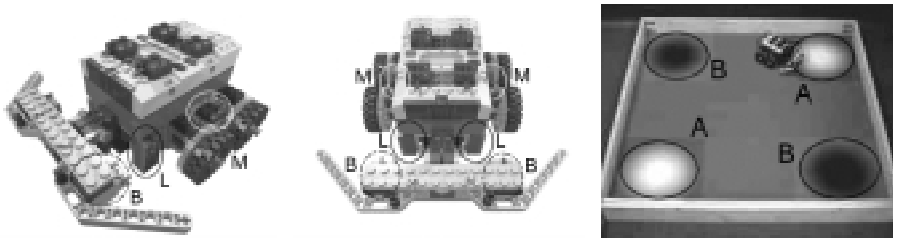

Fig. 3. Left and center: Tauron robot. Circles show the position of sensors (L: light, B: bumpers) and motors (M). Right: Environment. Circles surround resources: heat (A) and food (B).

4.2 Method

Five sets of experiments were performed to test both architectures with five different α (1, 0.7, 0.5, 0.3, 0). Each set consisted of 20 runs, 10 for each architecture, with a total of 100 runs (about 12 hours). Each run lasted 1600 steps of $260ms$ each, i.e., about 7 minutes. For each run we initialized randomly physiological variables (to values within their viability range) and location of resources in the environment.

4.3 Results

Figure 4 shows the average performance of both architectures for different α in terms of mean and minimum optimization errors. Life span is also shown as a reference indicator. We have used analysis of variance (ANOVA) to describe the results. The difference between architectures in terms of *life span* was not significant; however, there is a tendency indicating that VB performed better than WTA. For $\alpha = 0$, the life span of both architectures decreases considerably due to the lack of persistence in the execution of consummatory behaviors and therefore the dithering produced. In terms of *mean optimization error*, the difference in the performance of each architecture for different values of α is statistically highly significant (99%). The difference between architectures is also highly significant, VB outperforming WTA. In terms of *minimum optimization error,* the difference in the results obtained by VB for different values of α is statistically highly significant (99%), whereas it is not significant for WTA[6]. The difference between architectures is significant (95%), VB outperforming WTA.

Our results thus show that α strongly influences optimization—the smaller the values used for α, the bigger the optimization error, as predicted by our model (see Section 3.1). The position of the stable cycle varies between WTA and VB architectures—it gets closer to the origin in VB (Figure 5), reflecting the fact that VB outperforms WTA in terms of our two optimization indicators, as explained above. This can be illustrated taking into account the way in which both architectures consume resources as reflected

---

[6] It should be noted the high level of variance in this case, possibly due to the big error intervals in the results obtained by this architecture for life span.

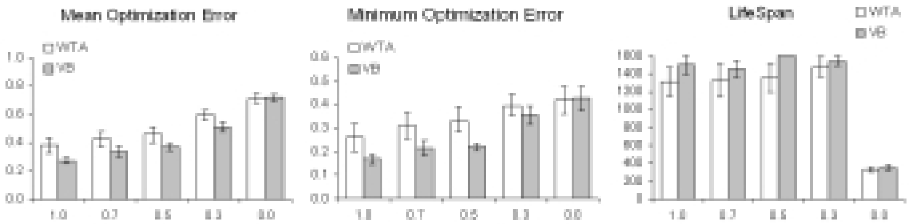

Fig. 4. Average performance of both architectures for different α (x-axis) in terms of mean (left) and minimum (center) optimization error, and life span (right). Standard error is shown.

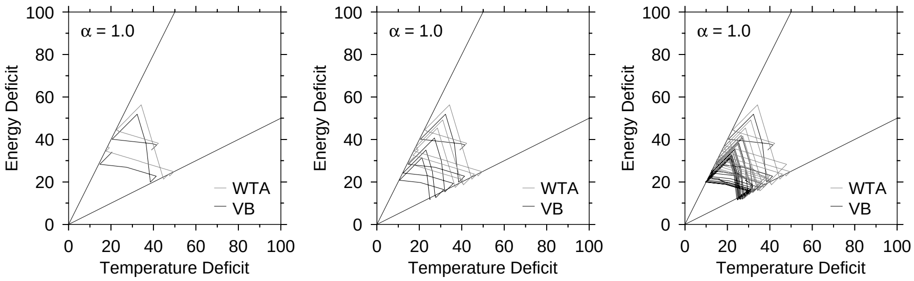

Fig. 5. Example of evolution of both architectures within the physiological space in a simulation run ((α = 1). Left: first 200 steps. Center: first 500 steps. Right: a complete run of 1600 steps.

in the shape of their respective cycles depicted in Figure 5. The fact that WTA only executes the behavior that corrects physiological imbalance to a greater extent is reflected in the straight shape of the consummatory phases of the cycles; on the contrary, the tendency of VB to use all the available behaviors can be seen in the convex shape of its consummatory phases. Therefore, the more "cost-effective" strategy used by VB causes that its basic cycles approach the origin more than those of WTA.

5 Conclusion and Future Work

We have compared the optimization policies of simple "winner-take-all" and "voting-based" action selection mechanisms using two corresponding robotic architectures within the framework of the two-resource problem. Our results indicate that VB outperforms WTA in the sense that its basic behavioral cycles approach more the ideal physiological state of absence of deficit. The more "cost-efficient" way in which VB consumes resources accounts for this result in this two-resource problem study. However, other factors need to be considered that we could not take into account this time (but that we considered in previous work) due to the simplified architectures used to adapt them to the two-resource problem. Some of these factors are the fact that a behavior can contribute to satisfy more than one motivation, or the fact that different behaviors can satisfy the same motivation, possibly using different resources. It is also important to bear in mind that the experiments we report in this paper have been performed in static worlds only; therefore, these results cannot be generalized to dynamic environments that produce non-deterministic changes to the physiology of the robot as a result of the effects of both, behaviors and environment. This study has also shown that the weight given to external motivational factors has a big impact on the location of basic cycles in the physiological space, and hence on how well this space is managed. Our introduction of a parameter (α) in the formula used to calculate motivational intensity has allowed us to control this influence, and to determine the location of stable basic behavioral cycles within the physiological space.

To continue this study, we envisage three main directions. First, we would like to do a thorough comparison of our study with other different implementations of the two-resource problem reported in the literature. Second, we will introduce different sources of dynamism (namely mobility of resources and an "enemy" species) to investigate the anticipation and reactivity features of both AS policies. Finally, we intend to study how the dynamic variation of α (the weight given to the influence of external stimuli) during the robot's lifetime can affect on its capability to adapt to the environment.

References

1. Ashby, W.R. (1952). *Design for a Brain*. London: Chapman and Hall.
2. Avila-García, O. and Cañamero, L. (2002). A comparison of Behavior Selection Architectures Using Viability Indicators. In *Proc. International Workshop on Biologically-Inspired Robotics: The Legacy of W. Grey Walter*. 14–16 August 2002, Bristol HP Labs, UK.
3. Blumberg, B. (1994). Action Selection in Hamsterdam: Lessons from Ethology. In *Proc. Third Intl. Conf. on Simulation of Adaptive Behavior (SAB94)*. Cambridge, MA: MIT Press.
4. Cañamero, L.D. (1997). Modeling Motivations and Emotions as a Basis for Intelligent Behavior. In W.L. Johnson, ed., *Proc. First Intl. Conf. on Autonomous Agents*, 148–155. New York: ACM Press.
5. Cañamero, L., Avila-García, O., Hafner, E. 2002. First Experiments Relating Behavior Selection Architectures to Environmental Complexity. In *Proc. 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS 2002), 3024–3029. IEEE Press.
6. Girard, B., Cuzin, V., Guillot, A., Gurney, K.N., and Prescott, T.J. (2002). Comparing a Brain-Inspired Robot Action Selection Mechanism with 'Winner-Takes-All'. In *Proc. Seventh Intl. Conf. on Simulation of Adaptive Behavior*. Cambridge, MA: MIT Press.
7. Maes, P. (1991). A Bottom-Up Mechanism for Behavior Selection in an Artificial Creature. In J.A. Meyer and S.W. Wilson, eds. *Proc. First Intl. Conf. on Simulation of Adaptive Behavior*, 238–246. Cambridge, MA: MIT Press.
8. McFarland, D. (1974). Experimental Investigation of Motivational State. In D. McFarland, ed., *Motivational Control Systems Analysis*, 251–282. London: Academic Press.
9. McFarland, D. (1976). Form and function in the temporal organization of behavior. *Growing Points in Ethology*, Bateson P., Hinde R. ed. Cambridge University Press.
10. McFarland, D. and Sibly, R. (1975). The behavioural final common path. *Philosophical Transactions of the Royal Society (Series B)*, 270, 265-293.
11. Spier, E. and McFarland, D. (1997). Basic Cycles, Utility and Opportunism in Self-Sufficient Robots. *Robotics and Autonomous Systems*, 20, 179-190.
12. Spier, E. and McFarland, D. (1997). Possibly Optimal Decision Making under Self-Sufficiency and Autonomy. *Journal of theoretical biology*, 189, 317-331.
13. Tyrrell, T. (1993). Computational Mechanism for Action Selection. *PhD. Thesis, Centre for Cognitive Science, University of Edinburg*.