

Using Hormonal Feedback to Modulate Action Selection in a Competitive Scenario

Orlando Avila-García and Lola Cañamero

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

Abstract

In this paper we investigate the use of hormonal feedback as a mechanism to modulate a “motivation-based,” homeostatic action selection mechanism (ASM) in a robot. We have framed our study in the context of a dynamic, multi-robot, competitive “two-resource” action selection problem. The introduction of competitors has important consequences for action selection. We first show how the interaction between robots introduces new forms of environmental complexity that affect their viability. Secondly, we propose a “hormone-like” mechanism that, modulating the input of the ASM, tackles these new sources of complexity.

1 Introduction

The choice of appropriate behavioral subsystems is one of the major problems of designing behavior selection architectures. While achieving flexible behavior adapted to complex environments requires a non-trivial set of subsystems, including too many leads to problems arising from overdesign. A balance must be found between the “machinery” used and the functionalities required from this underlying structure.

In biological systems, the bath of chemicals where neurons are placed, and that we will refer to here as neurohormones, is the major responsible for the plasticity of the nervous system. Neurohormones can act at multiple levels of the nervous system, from sensory to motor neurons, and can evoke a spectrum of different responses on different target neurons. For example, studies in invertebrates have shown that the same anatomical neural circuit can produce different motor patterns or behaviors in the presence of different neurohormones (see (Harris-Warrick *et al.* 1992) for a general overview). These modulatory inputs are thus seen as a set of instructions for building different functional networks from the same anatomically defined neural circuit (Kravitz 1988, Harris-Warrick *et al.* 1992).

Different mechanism inspired from the role of hormones have been proposed for behavioral learning and coordination (Brooks & Viola 1990, Grand *et al.* 1997), and emotion modelling (Cañamero 1997). In this paper, we take inspiration from the ability of neurohormones to rewire the same

anatomically defined neural circuit; we explore a mechanism that modulates an action selection architecture to achieve different functionalities adapted to different environmental conditions. We have framed our study in the context of a dynamic “two-resource” action selection problem (TRP) extended with competitors, in which two robots competing for the same resources must choose between different activities in order to survive. We first show how the interaction between robots introduces various types of environmental complexity that affects the viability of the homeostatically controlled robots. Secondly, we propose a “hormonal feedback” that, acting at the sensory level, modulates the network of interrelations within the action selection architecture, biasing the behavior of the robot toward more adaptive responses to the environment—and giving rise to interesting emergent behavioral phenomena when both robots interact.

The remainder of the paper is as follows. Section 2 describes our action selection architecture, which is framed within the TRP with competitors (Section 3). Section 4 provides background theory necessary to understand the analysis of results. Section 5 reports the preliminary robotic experiments within the TRP, while in Section 6 we describe the implementation of a hormonal system to improve the performance of the robots, and the subsequent experiments. Finally, Section 7 draws some conclusions.

2 Behavior Selection Architecture

In previous work we defined different behavior (or action) selection architectures sharing the same general framework and components but implementing different Voting-Based (VB) and Winner-Take-All (WTA) policies. In this study, we have only used the WTA version of this architecture since our goal was to investigate action selection in a dynamic environment, and previous results (Avila-García & Cañamero 2002) showed that a WTA policy outperformed VB ones in dynamic environments¹. Our architecture is neither strictly flat (parallel) nor hierarchical (structured), but a combination of both. It

¹This does not imply that WTA outperforms VB in general, since in another study (Avila-García *et al.* 2003) a VB policy outperformed WTA in a static two-resource problem scenario.

consists 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. The motivational layer is made of motivational states that set the goals of the system—the tendency to satisfy bodily (physiological) or internal needs. The behavioral layer implements different ways in which those bodily needs can be satisfied. This distinction between motivations and behaviors is essential when implementing more than one behavior satisfying the same motivation (Toates 1986). Although for the simplified implementation of WTA that we use in this paper such a distinction is not necessary (see Figure 1), we have preferred to use the same model as in our previous studies to be able to compare our results, and also to allow scalability when other sources of dynamism are considered in future studies.

The Physiology consists of a number of survival-related, homeostatically controlled 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 (Sibly and McFarland 1974) or viability zone (Ashby 1952, Meyer 1995) within which survival (continued existence) is guaranteed, whereas transgression of these boundaries leads to “death.”

Motivational states are abstractions representing tendencies to behave in particular ways as a consequence of internal and external factors (Toates 1986). Internal factors are mainly (but not only) bodily or physiological deficits or needs (d_i), also traditionally known as “drives,” that set urges to action to maintain the state of the controlled physiological variables within the viability zone. External factors are environmental stimuli or incentive cues (c_i) that allow to execute (consummatory) behaviors and hence to satisfy bodily needs. In our implementation, each motivation performs homeostatic control of one physiological variable. In that control, a feedback detector generates an error signal when the value of the controlled variable departs from its ideal value (set point). Thus, each motivation receives an intensity (activation level) proportional to the magnitude of the error. The motivational state is then the tendency to correct that error or deficit through the execution of behaviors. The homeostatic control of the physiological variable is also characterized by a limit or lethal boundary, the transgression of which leads to the robot “death.” In addition to physiological deficits (d_i) and the presence of environmental cues (c_i), *other factors* may influence motivational states, such as the quality of the stimulus (e.g., palatability of food), abnormal bodily states (e.g., “illness”), hormonal levels (Cañamero 1997), etc. To account for these factors we have introduced a parameter (α) in the computation of the motivations’ intensity (see Equation 1).

Different combination rules for external and internal fac-

Motivation	Physiological Drive	Limit	Set point	Ext. Stím.
m_{cold}	$\downarrow d_{temperature}$	0	100	c_{heat}
$m_{fatigue}$	$\downarrow d_{energy}$	0	100	c_{food}

Table 1: Motivations used. Physiological drive represents the tendency of the motivation to decrement the physiological deficit until the correspondent variable reaches a set point of 100.

Behavior	Type	Stimulus	Effects on physiology (f_{ij})
b_{avoid}	<i>Reflex.</i>	Obstacle	$+ 0.2 d_{temp}, + 0.2 d_{energy}$
b_{warmup}	<i>Consum.</i>	c_{heat}	$- 1.0 d_{temp}, + 0.3 d_{energy}$
b_{feed}	<i>Consum.</i>	c_{food}	$+ 0.3 d_{temp}, - 1.0 d_{energy}$
b_{search}	<i>Appet.</i>	None	$+ 0.2 d_{temp}, + 0.2 d_{energy}$

Table 2: Behaviors used by the WTA architecture.

tors when computing motivational states have been discussed extensively by ethologists, e.g. (McFarland 1976). Following ideas proposed by Tyrrell (1993) and McFarland & Spier (1997), we have adopted a *Deficit + Cue \times Deficit* model:

$$m_i = d_i + (d_i \times \alpha c_i) \quad (1)$$

Behaviors are coarse-grained subsystems (embedding simpler actions) that implement different competencies, similarly to those proposed in (Maes 1991, Cañamero 1997). Following the classical distinction in ethology (Hinde 1953) and more recent advances in neuroscience (Robbins & Everitt 1999), 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 incentive 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. Therefore they are a mechanism to maintain the state of the physiological variables within the viability zone. In its Winner-Take-All form, our architecture will always execute the behavior that best satisfies the highest activated motivation.

3 The Competitive Two-Resource Problem

Our framework is an extension of the scenario known in the action selection literature as the two-resource problem (TRP). In the traditional TRP, 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. The TRP 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 (see e.g., (McFarland 1974)) and in artificial creatures (see e.g., (Blumberg 1997, Spier & McFarland 1997, Girard *et al.* 2002)). 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 (Spier & McFarland 1997).

Our extension of this problem, the competitive two-resource problem (CTRP) consists in the introduction of two robots in the same environment simultaneously performing their own two-resource problem. The fact that they have to use the same resources to satisfy their needs introduces competition for those resources, as both robots might need access to the same resource at the same time.

3.1 Robotic Implementation of the TRP

Our competitive two-resource problem (CTRP) builds on a previous implementation of a typical static two-resource problem (TRP) (Avila-García *et al.* 2003), in which a single robot must choose between two types of static resources, i.e. the resources do not change location and are always equally accessible. Our particular implementation of the TRP, detailed in Tables 1 and 2, was as follows (see Figure 1 for a whole picture of the motivational system). Like any implementation of this problem, we used two environmental resources (incentive cues) that allow to satisfy two physiological needs, the deficits of which give rise to two motivational states. Motivations are satisfied by the execution of a consummatory behavior. The execution of each behavior reduces the deficit of one physiological variable, increasing the deficit of the other one.

In addition to consummatory behaviors (b_{feed} and b_{warmup}), our architecture also includes appetitive (b_{search}) and reflex (b_{avoid}) behaviors. It is important to point out that in this implementation only consummatory behaviors, but not appetitive or reflex, participate in the action selection process; the other two behaviors only are called upon to support their execution. Thus, b_{avoid} is a reflex behavior that is executed whenever an obstacle is present, and b_{search} is an appetitive behavior that is executed when the winner consummatory behavior cannot be executed because its related incentive stimulus is not present.

The particular robotic scenario we used to implement this TRP (see Figure 2) was as follows. Our robot, Tauron, was a modified version of the standard Lego Mindstorms “roverbot”, with a controller programmed in C implementing the synthetic physiology and architecture described above (see Figure 1). We chose this simple design model to facilitate replication of our experiments. The robot was equipped with: two bumpers for obstacle avoidance, two frontal light sensors to detect brightness and darkness on the floor of the arena, and two motors providing differential steering to the wheels. The environment was a $1m \times 1m$ arena surrounded by a wall. The floor is made of 9 tiles (of $33cm \times 33cm$ each) of three types: “empty space” (uniform gray), two “heat sources” (white gradient), and two “food sources” (black gradient). The use of tiles does not mean that our environment is a “grid world” since our tiles have continuous physics.

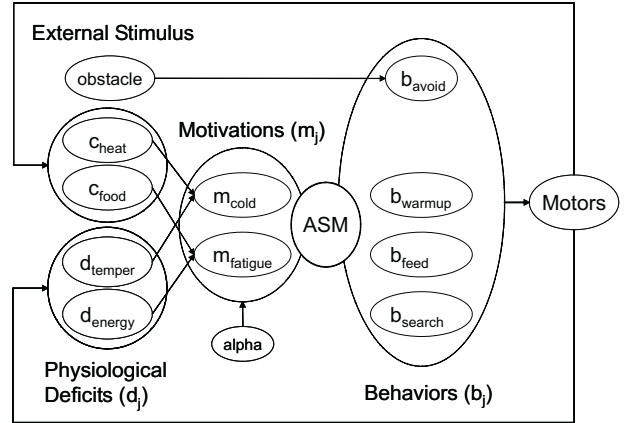


Figure 1: Motivational control system for our WTA architecture. The winner behavior modifies the physiology of the robot (negative feedback for the ongoing behavior) and sends commands to the motors, which can eventually change the perception of incentive cues (acting as a positive feedback).

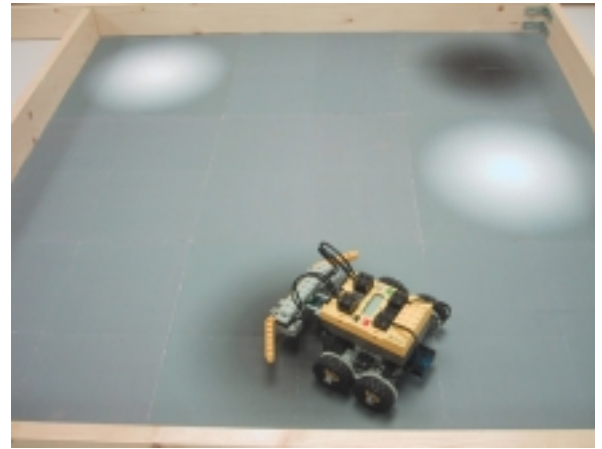


Figure 2: Our robotic TRP scenario.

3.2 The CTRP and Environmental Complexity

In this study we extend our previous implementation of the TRP to create a dynamic environment in order to introduce new forms of environmental complexity. Dynamism is achieved by the addition of another robot in the same arena. This element adds complexity to the environment by introducing competition between both robots—hence we call this new problem the competitive two-resource problem (CTRP). We can identify at least two new types of environmental complexity, that we have studied in this paper:

- (a) Due to the constraints of our implementation, two robots cannot have simultaneous access to the same resource, since a robot must be placed on top of a resource to consume it. This results in a continuous change in availability and accessibility of resources. For example, while one

robot is consuming one resource, this type of resource will be less available for the other robot. Likewise, the particular resource that is being consumed will not be accessible to the other robot, which means that, from the point of view of the other robot, that resource has “disappeared” from its original position.

- (b) When one robot is located on top of a resource—i.e. consuming it—the other robot might bump into it and push it out of the resource. This will result in the interruption of the ongoing consummatory activities, which for the robot that was executing them means that the resource has been extinguished.

4 Background Notions

4.1 Physiological Space

Physiological variables, controlled homeostatically, define a physiological space (McFarland 1974) or viability zone (Ashby 1952, Meyer 1995) within which survival (continued existence) is guaranteed, whereas transgression of these boundaries leads to death. In our TRP the physiological space has two dimensions, with each axis representing the deficit of each physiological variable (d_{temp} , d_{energy}). The origin of this space is the ideal state where both physiological needs are satisfied ($d_{temp} = 0$, $d_{energy} = 0$). Lethal boundaries are defined by $d_{temp} = 1$ and $d_{energy} = 1$ (Figure 3). We have used this physiological space to analyze the behavior and performance of our robots, with the help of the notions characterized in the remainder of this section.

4.2 Activity Cycles

Most work on action selection has focused on the choice of behavioral subsystems and individual behaviors resulting from different underlying selection mechanisms. However, McFarland and Spier (McFarland & Spier 1997) noted that a self-sufficient robot must perform a *basic cycle* of activities (and not only atomic behaviors) to maintain viability. The study of the physiological space of our robot should give us these cycles, which can then be used to analyze the behavior and performance of the robot.

Figure 3 (left) shows the basic cycle for our TRP implementation. During phase *a*, our robot executes the b_{search} behavior to find the resource c_{heat} . The phase labelled *b* corresponds to the execution of b_{warmup} , the consummatory behavior that decreases Temperature deficit (d_{temp}), also increasing the deficit of Energy (d_{energy}) to a lesser extent. When this need is satisfied, another episode of b_{search} is executed (label *c*), this time to obtain c_{food} . This is followed by the execution of b_{feed} , the consummatory behavior that helps to correct the Energy deficit (and also increments the deficit of 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; this is

due to the fact that both aspects of our problem are symmetric in all respects (physiology, motivations, behaviors, etc.).

4.3 Satiation boundaries.

The point at which the robot stops consuming one resource—i.e., it finishes satisfying one motivation—can be termed *satiation boundary* (Figure 3, left). This limit defines the amplitude of the activity cycles, 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 that leads to the satisfaction of m_{cold} , and hence the intensity of this incentive cue must be near the maximum ($c_{heat} \simeq 1$). In this situation, the incentive cue of the second motivation ($m_{fatigue}$) is disregarded ($c_{food} \simeq 0$). Therefore, using the *Deficit + Cue × Deficit* equation (1) that we have adopted to calculate motivational intensity, the final equation will be:

$$\begin{aligned} d_{ener} + (d_{ener} \times \alpha c_{food}) &= d_{temp} + (d_{temp} \times \alpha c_{heat}), \\ d_{energy} &\simeq (1 + \alpha)d_{temp} \end{aligned}$$

The line determined by this equation is the satiation boundary of m_{cold} (Figure 3, left, SB_T), that is, the point at which there is no more tendency to decrease d_{temp} . Note that the parameter α , determining the influence of the incentive cue, determines also the slope of the satiation boundaries.

4.4 Opportunism and Persistence

Opportunism—taking advantage of the opportunities offered by the environment to satisfy a lower-priority need while searching to satisfy a higher-priority one—and persistence—the ability to keep working towards the satisfaction of the current goal—are two of the main features that a good action selection architecture must show (Maes 1991, Tyrrell 1993). These notions are also used in ethology to explain phenomena commonly observed in animal behavior related to decision-making, see e.g., (Slater 1999).

In our architecture, both phenomena are supported by the same positive feedback mechanism: The intensity of incentive cues. This is a well-known mechanism to avoid dithering (McFarland 1974). When the robot approaches a resource, the incentive cue (c_i) associated with it produces a positive feedback that increases the motivation (m_i) to consume that resource. This causes that under certain circumstances, a motivation wins the competition to be satisfied even when its related physiological deficit (d_i) is not the highest. In other words, in a system without such a positive feedback, a motivation would always win the competition when the physiological state is at one specific side of the diagonal of such space—that is, m_{cold} would always win whenever $d_{temp} > d_{energy}$. Figure 3 (center) depicts such a diagonal and the regions A and B, where $m_{fatigue}$ and m_{cold} are

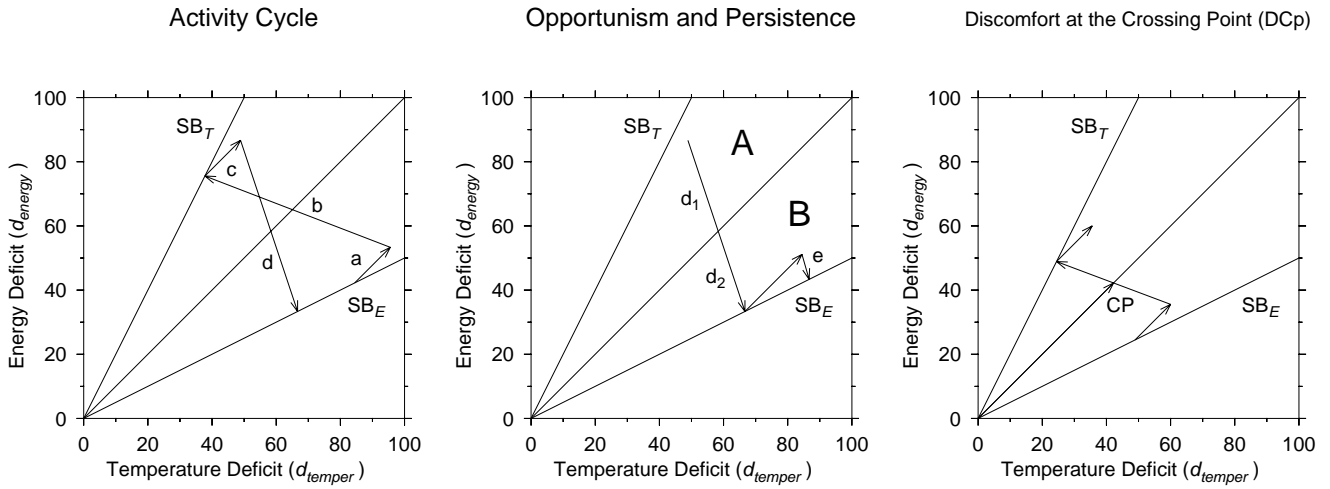


Figure 3: Left: Basic cycle of activities within the physiological space, with each axis representing the deficit of each physiological variable (d_{temp} , d_{energy}). Every activity cycle is depicted within the temperature and energy satiation boundaries (SB_T and SB_E respectively), and it consists of an appetitive phase (a) to look for c_{heat} (b_{search}); a consummatory phase (b) to decrease the Temperature deficit (b_{warmup}); followed by an appetitive phase (c) to look for c_{food} (b_{search}); and a consummatory phase (d) to decrease the Energy deficit (b_{feed}). Center: Situations of persistence (d_2) and opportunism (e) to satisfy $m_{fatigue}$ —reducing Energy deficit. Right: Discomfort at the Crossing point (DCp) as the Euclidean distance to the point at which the cycle crosses the diagonal.

dominant respectively—i.e., where they would be executed unconditionally. However, in our system, due to such a positive feedback, a motivation can be executed even when the physiological state is not within its dominance region—and that is always the case when its related incentive cue is higher than zero ($c_i > 0$).

Therefore, when the robot is next to a resource, the motivation to consume that resource increases, giving rise to two interesting phenomena: *opportunism* and *persistence*. We consider that our robot is being opportunistic when it starts satisfying a motivation—consuming its associated resource—when the physiological state is not in its dominance region. Figure 3 (center) depicts a typical opportunistic activity (e), when the robot decides to satisfy the motivation $m_{fatigue}$ —reducing energy deficit d_{energy} while the physiological state is in B, the dominance region of m_{cold} .

We define persistence also as the act of satisfying one motivation out of its dominance region, but in this case the robot starts consuming such a resource in the “right” dominance region. Figure 3 (center) depicts a typical situation (d_2) of persistence. The robot starts satisfying $m_{fatigue}$ in A, but at some point the physiological state crosses to the other dominance region B. Due to the presence of the incentive cue, $m_{fatigue}$ continues being the winner until the physiological state reaches its satiation boundary (SB_E). Thus, there will be two typical phases in every consummatory process: a normal one, where there is no persistence (d_1), and the following phase with persistence (d_2), where the execution of the same behavior is only due to the presence of the incentive cue.

4.5 Performance Indicators

In previous studies we showed that, in order to successfully compare and understand the performance of action selection mechanisms in different tasks and environments, we must use more than one performance indicator (Avila-García & Cañamero 2002). We have extended that study developing the novel idea of understanding the action selection process in terms of activity cycles rather than single decisions (Avila-García *et al.* 2003). In that sense, a study of activity cycles must be included in any comparison between architectures. In this paper, we use four performance indicators, three based on the notion of viability and one on activity cycles, to assess quantitatively the performance of the robots.

4.5.1 Viability Indicators

We have used the notion of viability (Ashby 1952) as first criterion to assess performances. The behavior of the robot is said to be viable when it keeps the values of the essential variables within the boundaries of the physiological space. In our opinion, however, this notion of viability is too vague to provide a direct criterion to measure the goodness or performance of a behavior selection architecture, as it leaves several possibilities open. For example, the performance of a behavior selection architecture can be simply assessed in terms of the time it allows the robot to remain viable (survive) in a given environment. Longer life spans usually indicate better behavior selection performance but this correlation is not necessarily straightforward, since the “life quality” of the robot can be very different during its life span depending on how viability is preserved during that period. A robot can live a long

life with poor “life quality” if the values of its essential variables are kept close to the critical zone (near the boundaries) of the physiological space for a long period. On the contrary, it can live a shorter life of “high quality” that ends due to accidental factors. “Life quality” can also have different interpretations. It can for example be measured in terms of global internal stability or “comfort” that takes into account the average level of satisfaction of all the essential variables simultaneously. It can also be seen in terms of how “balanced” the satisfaction of the different physiological needs is. Finally, the viability zone (physiological space) can be divided in different “regions”, ranging from an “ideal zone” around the ideal state in which the physiological variables have their ideal values, to a “critical zone” near the lethal boundaries that lead to death if transgressed. These regions can indicate the “Risk of Death” (RoD) due to parameter overflow.

We have used three² indicators of viability in order to reflect various ways in which good viability can be obtained: Life span, Overall Comfort and Physiological Balance.

Life Span. It is the time that the robot survived (remained viable) during each run:

$$LS = t_{life}/t_{run} \quad (2)$$

where t_{life} is the number of steps that the robot lived and t_{run} is the maximum number of steps of one run.

Overall Comfort. It measures the average level of satisfaction of both physiological variables during a run:

$$OvC = \sum_{i=1}^{t_{life}} (1 - \bar{d}_i) / t_{life} \quad (3)$$

where \bar{d}_i is the mean of both physiological deficits.

Physiological Balance. It measures the homogeneity with which both physiological needs are satisfied during a run:

$$PhB = \sum_{i=1}^{t_{life}} (1 - \sigma^2(d_i)) / t_{life} \quad (4)$$

where $\sigma^2(d_i)$ is the variance of both physiological deficits.

4.5.2 Discomfort at the Crossing point (DCp)

Our last performance indicator is related to activity cycles theory, and it exploits the fact that the cycles shift along the axis determined by the diagonal of the physiological space. Every cycle, in its way, crosses the diagonal of the physiological space in a certain point. The Euclidian distance from the origin to this point gives us a measure of how close the cycles get to the ideal state, and hence, how close to the goal the robot is. Our last indicator will thus be the average of Euclidean distances of the crossing points during the robot lifetime (Figure 3, right).

²The fourth viability indicator we have defined, Risk of Death, will be used to produce hormonal feedback in the experiments reported in Section 6.

5 Preliminary Experiments on the CTRP

In Section 3.2 we saw that the interaction between robots in our CTRP will produce at least two new forms of environmental complexity; in this preliminary experiments we show how they affect the performance of our robots. First, we report new experiments with two robots performing the CTRP, and then we compare the results with those obtained by a single robot in a static TRP, reported in a previous study (Avila-García *et al.* 2003). We discuss the results in terms of different performance indicators (described in the previous section) and behavioral phenomena. The arena and robot configuration used in these experiments are the same as those used in the static TRP (Section 3.1).

5.1 Method

We have tested the robots in a total of 16 runs of 1600 steps each. Each step represents a loop of the action selection mechanism, taking $260ms$ ³. This means that each run lasts about 7 minutes. For each run we initialized randomly physiological variables (to values within their viability range) and location of resources in the environment. The parameter α in the motivational intensity equation (Equation 1) is constant ($\alpha = 1$). We will compare our previous TRP’s results with those obtained by the robot that performs worse in every run of the CTRP. Therefore, when we speak of results obtained in the CTRP, we mean the results obtained by the robot that first died in every run.

5.2 Results and Discussion

Figure 4 shows the comparison of results obtained in the static two-resource problem (TRP) with those of the dynamic competitive two-resource problem (CTRTP). In terms of Overall Comfort, Physiological Balance, and Discomfort at the Crossing point, performance within the CTRP is worse than within the TRP. This difference is statistically highly significant, and it shows that the two new forms of environmental complexity (see Section 3.2) highly affect the viability and cycles of the robots. In terms of Life Span, the results are not statistically significant; however, there is a tendency indicating that the robots live less when there are competitors.

To understand the results, let us analyze the new forms of environmental complexity (dynamism) introduced by the interaction between robots. Firstly, the availability and accessibility of resources causes unpredictability in appetitive phases—that is, when the robot starts to look for a resource the time that it spends to find it is more variable. This also causes that the robot, in the search process to find a specific resource, gets more resources of the other type, falling consecutively in an opportunism that eventually could drive it to death (over-opportunism).

In Figure 5 (left) we can see how the robot, looking for the resource c_{food} (to satisfy $m_{fatigue}$) finds consecutively

³Lego Mindstorms robots use a 16MHz microcontroller

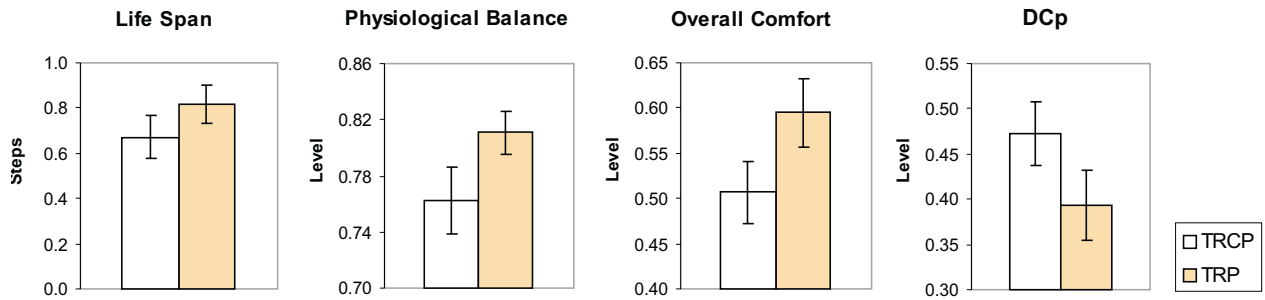


Figure 4: Average performance in terms of our four indicators. Standard error of the mean is shown within each bar.

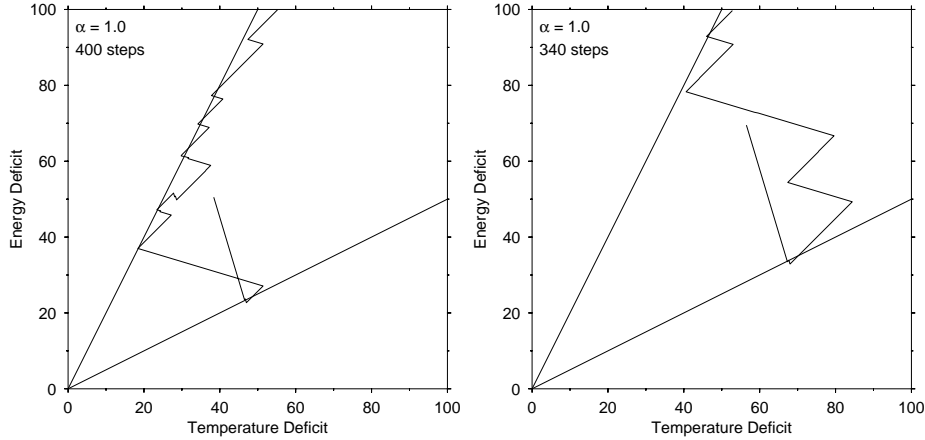


Figure 5: Left: During an appetitive phase looking for c_{food} to satisfy $m_{fatigue}$, Tauron finds consecutively several sources of c_{heat} . This makes it incur in opportunism in such a repetitive way that it finally dies (over-opportunism). Right: While consuming c_{heat} , Tauron is pushed out of the resource by the competitor, stopping its ongoing consummatory activity. Later, it gets again another c_{heat} source to continue reducing the Temperature deficit. It finalizes the activity cycles performing over-opportunism over c_{heat} .

the resources of the other type, c_{heat} , and this makes it incur in opportunism; however, it does so in such a way that the robot dies finally due to the execution of the sequence of opportunistic activities. Since the robot has no mechanism that allows it to detect when opportunism will drive it to death, it often performs it without any coherence.

The other source of complexity is the fact that consummatory activities can be suddenly interrupted. Figure 5 (right) reflects one of those situations; the first time that the robot is consuming c_{heat} (decreasing the level of Temperature deficit) it is interrupted by the other robot and has to look for the same resource again to carry on. The cycle of activities ends up with an opportunism that drives the robot to death.

The effect that both forms of dynamism have on the robot are: (a) the robot shows more bizarre activity cycles than in the static TRP; and (b) they reduce the viability of the cycles in every sense—distance to the origin, amplitude, etc. It seems now interesting to introduce some sort of mechanism that tackles those complexities to some extent, or that helps the robot to use them to compete advantageously. This is the purpose of our next set of experiments.

6 Experiments using Hormonal Feedback

To tackle the problems introduced by the new forms of dynamism described in the previous section, we investigated a mechanism to modulate action selection. As we saw in Section 4, the parameter α in the equation computing motivational intensity (Equation 1) determines the relevance given to external stimuli and also influences viability as reflected by the shape and position of the activity cycles in the physiological space (Avila-García *et al.* 2003). For this reason, we decided to act on (modulate) the perceptual level of the action selection mechanism in order to neutralize the problems produced by the introduction of dynamism. The idea that some phases of the action selection process can be performed at the sensory level is not new—see (Toates 1986, McFarland & Sibly 1975) for some references. In this paper we go beyond this idea and propose that, by acting at that level, we can modulate an action selection architecture to adapt it to internal and/or external environmental circumstances.

Drawing on the architecture described in (Cañamero 1997) regarding hormonal (“emotional”) modulation of action selection, we have implemented a “synthetic hormone” to mod-

ulate perception. Although in the original architecture different types of emotion-released hormones affected many different aspects of the architecture (e.g., synthetic physiology, arousal, attention, perception of the internal and external environment, etc), in this paper we only explore modulatory effects on the perception of external stimuli.

We therefore conducted new experiments in the CTRP scenario to test a modified version of the previous architecture upgraded with this hormonal mechanism. The robots have suffered minimum modifications with respect to the previous setting (see Section 3.1). The robot with the hormonal system, Uruk, is equipped with an extra light sensor to detect the presence of the competitor, a Tauron type robot. Likewise, Tauron is equipped with a sphere at its top that emits infrared light in all directions. Since Uruk’s extra light sensor is aimed forwards, Uruk must be facing Tauron in order to detect the light coming from its infrared ball.

We have performed the same number of runs as in the preliminary experiments, but this time lasting 1200 steps, i.e., each run lasts about 5 minutes.

6.1 Hormonal mechanism

Neurohormones interact with the nervous system to modulate behavioral output (Kravitz 1988), acting at different levels of the nervous system and evoking a spectrum of different responses on different target neurons. Based on these ideas, we have designed a “hormone-like” mechanism that modifies the perception of both incentive cues (c_{heat} and c_{food}) by acting on the parameter α of the motivational intensity equation (Equation 1). The same hormone also modifies the perception of the stimulus Obstacle that unconditionally fires the reflex behavior b_{avoid} . Its function is thus twofold:

- (a) To avoid falling under excessive opportunism and persistence in situations where the Risk of Death (see Section 4.5) is high, therefore neutralizing the problems due to dynamism regarding accessibility and availability of resources. This is achieved by controlling the parameter α : The smaller α , the smaller the positive feedback (see Section 4.4). Therefore, variations in the hormonal level will modify the levels of opportunism and persistence.
- (b) To potentiate the competition skills of the robot by enhancing its capacity to push the other robot out of the resources. This means, first, modifying accessibility and availability of resources—it will be easier to obtain a resource occupied by the competitor. Secondly, the complexity of the environment for the other robot increases. This is obtained by the cancellation of the perception of the stimulus Obstacle (bumpers), which produces the cancellation of the reflex behavior b_{avoid} . Therefore, hormonal levels determine whether the robot avoids obstacles or pushes them.

The question is now how to control that double functionality with only one hormone, and how to compute the level

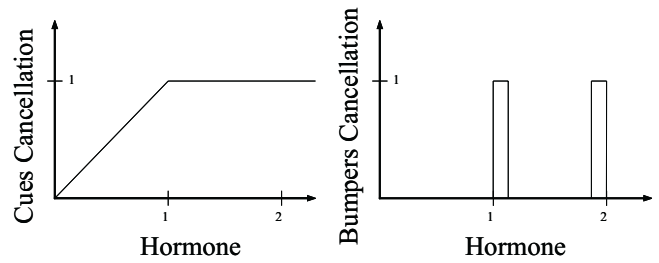


Figure 6: Cancellation functions of incentive cues (left) and bumpers (right) with respect to the hormonal level.

of such hormone. We do so by combining both external and internal “appraisals”⁴ of the risk of death. The internal assessment of Risk of Death (RoD) is given by the distance of the physiological state to the lethal boundaries, as follows:

1. If $d_{max} > 0.5$, then $RoD = d_{max} - 0.5/0.5$
2. Else, $RoD = 0$

where d_{max} is the maximum of both physiological deficits. This means that RoD is a number normalized between [0,1] that gives us a measure of the distance to the lethal boundary when the maximum deficit is higher than half that boundary, i.e., $d_i = 1$.

The external risk of death is given by the perception of the competitor ($0 \leq Competitor \leq 1$). The hormonal level will be an additive function of internal Risk of Death (RoD) and the perception of the competitor:

$$Hormone = RoD + Competitor, \quad (5)$$

therefore $0 \leq Hormone \leq 2$.

Finally, we have to define the relation between the hormonal level and the cancellation of the perception of incentive cues and Obstacles (Figure 6). In order to achieve the first of the functionalities (a), the cancellation of α is directly proportional to the increment of the hormonal level. Therefore, when the RoD rises, α is reduced.

The second functionality (b) is obtained by cancelling the perception of the stimulus Obstacle when the stimulus Competitor is in front of the robot. There are two conditions that must be fulfilled to make a coherent pushing of the other robot. First, the robot must avoid getting engaged in fights when it has RoD. Second, it must only bump blindly into the other robot, not against the walls of the arena. To produce that effect the cancellation of the bumpers is at hormonal levels $Hormone \simeq 1$ and $Hormone \simeq 2$.

6.2 Results and Discussion

Figure 7 shows the comparison of results obtained by both robots, Uruk and Tauron, in the CTRP. In terms of Overall Comfort and Discomfort at the Crossing point results are

⁴Although the term “appraisal” is often used to denote conscious evaluation, appraisal can also be unconscious and very automatic process—it is in this latter sense that we use it here.

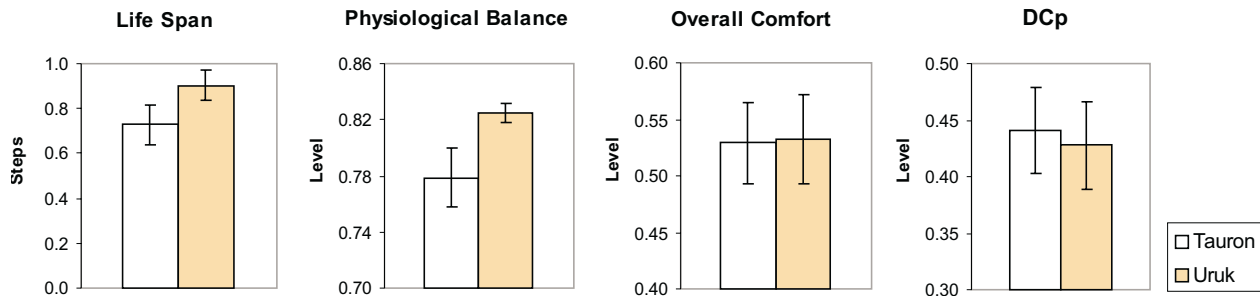


Figure 7: Average performance in term of our four indicators. Standard error of the mean is shown within each bar.

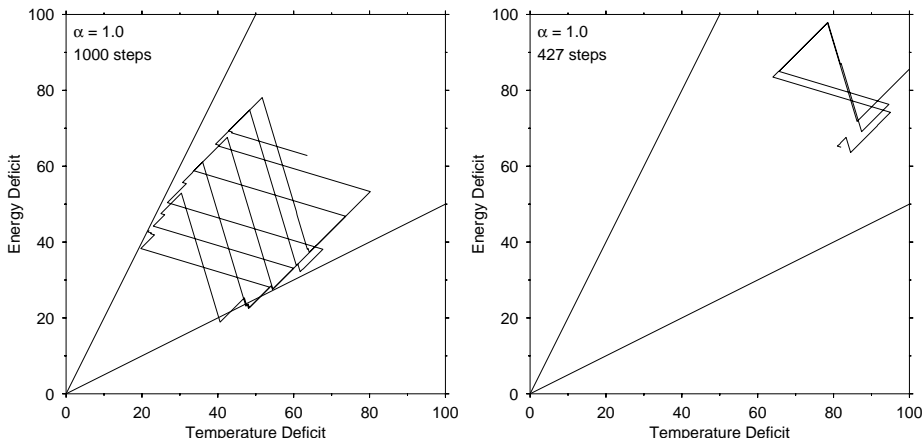


Figure 8: Activity cycles depicted by Uruk. Due to the hormonal system, satiation boundaries are reduced as the Risk of Death (RoD) increments. Both graphs depict satiation boundaries when there is not hormonal feedback, that is, with $\alpha = 1$. Left: Activity cycles with same amplitude due to the reduction of satiation boundaries—i.e., reduction of α . Right: Activity cycles evolving very close to the lethal boundaries through the previous reduction of α . Compare these cycles with those produced by Tauron in previous experiments (Figure 5).

alike. However, regarding Life Span and Physiological Balance the differences are statistically highly significant; Uruk outperforms Tauron.

Let us analyze the dynamics of the activity cycles in order to understand those quantitative results. Figure 8 (left) depicts cycles performed by Uruk in one of the runs. We can see how cycles keep the same amplitude when physiological deficits are higher than 50 ($d_{max} > 0.5$). This is because at that point the RoD starts to rise the hormonal level, which decreases the value of α . This results in a reduction of the satiation boundaries, and hence the persistence. In Figure 8 (right) we can see the drastic reduction of persistence at high levels of RoD. This process also counteracts the tendency to perform opportunism under risk of death. This is the main reason for such good results in terms of Physiological Balance.

In terms of Life span, the results are also the demonstration that the second functionality of the hormonal system improves the performance of the robot. This functionality leads it to push the other robot out of the resources, hence improving its competition skills.

This double functionality of the hormonal system gives rise to some emergent behaviors. The first one is stopping opportunism and persistence when Uruk is consuming a resource

and it detects its competitor approaching. This could be interpreted by an external observer as abandonment of a situation (waiting for the other robot at the resource) that is disadvantageous to compete. Instead, the robot will leave the resource and go straightforward towards the competitor until it reaches it; at that moment, if there is some level of RoD, the bumpers of Uruk will not be cancelled and it will avoid Tauron—showing a behavior that an observer could interpret as “fear” after evaluating the competitor. On the contrary, if there is not RoD, the hormonal system will cancel the bumpers and Uruk will push Tauron unconditionally—as if it showed some sort of “aggression” against it. If we study the whole picture as external observers, the previous behavioral phenomena could well be interpreted as some sort of “protection of resources”.

7 Conclusion

Neurohormones have the ability of rewiring the same anatomically defined neural circuit to build the appropriate network for the behavior needed at any moment. Inspired in this functionality, we have investigated the role of an “hormone-like” mechanism to produce adaptive behavior in an action selection architecture. We have carried this study in two steps.

In a first step, we have discussed the extension of a robotic implementation of a static two-resource problem (TRP) into a dynamic competitive two-resource problem (CTRP). In this extension, with only the introduction of a second robot—a competitor—new forms of environmental complexity are generated due to the interaction between robots. These forms of complexity effectively reduce the viability of the robots. Using activity cycles analysis, we have also shown that the action selection mechanism used within the TRP presented incoherences within the CTRP, performing pathological opportunism and persistence.

In a second step, we have proposed a first rudimentary “hormonal mechanism” to tackle these new forms of environmental complexity. Hormonal feedback modulates the action selection mechanism—the interconnections between the sensory elements and other elements of the ASM—adapting its output to different environmental conditions. This modulation gives rise to adaptive behavioral patterns to the changing environmental conditions, as well as interesting emergent behavioral phenomena. This mechanism can be seen in relation to autopoietic systems (Maturana & Varela 1980), which change their internal structure in response to environmental circumstances. A modulatory hormonal system improves the *structural coupling* between an agent and its environment in a way that is different from other mechanism such as learning or evolution, for which “past solutions” are “overwritten” by new ones. Hormonal modulation extracts different functionalities from the same architectural “skeleton” to produce adaptive behavior to any of the environmental conditions of the agent’s niche.

Acknowledgments

Orlando Avila-García is supported by a research scholarship of the University of Hertfordshire. We are grateful to René te Boekhorst for his helpful comments on an earlier version of this paper.

References

- Ashby, W.R. (1952). *Design for a Brain*. Chapman & Hall.
- 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*, pp. 86–93. 14–16 August 2002, Bristol HP Labs, UK.
- Avila-García, O., Cañamero, L., and Boekhorst, R. (2003). Analyzing the Performance of “Winner-Take-All” and “Voting-Base” Action Selection Policies within the Two-Resource Problem. In *Proc. Seventh European Conference in Artificial Life (ECAL03)*, pp. 733–742. Springer-Verlag.
- Blumberg, B. (1997). Old Tricks, New Dogs: Ethology and Interactive Creatures. *PhD. Thesis*, MIT Media Laboratory, Learning and Common Sense Section.
- Brooks, R. A. & Viola, P.A. (1990). Network Based Autonomous Robot Motor Control: from Hormones to Learning. In R. Eckmiller (Editor), *Advanced Neural Computers*, pp. 341–348. Elsevier Science Publishers.
- Cañamero, L.D. (1997). Modeling Motivations and Emotions as a Basis for Intelligent Behavior. In W.L. Johnson, ed., *Proc. First Intl. Conf. Autonomous Agents*, 148–155. ACM Press.
- Girard, B., Cuzin, V., Guillot, A., Gurney, K.N., 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*. MIT Press.
- Grand, S., Cliff, D., Malhotra, A. (1997). Creatures: Artificial Life Autonomous Software Agents for Home Entertainment. *Proc. First Intl. Conf. on Autonomous Agents*, 22–29. ACM Press.
- Harris-Warrick, R.M., Marder, E., Selverston, A.I., Moulins, M., eds. (1992). *Dynamic Biological Networks*. The MIT Press.
- Hinde, R.A. (1953). Appetitive Behaviour, Consummatory Act, and the Hierarchical Organization of Behaviour –with special reference to the great tit (*Parus Major*). *Behaviour*, 5, 189–224.
- Kravitz, E.A. (1988). Hormonal Control of Behavior: Amines and the Biasing of Behavioral Output in Lobsters. *Science* 241, September 30, 1175-1781.
- 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.
- Maturana, H.R., Varela F.J. (1980). Autopoiesis and Cognition. *The realization of the Living*. Dordrecht, Holland: D. Reidel.
- McFarland, D. (1974). Experimental Investigation of Motivational State. In *Motivational Control Systems Analysis*, D. McFarland, ed., 251–282. London: Academic Press.
- McFarland, D. (1976). Form and function in the temporal organization of behavior. *Growing Points in Ethology*, Bateson P., Hinde R. ed. Cambridge University Press.
- McFarland, D. and Sibly, R. (1975). The behavioural final common path. *Phil. Trans. Royal Society (Series B)*, 270, 265-293.
- McFarland, D. and Spier, E. (1997). Basic Cycles, Utility and Opportunism in Self-Sufficient Robots. *Robotics and Autonomous Systems*, 20, 179-190.
- Meyer, J.A. (1995) Artificial Life and the Animat Approach to Artificial Intelligence. In M. Boden, ed., *Artificial Intelligence*. Academic Press.
- Robbins, T.W. and Everitt, B. J. (1999). Motivation and Reward. In M. J. Zigmond, F. E. Bloom, S. C. Landis, J. L. Roberts, L. R. Squire, ed., *Fundamental Neuroscience*, 1245–1260. Academic Press.
- Sibly, R.M. and McFarland, D. (1974). A State-Space Approach to Motivation. In D. McFarland, ed., *Motivational Control Systems Analysis*, 213–250. London: Academic Press.
- Slater, P.J.D. (1999). Motivation. In *Essentials of Animal Behaviour*, 57–83. Cambridge University Press.
- Spier, E. and McFarland, D. (1997). Possibly Optimal Decision Making under Self-Sufficiency and Autonomy. *J. Theor. Biol.*, 189, 317-331.
- Toates, F. (1986). *Motivational Systems*. Cambridge Univ. Press.
- Tyrrell, T. (1993). Computational Mechanism for Action Selection. *PhD. Thesis*, Centre for Cognitive Science, Univ. of Edinburgh.