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

Andrea Soltoggio
University of Birmingham
UK

ECAL 2007 - 10-14 September. Lisbon, Portugal
Workshop on Dynamics of Learning Behaviour and Neuromodulation

ABSTRACT - The autonomous emergence of **neuromodulatory topologies** is considered by means of artificial evolution in reinforcement learning-like environments.

Two models of neurons (standard and modulatory) can be present in the network in arbitrary number and topological connections.

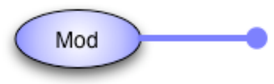
Feature selection on modulatory and standard neurons is performed to provide a valuable insight into which environmental characteristics and problems elicit the rise of neuromodulation.

1 A Model of Neuromodulation

Two types of neurons are modelled.

A **standard** type

and a **modulatory** type



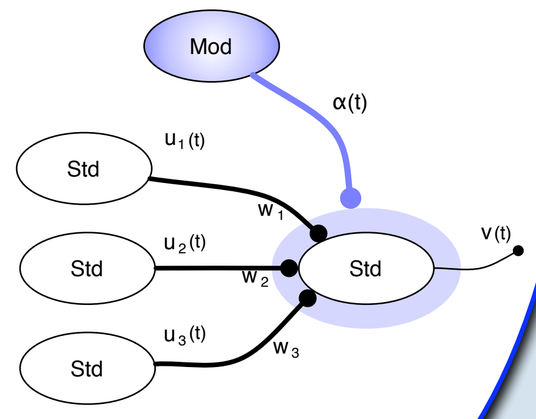
- Modulatory neurons activate plasticity at the target neuron.
- Plasticity can be implemented according to any rule.
- The effect of the rule will be **proportional to the modulatory signal**.

Example

$$\Delta \mathbf{w} = \alpha(t) \cdot f(\mathbf{u}(t), v(t))$$

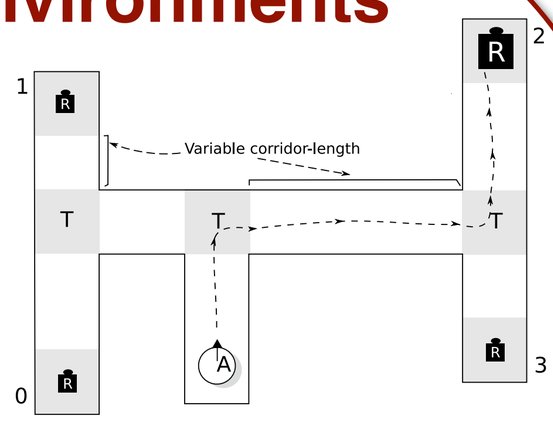
where

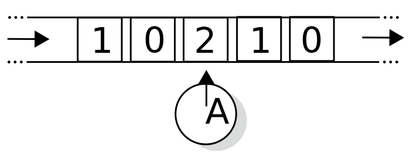
- $\Delta \mathbf{w}$ is the weight update
- $\alpha(t)$ the modulatory signal at the postsynaptic neuron
- $f(\mathbf{u}(t), v(t))$ is a plasticity rule implemented as $[A\mathbf{u}(t)v(t) + B\mathbf{u}(t) + Cv(t) + D]$ with A-D tuneable/evolvable parameters



2 Learning Problems : Uncertain Environments

- Non-stationary n-armed bandit problems represent uncertain rewarding environments.
- Variability of rewarding conditions during lifetime requires online learning to maximise the reward intake.



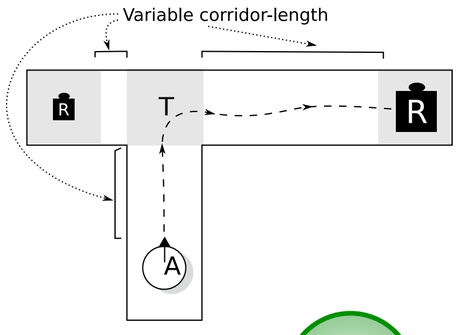


N-armed bandit problem.

- the agent selects an arm.

- each arm returns a different value of reward.

- in the non-stationary version, arms change their reward during lifetime: arms that were good choices initially become poor and vice versa.



T-maze.

- in the non-stationary version, the position of the high reward changes during lifetime.

Multiple T-maze. The high reward is located at one of the 4 maze-ends.
Here the agent needs to explore the maze, remember a sequence of two actions that lead to the high rewarding maze-end once found, and go back to exploration when the location change.

Conditions for the Emergence of neuro-modulation

3 Evolutionary Feature Selection

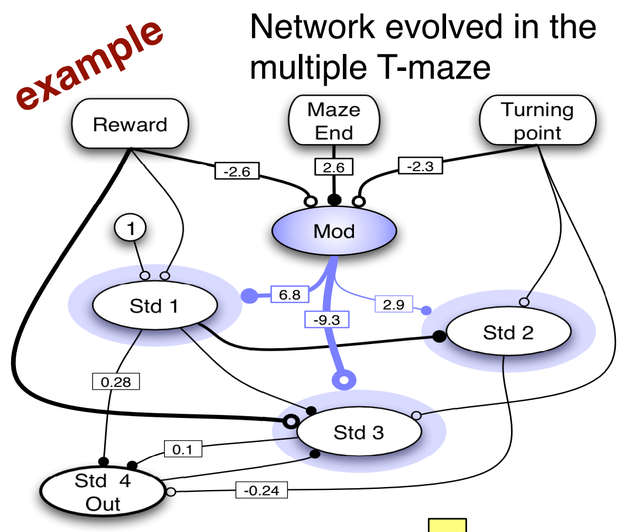
Random networks are initialised and then evolved by an Evolutionary Algorithm that **designs topologies and weights**. The **fitness function** is the total amount of **reward** collected during a lifetime.

...so does learning elicit neuromodulation? YES

- **If learning IS required: networks discover the use of neuromodulation throughout evolution and learn to adapt their strategy to the changing reward contingencies.**
  - **Modulation tells WHEN and WHAT to learn.**
  - **Remarkable performance in terms of learning, memory and adaptation.**
- **If learning IS NOT required (stationary n-armed bandit problems) modulation does not emerge.**

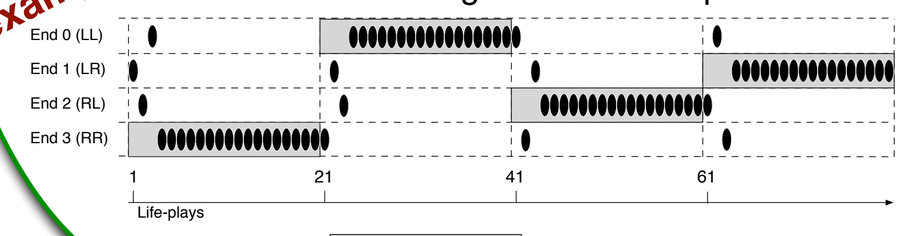
example

Network evolved in the multiple T-maze



example

Behaviour of an agent in the multiple T-maze



Analysis : the **modulatory signal** appears to be an error-prediction/ surprise signal, which mimics dopamine activation in biological brains

Acknowledgement
This work was inspired by the collaboration with Dario Floreano, Claudio Mattiussi and Peter Dürr at the Laboratory of Intelligent Systems at EPFL, Lausanne, CH.

References
Soltoggio, A., Dürr, P., Mattiussi, C. and Floreano, D. "*Evolving Neuromodulatory Topologies for Reinforcement Learning-like Problems*", In the Proceedings of the Congress on Evolutionary Computation 2007