# Does Learning Elicit Neuromodulation? Evolutionary Search in Reinforcement Learning-like Environments

Andrea Soltoggio

School of Computer Science
University of Birmingham
Birmingham B15 2TT, United Kingdom
a.soltoggio@cs.bham.ac.uk

**Abstract.** Although the importance of neuromodulation in neural substrates has been widely recognised, the computational role, characteristics and advantages of such models in Artificial Neural Networks are mostly unknown. To investigate this issue, here the autonomous emergence of neuromodulatory structures is considered by means of artificial evolution in reinforcement learning-like environments. By giving evolution the flexibility of selecting and employing modulatory neurons when needed, artificial selection could provide a valuable insight into which environmental characteristics and problems elicit the rise of neuromodulation.

### A Model for Neuromodulation

The model adopted in this study implements neuromodulation as a neuron-specific modulation based on the heterosynaptic plasticity concept [1] already employed in computational models described in [2, 3]. A synaptic update $\Delta w$ is equal to a plasticity term $f(V(t), P(t))$ [1] times a neuron-specific modulatory signal $m(t)$, where $V(t)$ and $P(t)$ are pre- and postsynaptic activities.

The neural model includes two types of neurons: a standard and a modulatory type. The output activity of modulatory neurons targets specific standard neurons whose plasticity (i.e. the rate of change in synaptic connections) is regulated accordingly to the intensity of the modulatory signal.

The emergence of neuromodulatory dynamics is investigated by allowing an evolutionary algorithm to build tentative neural architectures. Standard and/or modulatory neurons are randomly selected as building blocks to be used in conjunction with an Evolution Strategy that evolves the topology and weights of neural networks. The evolutionary algorithm relies on a reward-based fitness to select good solutions. The model and algorithm were devised on the hypothesis that adaptivity and learning can be achieved by means of neuromodulation when coping with the uncertain environments presented in the next section.

[1] The generic plasticity rule used in this study is expressed by $[A \cdot V(t)P(t) + B \cdot V(t) + C \cdot P(t) + D]$, where A..D are tuneable network parameters.

### Reinforcement Learning-like Problems

Non-stationary *N-armed bandit problems* [4] require adaptive behaviour to cope with variable reward contingencies during lifetime. In such environments rewards that are obtained as consequences of certain actions change over time, therefore causing in the long term beneficial actions to become detrimental (or less advantageous) and vice versa. Two instances of such problems are proposed: 1) $N$ arms are sequentially presented as input to the network. The agent chooses the current option (arm) by triggering an output. When a choice is taken, a reward is provided as an additional input and its value is added incrementally to the fitness of the individual. 2) In a T-maze task a control network chooses left or right at a turning point signalled by a high value of a sensory input. At the end of the passage of variable length, the network is rewarded with a delayed reward based on the previous choice at the turning point. In both cases a lifetime consists of a number of trials. The association action-reward (i.e. the quantity of reward that follows an action) is changed at least once during the lifetime. The environments described provide basic reinforcement learning-like problems.

### Evolved Networks

Experimental results show that networks started making use of modulatory neurons during evolution, and displayed modulatory dynamics to increase the performance (total amount of collected reward) in both problems 1) and 2). Modulatory signals are triggered by particular events (such as a reward or turning points) and regulate the synaptic update. The sign of synaptic update appears to be determined according to a measure of good/bad surprise upon reward collection. Remarkably, even with the delayed reward that follows the action in the t-maze task, evolution found well performing solutions. When the variability of action-reward associations was removed (i.e. rewards are fixed and lifetime learning is not required), the networks did not evolve neuromodulation.

These initial results indicate that the autonomous evolution of modulatory dynamics is not only possible but also induced when the environment provides uncertain reward contingencies and therefore requires online learning. The topic of future studies is the analysis of the evolved modulatory structures that emerge in such uncertain environments. The investigation should also address issues such as computational capabilities and evolvability.

## References

1. Bailey, C.H., Giustetto, M., Huang, Y.Y., Hawkins, R.D., Kandel, E.R.: Is heterosynaptic modulation essential for stabilizing hebbian plasticity and memory? Nature Reviews Neuroscience **1**(1) (October 2000) 11–20
2. Niv, Y., Joel, D., Meilijson, I., Ruppin, E.: Evolution of Reinforcement Learning in Uncertain Environments: A Simple Explanation for Complex Foraging Behaviours. Adaptive Behaviour **10**(1) (2002) 5–24
3. Soltoggio, A., Dürr, P., Mattiussi, C., Floreano, D.: Evolving Neuromodulatory Topologies for Reinforcement Learning-like Problems. In: Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2007. (2007)
4. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, USA (1998)