

# Learning Affordances of Consummatory Behaviours: Motivation-Driven Adaptive Perception

Ignasi Cos<sup>†‡\*</sup>

Lola Cañamero<sup>‡</sup>

Gillian M. Hayes<sup>†</sup>

<sup>†</sup> *Institute of Perception, Action and Behaviour,  
School of Informatics, University of Edinburgh,  
Informatics Forum, 10 Crichton Street,  
Edinburgh, EH8 9AB, Scotland, UK.*

<sup>‡</sup> *Adaptive Systems Research Group,  
School of Computer Science, University of Hertfordshire,  
Hatfield, Herts, AL10 9AB, UK.*

---

\*Current Address: Departement de Physiologie, Université de Montréal, 2960 Chemin de la Tour, Montreal, QC H3T 1J4, Canada. Email: ignasi.cos-aguilera at umontreal.ca

### **Abstract**

This paper introduces a formalization of the dynamics between sensorimotor interaction and homeostasis, integrated in a single architecture to learn object affordances of consummatory behaviours. We also describe the principles necessary to learn grounded knowledge in the context of an agent and its surrounding environment, which we use to investigate the constraints imposed by the agent's internal dynamics and the environment. This is tested with an embodied, situated robot, in a simulated environment, yielding results that support this formalization. Furthermore, we show that this methodology allows learned affordances to be dynamically redefined, depending on object similarity, resource availability and on the rhythms of the agent's internal physiology. For example, if a resource becomes increasingly scarce, the value assigned by the agent to its related effect increases accordingly, encouraging a more active behavioural strategy to maintain physiological stability. Experimental results also suggest that a combination of motivation-driven and affordance learning in a single architecture should simplify its overall complexity while increasing its adaptivity.

**Keywords:** Sensorimotor Interaction, Homeostasis, Affordance, Motivation, Reinforcement, Robotics.

# 1 Introduction

The goal of this paper is to formalise and test some of the principles required to learn object affordances, focusing on the relationship between sensorimotor interaction and the agent's internal physiological dynamics (the homeostatic loop). This work is mainly inspired by Gibson's theory of affordances (Gibson, 1966), some of which ideas we present next. Although the notion of affordance has been controversial, the appeal of the ecological formulation to design the sensorimotor loop for artificial agents has certainly not been affected (Morineau and Frénod, 2009; Chemero and Turvey, 2007; Sahin et al., 2007; Cos-Aguilera et al., 2004, 2003a). Gibson's affordances introduced the notion of directed perception, therefore suggesting the possibility of viewing an environment as a set of potentialities for action. Most of the theory's appeal for roboticists lies precisely in this *directedness* (Gibson, 1966), i.e., the potentialities for action arise/disappear as elements offering the potentiality of executing a behaviour from the animal's repertoire enter/leave its perception range. The simplicity of this formalisation has been accepted by some in the autonomous robotics community, who have sought inspiration in ethological studies (Spier and McFarland, 1996; McFarland, 1993). In support to this formalisation, recent neuro-physiological recordings have shown that the cortical fronto-parietal loops can simultaneously represent several potential options and that these compete even after a decision has been made and its related motor command has been initiated (Song and Nakayama, 2007; Cisek and Kalaska, 2005).

Therefore, a vast majority of studies on adaptation have focused on behaviour planning and action selection as the main processes underlying adaptive behaviour, yielding models of autonomous robots (Konidaris and Barto, 2006), cognitive science and physiology (Cisek, 2007), reproducing different aspects and views of behavioural planning and behaviour selection (Khamassi et al., 2006). In a complementary fashion, this paper introduces a formalisation of the fundamentals enabling adaptation to the environment by associating the physiological effect exerted by behavioural interactions with certain elements in the environment on the homeostatic processes, to the perception of those same elements. In other words, if the object affords to the agent the behaviour currently being executed, interacting with it may exert some compensatory effect on the agent's physiology. The resulting fluctuation on the agent's internal physiological variables may then be used to gain grounded behavioural knowledge about that particular element of the environment. The continuous nature of this learning process and the ability to quantify differences and similarities between the morphology of different objects is what confers perception some of its adaptivity. Therefore, fruits that have never been seen before may be recognisable as edible in anticipation to a first interaction by similarity. Furthermore, changes on the physiological effect exerted by a particular behaviour in the presence of a given element of the environment, due to changes on the environment (e.g., seasonal changes) or to the rhythms of the animal internal physiological processes, could be rapidly accounted for. For example, if food is no longer so available, the physiological effect resulting from eating may remain unchanged. In contrast, the value of this effect may increase significantly, since there is a mismatch between the internal pace of energy consumption and the previous rate of food intake. This dynamic relates grounded knowledge from the environment to potential changes of behaviour and alters the natural habits or behaviours of the individual maintaining the internal physiological variables within their viability zone (Ashby, 1965) in a manner more efficient than if perception were a fixed input process only.

This view of affordances is inspired by perception as a process resulting from developmental and genetic adaptation, attuned by the mutuality between an animal and its surrounding environment, where the animal lives and interacts. This reflects on the elements of the environment which have become meaningful to animals of different species, always dependent on their abilities and morphology. For example, while a tree may afford us shadow and its fruits some food, it may also offer different affordances to an animal with different morphology. Furthermore, the average conditions on an environment, e.g., extreme drought or humidity shape the internal physiology and therefore the general behavioural organisation with which animals face the environment. The physiology of a species has developed in a given environment, the same environment where that species belongs. Altering this may severely diminish their chances of survival. Reformulating this in a more mechanistic fashion, only if the animal possesses grounded knowledge about the environment will the agent have the necessary background to manage its behaviours efficiently to satisfy its internal deficits. As formalised by Ashby, physiological stability is a necessary condition for survival (Ashby, 1965).

Additionally, the value of a potential resource depends on the generalisation an animal can make of the affordances offered by similar resources. Therefore, animals use metrics of similarity between

objects allowing them to extend their learned affordances from the original object they interacted and learned with to any other similar object. Furthermore, although this mechanism may not be universal across species, animals are also endowed with the additional ability to interpret the physiological effect exerted by a given behaviour: the often referred to as the sense of valency (Ackley and Littman, 1991). This may be explained as the ability to assess the physiological effect during and after interaction with elements from the environment, and to modulate the resulting reward. This assessment may then be used to promote or discourage the behavioural interaction with that particular object, depending on whether the interaction is rewarding or not. Therefore, reward may be internally modulated by the animal if required (Damoulas et al., 2005). For example, it may be possible to revert the rewarding effect of eating if the animal is sated, consequently discouraging its exercise.

We have formalised these principles in a agent architecture, embedded in a situated autonomous robot, tested in simulated environment. Since we have focused on the interaction between internal physiology and sensorimotor interaction, the agent architecture we present here encompasses three related processes: the internal *homeostatic processes*, driving the animal’s internal physiology, the *sensorimotor loop* relating the agent and its environment, and the *internal sense of valency* the agent is endowed with, assessing the physiological effect of the execution of a certain behaviour involving certain elements of the environment. This formulation introduces the natural processes involved in affordance learning, in an abstract recreation of the process of natural, developmental adaptation to the environment.

## 2 Background and Related Research

The term affordance has been discussed and interpreted in several ways in the psychological literature. In its original, Gibsonian conception, the affordance was conceived as an evolution of the notion of valence and of the notion of *Aufforderungscharakter* —cf. Marrow (1969). However, unlike both of them, the affordance is grounded on the principle of ecology, in the context of the interaction of an animal and its surrounding environment, therefore cutting across the phenomenological and the physical dimensions, based on the perception-action loop any animal is endowed with. Citing Gibson himself: “*The object offers what it offers, because it is what it is. Duality does not exist, the physical object (the only one), possesses meaning and value to begin with*” (Gibson, 1966). Furthermore, this value is independent of the animal’s internal state, e.g., a piece of bread affords eating to me, whether I am hungry or not. Therefore, the Gibsonian affordance is a notion defined in the context of the interacting dynamics of an animal and its surrounding environment.

In its subsequent definitions, with the exception of Turvey (1992), who considers affordances as properties of the environment the animal interacts with, most definitions of affordances in ecological psychology have considered an animal and its environment as the context wherein the notion of affordance is meaningful (Chemero and Turvey, 2003; Stoffregen, 2003). A review of the definitions of affordance, together with a novel interpretation of the same notion, can be found in Sahin et al. (2007). The main goal of this formalisation is the development of an affordance-based framework for robotic architectures. Towards this end, Sahin et al. (2007) integrates previous definitions in a single formalisation of affordances, composed of three perspectives: the agent’s perspective, the environmental perspective and the observer’s perspective. In a way, their merit is the integration of previous definitions of affordances in a single hierarchy. The agent’s perspective focuses on the affordances between an animal/agent and its surrounding environment. Because of this same reason, it has been probably utilized by most studies using affordances in mobile robotics (Doğar et al., 2008; Cos-Aguilera et al., 2004). Unlike this, the observer’s perspective is as a step towards establishing a common semantic of grounded knowledge, whereon symbols could be added (Harnad, 1990) and communication across several individuals may be established. After the same principles, Nehaniv and Dautenhahn (1998) introduced the possibility of using affordances as the unit of perception to be used to facilitate imitative processes between agents of different morphology.

Despite the numerous interpretations and the formalisation of Sahin et al. (2007), there are still fundamental elements of disagreement on the very essence of the notion of affordance. In particular, the use of internal representations disputes the compliance of these interpretations with Gibson’s original theory (Gibson, 1986). Chemero and Turvey (2007) classified affordance interpretations into two main streams: representationalists and Gibsonians. The view of representationalists could be summarised as the necessity of having at least partial representations of objects in the environment, either representing objects from the environment or “functional representations” relating sensory patterns and behavioural responses. Conversely, a Gibsonian view suggests a direct relationship

between sensory flow and behavioural responses, governed by elementary principles, often referred to as laws of control (Duchon and Warren, 1994), e.g., the balance strategy or the avoid-closest used by Duchon et al. (1995) for navigation. Although it has been suggested that robotic architectures should try to extend from the Gibsonian perspective (Chemero and Turvey, 2007), the fact is that most elaborations of affordances have used some sort of representation. From our modest perspective, we cannot argue in favour of a single view on affordances, since one of the aspects not understood about affordances lays precisely on the scalability from basic rules governing behaviour using laws of control to navigate the environment, more typical from insects or animals in the absence of a centralised decision center, to the wide range of cognitive affordances exhibited by higher mammals. Some of the principles in the context of decision making are introduced by Konidaris and Hayes (2004).

From our perspective, the notion of affordance is only meaningful in the context of an agent and its environment, viewed as two parts of a single unit, characterised by the agent's internal dynamics, the environment's natural rhythms and their interacting dynamics. This is the context wherein we focus our research on the ecological procedure of learning object affordances. In particular, we have studied the integration between the homeostatic processes and the perception-action loop, based on the effect exerted by behavioural execution. In fact, learning affordances requires the agent to be sensitive to that particular effect, either because a behaviour modifies the environment in some way or because it reflects on the agent's internal state. In other words, the effect of performing a behaviour provokes a change on the environment, e.g., when we hit a soccer ball, it moves opposite to the kick. At a higher description level, when we perform a consummatory behaviour such as eating an apple, this has a reflect on the internal level of sugar in blood, which rises as a consequence of its digestion. This same principle was first proposed by Spier and McFarland (1996), in the context of learning without cognition, which we relate to learning affordances and grounded knowledge about the environment. We argue that this method establishes the basics of the semantic relationship with the environment, integrating an agent's internal physiology with the perception-action loop in a single unit. This is further analysed in the next section.

Most approaches are still not usable to build complex robots, indicating that affordance based robotics may still require significant research effort to scale up to complex tasks in a satisfactory manner. Nonetheless, it is fair to mention that other fields such as cognitive psychology have developed working models to learn affordances, inspired after similar principles. For example, the context created by Cooper and Glasspool (2002) explicitly built a model using affordances as attentional mechanisms to filter significant information from the environment. Again, their definition is not entirely Gibsonian, since it is based on symbolic features perceived from objects, defined and engineered by the designer.

In this context the contribution of neurophysiology has significantly contributed to postulate a potential cortical representation of multiple courses of action on the fronto-parietal loop in the cortex of higher mammals (Cisek and Kalaska, 2005). Furthermore, Fagg and Arbib (1998) introduced a model for the extraction of affordances from the sensory input in the context of navigation and manipulation tasks, using affordances as the link between perception and the potential action offered by every situation in the environment. More recently, Cisek (2007) introduced a model of the fronto-parietal cortical loops, suggesting the simultaneous representation of several potential actions (or affordances) to be executed next.

Also, the notion of affordance has been used in related fields such as Human-Computer Interaction (HCI) (St. Amant, 1999) emphasizing the situated nature of human behaviour. In this context, designers view affordances as the perceived properties of a software artefact, which indicates how it can be used (Baecker et al., 1995). Amant also defined them as: *"a mechanism that allows or facilitates the execution of some operator. More specifically, an affordance preserves the conditions necessary for the successful completion of the operator by reducing the execution cost of other appropriate operators or by increasing the execution cost of inappropriate operators."* Together with Sahin et al. (2007), St. Amant (1999) may be one of the most complete definitions of affordance, including four separate dimensions for the concept: first, Gibsonian affordances (relationships or properties of relationships), second our perception of several properties, the surfaces, distances, areas, textures, relationships between parts (on this fact relies the design of ecological HCI), third the mental interpretation derived from the perceptions, fourth the act of performing an action itself.

### 3 Motivation-Based Affordances

The fundamentals of the relationship between an individual (natural or artificial) and its environment are formalised by the ecological principle (Gibson, 1966). This principle states the mutuality between the animal and the niche where it lives and evolves, since both the animal and its surrounding environment are the result of their mutual interaction and should be considered as a single entity. Therefore, when later on in the text we mention the notion of object affordance, we will refer to the behavioural relationship between the perception of an object and the potentiality of executing a behaviour, only meaningful in the context of the agent and its surrounding environment, the only context wherein the affordance exists. The next paragraphs of this section describe the elements playing a significant role in learning those affordances related to consummatory behaviours: homeostasis, drives, motivation, reinforcement, reward and physiological stability.

*Homeostasis* has been defined as “The condition of a system when it is able to maintain its essential variables within limits acceptable to its own structure in the face of unexpected disturbances” (Cannon, 1929). This notion has been often referred to in the psychological literature, e.g., Maturana and Varela’s process of dynamic self-regulation (Maturana and Varela, 1980). In agreement with these definitions, we view homeostasis as the set of bodily self-regulation processes, which could be classified into two categories: those *self-regulating* the body’s internal resources independently of external stimuli, e.g., the digestive process after food intake, and those relating to the *external world*, either by influencing the animal’s behaviour or by the influence interacting with the environment on the internal physiological dynamics. Therefore, homeostasis lies at the very heart of the processes driving behaviour.

*Drives and Stability.* Homeostasis is the context used by Hull (1943) to frame the notion of drive, capturing the status of deficit or excess of its related homeostatic variables, expressing urges for compensation, either by internal self-regulation or by external interaction. Therefore, drives are internal behaviour initiators, as opposed to external stimuli. Although homeostasis has been traditionally classified as a purely internal process (Cannon, 1929), there is little doubt that both processes of internal self-regulation and external compensation can be viewed as two interacting elements of a single regulatory system. In fact, an animal’s behavioural repertoire directly relates its urges for compensation to the agent’s environment, each of which requires a certain set of elements in the environment to have some effect on the internal variables, hence to complete the regulatory cycle. The capacity of regulation, external or internal, relates directly the organism stability and to its ability to survive in a changing environment (Ashby, 1965).

*Motivation/Reinforcement.* Motivation was introduced in its different flavours by McDougall (1913); Freud (1940); Tinbergen (1951); Lorenz (1966). All of their definitions “share the idea of a substance, capable of energizing behaviour, held back in a container and subsequently released in action” (Hinde, 1960). According to this, motivation intrinsically relates internal physiological drives to the behaviours, whose execution affects the intensity of their related motivations themselves. In a complementary manner, *reinforcement* affects the future probability of choosing a certain behaviour, depending on the outcome of the interactions, influencing internal physiology. Therefore, one can view motivation (or response instigation) acting in the direction from perception to action and reinforcement from action to perception. Both processes modify future behavioural preferences. This mutual interaction together with the hypothesis of common neural substrate (Bindra, 1969), suggest that both, motivation and reinforcement, may be measured in the same unit. This inter-relation between the homeostasis and the perception-action, or between motivation and reinforcement, was first made explicit by Bindra: “the effects on behaviour produced by reinforcement and motivation arise from a common set of neuro-psychological mechanisms, and the principle of reinforcement is a special case of the more fundamental principle of motivation” (Bindra, 1969). Bindra also suggested a physiological and anatomical substrate relating motivation, behaviour and reinforcement as processes.

*Reward.* Related to this, the common currency problem arises from the necessity to compare different motivational states during the process of action selection (Redgrave et al., 1999; McFarland and Sibly, 1975). We argue that the common currency is reward. On the one hand, this would be consistent with the aforementioned notions of motivation and reinforcement, since reward can be viewed as the strengthening principle of the common neural substrate that reinforces or weakens the relationship between the perception of a certain object and the selected behaviour. Furthermore, decision making could then be conceived as an elaboration of motivational states in terms of expected reward. On the other hand, this definition of common currency is consistent with current neurological

views of hypothetical roles for the basal ganglia (Houk et al., 1995), suggesting reward to be the reinforcer of behavioural tendencies. Learning can be naturally integrated in this context if neuro-modulatory phenomena (Fellous, 2004; Fellous and Suri, 2003; Usher and Davelaar, 2002; Fellous, 2001; Hebb, 1949) are considered in the context of reinforcement comparison algorithms (Sutton and Barto, 1981), the reason being that these algorithms have been hypothesised to drive the learning of behaviour selection in circuits in the basal ganglia (Houk et al., 1995). Hence, nigro-striatal projection and striatal cells in the basal ganglia may be the “common circuitry for motivation and reinforcement” suggested by Bindra’s hypothesis (Bindra, 1969), and could also be directly related to the process of learning affordances.

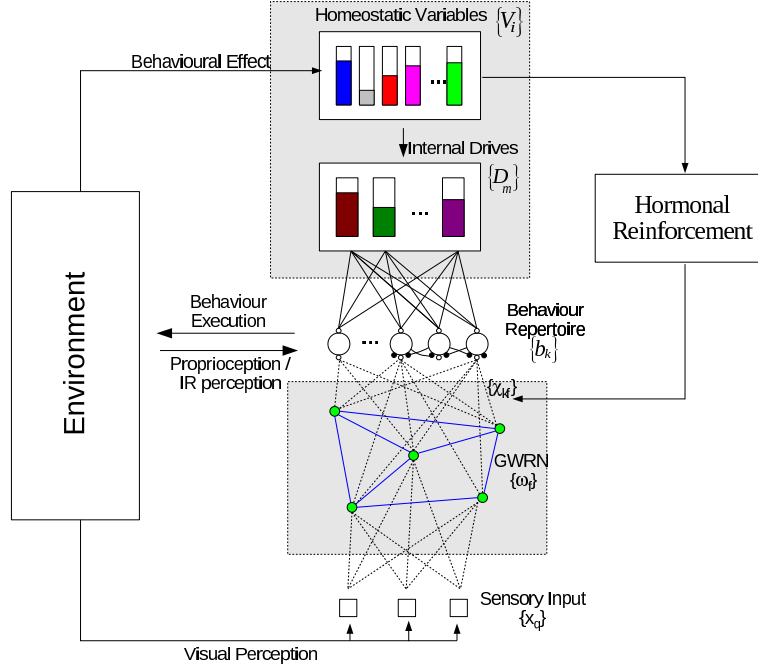


Figure 1: Illustration of the agent’s architecture. The physiological variables and the internal drives are shown in the centre-top of the figure, projecting directly to the agent’s behaviours (centre). The bottom blue network is the GWR network, which clusters visual input from the objects via the robot’s camera. The *Hormonal Reinforcement* box on the right stands for both the hormonal reinforcement used to learn object affordances. The circles in the centre stand for the agent’s behaviours, influenced by the agent’s internal physiology and objects nearby, represented by the most active node of the SOFM.

*Physiological Stability.* The method proposed to learn affordances should provide a stable representation between the perception of certain objects and the salience of a certain behaviour, meaning its potentiality for execution. Once this grounded knowledge has been made available, behavioural responses when certain objects enter perception range should lead to responses demonstrating a better adaptation. In the light of this reasoning, it seems sensible to use behavioural experiments to assess the quality of the learned knowledge. During a first batch of experiments (Cos-Aguilera et al., 2003a), we compared the lifespan (time until one of the internal variables escapes its viability zone) between agents selecting behaviours at random and agents selecting the most salient behaviour afforded by the object nearby. The results suggested that a behavioural pattern that considers the potentialities of action afforded by the objects in the nearby environment yielded a longer lifespan. Lifespan is a behavioural measure that reflects the amount of time the agents can maintain their internal resources within their viability zone. However, it disregards additional behavioural assessments about the overall physiological stability during the life of the agent. A temporal analysis of balance between variables should yield a more accurate description of its performance in terms of *quality of life*. Based on this, we have suggested the use of a physiological metric (see next section) which reflects the learning performance of the affordances of objects in their surrounding environment.

Next we describe an artificial model of internal physiology, which integrates the different elements and requirements described in this section. This model is also the test bed we have used to learn affordances about consummatory behaviours in the experiments described in the following sections.

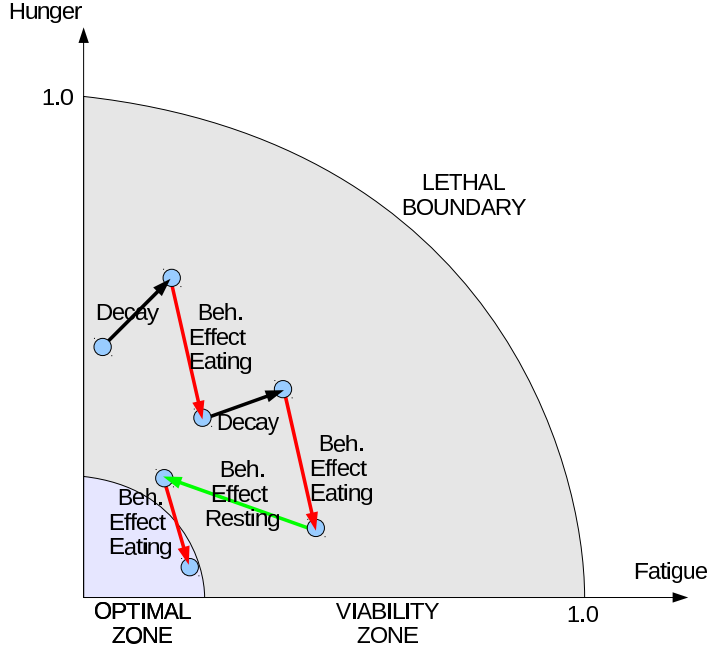


Figure 2: Depiction of a schematic representation of a 2D physiological space. The blue dots indicate the different states of the agent’s internal physiology. As time goes by, the state exhibits a decay, shifting towards the lethal boundary (black transitions). However, as consummatory behaviours are executed in the presence of an object affording that behaviour, the state shifts towards the optimal zone. The optimal zone is the area where the drives are minimal. The viability zone is defined the tolerance of the drives. This is enclosed by the lethal boundary.

## 4 The Model

### 4.1 Overview

Next we present the model implementing the aforementioned principles that relate homeostasis and sensorimotor interaction. The model consists of three main elements (cf. figure 1): an artificial physiology (top part of the figure), which is an abstraction of physiological processes of self-regulation; a modified version of the Growing When Required (GWR) network (bottom part of the figure), aimed at clustering the agent’s sensory input; and the affordance *Learning Module* (reinforcement module), that uses behavioural reinforcement to learn the potentiality of executing a behaviour for a given sensory input. The running time-line of this model may be described as the sum of different concurrent processes: first, the robot interacts with the objects it encounters at random in its environment. Visual perception of the objects is used by the self-organising feature map (GWRN) to build the world map. The closest topological node to the nearby object increases its activity. If the object is under-represented, a new node may be added. In a concurrent fashion, as times goes by the agent’s homeostatic variables decay, thus increasing the agent’s drives. The perception-action loop is closed by executing a behaviour within the agent’s repertoire, depending on the node most active in the GWR and on the internal physiological state. If the execution is successful, one or more of the agent’s internal variables may increase, and the behavioural pattern may be reinforced. Each step of these processes is described in detail in the following sections.



## 4.2 Artificial Physiology

A model of artificial physiology provides the necessary context to study how behavioural feedback is integrated into the agent's internal dynamics, influencing the agent's adaptivity. The model of artificial physiology is shown in the top part of figure 1. It consists of a set of homeostatic, survival-related variables, a set of drives that depend on the internal variables, a repertoire of behaviours, and an arbitration mechanism to resolve conflicts among competing motivations/behaviours and to choose the appropriate behaviour to satisfy internal demands. The structure of this physiological model keeps a close similarity to the one used in previous studies (Cos-Aguilera et al., 2003a,b, 2004).

The controlled *homeostatic variables* vary due to internal bodily dynamics and to the interactions of the agent within its environment. We model the derivative of every homeostatic variable  $\dot{V}_i$ , as shown by equation 1. The first term of the equation captures the variable typical decay (at rate  $\tau_i$ ), and the second the increase by  $\alpha(b_{ki})_j$  due to behavioural interactions.  $\delta(t)$  is the delta function, associated to the occurrence of a successful behavioural execution, yielding some physiological effect on the homeostatic variable.

$$\dot{V}_i = -V_i/\tau_i + \sum_k \alpha(b_{ki}) \sum_{j(k)} [u(t - t_{j(k)}) \exp(t - t_{j(k)}) + \delta(t - t_j(k))] \quad (1)$$

Equation 2 shows the analytical solution to equation 1, where  $V_i(t)$  is the homeostatic variable  $i$  and  $\tau_i$  the constant describing its natural decay.  $\alpha(b_{ki})$  indicates the amount by which the homeostatic variable increases due to successful execution of behaviour  $b_k$  (by interacting with an appropriate object) at times  $t_j(k)$ .

$$V_i(t) = V_0 e^{-t/\tau_i} + \tau_i \sum_k \alpha(b_{ki}) \sum_{j(k)} u(t - t_{j(k)}) \exp(t - t_{j(k)}) \quad (2)$$

The homeostatic variables are abstractions representing the level of internal resources that the agent needs in order to survive. Previous studies have been constrained to the two resource-problem (Avila-García and Cañamero, 2002; Spier and McFarland, 1997), since this suffices to study some of the foundations of action selection or behaviour arbitration. However, since the main topic of this paper is based on the effect of behavioural effect on the homeostatic variables, we have considered appropriate to add an additional variable. This should introduce some desirable variability to the system while permitting an approach simple enough to study this phenomenon. We have used three basic homeostatic variables: nutrition, stamina and restlessness. Nutrition is an abstraction for a number of elements controlling food intake (e.g., glucose level in blood). As time goes by, its level decreases, as described by equation 1, if food/rest is not periodically consumed (the delta term indicates the moment at which the effect of successful behaviour execution occurs). Stamina stands for the level of energy of the agent, which decreases over time if the agent does not regularly rest. Restlessness is an abstraction of the level of inactivity, which increases over time. Each of these variables has an optimal set point, and a range of values —their viability zone (Ashby, 1965) for the agent to remain viable in its environment (the equivalent to remaining alive in the natural context). If the values over/underflow the upper/lower boundaries, the agent dies. Furthermore, each homeostatic variable exhibits a status of normality (when its current value approaches its ideal value or set point), deficit or excess. The homeostatic variables may express their status of deficit or excess via a set of *drives* (Hull, 1943). Drives are abstractions denoting urges to action based on bodily needs related to self-sufficiency and survival. They monitor the divergence of each homeostatic variables with respect to their set point, and initiate a process of correction. Equation 3 defines the general relationship between a drive and its related homeostatic variable/s.

$$D_m(t) = \sum_l a_{lm} (V_{op_l} - V_l(t)) + \sum_l b_{lm} \dot{V}_l. \quad (3)$$

$D_m$  is drive  $m$  and the  $V_l$  the value of the  $l^{th}$  related homeostatic variable and  $V_{op_l}$  its the optimal (or set) value.  $a_{km}$  are the coefficients relating the variable  $k$  to the drive  $m$ ,  $b_{km}$  are the coefficients relating the rate of consumption of variable  $k$  (their derivative) to the drive  $m$ . As an initial simplification in our experiments, we have restricted this general expression to the case where each drive depends on a single homeostatic variable, furthermore making their dependence linear. To this end, the summation for the first expression will solely contain a single term, and the  $b_{km}$  coefficients will

be zero. When a homeostatic variable diverges from its optimal point, an appropriate mechanism of compensation is triggered. In our case, the mechanism of compensation is the selection and execution of a behaviour. In the general case, both the value and the rate of change of the related homeostatic variables may affect the drive, thus capturing an urgency inherently related to each internal resource. In the study presented here, we have used three motivations: *hunger* (which controls nutrition), *fatigue* (controlling stamina), and *curiosity* (controlling restlessness). At each time step, the drives are assigned an intensity (activation level) proportional to the magnitude of the error of their controlled variable. Several drives can be active at the same time, thus establishing a competition for their related behaviours to be compensated. If in the general case every homeostatic variable may be compensated by several behaviours, in this initial simplified configuration, we have related each internal variable to a particular behaviour, which may be executed for compensation. For example, if food is ingested, the level of its related variable (nutrition) increases, as captured by the  $\alpha_{ik}$  factor of equations 1 and 2.

The overall goal for any living organism is to survive in a stable manner. Because of this, we have used Ashby’s definition of *physiological balance* (Ashby, 1965) as the metric of performance. For the case of our agent, the physiological state is defined by the agent’s internal homeostatic variables and drives, which can be described as a dynamical system to facilitate its study. Figure 2 displays an example of two-dimensional physiological space that illustrates the evolution of the agent’s internal state as a result of the internal resources decays and of the execution of several behaviours. The figure also illustrates two regions within the physiological space: the viability zone, where the agent has to maintain its state to continue living, and the optimal zone, where the agent is most satisfied and its drives are minimal. In this context, the execution of a behaviour can yield two different effects: it does not modify the agent’s internal state (failed execution) or it diminishes the level of one or more drive (successful execution). This sort of representation is inspired by Spier’s work (Spier and McFarland, 1996). The model also includes a *hormonal response*, which is internally triggered when there is a sudden variation of the internal homeostatic variables and decays exponentially afterwards. This plays a fundamental role in learning object affordances by strengthening or weakening the synapses of their neural representations (cf. section 4.4).

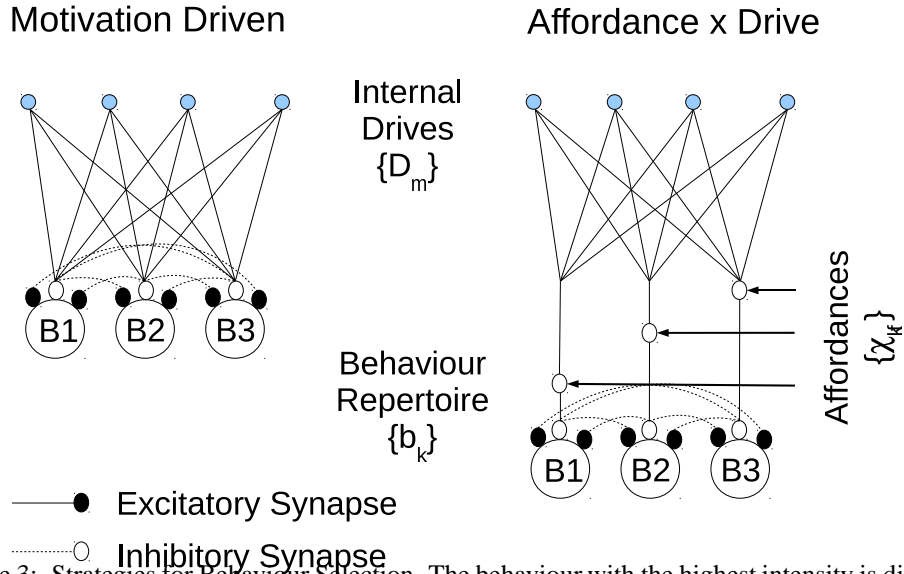


Figure 3: Strategies for Behaviour Selection. The behaviour with the highest intensity is dis-inhibited to be executed. Two procedures to calculate its activation are considered: motivation driven and *affordance*  $\times$  *drive*.

*Behaviours* in this model are coarse-grained sub-systems (embedding simpler actions) that implement different competencies. The execution of a behaviour modifies (increases or decreases) the levels of particular internal variables, therefore affecting the satisfaction of drives. In the general case, different behaviours can contribute to satisfy the same drive, but in our simplified model each drive can be satisfied by one behaviour only — “eat” (approach edible objects) satisfies hunger, rest in a shelter satisfies fatigue, and interaction with the environment satisfies curiosity. For the behaviour

to have a *compensatory effect* (increment or decrement of the internal homeostatic variables), its execution must happen in an agent-environment context where that affordance is defined.

Although this is not a study on *action selection* per se, it is aimed at assessing the working mechanisms for learning grounded knowledge about the environment in terms of physiological stability. Therefore, it is necessary to at least establish some basic rules for how the agents will resolve conflicts among competing behaviours. Two strategies have been considered to this end: motivation-driven and *affordances*  $\times$  *motivation* (cf. figure 3 for a visual description). The former consists of selecting the behaviour whose related drive exhibits the highest urge, the latter combines the averaged drives' urge and the affordances offered by the object nearby to select the next behaviour. The selection mechanism works by dis-inhibiting the behaviour whose activation is the highest while it inhibits the others. The outcome of the interaction depends on the robot interacting with an object affording the behaviour selected for execution. The agent will have to learn which object offers the right affordance depending on the behaviour it is trying to perform.

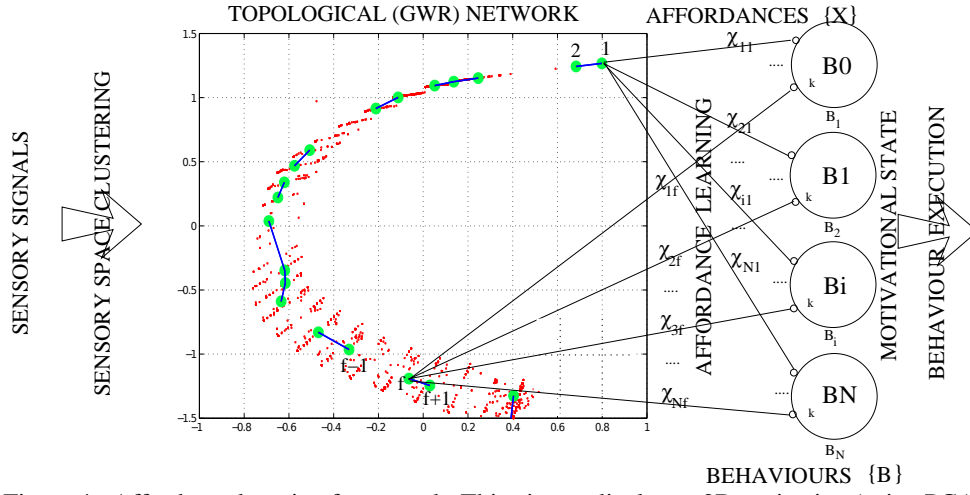


Figure 4: Affordance learning framework. This picture displays a 2D projection (using PCA) of the GWRN (medium size dots are the nodes, connecting lines the synapses), the data used to train it (small dots) and the *affordance* synapses relating each topological node to every behaviour ( $\chi_{kf}$  synaptic activation relating node  $f$  to behaviour  $k$ ).

This section has described the elements and processes integrating the agent's internal physiology (cf. figure 1, centre-top). The next section describes the use of an unsupervised network to cluster sensory space.

### 4.3 A GWR for Sensory Perception

The very low percentage of task-relevant information in the sensory space suggests that a definition of states directly based on sensory input may not be at the right level of description to efficiently capture environment features. We have approached this problem in several related studies in a gradual manner. In a first study, we used direct feature perception, thus the agent perceived the features of objects as long as they were within range, and a two layered architecture (Cos-Aguilera et al., 2003a). The method presented in this paper intends to be a first step to approach the principles of learning grounded knowledge for their use in real robots. Therefore dealing with raw sensory data and using the right level of abstraction is a requirement. Related to this, behaviour based robotics has demonstrated that the choice of behaviour primitives is critical for an efficient control strategy (Konidaris and Hayes, 2004; Matarić and Brooks, 1990). Although this escapes the reach of this paper, we suggest that it may also be advantageous to structure perception and behaviours as primitives that make both perception and behaviour more tractable. Although this perception mechanism should ideally be based on processing sensory flow, to make a test of principles we have preferred to use the more modest approach of processing series of snapshots of nearby objects once these are in range of the robot's IR sensors.

We have designed the agent's behaviour as a series of motor commands, which allow the agent to approach the object for its identification, and to execute a consummatory behaviour. The snapshots,

organised as  $64 \times 1$  horizontal pixel images of light illumination, are normalised between 0 and 1, and fed into a variation of the Growing When Required Network (GWRN) (Marsland et al., 2002). In essence, the GWRN is an unsupervised, topological network composed of nodes and synapses, which expands or contracts depending on how well they fit sensory space. The algorithm to grow the GWRN is described next, see (Marsland et al., 2002) for further detail.

- The first and the second closest nodes to the sample, nodes  $i$  and  $j$ , respectively, are selected—the metric is the Euclidean distance<sup>1</sup>.
- A synapse between  $i$  and  $j$  is grown (if not yet existing).
- The activity is calculated for each node in the network according to equation 4.

$$a_n = \exp\|\vec{\omega}_n - \vec{x}_q\|^2, \quad n \in \{1..N\}. \quad (4)$$

where  $\vec{x}_q$  is a sample from the sensory input and  $\vec{\omega}_n$  the vector of node  $n$ . We have replaced the original linear metric by the euclidean quadratic metric. In this manner, similar nodes in the original metric will increase their similarity, thus tending to collapse into a single node. Furthermore, with the quadratic metric, distant nodes will appear even further away. Since these nodes correspond to different object, it should facilitate a separate representation for them. The two main parameters of the network are: the activity of the node, as defined by equation 4, and the habituation threshold ( $h_T$ ), which is the maximal time allowed for a single node to find the best fitting location. If the habituation threshold is reached and the activity of the closest node is lower than the  $a_T$  threshold, it is assumed that the area neighbouring that node is under-represented, and a new node 'o' is inserted between the best matching node 'r' and the sample  $\vec{x}_q$ . The new node o is connected by two new synapses to nodes r (best matching) and g ( $2^{nd}$  best matching). The original synapse between nodes r and g is pruned.

- If no new node is added (the activity of the node closest to the sensory input ( $a_r$ ) is smaller than  $a_T$  and/or the habituation  $u_r$  is larger than the threshold  $h_T$ ), the winner node r and the nodes immediately adjacent in the topology ( $b$  nodes,  $b \in \{1, \dots, L\}$ ) are *dragged* towards the location of the sample by an amount ( $\epsilon_b$  and  $\epsilon_r$ ) proportional to the mismatch between them. The winner node is shifted by

$$\vec{\Delta}_r = \epsilon_b \times u_b \times (\vec{x}_q - \vec{\omega}_r) \quad (5)$$

where  $\vec{\omega}_r$  is the vector representing the position of node r and  $vecx_q$  is the input sample. The nodes immediately adjacent are shifted by

$$\vec{\Delta}_b = \epsilon_b \times u_b \times (\vec{x}_q - \vec{\omega}_b), \quad \forall b \in \{1, \dots, L\}, \quad (6)$$

where  $L$  is the number of adjacent nodes,  $\vec{x}_q$  a sample from the sensory space,  $u_b$  and  $u_r$  two normalisation constants and  $\vec{\omega}_b$  the vector of node b.

- The age of all synapses connecting to the winner node is decremented by 1.
- Reduce the habituation according to

$$u_r(t) = u_0 - \frac{1}{\kappa_r} (1 - \exp(-\kappa_r t / \tau_r)) \quad (7)$$

where  $\kappa_r$  is a normalisation constant and  $\tau_r$  the decay constant for the winner node. Likewise, the counters of its  $L$  closest neighbours obey equation 4.3.

$$u_b(t) = u_0 - \frac{1}{\kappa_b} (1 - \exp(-\kappa_b t / \tau_b)) \quad \forall b \in \{1, \dots, L\}. \quad (8)$$

where  $\kappa_b$  is a normalisation constant and  $\tau_b$  the decay constant.

- If there are nodes with no synapses or synapses the age of which is larger than  $age_{max}$ , these are deleted.

---

<sup>1</sup>This is defined as  $\sum_{i=0}^N (\vec{x}_i - \vec{y}_i)^2$ .

To summarize, the *parameters* of the network are: *activity threshold*  $a_T$ , *habituation threshold*  $h_T$ , the maximal *synapse age*,  $age_{max}$ . The age of synapses connected to frequently active nodes are re-set to 0. Conversely, synapses seldom active diminish their strength by increasing their age. If the age is larger than  $age_{max}$ , the synapse is pruned. Nodes with no synapses are also pruned. Finally, the *shifting coefficients*  $\epsilon_r$  and  $\epsilon_r$  specify the dragging speed of the nodes towards the new sample.

This section has described the method applied to reduce the robot’s sensory input to a tractable level of abstraction, consisting of a topological network of nodes and synapses ( $\vec{\omega}_f$ ), which characterise the different states of the environment. In a complementary fashion, the next section introduces the structure that relates the topological nodes to the potentiality of performing a behaviour and method used to learn object affordances.

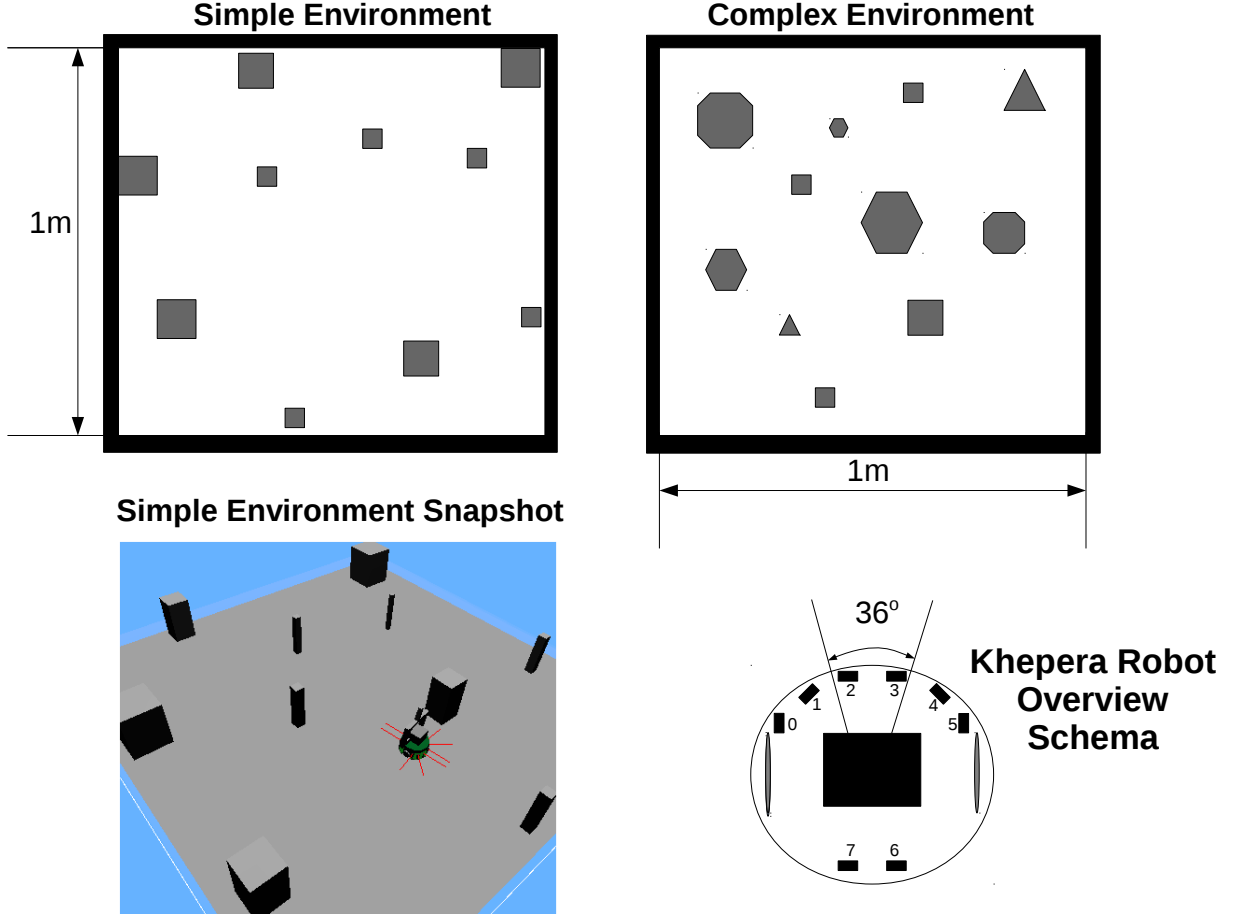


Figure 5: Schematic depiction of the simple and complex environments used for the experiments and Khepera robot overview. IR sensors are numbered from 0 to 7. The robot is also endowed with a normal vision camera (drawn as a black square) and with a gripper (not drawn) to apprehend objects. Left image shows a snapshot of the actual simulation environment and of the robot in the simple environment.

#### 4.4 Learning Affordances

Learning consists of relating the perceived sensory input to the potentiality of performing a behaviour. To this end, we have grown a set of *functional* synapses ( $\chi_{kf}$ , cf. figure 4 synapses extending from GWRN topological nodes to the agent’s behaviours and figure 1, bottom-centre neural structure). Each of these synapses originates from a topological node within the GWRN, and relates to a behaviour within the agent’s repertoire. It is important to remark that the neural structure dealing with perception contains two different sorts of synapses: the topological synapses relating nodes  $\vec{\omega}_f$ , described in the previous section, and the functional (or affordance) synapses  $\chi_{kf}$ . The activation of

these latter synapses indicate the potentiality of performing behaviour  $k$  when an object represented by the node  $f$  is within the agent's range of perception. Therefore, if the agent is able to establish a stable synapse, it should also be able to predict the potential outcome of performing a behaviour when the  $i^{th}$ -node is active. The learning process is illustrated in figure 4 and is described next:

- At the detection of an object, the closest node is identified as determined by equation 4.
- If the execution of the behaviour succeeds, there will be an increase of one or more homeostatic variables, which will trigger the hormonal response. In general, the amount of effect depends on the behaviour and on the resource. For the experiments we present next, the effects are: grasping (or eating) on nutrition 0.3, resting on stamina 0.2, interacting on restlessness 0.1. The release of the hormone following the delivery of effect influences the reinforcement of the functional synapse relating the active node at that moment to the behaviour just executed as described by equation 9. All synaptic values experience a very slow decay, aimed at pruning synapses rarely used.

$$\chi_{kf} \leftarrow \chi_{kf} + \gamma S \quad (9)$$

$\chi_{kf}$  is the weight of the synapse between node  $f$  and behaviour  $k$ .  $S$  is the hormonal response resulting at a physiological level from the interaction of behaviour  $b_k$  with an object affording its execution (see next section).  $S$  is ruled according to equation 10 if  $k$  is the behaviour previously executed. For any other behaviour, it is zero. Therefore, only the activation relating the active node  $f$  to behaviour  $k$  is updated. The value of  $\gamma$  is a small positive value. The synaptic values also experience a very slow decay, aimed at weakening those synapses rarely reinforced.

- The learning results in a set of synapses relating each node in the GWRN to every behaviour within the agent's repertoire. We interpret their associated weights as *affordance values*.

At every interaction with its surrounding environment, the synaptic weights relating topological nodes to behaviours ( $\chi_{kf}$ ) are modified. This learning mechanism has been tested in two environments: a simple and a complex environment.

## 4.5 Hormonal Reinforcement

As described by equation 1, the execution of behaviour  $k$  in the vicinity of an object offering that affordance will exert an effect equal to  $\alpha_{ik}$  on the homeostatic variable  $i$ . We have proposed to use this behavioural effect, which closes loop of interaction between the agent's internal drives and the environment, to learn grounded knowledge from the environment. As aforementioned in previous sections, we have related the agent's survival to the notion of physiological stability introduced by Ashby (1965), which imposes boundaries for the agent's internal homeostatic variables for the animal to survive. The level of the agent's internal variables can be controlled by behaviour execution. In agreement with these principles, we have endowed our agent with an egocentric notion of valency, implemented by the hormonal release that follows the physiological effect of executing a behaviour on a homeostatic variable. This internal signal reinforces those synapses whose related behaviour yields a compensatory internal effect. In other words, the role of the hormonal response is to reinforce the functional synapses between the GWRN nodes and the behaviours afforded to the agent in that context. The dynamics of the hormonal release are defined by equation 10.

$$\dot{S} = -S/\tau_s + \beta_i \sum_i \sum_n u(t - t_n) e^{-(t-t_n)/\tau_s}, t_n = t | \dot{V}_i > X^* \quad (10)$$

$S$  is the hormonal value reinforcing the affordance synapse  $\chi_{kf}$  connecting the winner node  $f$  to the behaviour  $k$  just executed and  $V_i$  is the  $i^{th}$  homeostatic variable and  $\beta_i$  is a normalisation constant for that homeostatic variable, and  $t_n$  are the instants at which the homeostatic variable  $i$  experiences a sudden compensatory change (its derivative is larger than a large value  $X^*$ ).  $u(t)$  is the Heaviside step function. Since the rate of change of the homeostatic variables is several orders of magnitude smaller than that of the hormonal response ( $\tau_s \gg \tau_i$ ), the value of the hormone will rapidly decrease to zero after the behavioural effect has been exerted. During the time  $S$  is larger than zero, the related synapse will be reinforced.

## 5 Experiments

The overall goal of the experiments is to test the architecture described in the previous sections to learn object affordances as a method to facilitate adaptivity by increasing the agent’s internal physiological stability. To this end, the robot has been endowed with a layered behaviour-based architecture, consisting of an abstract, high-level layer and of a low-level action layer. The low-level action layer controls the simpler motor commands that compose the procedural interactions to perform a consummatory behaviour with a particular object. The high-level layer controls the execution of coarse-grained behaviours and is partly controlled by the agent. The wandering of the agent is driven by random navigation, switching to a consummatory behaviour (eat, shelter, interact) when an object nearby is encountered. However, the agent is in control of the consummatory behaviour to be executed. Once the interaction with the object has finished, the object is abandoned and the wander behaviour is engaged again until another object is encountered. This set of experiments were performed with a custom designed C/C++ code, embedded in a simulated Khepera robot with a *Webots<sup>TM</sup>* simulator 4.0, K-Team (Yverdon-les-bains, Switzerland), licensed to the Institute of Perception, Action and Behaviour (IPAB), University of Edinburgh.

The main focus of these experiments is to test adaptiveness of the agent’s perception, which we have divided into two concurrent processes: *GWRN training* and *learning affordances*. Both processes use the same sensory input to work, consisting of images obtained during the approach to the object. The training develops a topological representation of the different angles of view of the objects of different shape and size encountered. Also, an additional neural structure, which extends its synapses off each node towards every consummatory behaviour is used to learn the potentialities of performing each behaviour for every object represented by that node. We refer to this overlapping neural structure as the *affordance network*, since the activation level of its synapses ( $\chi_{kf}$ ) indicates the potentiality of performing the related behaviours. Their level of activation is reinforced or weakened by the hormonal response following after every successful object interaction. Next we describe the *experimental method* for both concurrent perception processes to adapt the agent’s perception to the environment in a dynamic manner.

1. *Training the GWRN.* The robot behaves in a scenario where a set of objects of different shapes and sizes have been randomly distributed (cf. fig. 5). While no object is first detected by the robot’s infra-red sensors (IRs), the robot searches new resources by executing the behaviour wander. When an object is detected, a consummatory behaviour is executed (eat, rest, interact). Each of these behaviours consists of a set of motor primitives, which allow the agent to execute a consummatory behaviour with the nearby object. Snapshots of the object are taken near the object (within IR range), and data vectors of 64 horizontal vector of light intensity, normalised between 0 and 1, are used to train the GWRN. Based on this data, the GWRN algorithm monitors the distance between the input data and the most similar node in the topology. If the sensory input is close enough to the node, its activity increases, otherwise a new node may be added. Two main parameters control the structure of the GWRN: the activity parameter  $a_j$  of each node and the habituation threshold  $h_i$ . The *accuracy* of the topology is balanced between the habituation  $h_i$  of its nodes, controlled by its decay constant  $\tau_i$  and by the maximum activity  $a_T$  (see section 4.3). Therefore, the GWRN can dynamically increase or reduce the number of nodes, depending on the complexity of the input signal, which makes this topological network appropriate for dynamic environments.
2. *Affordance Learning.* The GWRN consists of a set of nodes ( $\omega_f$ ) connected by a set of synapses. Each node represents a set of objects of similar size. Therefore, in order to represent object affordances, a set of functional synapses ( $\chi_{kf}$ ), relate each node to every potential consummatory behaviour of the agent’s repertoire. We refer to these synapses as to the *affordance network*. These synapses represent the affordances of the set of objects represented by each GWRN node and the behaviours the agent can perform. Next we describe the procedure to establish the activity of these functional synapses, which are initialised at a random small value.
  - A series of snapshots are taken within IR range from the object. The closest GWRN node to the sensory input is identified.
  - A consummatory behaviour is selected at random and executed.
  - The hormonal response is initiated if the effect of behaviour execution reflects internally on the level of the homeostatic variables. Otherwise the hormone level remains

unchanged.

- The hormonal release strengthens (or its absence weakens) the synapse of the affordance network relating the behaviour just executed to the node currently active as described by a Hebbian inspired algorithm, see equation 9.
- The simulation stops when the number of nodes is stable and a minimum of 50 interactions per node are executed.

To measure the performance of both perception processes we have used two sets of metrics: *purely statistical* and *behavioural/physiological*. The fitting performance of the GWRN is measured for all nodes in the network, computed as defined by equation 11.

$$\sigma_k = E\left\{\sum_{f=0}^{M-1} \sigma_{\chi_{kf}}\right\}. \quad (11)$$

where  $E\{\}$  is the mean value of the summation of error variances for every functional synapse relating nodes within the GWRN to behaviour  $k$ . Additionally, to assess the learning of the affordance network, we have calculated the stability of the functional synapses ( $\chi_{kf}$ ). Their values depend on the behaviours the objects afford to the agent and on the agent’s skill, and are established by the hormonal reinforcement mechanism defined above. For example, depending on whether the orientation has been correctly matched, the object may be consumed or not. The proximity of the synaptic value to 1.0 indicates the potentiality of executing the behaviour. If not, the value approaches zero.

The metrics of fitting performance and synaptic stability indicate the performance of the learning method. However, a complementary set of *behavioural metrics* has been designed to assess the adaptiveness of this perception method on the agent’s internal milieu. In particular, we have used a *viability indicator* to characterise the overall physiological dynamics during the learning process and during the ‘life’ of the agent, namely *physiological stability*—see equation 12.

$$Physiological\ Stability = E\left\{\frac{1}{N} \sum_{m=0}^{N-1} D_m\right\} \quad (12)$$

$N$  is the total number of drives,  $D_m$  is the  $m$ -th drive, and  $E\{\}$  indicates calculation of average value. This indicator is inspired by the indicators of stability introduced by Avila-García and Cañamero (2002), relating behavioural performance to the dynamics of internal physiology. The statistical and the physiological metrics provide complementary assessments about the *training the GWRN* and *affordance learning* processes. It is important to notice that both learning processes are driven by interaction between the objects and the robot’s sensors and effectors. Therefore, errors due to the variability of real interaction are expected to significantly influence their performance. The remaining of the section describes the results of two sets of experiments, performed in a basic and in a more complex environment, aimed at testing the performance of the proposed learning methods.

## 5.1 Learning Affordances in a Simple Environment

We have previously described the principles underlying affordance learning of consummatory behaviours and the learning method we have implemented. Despite its simplicity, the learning method captures the fundamentals of these principles and allows to explore the interactions between the different elements that compose the dynamics of interaction between the environment and the agent’s internal physiology. As a first step, we have tested the learning algorithm in an environment containing two sorts of objects only: smaller than the robot’s gripper or larger than that. The affordances of these objects will be distributed according to their physical size (diameter), since only “thin” objects can be grasped or shape, since only some of the objects’ shape afford the behaviours rest and interact. Within this framework, this preliminary experiment is aimed at elucidating the right parameters (node activity  $a_T$ , habituation threshold  $h_T$  and maximal synaptic age  $age_{max}$ ) to adapt to the given environment, and to test the performance of the hormonal mechanism for the synapses of the affordance network. As a rule of thumb, we have initially set range for each parameter from the statistics



of the sensory input. Thus, we have collected sensory data by allowing the robot to wander in its environment during several minutes. A PCA 2D projection of the data is shown in figure 4, where the different views of each object appear as clusters of small dots. Using the minimal distance between clusters as rule of thumb, we have set the initial range values for the  $a_T$  threshold, between 0.5 and 0.9. In a similar manner, we have used the rate of object encounter to set the  $h_T$  threshold range between 0.01 and 1.0 and of  $age_{max}$  between 5 and 40. Some of the resulting GWRN topologies are shown in figure 6.

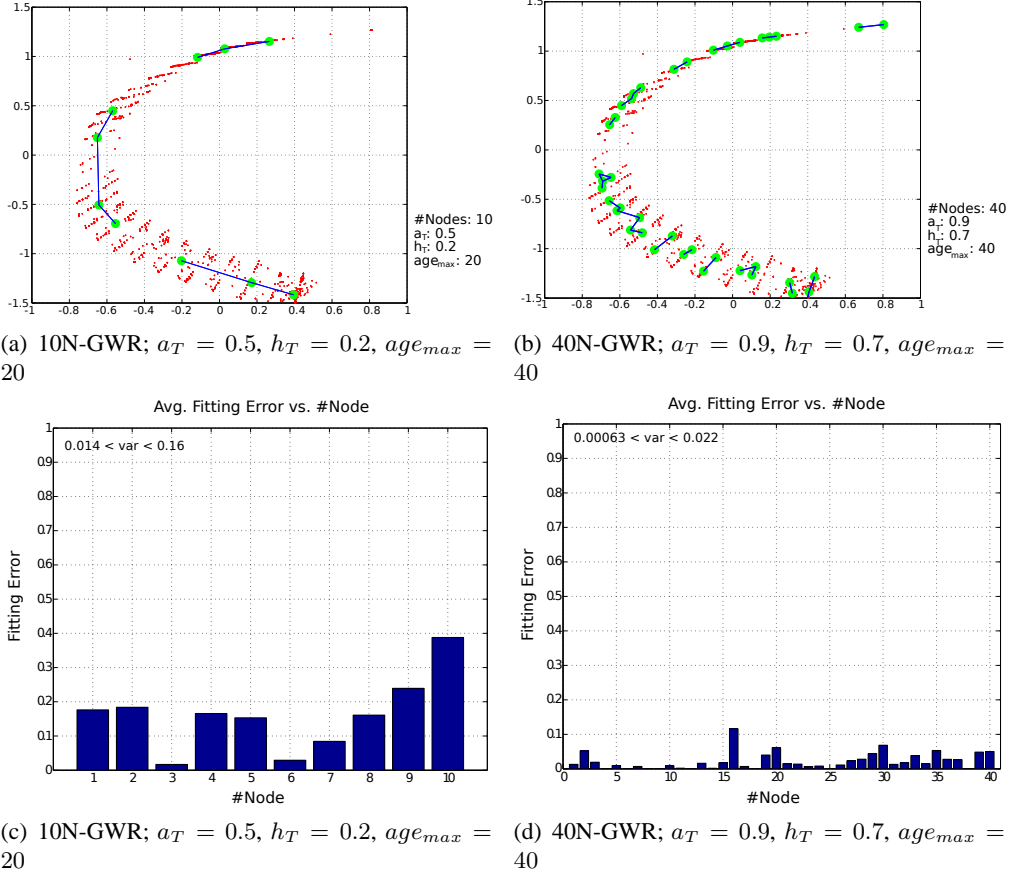


Figure 6: Depiction of two GWRN networks for the simple environment with 10 and 40 nodes, left and right, respectively. The small dots are the samples used for training. The resulting GWRN consists of a set of nodes (large circles) inter-connected by synapses. The histograms underneath show the mean fitting error of the sensory space for each of node.

Plots A and B in figure 6 show a 2D-PCA projection of the data (red dots) and the network topology (green dots, blue synapses). These simplified representations capture 75% of the data variance. Plots C and D show the mean fitting error for each affordance synapse ( $\chi_{kf}$ ) for the two networks shown above, calculated according to equation 11 for each individual node. Top left we show the ranges of variance for all nodes in the network. There are three parameters influencing the network topology: activity  $a_T$ , controlling the fitting network accuracy; habituation threshold  $h_T$ , controlling the limit of time a node is allowed to reach its final position; and the  $age_{max}$ , setting the maximum time a synapse can exist in the absence of excitation. The position of a node is final when its habituation is larger than the allowed threshold  $h_T$ . In a complementary fashion, if the activity of the node closest to the sensory input is smaller than  $a_T$  a new node may be added, since the sensory space is under-represented. In consequence, by increasing the habituation threshold  $h_T$  we are reducing the time given for the nodes to fix their final location. Thus, if the specified accuracy has not been reached yet, some nodes will be rapidly inserted. In agreement with this, experimental results show that by increasing  $h_T$  from 0.5 to 0.7 the total number of nodes increases by five nodes. The network connectivity is mostly controlled by the  $age_{max}$  parameter. Our results show that as we increase the maximal age from  $age_{max}$  from 5 to 20, the number of synapses increases accordingly,

and therefore the accuracy of the representation.

In order to assess the *affordance learning* method, we have calculated the mean fitting error of the GWRN during the last 25% of the simulation, averaging over fifteen simulations (20,000 steps each). The results have been represented as surfaces to cover a larger range of parameter space (cf. figure 7). The x-axis stands for the forgetting factor  $h_T$ , the y-axis for the maximal synaptic age allowed  $age_{max}$ . The z-axis shows the mean and the variance of the functional synapses. In general, the error decreased from 0.03 to 0.01, meaning a larger stability when  $a_T$  shifted from 0.5 to 0.9. For both  $a_T$  values shown, there is a plateau for synapse  $age_{max}$  larger than 20. Furthermore, the stability is not strongly dependent on the habituation threshold  $h_T$ . We have based the estimates for the learning algorithm in future simulations on the observations obtained from this preliminary experiment. In general,  $a_T$  has to be large enough to ensure sufficient accuracy (larger than 0.7), and a  $h_T$  low enough to allow a long habituation that prevents over-representation. Furthermore, this must be encompassed with an age  $age_{max}$  larger than 20 to allow this to happen before the synapses are deleted and the node is pruned. Given these parameter constraints, we have calculated the mean fitting error vs. the number of nodes. Although figure 7 shows that there are consistent variations of the fitting error and its variance with the parameters of the network, the fact is that the average error does not significantly decrease as the number of nodes increases. This suggests that, for this simple environment, a distinction between small and large objects, obtained with a network of two nodes only, would suffice and would allow the robot to learn the affordances offered by the objects in the environment.

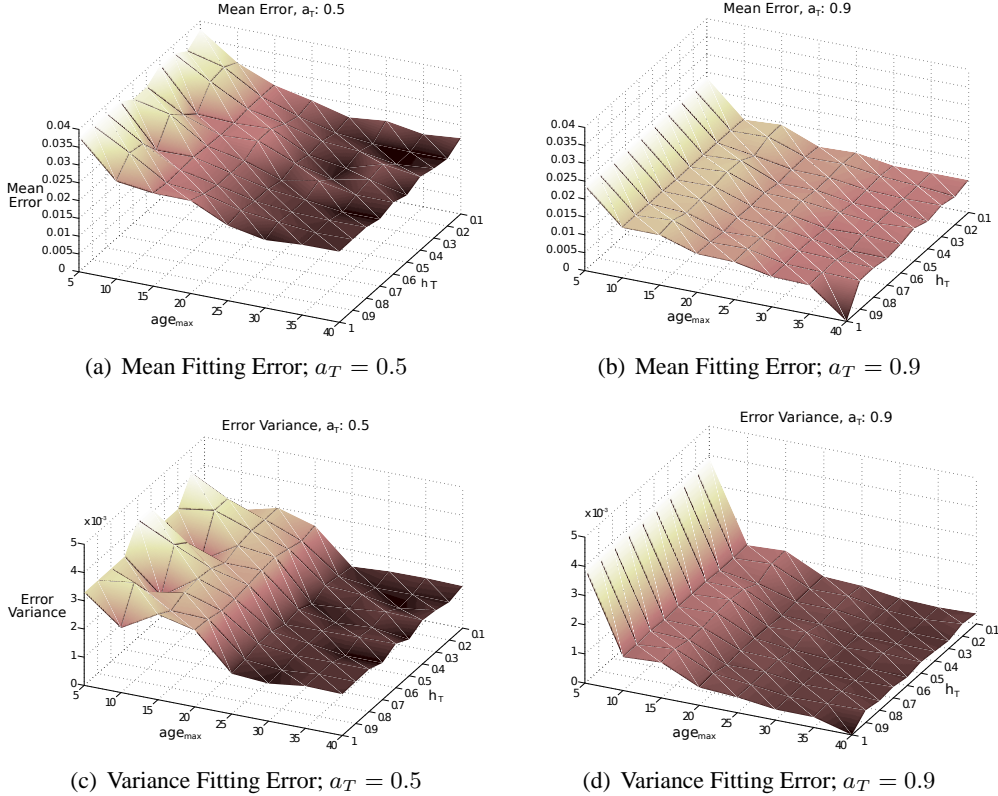


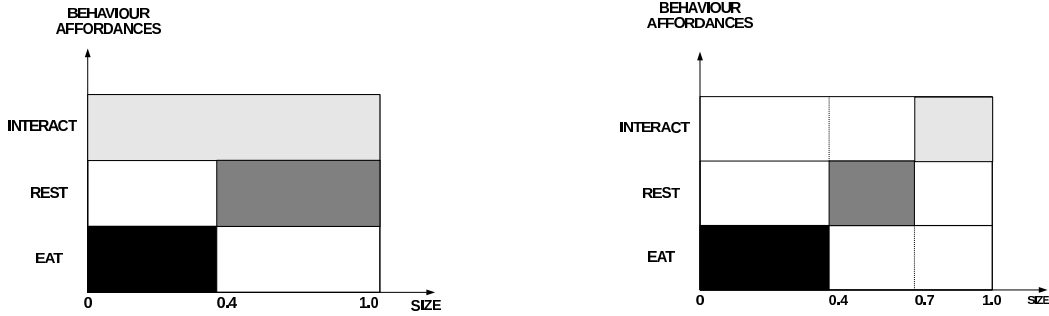
Figure 7: Fitting error for the GWRN for the case of the simple environment, mean and variance, top and bottom surfaces, respectively. The x and y axis stand for the  $h_T$  and the  $age_{max}$  parameters.

This may suggest that the size of the network is relatively unimportant for simple scenarios. However, next section will demonstrate that the complexity of the network is duly related to the complexity and variability of the sensory input and the given distribution of resources of the environment. Based on this, the size and the parameters utilized for the GWRN may significantly facilitate or constrain the agent's final behaviour.

## 5.2 Learning Affordances in a Complex Environment

The preliminary experiments in the previous scenario have demonstrated that the algorithm described previously is capable of learning grounded knowledge from the environment. This second set of experiments aims at exploring the interaction between several parameters relating the agent’s internal physiology to the sensorimotor loop in a more complex set of environments. As mentioned in the introduction of this section, we have considered two main parameters influencing the interaction between the agent’s physiology and its environment: the availability of resources in the environment and the rate of decay of the agent’s internal variables. Two sorts of environments have been considered and labelled as scarce or abundant depending on the *availability of affordances* offered by their objects. Since the affordance eat (grasp) directly depends on the width of the object, we have extended the same relationship to the other two behaviours in a similar manner, as described in figure 8. Every object in the *abundant* scenario will afford interaction, objects smaller than 0.04 will afford eating and objects larger than 0.04 will afford shelter. In the *scarce* scenario this is significantly restricted, since only objects whose size is between 0.04 and 0.07 will afford shelter and only objects of sizes between 0.07 and 0.1 will afford interaction. Also, an additional parameter of this experiment is the *rate of decay* of the agent’s internal physiology, characterised as rapid and slow ( $\tau=10\text{E-}3\text{s}$ ,  $\tau=10\text{E-}4\text{s}$ , respectively, for all homeostatic variables).

Also, we have proposed two strategies of behaviour selection: selection of the behaviour related to the drive exhibiting the highest value, selection of the most active behaviour, according to the *affordance  $\times$  drive* formula (see section 4.2 and figure 3 for further detail). The performance of the learning mechanisms we have used two different metrics: first the fitting error of the GWRN nodes, second the overall physiological stability (cf. equation 12) of the agent’s internal milieu.



(a) Abundant distribution of affordances.

(b) Scarce distribution of affordances.

Figure 8: Abundant and scarce distributions of affordances. The affordance eat is related to the physical width of the gripper, which is 0.04. Therefore, objects smaller than this width do afford this behaviour to our robot. For simplicity, we have preferred to related object size to the other behaviour affordances offered to our robot. The area indicates the range of object sizes that offer that behaviour. Each grey-scale indicates one behaviour: eat – black, rest – dark grey, interact, light grey.

*Preliminary test.* As for the previous set of experiments, the first goal has consisted of assessing the quality of the fitting to the input data. To this end, we have run a preliminary test consisting of ten simulations (20,000 steps each) to explore the state space of the resulting topological networks (GWRNs). At each simulation, the agent wandered in the virtual environment and interacted with the objects nearby by selecting the behaviour to execute next at random from its behavioural repertoire (see experimental method in section 5). As for the previous set of experiments, the ranges of the parameters have been equally set to:  $a_T$  between 0.5 and 0.9,  $h_T$  between 0.01 and 0.1, and  $age_{max}$  between 5 and 20. Figure 9 shows some of the resulting topological structures (2D-PCA projections of data and network nodes), accompanied by the mean fitting error for each node underneath. The average error per node between both networks indicates that the number of nodes plays a significant role as the complexity of the scenario increases. We have also averaged the mean and variance of the activity of the functional synapses relating the topological nodes to the agent’s behaviours ( $\chi_{kf}$ ). Figure 10 shows the values obtained for two networks of sizes 14 and 40 nodes, respectively. The

average fitting error has decreased from 0.024 to 0.012, indicating that the larger network provides a significantly more accurate representation of the object views. In general terms, the fitting error decreases as the size of the network increases, yet reaching a plateau around twenty-five nodes (data not shown). The effects of the parameter adjustment observed are similar to those obtained from the experiments in the simple environment. As shown in figure 10, the fitting error diminishes sensibly when the node activity rises over the threshold value  $a_T$ . Similarly, the larger  $age_{max}$ , the smaller the fitting error. As for the previous experiments, the habituation threshold  $h_T$  (y-axis) exerts little influence.

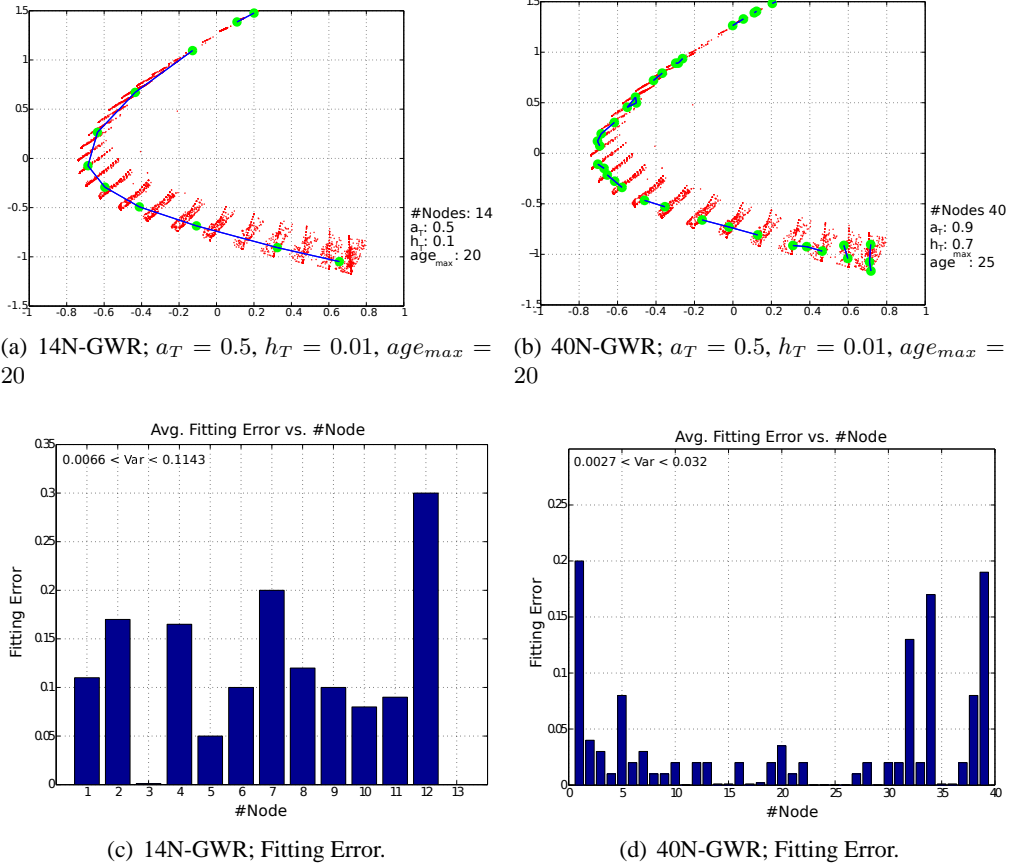


Figure 9: Top: 2-D Principal Component Analysis (PCA) of two GWR networks built for the case of the complex environment, with 2 and 24 nodes, left and right, respectively. The small dots are the samples used to train the GWRN. The GWRN nodes are shown as circles, connected by synapses. Bottom: Histogram representing the mean fitting error for every node of the GWR network represented on top for every node of them.

*Physiological and behavioural assessment.* To test the behavioural and physiological performance of the learning method proposed, we have assessed the physiological stability of the agent's internal milieu during the final 25% of every simulation. As mentioned in the introduction of this section, we have considered two main parameters influencing the interaction between the agent's physiology and its environment: the availability of resources, characterised by the abundant or scarce scenario and the internal physiology, characterised by its rate of decay. Experiments exploring the parameter space for the two aforementioned parameters have been performed to measure the physiological stability of the agent's internal milieu. Also, two behaviour selection strategies for each set of experiments have been tested.

1. The results for the first set of experiments performed with agents endowed with a *slow decaying physiology* in a *scarce distribution* of affordances are displayed as a function of the number of nodes and of the arbitration mechanism in figure 11. The left graph shows that the fitting error decreases rapidly after the GWRN expands over a few nodes. However, each

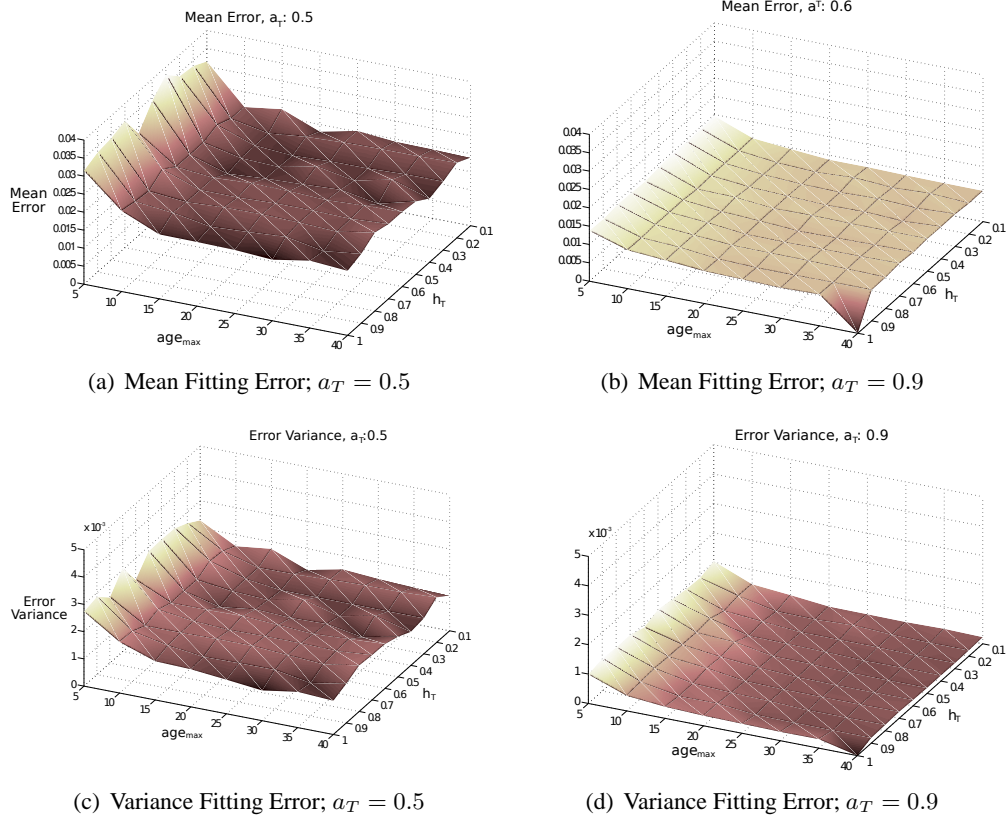


Figure 10: Mean and Variance of the fitting accuracy of the GWR networks, top and bottom surface graphs, respectively. The surfaces are parametrised after the activity of the GWR. It displays data for  $a_T=0.5, 0.9$ , from left to right, respectively. On x and y axes are plotted the  $age_{max}$  and  $h_T$  parameters.

individual drive (central bar graphs) and the physiological stability indicator (graphs on the right) exhibit a strong dependence on the network size and on the behaviour selection strategy. In general, the drives decrease and hence the physiological stability increases as the network size expands, until a plateau is reached for networks of 15 nodes or more, for the case of the *affordance*  $\times$  *drive* selection strategy. As for the values obtained using the motivation-driven selection strategy, these should be considered as a performance baseline, since this strategy consists of selecting and executing the behaviour associated to the drive expressing the highest urgency, disregarding the potentialities of action offered by nearby objects. Although this strategy closely monitors the agent’s internal stability, it should strongly depend on the distribution of affordances. Therefore, its performance would only be expected to rival that of the multiplicative strategy when the distribution of affordances is very abundant.

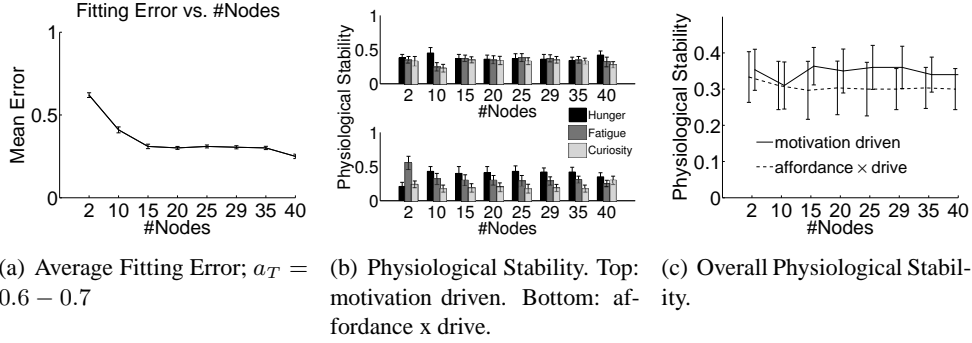


Figure 11: Statistical and behavioural metrics for agents of *slow decaying* physiology ( $\tau = 10^{-4}$ ) in an environment with *scarce distribution* of affordances. The **left plot** shows the fitting error of the GWRNs obtained for networks of size ranging from 2 to 40 nodes. The **central plot** shows the physiological stability for experiments using two behaviour selection strategies: motivation driven and *affordance*  $\times$  *drive*, top and bottom, respectively, as a function of the number of nodes. The **right plot** shows the overall physiological stability for both behaviour selection strategies as a function of the number of nodes. The values shown are the average over 20 simulations.

- Figure 12 shows the results for a second set of experiments in the same environment containing a *scarce* distribution of affordances, but with agents endowed with a *fast varying* internal physiology. As a consequence, the adaptation to this scenario should be driven by the urge to perform behaviours in a more efficient manner, as the rate of growth of the drives will demand compensation quicker than previously. Firstly, the fitting error (left graph) in figure 12 shows that the fitting error decreases as the number of nodes increases in two phases: first, within the range from 2-15 nodes and again from 30-40 (and possibly onwards). From the physiological metrics we have calculated, we observe that the motivation-driven strategy performs poorly, since most interactions with the objects for the environment are unsuccessful and the urges for action vary significantly faster than in the previous experiments. Therefore, both the individual drive values (central bars in figure 12) and the overall stability indicator (right graphs) are close to 1.0 (the maximal drive value), meaning that this selection strategy can hardly maintain the agent within its viability zone. Unlike this, the results obtained for the *affordance*  $\times$  *drive* strategy, which combines the learned grounded knowledge from the environment potentialities of action with the internal needs of the agent, shows an increase of performance along with the size of the GWRN. In particular, as it occurs with the fitting error, the overall physiological stability (graph on the right) decreases first between 2-15 nodes and again as the GWRN size increases from 29 to 40 nodes.
- Finally, we have also performed a set of experiments for the *abundant* environment with agents, both with a *slow* and *fast* decaying physiology. We only show the results from the latter case in figure 13. As for the two former cases, the left graph indicates that the GWRN fits the objects in the environment in two stages, as the number of nodes increases from 2-15 nodes, and then from around 29 to 40. As the availability of resources in this environment is larger than for the scarce case, the requirement of efficiency will decrease accordingly. Therefore, the motivation-driven strategy can already provide a stable physiology in this case (see central-top bar graph and continuous line on the right graph), although the rate of successful interaction is

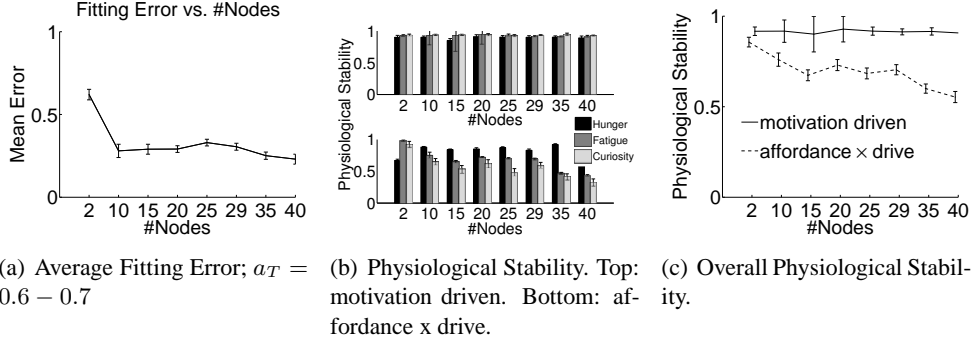


Figure 12: Statistical and behavioural metrics for agents of *fast decaying* physiology ( $\tau = 5 \times 10^{-4}$ ) in an environment with *scarce distribution* of affordances. The **left plot** shows the fitting error of the GWRNs obtained for networks of size ranging from 2 to 40 nodes. The **central plot** shows the physiological stability for experiments using two behaviour selection strategies: motivation driven and *affordance  $\times$  drive*, top and bottom, respectively, as a function of the number of nodes. The **right plot** shows the overall physiological stability for both behaviour selection strategies as a function of the number of nodes. The values shown are the average over 20 simulations.

worse than for the other selection strategy. The overall performance for agents endowed with the *affordance  $\times$  drive* behaviour selection strategy indicates that using grounded knowledge is still advantageous in this abundant environment.

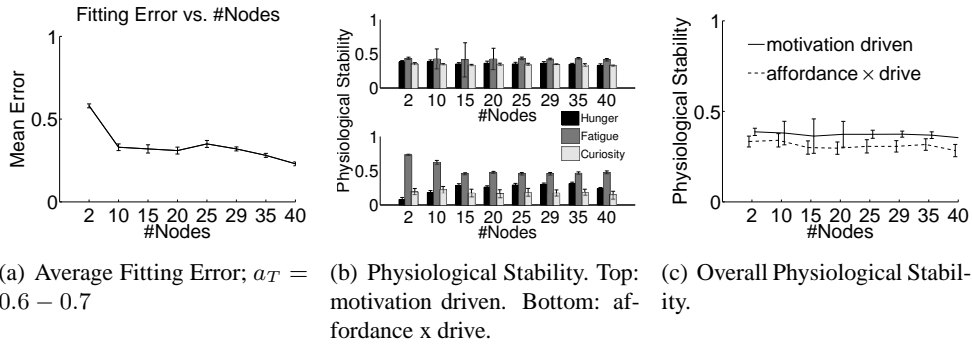


Figure 13: Statistical and behavioural metrics for agents of *fast decaying* physiology ( $\tau = 10^{-3}$ ) in an environment with *abundant distribution* of affordances. The **left plot** shows the fitting error of the GWRNs obtained for networks of size ranging from 2 to 40 nodes. The **central plot** shows the physiological stability for experiments using two behaviour selection strategies: motivation driven and *affordance  $\times$  drive*, top and bottom, respectively, as a function of the number of nodes. The **right plot** shows the overall physiological stability for both behaviour selection strategies as a function of the number of nodes. The values shown are the average over 20 simulations.

Therefore, these results demonstrate that the availability of resources in the environment and the rate of decay of the agent's homeostatic variables strongly depend on one another, and may dramatically influence the agent's chances of survival. As the variables' rate of decay increases, the demand for effective and frequent behavioural interactions will gradually become more urgent. This is reinforced by the results shown in figure 12, which result from the case of a scarce distribution with fast decaying variables. These highlight that the physiological stability strongly depends on the rhythms of the agent's internal physiology and that having a mechanisms of adaptive perception, integrating perception, physiology and action in a single dynamics provides an advantageous mechanism for adapting to the environment. These results also suggest that the mechanisms proposed can be used to learn the affordances of consummatory behaviours offered by objects in the environment. In other words, to learn grounded knowledge from the objects in the environment.

### 5.3 Theoretical Formalisation

As a whole, the results of the previous section demonstrate that there is a close relationship between the dynamics of the agent's internal physiology and the resources of the environment. This final section aims at providing an analytical formulation of this inter-dependence. To this end, we have formalised the dynamics between the agent and its environment by formulating the lethal boundary as a function of the decay rate of the agent's homeostatic variables and the distribution of resources of the environment in a probabilistic manner. As suggested by Ashby's principle of physiological stability, the survival of an agent will depend on its ability to maintain every homeostatic variable within its viability zone. If any of these variables decays under its critical value, the agent will die. Therefore, for every agent there will be a variable that is most critical, either due to its fast decaying rate or due to the difficulty of finding objects in the environment that afford the behaviour contributing to its compensation. For that particular variable  $i$ , we can derive the critical time it takes to reach the lethal boundary. Elaborating from equation 2, the critical time  $t_{max}$  equals the expression shown by equation 13.

$$t_{max} = -\tau_i \log\left(\frac{V_i^*}{V_i^* + b_{ki}}\right) \quad (13)$$

Where  $t_{max}$  is the time variable  $i$  takes to reach its lethal boundary  $V_i^*$  in the absence of successful behaviour interactions.  $b_{ki}$  is the effect exerted by a successful behaviour interaction and  $\tau_i$  the homeostatic variable decay rate. Therefore, for the agent to survive, the interactions with an object that affords its compensatory behaviour must occur more often than  $t_{max}$ .

If we assume that the availability of resources of an environment is uniform, it is possible to model the statistics of successful interactions as a poisson process with parameter  $\lambda_i$ . If we also assume that the interactions are always successful, the average time to compensate the variable  $i$  will equal  $1/\lambda_i$ . Based on these two assumptions, we may derive an analytical estimate of the lethal boundary for that particular agent. In other words, the probability of survival, as described by equation 14, equals the probability that the time between compensatory interactions  $T$  were smaller than  $t_{max}$ , in other words, the probability that the number of compensatory interactions within a  $t_{max}$  interval were larger than 0.

$$P_{Survival} = P\{T < t_{max}\} = P\{n > 0\} = 1 - P\{n = 0\} = 1 - e^{-\lambda_i t_{max}} \quad (14)$$

Where  $T$  is the interval between interactions, and  $n$  the number of interactions during that interval. If we fix the probability that at least a compensatory interaction occurs to an arbitrary value  $p$  (probability of survival), it becomes straightforward to derive an analytical boundary for the distribution of resources the environment has to offer (characterised by the  $\lambda_i$  parameter of the poisson process) for the agent to survive (see equation 15).

$$\lambda_i = -\frac{1}{t_{max}} \log(1 - p) = \frac{\log(1 - p)}{\tau_i \log\left(\frac{V_i^*}{V_i^* + b_{ki}}\right)} = \log_{\left(\frac{V_i^*}{V_i^* + b_{ki}}\right)^{\tau_i}} (1 - p) \quad (15)$$

Therefore, the agent will survive in environments in which the average number of interactions is smaller than  $1/\lambda_i$ . In this light, the distributions used in the previous experiments, scarce and abundant, may be related, respectively, to a large and a small  $\lambda_i$  values. Certainly, this simple parametrisation does not include an explicit characterisation of the abilities of the agent, such as its moving speed, the agent's behaviour selection strategy or the reliability of its interactions. However, it does illustrate the limitations of the relationship between the environment and the agent in a quantitative manner, and therefore their inter-dependence.

## 6 Discussion

Although adaptation is probably the most studied topic in the adaptive behaviour community, most attention has been devoted to study and design increasingly complex algorithms of decision-making. Certainly, the lifespan and the quality of life of an agent will depend on its ability to plan and decide appropriately. However, in order to do that, any living being needs to also know how to prioritise responses to compensate its internal needs. To this end, it also has to know which elements to interact with to obtain the desired effect. This is the fundamental relationship between the agent and its surrounding environment that we have formulated as a process that dynamically learns *grounded*



knowledge about the environment. In other words, this paper has proposed a formulation of the dynamics agent-environment to learn affordances of consummatory behaviours. The learning is driven by a hormonal process controlled by fluctuations in the agent's internal physiology, which result from interactions with objects in the surrounding environment. A similar process had been previously introduced in the *drk* model (Spier and McFarland, 1997, 1996) to explain the devaluation effect (Kriekhaus and Wolf, 1968) in the absence of cognition by learning the compensatory effect of several stimuli (*d-values* in their notation) using a variation of the delta rule (Widrow and Hoff, 1960). However, our architecture introduces several elements fundamentally different. First, the control of the learning process is *indirectly controlled* by the hormonal release. The reason for us to separate the homeostatic processes and the actual learning process is that the physiological effect due to behaviour execution requires some additional assessment. For example, while the effect of consuming food is beneficial if the animal is hungry, it may be harmful if sated. This assessment process is commonly referred to as the sense of valency (Ackley and Littman, 1991), which could be loosely defined as the internal process of that makes us perceive the effect of executing a behaviour as beneficial or harmful. The idea of a hormonally controlled learning method is inspired from this principle. The *hormonal* release is controlled by fluctuations of the agent's internal physiological variables, and only responds to positive compensations of any internal homeostatic variable, thus to contributions to physiological stability. Although we have not altered this basic rule, we suggest that it may be possible to modify both perception and behaviour in an overall beneficial manner by evolving the valency criteria to facilitate adaptation in a larger variety of environments. The formulation of the hormonal response intends to be a first step towards this end. Also, this separation may facilitate the generalization of behavioural effects involving more than one homeostatic variable. Second, the learning process and therefore the assessment of the behavioural effect is framed by the notion of *internal physiological stability* formulated by Ashby (1965), formalized as an indicator of physiological stability. This notion establishes a basic condition for an agent to live, e.g., an agent's overall behavioural pattern must perform to avoid a lethal boundary. Therefore it establishes a quantifiable metric to assess the agent's performance. The hormonal response of our agent is also inspired after this notion, and provides a general context for the agent's perception and also for its behaviour. Third, unlike the *drk* model, which disregards real perception, the learning architecture here proposed has a perception layer consisting of two concurrent processes: the adaptive topological network (GWRN) and the Hebbian learning of the affordance network. The unsupervised network can dynamically adapt its topology to the objects in the surrounding environment. If these changes are durable, their sensory perception will be captured by the GWRN and their related affordances will be learned by the hormonally controlled affordance learning. As the results demonstrate, if the representation is sufficiently accurate, the agent's perception adapts to the system, and provides grounded knowledge for the agent to maintain its physiological stability within the viable area.

*Internal physiology and sensorimotor interaction.* The results presented reflect the close link between the agent's internal physiology, characterised by the internal physiological decay and their environment, characterised by the distribution of resources. Although this does not intend to reject any other agent architecture, it does reinforce the view of a single, ecological dynamics for an agent and its environment. In other words, this provides an implementation of the relationship between the resources afforded by the environment and the agent's internal rhythms. Importantly, this also relates to two major problems of robotics: building sensory representation and building a motivation driven architecture. These two problems share the common ground of being very difficult when addressed separately. However, the results obtained from our experiments suggest that this difficulty may lessen when the knowledge learned is grounded by using a motivation driven architecture, i.e., when both problems are addressed simultaneously.

*External vs. internal effect.* This robotic environment has allowed us to investigate those behaviours and that part of the behavioural effect that can be perceived by changes in the agent's internal physiology as a method to learn object affordances of consummatory behaviours. Nevertheless, many, if not every motor behaviour in the animal realm may also provoke an alteration of the physical environment, which may be perceived via visual or proprioceptive feedback. Thus, it seems reasonable that a possible extension of the architecture we presented here may include mechanisms such as the hormonal response to learn affordances by combining external and internal effect by using multiple sensory modalities. For example, the effect of shifting the position of an object reflects on changes in the optic flow, as it also correlates with our increase of fatigue as the task gradually extends over time. In a way, this future extension encompasses a potential manner to extend affordance learning from its simplest expression, maybe the *drk* model (Spier and McFarland, 1996) towards a

multi-sensory model encompassing all sorts of behaviours.

*Limitations.* Most boundaries of this study are related to the divergence between the model and the animal models of adaptation. In this light, one of the main conceptual limitations is that the representation of the sensory is not directly based in optic flow, but on a series of snapshots taken from objects. We argue that although this procedure is not ideal, it has significantly reduced the complexity of the experimental set-up, while preserving the overall goal of the experiments, which is testing the adaptivity of the perception layers and the grounding principle. We would also like to stress that we do not claim a faithful biological compliance of the processes described, although the model has been strongly influenced by ethological and biological data. In the same manner, although the results relating behaviour and physiological stability resulting from the experiments model physiological processes, they ought to be viewed in the context described and by the experimental set-up. A possible future extension of this work may consider the study of internal stability for the agent's physiological dynamics when the relationships between variables, drives and behaviours are not univocal. Keeping in mind the goal of this article, this would not add any additional conceptual element to the procedure to learn grounded knowledge, but it would facilitate a complementary study on the internal physiology, which may in turn demand a more elaborated procedure to arbitrate the agent's behaviour repertoire. Furthermore, some additional limitations may result from the use of a simulated robot. The perception and the interaction of a real robot with the real environment, even for the case of a relatively humble robot as the Khepera, has a component of unreliability that simulators tend to overlook. As for the main elements of this study, they may be separated into two different groups: the homeostatic loop and sensorimotor interaction. The former depends on the intrinsic rate of consumption of the internal variables only, and would be transferred to a real robot without any further changes. As for the latter, some additional fine tuning may be required, since it would be expected perception to be noisier and the interaction with the objects to be less reliable. Although no conceptual changes would be required in any case.

*Behaviour selection.* As mentioned in the introduction, this study does not address behaviour selection per se. However, because we have selected a metric of performance based on physiological stability, it is necessary to endow the agent with a strategy to arbitrate among behaviours. Because the goal is to measure the performance provided by the perception mechanisms, we have chosen two very basic rules: motivation-driven and *affordance*  $\times$  *drive* (McFarland, 1993). The former disregards any knowledge from the environment, while the second one utilizes the grounded knowledge learned about the surrounding environments.

## 7 Conclusion

This paper has presented a method to learn object affordances based on the interaction between the sensorimotor and homeostatic loops. The learning principle is based on the internal reward generated by effect of a consummatory behaviour on the agent's homeostatic variables. The architecture proposed to learn the affordances consisted of the combination of a topological network, grouping together objects exhibiting similar perception patterns, and of an affordance network, connecting each node to every behaviour. This second set of synapses, connecting to every behaviour, encode the affordances offered to the agent by the objects represented by every topological node. The learning procedure consists of two concurrent processes: first, a sensory mapping process, which assesses the distance between the node closest and the sensory input. If the distance is larger than a certain threshold, the sensory pattern is considered under-represented, and a novel node is added to the topological network. The synaptic weights connecting to the topological network are initialised to the average values of the two closest nodes and are fine-tuned as new interactions with similar sensory patterns occur. Also, the nodes used less frequently are pruned from the topological map on the long run. Secondly, the reinforcing or weakening of the affordance synapses relating the topological nodes to the behaviours. This process is initiated by the hormonal release, which follows an interaction with an object in the environment if the object affords the behaviour the agent executed. A successful interaction typically produces an increase of one or more of the agent's internal homeostatic variables, and consequently a hormonal release, which reinforces the synapse relating the active node to the behaviour executed. Also, the nodes and synapses of the network, after the ecological principle, degrade if not used frequently.

We have tested this architecture in two different scenarios to investigate if the proposed sequence of processes relating sensorimotor interaction to the homeostatic loop can yield a neural substrate that learns affordances from objects in its environment. We have assessed both the statistical and be-

havioural performance of the proposed learning method. Firstly, we have used statistical metrics to assess the accuracy required to represent the sensory patterns of each scenario by gradually increasing the network complexity and assessing the resulting fitting error. Secondly, we have used the physiological stability as an ecological criterion to assess whether the learned affordances do significantly affect the agent’s behavioural patterns and consequently its physiological stability. In particular, we tested the physiological variability indicator to assess the stability of the agent’s internal resources under two behaviour selection policies, extracted from the ethological literature (McFarland, 1993). The results obtained suggest that the learning method effectively learns the behaviours afforded by every object and that the *affordance*  $\times$  *drive* selection strategy produces behavioural patterns such that the agent’s internal milieu is most stable. Therefore, the affordances learned from the environment can be used to arbitrate the agent’s behaviour repertoire in a way that maintains the internal physiological variables within the range of viability for that agent to survive. The learning method is relatively effective, since it can infer the affordances of new objects from the similarity to nearby nodes, representing similar sensory patterns. This greatly improves the efficiency of the learning algorithm, since the synaptic update, driven by a Hebbian process would require several interactions for fine-tuning.

This paper presents a theoretical formalisation of the dynamics between two interacting processes, governing developmental adaptation in the natural world: sensorimotor interaction and homeostasis. Furthermore, it presents the interaction between internal physiology and sensorimotor interaction as the foundation of grounded knowledge about certain elements of the environment. In other words, a foundation to learn the affordances in the context of artificial, embodied agents. We have investigated the nature of that interaction and the role internal motivations and external behavioural effect play in weakening or reinforcing the relationship between the agent’s internal motivations and the environment. Therefore, this paper is not addressing behaviour selection or planning per se, but only perception as an agent’s adaptive process to the environment. An adaptiveness that depends both on the change of the environment as well as on the internal assessment the agent makes of the behavioural effect (valency function/hormonal response). Affordances are learned from the relationship between the selection and execution of a behaviour and the resulting physiological effect. We argue that the role of adaptive perception, as the capability of adapting the behavioural patterns by the knowledge newly acquired from the environment through this method provides an additional mechanism of adaptation to the environment which has not yet been properly harnessed.

As mentioned in the discussion, to build a representation of the world and to build a motivation driven architecture separately is a difficult task. We believe that this paper has provided the background and the experimental evidence to suggest that both aforementioned problems should be addressed simultaneously using a motivation driven architecture. If there is some consensus on the fact that these processes occur simultaneously in the animal realm and that value is learnt on the grounds of animal urges for action, this may also suggest that an agent design that considers both internal motivations and affordance learning simultaneously, may also benefit from the same inspiration than those using rigid schemas viewing perception and action as sequential processes.

## Acknowledgements

This research has been partly supported by a fellowship of the Graduate School at the University of Edinburgh. We thank George Konidaris for his support and valuable feedback, Andrew Barto at the University of Massachusetts Amherst for his generous hospitality, and Paul Cisek for his valuable feedback on an earlier version of this manuscript.

## References

- Ackley, D. and Littman, M. (1991). Interactions between learning and evolution. In Langton, C., Taylor, C., Farmer, J. D., and Rasmussen, S., editors, *Artificial Life II, SFI Studies in the Sciences of Complexity*, volume X, pages 487–509. Addison-Wesley.
- Ashby, W. (1965). *Design for a Brain: The Origin of Adaptive Behaviour*. Chapman & Hall, London.
- Avila-García, O. and Cañamero, L. (2002). A comparison of behaviour selection architectures using viability indicators. In *Proc. of International Workshop on Biologically-Inspired Robotics: The Legacy of W. Grey Walter*. Bristol HP Labs, UK.
- Baecker, R., Grudin, J., Buxton, W. A., and Greenberg, S. (1995). *Readings in Human-Computer Interaction: Towards the Year 2000*. Morgan Kaufmann.
- Bindra, D. (1969). The interrelated mechanisms of reinforcement and motivation, and the nature of their influence on response. In Arnold, W. J. and Levine, D., editors, *Nebraska Symposium on Motivation*, pages 1–33. University of Nebraska Press.
- Cannon, W. B. (1929). *The wisdom of the body*. London : Kegan Paul, Trench, Trubner & Co.
- Chemero, A. and Turvey, M. T. (2003). Adaptive behavior. *Adaptive Behavior*, 15(4):473–480.
- Chemero, A. and Turvey, M. T. (2007). Gibsonian affordances for roboticists. *Adaptive Behavior*, 15(4):473–480.
- Cisek, P. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society B*, 362:1585–1599.
- Cisek, P. and Kalaska, J. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron*, 45(5):801–814.
- Cooper, R. and Glasspool, D. (2002). Learning affordances and action schemas. In French, R. and Sougne, J., editors, *Connectionist Models of Learning, Development and Evolution*, pages 133–142. Springer-Verlag: London.
- Cos-Aguilera, I., Cañamero, L., and Hayes, G. (2003a). Motivation-driven learning of object affordances: First experiments using a simulated khepera robot. In Detje, F., Dörner, D., and Schaub, H., editors, *The Logic of Cognitive Systems. Proceedings of the Fifth International Conference on Cognitive Modelling, ICCM*, pages 57–62. Universitäts-Verlag Bamberg.
- Cos-Aguilera, I., Cañamero, L., and Hayes, G. M. (2003b). Learning object functionalities in the context of behavior selection. In *Proceedings of the 3rd. Conference Towards Intelligent Mobile Robotics*.
- Cos-Aguilera, I., Cañamero, L., and Hayes, G. M. (2004). Using a sofml to learn object affordances. In *Proceedings of the Workshop on Physical Agents (WAF04)*. University of Girona.
- Damoulas, T., Cos-Aguilera, I., Hayes, G., and Taylor, T. (2005). Valency for adaptive homeostatic agents: Relating evolution and learning. In *Proceedings of the 8th European Conference on Artificial Life (ECAL2005)*, Canterbury, UK., volume 3630 of *Lecture Notes in Artificial Intelligence (LNAI)*. Springer Berlin/Heidelberg.
- Doğar, M., Ugur, E., Şahin, E., and Çakmak, M. (2008). Using learned affordances for robotic behavior development. In *IEEE International Conference on Robotics and Automation*. IEEE.
- Duchon, A. P. and Warren, W. H. (1994). Robot navigation from a gibsonian viewpoint. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 2–5. IEEE.
- Duchon, A. P., Warren, W. H., and Kaelbling, L. P. (1995). Ecological robotics: Controlling behaviour with optical flow. In Moore, J. and Lehman, J., editors, *Proceedings of the 17th Annual Cognitive Science Conference*, pages 164–169. Lawrence Erlbaum Associates.

- Fagg, A. H. and Arbib, M. A. (1998). Modeling parietal–premotor interactions in primate control of grasping. *Neural Networks*, 11(7-8):1277–1303.
- Fellous, J.-M. (2001). Dopamine modulation of prefrontal delay activity-reverberatory activity and sharpness of tuning curves. *Neurocomputing*, (38–40):1549–1556.
- Fellous, J.-M. (2004). From human emotions to robot emotions. In Cañamero, L. and Hudlika, E., editors, *Architectures for Modelling Emotion: Cross-Disciplinary Foundations. Papers from the 2004 AAAI Spring Symposium*. AAAI Press.
- Fellous, J.-M. and Suri, R. E. (2003). *The Handbook of Brain Theory and Neural Networks*, chapter The Roles of Dopamine. MIT Press.
- Freud, S. (1940). *An Outline of Psychoanalysis*. Hogarth Press.
- Gibson, J. (1986). *The Ecological Approach to Visual Perception*. Hillsdale, N.J. ; London.
- Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Houghton Mifflin Company, Boston.
- Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42:335–346.
- Hebb, D. (1949). *The organization of behavior. A neuropsychological theory*. Wiley, New York.
- Hinde, R. A. (1960). Energy models of motivation. In *Symposium of Society of Experimental Biology*, volume 14, pages 199–213.
- Houk, J. C., Adams, J. L., and Barto, A. G. (1995). Models of information processing in the basal ganglia. In Houk, J. C., Davis, J. L., and G., B. D., editors, *A Model of How the Basal Ganglia Generate and Use Neural Signals That Predict Reinforcement*, A Bradford Book, chapter 13, pages 249–270. MIT Press, 2nd. edition (1998) edition.
- Hull, C. (1943). *Principles of Behaviour: an Introduction to Behaviour Theory*. D. Appleton-Century Company, Inc.
- Khamassi, M., Lachèze, L., Girard, B., and Guillot, A. (2006). Actor-critic models of reinforcement learning in the basal ganglia: From natural to artificial rats. *Adaptive Behavior*, 13(2):131–148.
- Konidaris, G. and Barto, A. (2006). An adaptive robot motivational system. In Nolfi, S., Baldassarre, G., Calabretta, R., Hallam, J. C. T., Marocco, D., Meyer, J.-A., Miglino, O., and Parisi, D., editors, *Animals to Animats 9: Proceedings of the 9th International Conference on Simulation of Adaptive Behavior*, volume optional of optional, pages 346–356, optional. Springer Verlag. optional.
- Konidaris, G. and Hayes, G. M. (2004). An architecture for behavior-based reinforcement learning. *Adaptive Behavior*.
- Kriekhaus, E. and Wolf, G. (1968). Acquisition sodium by rats: Interaction of innate mechanisms and latent learning. *Journal of comparative and Physiological Psychology*, 65:197–201.
- Lorenz, K. (1966). *Evolution and Modification of Behaviour*. Methuen & Co Ltd, London.
- Marrow, A. (1969). *The practical theorist: The life and work of Kurt Levin*. New York: Basic Books.
- Marsland, S., Shapiro, J., and Nehmzow, U. (2002). A self-organising neural network that grows when required. *Neural Networks*, 15(8-9):1041–1058.
- Matarić, M. and Brooks, R. (1990). *Cambrian intelligence: The early history of the new AI*, chapter Learning a distributed map representation based on navigation behaviors. The MIT Press.
- Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition: the Realization of the Living*, volume 42 of *Boston Studies in the Philosophy of Science*. Dordrecht; London: Reidel.
- McDougall, W. (1913). The sources and direction of psychophysical energy. *American Journal of Insanity*.

- McFarland, D. (1993). *Animal Behaviour*. Prentice Hall, third edition edition.
- McFarland, D. J. and Sibly, R. M. (1975). The behavioural final common path. *Proceedings of the Royal Society of London*, 270:265–293.
- Morineau, T. and Frénod, E. (2009). Affordance modelling and simulation in virtual environment. *European Journal of Scientific Research*, 33(1):130–143.
- Nehaniv, C. and Dautenhahn, K. (1998). Mapping between dissimilar bodies: Affordances and the algebraic foundations of imitation. In Demiris, J. and Birk, A., editors, *Proceedings of the European Workshop on Learning Robots 1998 (EWLR-7)*, Edinburgh, Scotland.
- Redgrave, P., Prescott, T., and Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89:1009–1023.
- Sahin, E., Çakmak, M., Dogar, M. R., Ugur, E., and Üçoluk, G. (2007). To afford or not to afford: A new formalisation of affordances toward affordance-based robot control. *Adaptive Behavior*, 15:447–472.
- Song, J. and Nakayama, K. (2007). Automatic adjustment of visuomotor readiness. *Journal of Vision*, 7(5):1–9.
- Spier, E. and McFarland, D. (1996). A finer-grained motivational model of behaviour sequencing. In Spier, E. and McFarland, D., editors, *A Finer-Grained Motivational Model of Behaviour Sequencing. From Animals to Animats 4: Proceedings of SAB96*.
- Spier, E. and McFarland, D. (1997). Possibly optimal decision-making under self-sufficiency and autonomy. *Journal of Theoretical Biology*, (189):317–331.
- St. Amant, R. (1999). User interface affordances in a planning representation. *Human-Computer Interaction*, 14(3):317–354.
- Stoffregen, T. (2003). Affordances as properties of the animal environment system. *Ecological Psychology*, (15):115–134.
- Sutton, R. and Barto, A. (1981). Towards a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 198(88):135–170.
- Tinbergen, N. (1951). *The Study of Instinct*. Oxford University Press.
- Turvey, M. (1992). Affordances and prospective control: an outline of the ontology. *Ecological Psychology*, (4):271–281.
- Usher, M. and Davelaar, E. J. (2002). Neuromodulation of decision and response selection. *Special Issue on Computational Models of Neuromodulation*, 15(635-645).
- Widrow, B. and Hoff, M. (1960). Adaptive switching circuits. In *1960 IRE WESCON Convention Record (pt. 4)*, pages 96–104.