HOMEOSTATIC PLASTICITY IN ROBOTS

FROM DEVELOPMENT TO OPERANT CONDITIONING TO HABIT FORMATION

Xabier E. Barandiaran*
Ezequiel Di Paolo*#
** Center for Computational Neuroscience and Robotics,*
Dept. of Informatics, University of Sussex, Brighton, UK.
\# Ikerbasque, Basque Science Foundation,
Information, Autonomy and Systems Research Group,
Dept. of Logic and Philosophy of Science, UPV-EHU, Spain

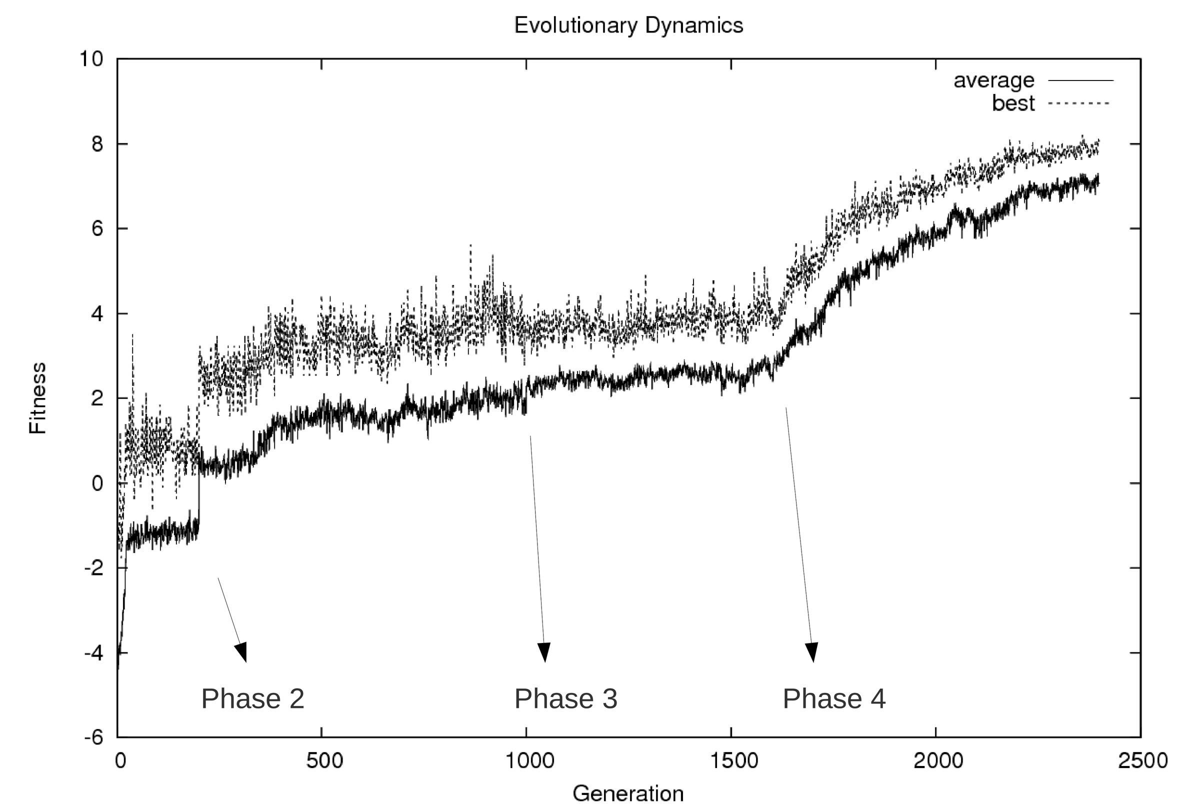
University of Sussex
Informatics
CCNR Centre for Computational Neuroscience and Robotics
ias-research Information and Autonomous Systems
Universidad del País Vasco
Euskal Herriko Unibertsitatea
The University of the Basque Country

Abstract

We present an evolutionary robotics model of an agent controlled by a homeostatically plastic neural network where the connectivity is modified by a pre-synaptic Hebbian rule when the firing rate of neurons goes out of pre-specified bounds. The very same underlying architecture supports a developmental process, an operant conditioning task and the spontaneous formation of sensorimotor habits (never selected during the evolutionary phase). The task consists on distinguishing two different colored food sources with changing profitability (food source types alternate in random intervals from “profitable” to “poisonous”). The agent has two arrays of sensors (one for each color) and an additional sensor to evaluate the profitability of the food-source. The control architecture is a fully connected Continuous Time Recurrent Neural Network with presynaptic Hebbian rule that is activated if the activity of pre or post-synaptic neuron is either to high or to low. Learning rule parameters are evolved to optimize the operant conditioning task with an additional fitness reward for internal stability. The agents are initialized with synaptic connections set to zero, a phase of development is observed after which the synaptic architecture is stabilized and the robot performs adequately on the operant conditioning task (distinguishing between profitable and toxic food). In addition, the robot can develop new habits in circumstances that were never encountered during evolution: e.g. the presentation of the same colored food source at the right produces a preference for turning to the right or alternating preference to different colors.

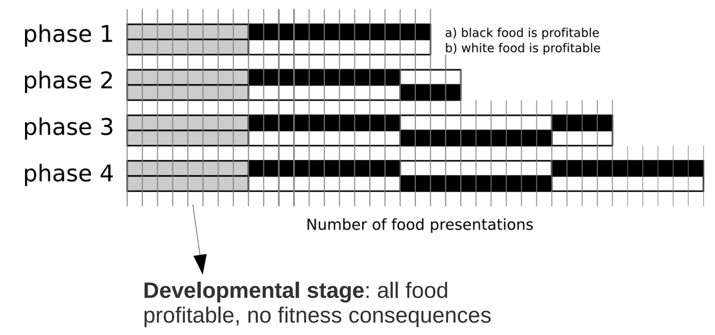
Artificial Evolution

Incremental Evolution, different phases



A Genetic Algorithm is used to evolve the learning rule parameters for the agent. A population of 30 individuals, with mutation rate of 0.05, standard deviation of 0.02, elitism of 2 and crossover.

Incremental evolution is used with 4 different phases with increasing number of food presentations and changes of profitability (see fig. Right)

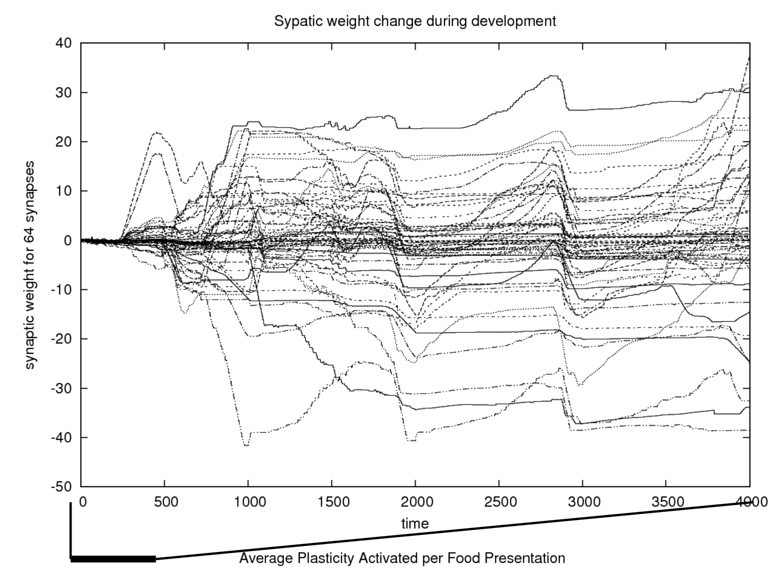


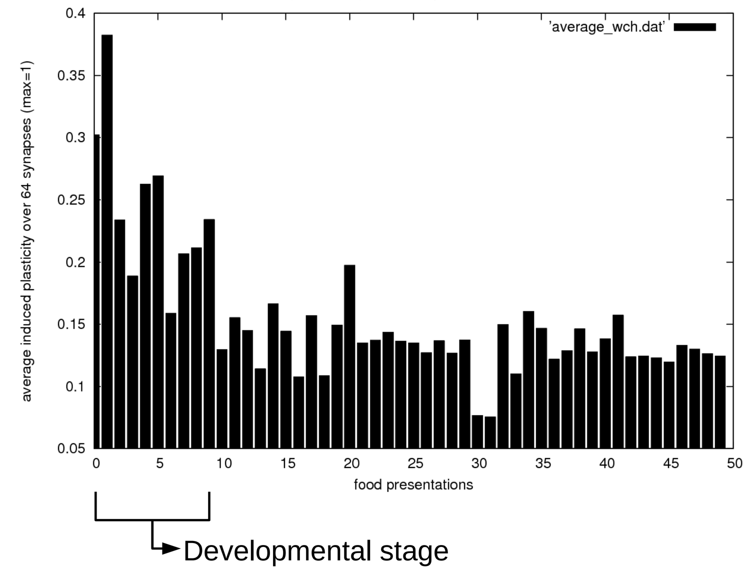
Fitness function: directly proportional to the energy accumulated by the agent during trials, synaptic plasticity consumes energy thus selecting for stability of neural dynamics between homeostatic bounds. The agent is also penalized if it doesn't choose any food.

Results 1: development and operant conditioning

Homeostatic plasticity generates and stabilizes a network configuration during development (left), this configuration is capable to sustain operant conditioning (right)

Development and Plasticity

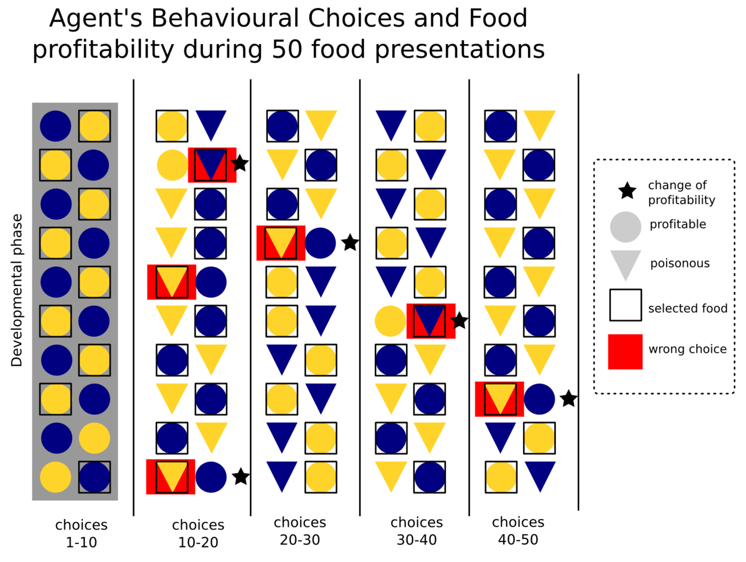




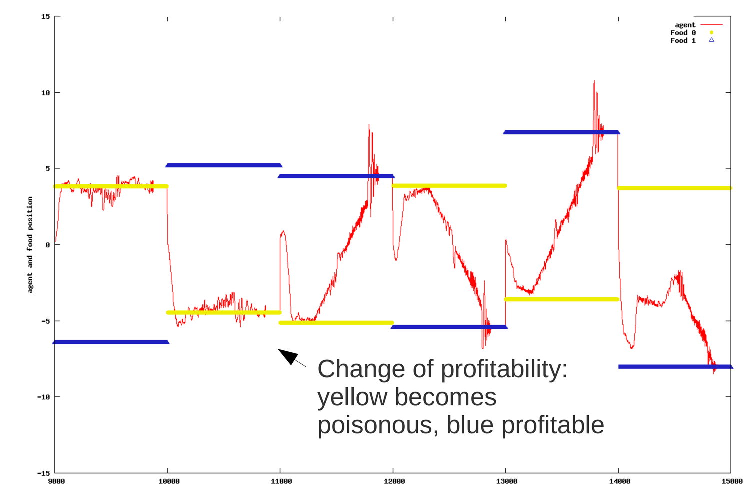
The Agent is presented with two food choices (yellow and blue), profitability of food sources (-10 or 10 energy units) changes every 6-10 presentations. The agent is capable of learning which food is beneficial and acts accordingly. The diagram on the right shows a characteristic chart of the agent's options during a 50 food presentation trial. The agent's behaviour is shown bellow. Statistical average over 20 samples of 50 trials each for food profitability change every 10, 8 and 6 presentations is shown on the bottom-right side.

Operant Conditioning

Agent's Behavioural Choices and Food profitability during 50 food presentations

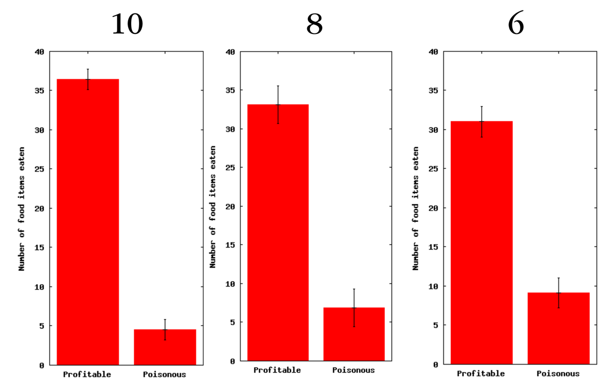


Characteristic Behaviour



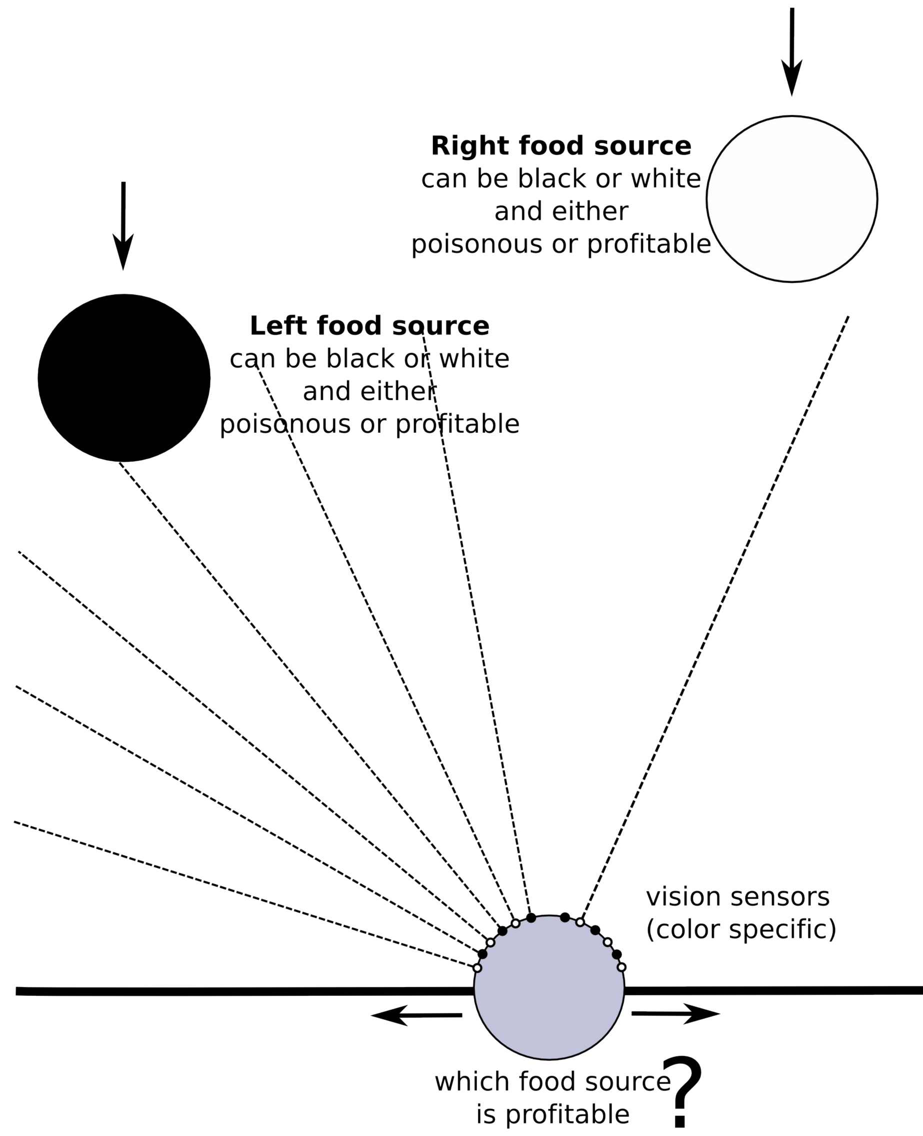
Average Choices over Trials

Average good and bad choices for food profitability changing every 10, 8 and 6 food presentations. Average of 20 samples of 40 food presentations each (+10 of development not counted)



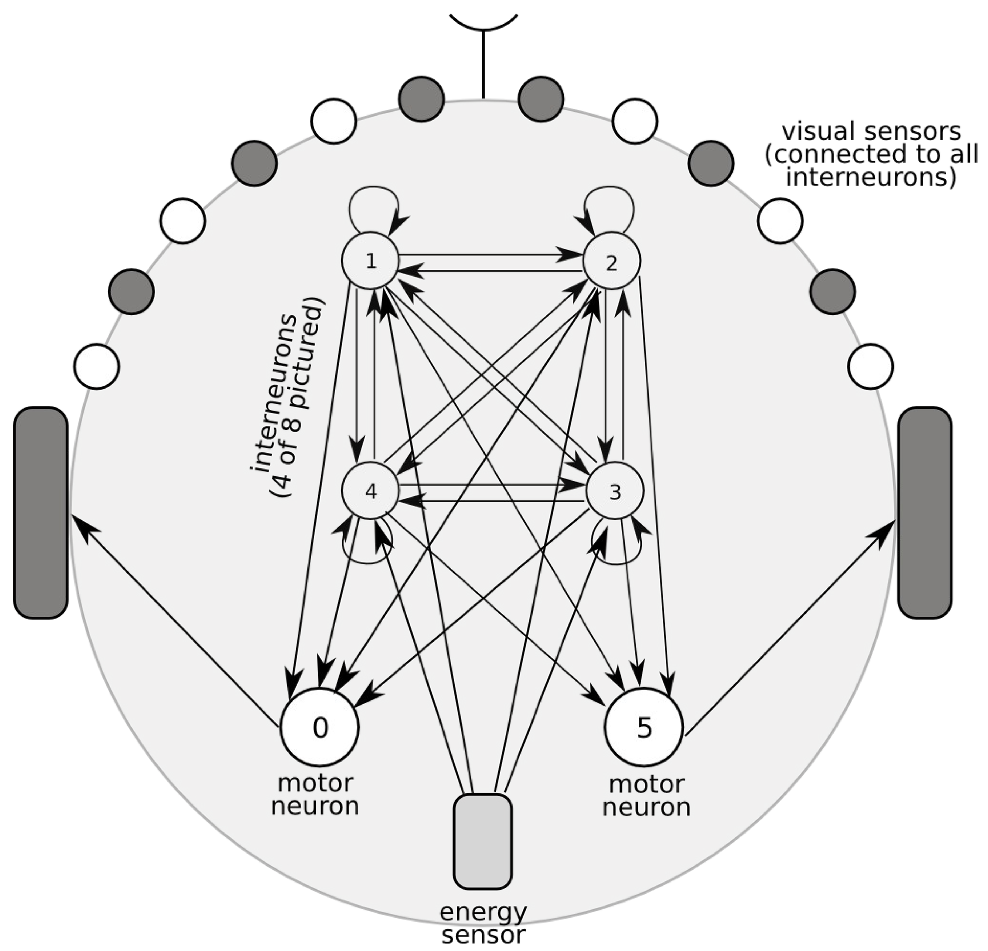
The Task

The agent has to select between two (differently colored) falling food sources. Each food color has an associated energy profitability that the agent can sense (poisonous -10 energy, profitable +10 energy). The agent has to learn this association that changes during its lifetime (typically every 8 food presentations)



The Agent

CONTROLLER: Continuous Time Reucrrent and Totally Connected Neural Network with Homeostatic Plasticity
SENSORS: One energy sensor and 6 laser-like visual sensor for each color.



Neuronal Activation Equation

$$\tau_i \dot{y}_i = -y_i + \sum_{j=1}^{n}(w_{ij} z_j) + g_i \sum_{k=0}^{5} s_{ki} I_k \;;$$

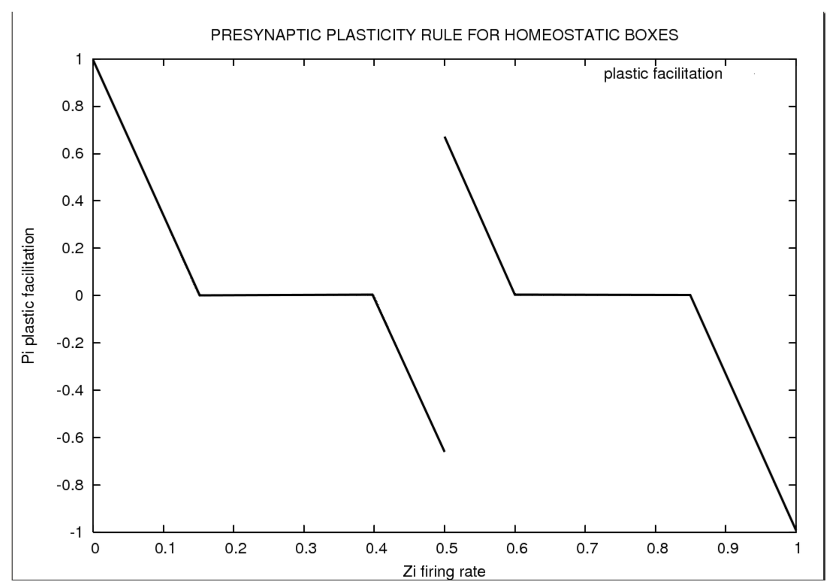
$$\text{where} \quad z_j = \frac{1}{1 + exp(-(y_j + b_j))}$$

where y is the state of each neuron, τ is the time constant, w_{ij} is the connection weight between neuron i and j, z_j is the activation of neuron j, y_j is j's state and b_j a bias term; g_i is a gain (-5,5) applied to the overall sensory input to the neuron, s_{ki} (-5,5) is the input weight from sensor k to neuron i and I_k is the input value of sensor k.

Hebbian Presynaptic Rule and Homeostatic Plasticity Rule

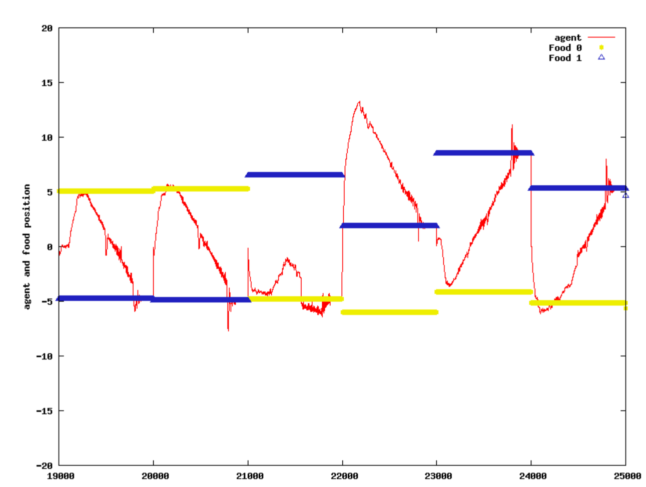
Presynaptic Hebbian: $\Delta w_{ij} = \eta_{ij}\, p_j\, (z_i - \mu_{ij})\, z_j,$

where p_j is the degree of plastic local facilitation explained below, μ_{ij} depends linearly on the value of w_{ij} so that $\mu_{ij} = 0$ if $w_{ij} = max$ and $\mu_{ij} = 1$ if $w_{ij} = min$, η_{ij} is the genetically specified learning rate

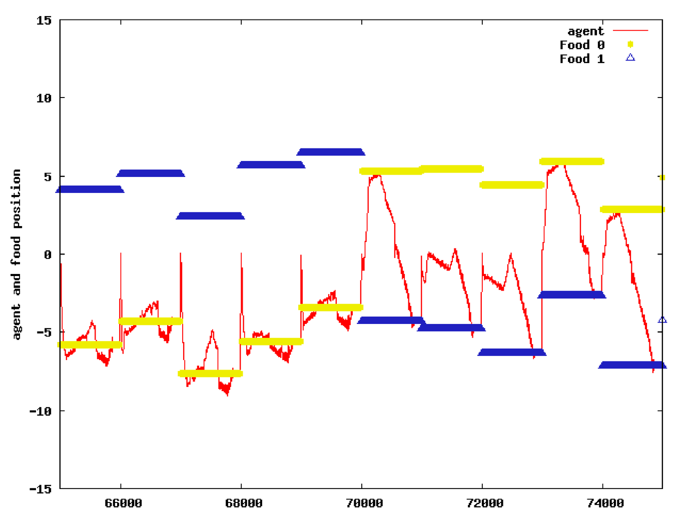


Results 2: spontaneous habit formation

When both food sources are profitable the agent develops a “preference” to choose the food on its right (that this behaviour is not an innate preference can be shown on early trials not biased systematically towards the right)



... 40 food presentations after...



Under different conditions, when no optimality criteria applies in terms of profitability, the history of interactions leads to the formation of particular preferences or habits on the agent, the exact nature of such patterns and their robustness needs to be tested and explored in more detail. Our working hypothesis is that recurrent interactions stabilize a neural configuration that sustains the interactions that generated it.

References

Iizuka, H., & Di Paolo, E. A. (2007). Toward Spinozist Robotics: Exploring the Minimal Dynamics of Behavioral Preference. *Adaptive Behavior*, 15(4), 359-376. doi: 10.1177/1059712307084687

Izquierdo, E., Harvey, I., & Beer, R. D. (2008). Associative Learning on a Continuum in Evolved Dynamical Neural Networks. Adaptive Behavior, 16(6), 361-384. doi: 10.1177/1059712308097316

Phattanasri, P., Chiel, H. J., & Beer, R. D. (2007). The Dynamics of Associative Learning in Evolved Model Circuits. Adaptive Behavior, 15(4), 377-396. doi: 10.1177/1059712307084688

Acknowledgements

Ezequiel Di Paolo would like to acknowledge the financial support of the Ikerbasque Foundation. Xabier Barandiaran is funded by *Programa Nacional de Movilidad de Recursos Humanos del MEC, Plan I-D+I 2008-2011*, Spanish Ministry of Science and Innovation.