# School of Informatics, University of Edinburgh

**Institute of Perception, Action and Behaviour**


## Learning Object Functionalities in the Context of Behaviour Selection

by


Ignasi Cos-Aguilera, Lola Canamero, Gillian Hayes



**Informatics Research Report EDI-INF-RR-0234**

**School of Informatics** **August 2003**

http://www.informatics.ed.ac.uk/

# Learning Object Functionalities in the Context of Behaviour Selection

Ignasi Cos-Aguilera, Lola Canamero, Gillian Hayes

Informatics Research Report EDI-INF-RR-0234

SCHOOL *of* INFORMATICS
Institute of Perception, Action and Behaviour

August 2003

appears in Proceedings of the 4th British Conference Towards Intelligent Autonomous Robots (TIMR'03)

**Abstract :**
In this paper, we propose a model that integrates behaviour selection and learning of affordances through the monitoring of internal motivations in the context of the interactions that an autonomous robot undertakes with objects in its environment while working towards its survival. To test the validity and effectiveness of this model, we have performed two different sets of experiments. The first one aimed to elucidate the question whether our robot could learn affordances by exploiting the outcome of interaction experiments through the monitoring of its internal motivations. The purpose of second one was to assess to what extent the use of previously learnt affordances in novel interactions could improve the survival of the robot.

**Keywords** :

# Motivation-Driven Learning of Object Affordances: First Experiments Using a Simulated Khepera Robot

**Ignasi Cos-Aguilera (ignasi@dai.ed.ac.uk)**
IPAB, School of Informatics, The University of Edinburgh
5 Forrest Hill, EH1 2QL Edinburgh, Scotland, UK

**Lola Cañamero (L.Canamero@herts.ac.uk)**
ASRG, Department of Computer Science, University of Hertfordshire,
College Lane, Hatfield, Herts, AL10 9AB, UK

**Gillian M. Hayes (gmh@inf.ed.ac.uk)**
IPAB, School of Informatics, The University of Edinburgh
5 Forrest Hill, EH1 2QL Edinburgh, Scotland, UK

**Abstract**

In this paper, we propose a model that integrates behaviour selection and learning of affordances through the monitoring of internal motivations in the context of the interactions that an autonomous robot undertakes with objects in its environment while working towards its survival. To test the validity and effectiveness of this model, we have performed two different sets of experiments. The first one aimed to elucidate the question whether our robot could learn affordances by exploiting the outcome of interaction experiments through the monitoring of its internal motivations. The purpose of second one was to assess to what extent the use of previously learnt affordances in novel interactions could improve the survival of the robot.

## Introduction

One of the main challenges for autonomous agents that have to survive in unknown and changing environments is to make the "right" decisions in their interactions with the environment. This problem is known in the adaptive behaviour literature as *behaviour or action selection* – deciding "what to do next" (what behaviour to execute in a particular situation) so that survival is guaranteed. The appropriateness of a decision for a particular situation may depend on many factors; two of the more fundamental ones are: (a) the degree to which an interaction (such as the execution of a behaviour involving an object) contributes to the satisfaction of the agent's survival-related internal needs, and (b) the choice of an object suited to the interaction. Although these problems are highly interdependent, their interrelation has not so often been investigated in the adaptive behaviour literature.

Different behaviour selection architectures have been proposed that make depend the agent's survival on its capability to maintain the stability of its internal milieu (survival-related internal variables defining its needs), drawing on Ashby's notion of *viability* (Ashby, 1952). Some examples are (Meyer, 1995; Cañamero, 1997; Velásquez 1998). These architectures have addressed the problem of designing mechanisms to select the behaviour that best contributes to the satisfaction of the agent's internal needs in particular situations; however, they have overlooked the problem of deciding whether a particular object present in the environment is suited to the interaction (and hence to the satisfaction of internal needs) by "hard-wiring" information about the functionality of objects (the potential they offer for action) in the architecture. However, endowing agents with the capability to build their own functional descriptions of the objects in the environment is crucial for efficient behaviour selection, in particular in unknown and dynamically changing scenarios, where new objects are often encountered. Only this second option would enable an agent to cope with a variety of scenarios, unforeseen by the designer.

Perception of function is related to the notion of *affordance* (Gibson, 1986), which can be broadly defined as the functionality that an object offers to an agent interacting with it, and this mainly depends on the perceived features of the object and on the morphology of the agent. In the adaptive behaviour community, the notion of affordance has been for example used to couple body and environment in the problem of learning by imitation (Nehaniv and Dautenhahn, 1998). More related to our work, (Guazelli *et al.*, 1998) proposed a behaviour selection model to simulate the behaviour of rats navigating a T-maze that integrates drives and affordances learnt for navigation.

In our study, we focus on the (motivation-driven) interactions of an agent with objects in its environment and how the feedback of their outcomes can be exploited in a model that integrates behaviour selection and the learning of affordances to support the survival of the agent. Toward this aim, we integrate and build upon previous work on behaviour selection, reinforcement learning, and affordance learning (Cañamero, 1997; Gadanho and Hallam, 2001; Cos-Aguilera and Hayes, 2002). To test our model, we have conducted experiments to address these questions: (a) could an agent learn affordances by exploiting the outcome of interaction experiments through the monitoring of internal motivations of the agent? And (b) to what extent the use of previously learnt affordances would enhance the robot's capability to survive?

## Model Architecture

The overall goal of our model is to support the robot's survival in a dual way: (a) by providing the right decisions to satisfy its internal needs, and (b) by learning and exploiting the affordances that the objects in its environment can offer to do so in an efficient way. Accordingly, our model consists of two tightly coupled parts, to ensure that the appropriate decision is made to choose the "right" behaviour: an architecture for behaviour selection and an affordance extractor.

### Architecture for Behaviour Selection

The behaviour selection architecture is a simplified version of that proposed by Cañamero (1997). It consists of a set of homeostatic, survival-related internal variables, a set of motivations that control the internal variables, a repertoire of behaviours, and an arbitration mechanism to resolve conflicts among competing motivations and to choose the appropriate behaviour to satisfy the internal needs.

*Controlled homeostatic variables* vary due to internal body dynamics and to the interactions of the robot in its environment. They are abstractions representing the level of internal resources that the agent needs in order to survive. In this case, we have used three variables – nutrition, stamina and restlessness. For example, *nutrition* is an abstraction for a number of elements controlling food intake (e.g. the level of glucose in the blood), and its level decreases with time if no food is periodically ingested. Homeostatic variables must be kept within a range of permissible values or *viability zone* (Ashby, 1952) for the robot to remain alive (viable in its environment); conversely, if the values overflow/underflow the upper/lower boundaries that define the variable's viability range, the robot dies. The value of each internal homeostatic variable reflects its status of "normality" (its current value approaches its ideal value or setpoint), deficit or excess. The homogeneous behaviour of these variables is given by equations 1 and 2, where $V^i(t)$ is a generic homeostatic variable, $\tau_i$ its time constant, and $V^i_M$ and $V^i_m$ its upper and lower boundary values, respectively.

Equation 1:

$$\frac{dV^i(t)}{dt} = \frac{V^i(t)}{\tau_i}$$

Equation 2:

$$V^i(t) = V^i_m + (V^i_M - V^i_m)\exp(\frac{-t}{\tau_i})$$

*Motivations or drives*[1] are abstractions denoting urges to action based on bodily needs related to self-sufficiency and survival. They monitor the levels of the homeostatic variables and, when an error is detected, they initiate a process to correct this error – in our case, the selection and execution of an appropriate behaviour, which involves interactions with an object. That behaviour will be executed only if the relevant object is present; otherwise, a search for that object will be initiated. In this study, we have used three motivations – hunger (which controls nutrition), fatigue (controlling stamina), and curiosity (controlling restlessness). At each time step, each motivation is assigned an intensity (activation level) proportional to the magnitude of the error of its controlled variable. Several motivations can be active at the same time (and therefore competing to be satisfied) with varying degrees of intensity. In the general case, drives will depend on more than one homeostatic variable, e.g., hunger may depend on the levels of nutrition and stress. Equation 3 gives the generic behaviour of a motivation $M^i(t)$ in terms of the homeostatic variables and equation 4 the particular case where a motivation depends on a single variable. $V^i_{op}$ is the optimal value for the generic homeostatic variable, and $a_i$ and $b_i$ are constants set by the designer.

Equation 3:

$$M^i(t) = \sum_i a_i(V_{op}{}^i - V^i(t)) + \sum_i b_i(\frac{dV^i(t)}{dt})$$

Equation 4:

$$M^i(t) = a_i(V_{op}{}^i - V^i_m) + \\ + (a_i(V^i_M - V^i_m) + \frac{b_i}{\tau_i})\exp(\frac{-t}{\tau_i})$$

*Behaviours* are coarse-grained subsystems (embedding simpler actions) that implement different competencies. The execution of a behaviour modifies (increases or decreases[2]) the levels of particular internal variables, therefore affecting (contributing to or hampering) the satisfaction of drives. In the general case, different behaviours can contribute to satisfy the same drive, but in our simplified model each motivation can be satisfied

[1] Motivations and drives are not equivalent. Motivations are complex abstractions involving processes of different types, drives being one of them. In this simplified model, we only take into account drives arising from internal survival-related needs. Therefore, we will use these terms interchangeably.

[2] In the experiments reported in this paper, a constant increment/decrement (of 0.2) has been chosen, but it could also depend on the nutritional value of the food, on the quality of the shelter or on the interestingness of the object.

by one behaviour only – "eat" (grasp[3] an object) satisfies hunger, shelter satisfies fatigue, and interact satisfies curiosity.

The *arbitration mechanism* in charge of resolving conflicts among competing motivations and choosing the appropriate behaviour to satisfy internal needs follows a winner-take-all approach. Conflicts are thus solved by trying to satisfy only one motivation at a time – that with the highest intensity, i.e., the most urgent need. The "winner" motivation will then select the behaviour that can best satisfy it – in this simplified architecture, this is trivially the only behaviour that can satisfy the drive. For example, if the most urgent drive were hunger, the behaviour to follow would be to eat. Since our behaviours involve interactions with objects, they can only be executed when the appropriate object is present; otherwise, the robot will perform a search for that object.

The outcome of the interaction depends on the object the robot interacts with. Therefore, the robot will have to find the object with the right functionality. In the architecture of (Cañamero, 1997), object functionalities were "hard-wired" in the different behaviours. In this study, the robot will have to learn which object offers the right affordance depending on the behaviour it is trying to perform (for the case of eating, eatable objects will have to be met), and this is carried out by the Affordance Extractor.

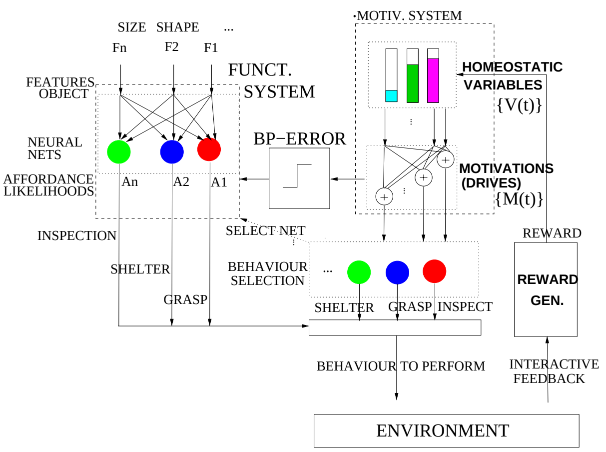


Figure 1: Model Schema

## Affordance Extractor

The role of the affordance extractor (top left of figure 1) is to learn the functionalities that objects in the environment afford to that particular agent interacting with them. This means, from a more engineering point of view, *to relate perceived features of objects with the morphology of the agent to extract the functionality it affords in the context of a particular interaction.* In our model, the mechanism underlying this process relies on the fact that the execution of a particular action with an object varies the level of one or several homeostatic variables, therefore increasing or decreasing its error (its distance with respect to the ideal value or setpoint). For example if the agent is trying to satisfy its hunger by consuming an eatable object, the interaction with that particular object will exert an increasing effect on the level of its nutrition variable. Thus, a positive reward will result from this interaction with the environment. Otherwise, if the object does not *afford* edibility, a negative reward will be encountered by the homeostatic variable, drifting it toward more critical values.

This problem is far from trivial, since even for a fairly limited number of object features, the number of dimensions of the feature space expands dramatically. This discards the use of tables matching combinations of perceived features with their corresponding set of affordance likelihoods, and made us opt for a neural network to estimate these affordances. Neural networks have the advantage that they can easily be trained via back-propagation on the basis of experiences with the environment if a suitable error measure can be formulated. They also save memory and endow the agent with the ability to generalise, i.e., to infer on the basis of previous perceived feature sets whether an object with similar features can afford that particular action to the agent or not. Therefore, we used a set of three 2-2-1 feed-forward neural networks, each of them devoted to estimating the *likelihood*[4] that a particular object, perceived as a set of features, affords the robot to perform a particular action from its behavioural repertoire – "eat" (grasp), shelter and interact. A simple back-propagation algorithm has been implemented to train the three neural networks.

The *learning principle* consists of exploiting the outcome of the interaction episodes by back-propagating, for a small number of cycles, an error signal whenever a motivation experiences a sudden variation, either positive or negative. The error to back-propagate is thus derived by monitoring the homeostatic variables and looking for sudden variations. This error is only back propagated after the end of each interaction episode. If the intensity of the motivation decreases, the agent carried out the correct action, therefore the corresponding affordance network should be trained towards an output of 1.0. Conversely, if the intensity of the motivation increases, the agent performed the wrong action and the network should be trained towards an output of 0.0. The error will then be the difference between the network's output and the aforementioned values. For example, a positive outcome of an **eating** episode causes the nutrition variable to increase and hunger to decrease; therefore, the estimation of grasp affordance likelihood of the object involved in that interaction should be 1.0, otherwise 0.0. The learning process is as follows:

---

[3] Since the robot cannot "eat", we have associated this action with grasping an object with the gripper.

[4] Each neural network estimates the probability of successfully performing that particular action with that object.

1. When an object is encountered, the three neural networks estimate, on the basis of perceived features (size and shape), the likelihood that it affords grasping, shelter and interacting, respectively.
2. Depending on what is the most urgent drive, the appropriate behaviour to compensate the drive is chosen and executed.
3. Depending on what the object affords, the interaction will succeed or fail.
4. Since the execution of a behaviour affects (increases or decreases) the value of the homeostatic variable controlled by the motivation that selected that behaviour, the outcome of this interaction has a (positive or negative) impact on the viability of the robot (the stability of its internal milieu).
5. The change in the values of the drives is next computed.
6. Once a sudden variation of the drive has been detected, the error to back-propagate to the neural network for that behaviour is computed, and the weights of the network updated.

## Experiments

To test our model, we have designed two sets of experiments. The purpose of the first set was to answer our initial question: *could an agent learn affordances by exploiting the outcome of interaction experiments through the monitoring of internal motivations of the agent?* The purpose of the second set was to test whether exploiting previously learnt affordances would enhance the robot's capability to survive.

### Simulation Environment

For these experiments, we have used a freely available 2D Khepera simulator, modified to allow the robot to perceive directly the features of objects in its vicinity. Two features of objects – size and shape – are perceived as a number within the range 0 … 1, as shown in the Shape and Size columns of table 1.

Table 1: Features-Affordances relationship. TS: Triangular Small; TL: Triangular Large; CS: Circular Small; and CL: Circular Large.

| QD | Shape | Size | Shelter | Grasp | Interact |
|---|---|---|---|---|---|
| TS | 0.0 | 0.0 | 0.0 | 1.0 | 1.0 |
| TL | 0.0 | 1.0 | 1.0 | 0.0 | 1.0 |
| CS | 1.0 | 0.0 | 0.0 | 1.0 | 1.0 |
| CL | 1.0 | 1.0 | 1.0 | 0.0 | 1.0 |

The arena where the simulated robot runs measures $1m^2$. It contains objects of several types, each affording particular actions (grasp, shelter or interact) with different likelihoods, as shown in table 1: large objects afford shelter, small objects afford grasping, and any object can be used to interact (see figure 2). The shapes have been approximated; therefore, a small circular object is rectangular, and a large circular is octagonal. The robot is equipped with 8 IR sensors disposed radially, each of them with a sensitivity range of approx. 5cm.

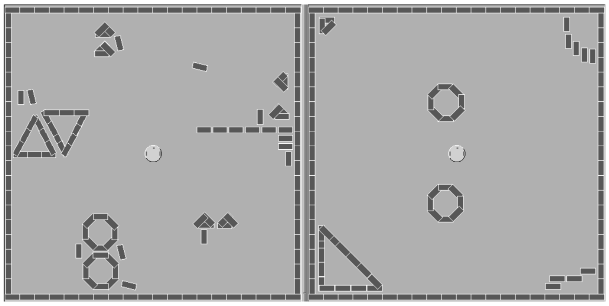

Figure 2: World 1 (left) and World 2 (right).

Scarcity and abundance of resources have been reflected in two different scenarios (figure 2). In the scarce scenario (World 2), it is highly likely that long distances would have to be covered before a suitable object is met, and therefore survival is difficult. In the abundant world (World 1) the converse is the case.

### First Experiment Set

This set of experiments investigates whether simple affordances like grasp, interact or shelter can be learnt by relating fast variations of internal homeostatic variables and drives with the outcome of interactions with objects. As previously mentioned, learning of affordances takes place in three 2-2-1 feed-forward neural networks, each one of them devoted to estimating a single affordance – either grasp, shelter or interact.

For this set of experiments, the set of homeostatic variables have been initialised to their mid-value (0.5). Also, since each drive has been related to a single homeostatic variable (nutrition to hunger, fatigue to restlessness and curiosity to interaction), the coefficients $a_i$ in equation 3 that relate the drive $M^i(t)$ to the set of homeostatic variable $V^i(t)$ are 0.0 except for the aforementioned cases, for which $a_i$ is 1.0. The coefficients $b_i$ have been initially set to 0.0 to make the expression of the drive the simplest possible. Finally, the weights of the neural networks have been initialised to random values between 0 and 1, and the simulation parameters for the homeostatic variables are shown in table 2. The results for the case of **grasping** and **shelter** affordance likelihood estimation are shown in figures 3 and 4 respectively. Both of them represent the last 500 steps of a 300,000 step simulation.

We observe that at a late stage of the learning experiment, the perception of each object is matched to the right affordances according to table 1. Relating the object type and the grasping affordance graphs (top and bottom) in figure 3, we observe that whenever a small object is encountered, the affordance likelihood is close to 1.0, otherwise close to 0.0. Some exceptions can be noticed, and are due to the fact that the output of the grasping network is frozen whenever the grasping behaviour is not being run.

Table 2: Parameters for the homeostatic variables.

| Param. | Nutrition | Stamina | Restlessness |
|---|---|---|---|
| $\tau_i$ | 1E-4 | 1E-4 | 1E-3 |
| $Vop$ | 0.9 | 0.2 | 0.1 |
| $V_M$ | 1.0 | 1.0 | 1.0 |
| $V_m$ | 0.0 | 0.0 | 0.0 |

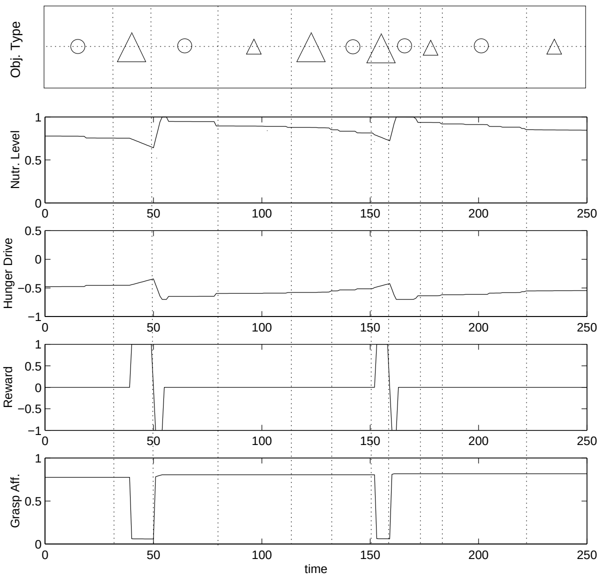


Figure 3: Grasp likelihood learning episode. The signals are (top to bottom): object type, nutrition variable, hunger drive, learning reward and grasp likelihood.

Similarly, from the object type and sheltering affordance graphics in figure 4, it is evident that only large objects afford shelter. Therefore, objects that afford shelter, grasping and/or interaction, can then be correctly detected without interacting with them, by simply relating their features with the successful or failed previous interactions with the same or with similar objects.

## Second Experiment Set

The second set of experiments aims to test whether agents that have already learnt what the different objects can afford them, based on their previous interactions with them, can perform better than agents lacking this knowledge. To this end we have run 20 simulations in two different environments, one with scarce resources and one with abundant resources. The same parameters as for the first set of experiments were used.

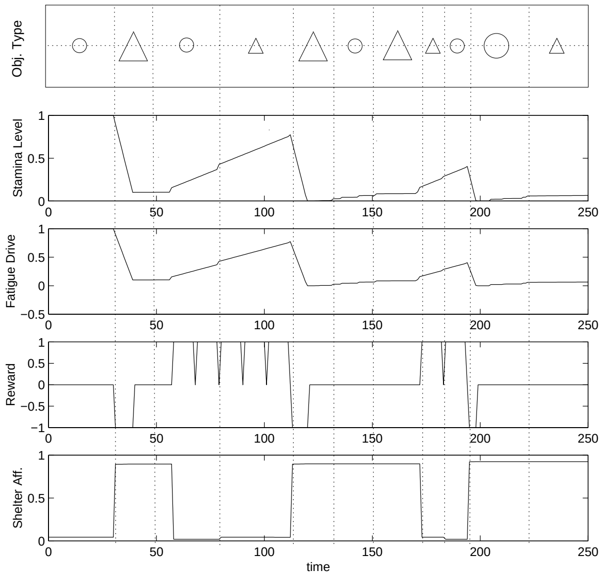


Figure 4: Shelter likelihood learning episode. The signals are (top to bottom): object type, stamina, fatigue drive, learning reward and shelter likelihood.

One set of agents chooses their behaviour only by monitoring their internal variables and by selecting the behaviour that best satisfies the most urgent need (the motivation with highest intensity), regardless of the type of object present in the environment. In other words, these agents do not have any knowledge of the objects they interact with, nor what they can afford; therefore, they perform a *blind* interaction that may result in either success or failure, depending on whether the object affords the required functionality or not. Therefore, their survival in this case depends on there being frequent fortuitous meetings with objects, which offer the right affordance whenever it is needed.

Conversely, the second type of agents does use the results of the three previously trained neural networks (in addition to monitoring their internal variables and selecting the behaviour that best satisfies their most urgent need) to decide whether it will be of some advantage to interact with a perceived object or not. The networks provide the agent with the affordance likelihoods of the objects. If they are not appropriate to satisfy the most urgent drive of the agent, the object will be immediately disregarded and avoided, and a new one will be searched for.

Figure 5 shows the average lifespan for both types of agents. The simulations have been run in the two worlds mentioned above (see figure 2). World 1 is a very rich world, where the agent can easily find food and shelter, and therefore satisfy its own necessities quite easily. World 2 instead is very scarce; therefore, repeated events of finding food and resting places become more infrequent. In consequence, agents decease sooner. Both these expectations are evident in figure 5. Bars are paired by the world where the simulations took place, the left are the results of the simulations run in world 1 (abundant resources), and the right are the results of simulations run in world 2 (scarce resources). BLIND are the first type of agents (those without knowledge of the objects and their functionalities), and AF the second (having previously learnt what the objects can afford). The use of functional knowledge of objects has increased the statistical lifespan of the agents by nearly a 50% in the scarce world, and nearly a 100% in the abundant.

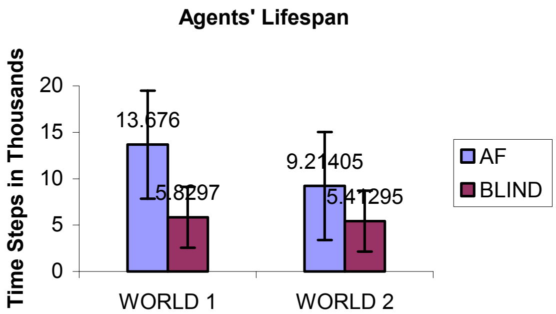


Figure 5: Agents' Lifespan.

## Discussion and Future Work

In this paper, our goal has been to learn affordances in order to exploit this knowledge for action selection, building a functional memory of objects in the environment. Agents endowed with the capability to know that an object may serve to fulfil a particular need can use this knowledge to decide whether it may be worth interacting with a particular object *prior* to any interaction.

Our experiments suggest that functional information about objects can be related to the agent's internal goals (drives) and used to satisfy these. The procedure we have proposed to compensate the agents' internal variables consists of executing a behaviour, which involves an object. If the object the agent interacts with offers the particular affordance that allows the successful execution of the behaviour, there is a high probability that the motivation (drive) will be satisfied, and therefore the needs of the agent will be diminished. Otherwise the effect will be the opposite. As shown by our second set of experiments, agents that already know the function objects afford survived much longer. This is due to a better interaction policy resulting in an increased efficiency of any object encounter and an increased frequency of successful interactions.

Moreover, the neural technique used to relate stimuli to affordance likelihoods should also endow the agent with more flexible and therefore more adaptive behaviour capabilities. Neural networks have the power to generalise over similar sets of stimuli, which may be caused by objects never perceived before, which however afford functionalities similar to those previously estimated. Our next step will therefore be to extend our environment with new types of objects and to measure the capability of an agent to exploit other (similar) objects to satisfy its internal needs.

Lastly, will embed this schema in a progressively more biologically-plausible behaviour selection architecture, starting with experiments in which motivations will depend not only on the values of the drives, but also on the influence of external stimuli – the values of the affordances themselves. We expect that an integration of the aforementioned systems under biological inspiration would also permit us to perform some predictions of real behaviour and of some of the disorders related to motivations and to behaviour selection.

## Acknowledgements


We are grateful to Andrew Gillies, Orlando Avila-García and the members of the IPAB for discussions on related topics, and to Yuval Marom for helping to solve technical problems with the simulator and the experiments. This research is funded by the Graduate School, University of Edinburgh.


## References


Ashby, W. R. (1952). *Design for a Brain. The Origin of Adaptive Behaviour*. London: Chapman & Hall, Ltd.

Cañamero, L.D. (1997). Modeling Motivations and Emotions as a Basis for Intelligent Behavior. In W.L. Johnson, ed., *Proc. First Intl. Conf. on Autonomous Agents*, pp. 148-155. New York: ACM Press.

Cos, I. and Hayes, G. (2002). Behaviour Control Using a Functional and Emotional Model. In Proc. 7th International Conference on Simulation of Adaptive Behavior. Cambridge, MA: MIT Press.

Gadanho, S.C., and Hallam, J. (2001). Emotion-Triggered Learning in Autonomous Robot Control. *Cybernetics and Systems*, 32(5): 531-559.

Gibson, J. J. (1986). The Ecological Approach to Visual Perception. Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ.

Guazzelli, A., Corbacho, F. J., Bota, M. and Arbib, M.A. (1998). Affordances, Motivation, and the World Graph Theory. *Adaptive Behavior*, 6(3/4): 435-471.

Meyer, J.A. (1995). Artificial Life and the Animat Approach to Artificial Intelligence. In M. Boden, ed., *Artificial Intelligence*. Academic Press.

Nehaniv, C. and Dautenhahn, K. (1998). Mapping between Dissimilar Bodies: Affordances and the Algebraic Foundations of Imitation. In J. Demiris and A. Birk, eds., *Proc. European Workshop on Learning Robots*, pp. 64-72. Edinburgh, UK, 20 July 1998.

Velásquez, J. D. 1998. Modeling Emotion-Based Decision-Making. In *Emotional and Intelligent: The Tangled Knot of Cognition. Papers from the 1998 AAAI Fall Symposium*, pp. 164-169. Technical Report FS-98-03. Menlo Park, CA: AAAI Press.