Evolutionary Advantages of Neuromodulated Plasticity in Dynamic, Reward-based Scenarios

Andrea Soltoggio[1], John A. Bullinaria[2], Claudio Mattiussi[3], Peter Dürr[4], and Dario Floreano[5]

[1,2]School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK
{a.soltoggio,j.a.bullinaria}@cs.bham.ac.uk
[3,4,5] Laboratory of Intelligent Systems, EPFL, Lausanne, CH
{claudio.mattiussi,peter.duerr,dario.floreano}@epfl.ch

Abstract

Neuromodulation is considered a key factor for learning and memory in biological neural networks. Similarly, artificial neural networks could benefit from modulatory dynamics when facing certain types of learning problem. Here we test this hypothesis by introducing modulatory neurons to enhance or dampen neural plasticity at target neural nodes. Simulated evolution is employed to design neural control networks for T-maze learning problems, using both standard and modulatory neurons. The results show that experiments where modulatory neurons are enabled achieve better learning in comparison to those where modulatory neurons are disabled. We conclude that modulatory neurons evolve autonomously in the proposed learning tasks, allowing for increased learning and memory capabilities.

Introduction

The importance of modulatory dynamics in neural substrates has been increasingly recognised in recent years. The notion that neural information processing was fundamentally driven by the electrical synapse has been replaced by the more accurate view that modulatory chemicals play a relevant computational role in neural functions (Abbott and Regehr, 2004). Experimental studies on both invertebrates and vertebrates (Burrell and Sahley, 2001; Birmingham and Tauck, 2003) suggest that neuromodulators such as Acetylcholine (ACh), Norepinephrine (NE), Serotonin (5-HT) and Dopamine (DA) closely affect synaptic plasticity, neural wiring and the mechanisms of Long Term Potentiation (LTP) and Long Term Depression (LTD). These phenomena are deemed to affect both short and long term configuration of brain structures, and therefore have been linked to the formation of memory, brain functionalities and considered fundamental in learning and adaptation (Jay, 2003).

The realisation that the Hebb's synapse (Cooper, 2005) does not account entirely for experimental evidence on synaptic modification has brought growing focus on modulatory dynamics. Associative learning as classical and operant conditioning, and various forms of long-term wiring and synaptic changes seem to be based on additional mechanisms besides the Hebbian synapse. Studies on mollusks like the *Aplysia californica* (Roberts and Glanzman, 2003) have shown neuromodulation to regulate classical conditioning (Carew et al., 1981; Sun and Schacher, 1998), operant conditioning (Brembs et al., 2002) and wiring in developmental processes (Marcus and Carew, 1998).

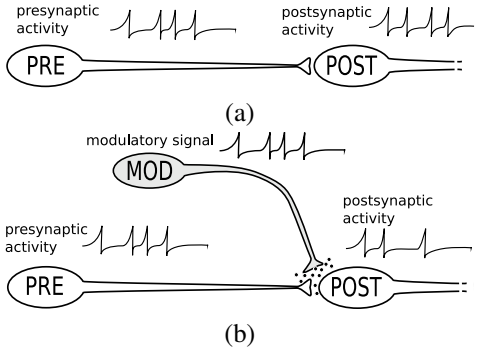

Figure 1: (a) Hebbian plasticity: the connection strength is updated as function of pre- and postsynaptic activity only. (b) Heterosynaptic mechanism, or neuromodulation: the connection growth is mediated by neuromodulation, i.e. the amount of modulatory signal determines the response to Hebbian plasticity. The dots surrounding the synapse represent the concentration of modulatory chemicals released by the modulatory neuron.

Classical Hebbian plasticity refers to synapse modification based on pre- and postsynaptic activities. A presynaptic and a postsynaptic neuron are involved in the process. Neuromodulation, on the other hand, involves a third modulatory neuron that diffuses chemicals at target synapses as illustrated in Figure 1. A unique working mechanism for neuromodulation has not been identified due the large variety of modulatory dynamics involving different chemicals, stimuli, brain areas and functions. However, Bailey et al. (2000a) suggest that heterosynaptic modulation is essential for stabilising Hebbian plasticity and memory. That review paper outlines the nonlinear effect of modulatory signals; when neuromodulation is coupled with presynaptic stimuli, it results in the activation of transcription factors and pro-

tein synthesis during synaptic growth. This in turn leads to durable and more stable synaptic configuration (Bailey et al., 2000b). The underlying idea is that the synaptic growth that occurs in the presence of modulatory chemicals is long lasting, i.e. has a substantially longer decay time than the same growth in absence of modulation.

At a system level, the release of modulatory chemicals has been linked to learning. In (Schultz et al., 1993), dopamine activation patterns recorded in monkeys' brains followed a measure of prediction-error during learning tasks in classical conditioning. Following studies have linked modulatory activity with learning, reward and motivation (Schultz et al., 1997). How cellular mechanisms of synaptic growth and global patterns of neural activation relate has not been unveiled yet, however, growing evidence indicates a direct link between cellular and system level.

Advances in biology have resulted in the formulation of computational models (Fellous and Linster, 1998; Doya, 2002), which try to capture the computational role and significance of neuromodulation. Artificial Neural Networks (ANNs) have also been extended to include forms of neuromodulation. Short term memory by means of neuromodulation was investigated in (Ziemke and Thieme, 2002) where a robot navigated in a T-maze and remembered turning directions according to visual clues in the maze. Improved evolvability in neural controller was shown with the use of GasNet (Smith et al., 2002), although these networks have modulated output functions rather than synaptic plasticity. Learning and adaptivity were shown in navigation tasks in (Sporns and Alexander, 2002) where a neural architecture was manually designed to update weights according to reinforcement signals. Improved performance and adaptation by means of neuromodulation were shown on a real robot in (Kondo, 2007).

Because synaptic plasticity is often considered a way to achieve adaptation and learning, many benchmark problems for neuromodulation are based on uncertain environments. A single modulatory neuron was used to evolve learning behaviour for a simulated foraging task in uncertain environments (Niv et al., 2002). In that study, a simulated flying bee was capable of choosing the higher rewarding flower in a flower-field with changing reward conditions. This experimental setting was chosen also by Soltoggio et al. (2007) to show that modulatory architectures could freely develop throughout evolution to achieve higher performance than in (Niv et al., 2002). These previous two studies (Niv et al., 2002; Soltoggio et al., 2007) support the idea that neuromodulation plays a central role in regulating plasticity when variable environmental conditions require a change in policies of control. However, despite the recent computational models, studies on the precise computational advantages of neuromodulation are very limited. In addition, there are few working models of learning in networks.

This work addresses the issue by analysing the spontaneous evolution of neuromodulation in T-maze navigation tasks, and assesses the advantage of modulatory over traditional networks when dealing with learning problems. In these environments, an agent navigates a T-maze to discover the location of a reward. The location of the reward is not kept fixed, but changes during the agent's lifetime, fostering the development of adaptive and learning behaviour. A comparison of modulatory and non-modulatory networks is presented, where the results suggest an evolutionary advantage in the use of neuromodulation.

The next section describes the computational model of modulatory neurons. Following, the T-maze learning problems are presented before illustrating the evolutionary algorithms by means of which networks are evolved. The Results section presents experimental results and discussion. The paper ends with final remarks in the Conclusion.

Modulatory Neurons

In Artificial Neural Networks (ANNs) with only one type of neuron, each node exerts the same type of action on all the other nodes to which it is connected. Typically this consists in the propagation of activation values throughout the network. However, given the variety of neurons and chemicals in the brain, it is conceivable to extend ANNs by devising different types of neurons. Here, we introduce a special kind of neuron that we define *modulatory neurons*: accordingly, nodes in the network can be either *modulatory* or *standard* neurons (Soltoggio et al., 2007). In doing so, the rules of interactions among neurons of different kinds need to be devised. Assuming that each neuron can receive inputs from neurons of both types, each node in the network will store the intensity of inputs deriving from each sub-system, i.e. from the sets of neurons belonging to different kinds. This principle is comparable to the presence of many kinds of receptors in biological neurons.

Because two types of neurons are considered here, *standard* and *modulatory*, each neuron i regardless of its type has an internal value for a *standard activation* a_i and a value for a *modulatory activation* m_i. The two activations are computed by summing the inputs from the two subsets of neurons in the network

$$a_i = \sum_{j \in Std} w_{ji} \cdot o_j \quad , \tag{1}$$

$$m_i = \sum_{j \in Mod} w_{ji} \cdot o_j \quad , \tag{2}$$

where w_{ji} is the connection strength from neuron j to i, o_j is the output of a presynatic neuron computed as function of the standard activation $o_j(a_j) = tanh(a_j/2)$.

The novel aspect in the model is the modulatory activation that determines the level of plasticity for the incoming connections from standard neurons. Given a neuron i, the

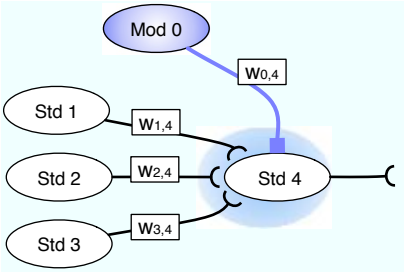

Figure 2: Ovals represent standard and modulatory neurons labeled with *Std* and *Mod*. A modulatory neuron transmits a modulatory signal – represented as a coloured shade – that diffuses around the incoming synapses of the target neuron. Modulation affects the learning rate for synaptic plasticity on the weights $w_{1,4}$, $w_{2,4}$ and $w_{3,4}$ that connect to the neuron being modulated.

incoming connections w_{ji}, with $j \in Std$, undergo synaptic plasticity according to the equation

$$\Delta w_{ji} = tanh(m_i/2) \cdot \delta_{ji} \quad (3)$$

where δ_{ji} is a *plasticity term*. A graphical interpretation is shown in Figure 2. The idea in Equation 3 is to model neuromodulation with a multiplication factor on the plasticity δ of individual neurons being targeted by modulatory neurons. A modulation of zero will result in no weight update, maintaining the weights at the current value; higher levels of modulation will result in a weight change proportional to the modulatory activity times the plasticity term.

In this work, the synaptic plasticity is described by the rule

$$\delta_{ji} = \eta \cdot [Ao_j o_i + Bo_j + Co_i + D] \quad (4)$$

where o_j and o_i are the pre- and postsynaptic neuron outputs, η is the learning rate, and A, B, C, and D are tunable parameters. Equation 4 has been used in previous studies of neuromodulation (Niv et al., 2002; Soltoggio et al., 2007). Its generality is given by the presence of a correlation term A, a presynaptic term B, a postsynaptic term C and a constant D. D allows for strict heterosynaptic update, meaning synaptic update in absence of pre- or postsynaptic activity. The use and tuning of one or more of these terms allow for the implementation of a large variety of learning rules. The modulatory operation of Equation 3 can be applied to any kind of plasticity rule δ and neural model, e.g. Hebbian correlation rules with discrete time dynamics, spiking neural networks, or other. From this view, the idea of modulating, or gating, plasticity is independent of the specific neural model chosen for implementation: its role consists in the activation of local plasticity upon transmission of modulatory signals to specific neurons.

When applied to a suitable neural architecture, this form of gated plasticity can selectively activate learning in specific parts of the network and at the onset of specific events.

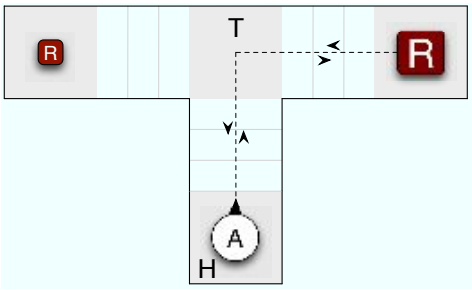

Figure 3: T-maze with homing. The agent navigates the maze returning home (H) after collecting the reward. The amount of reward is proportional to the size of the token.

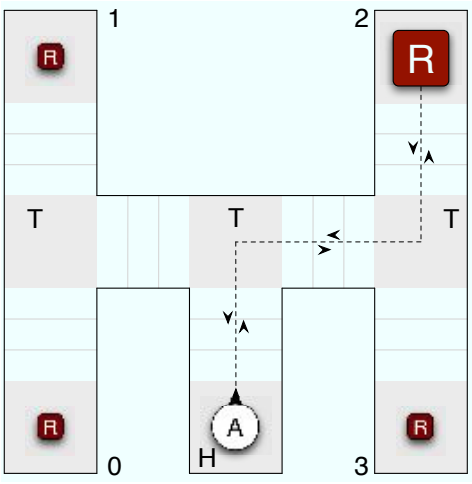

Figure 4: Double T-maze with homing.

This may prevent catastrophic forgetting that often results from continuously updating networks and lead to more efficient learning.

Learning in the T-maze

T-mazes are often used to observe operant conditioning (Britannica, 2007) in animals that are required to learn and remember – for instance – whether a reward in the form of food is located either on the right or on the left of a T-maze. This makes an ideal scenario for testing the effect of neuromodulation.

We simulated two T-mazes represented in Figures 3 and 4. In the first case (Figure 3), an agent is located at the bottom of a T-maze. At the end of two arms (left and right) there is either a high or a low reward. The task of the agent is to navigate the corridors, turn when it is required, collect the reward and return home. This is repeated many times during a lifetime: we call each trip to a maze-end a *trial*. A measure of quality in the agent's strategy is based on the total amount of reward collected. To maximise this measure, the agent needs to learn where the high reward is located.

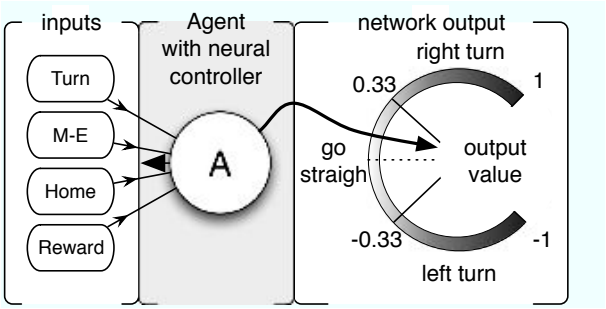

Figure 5: Inputs and output the neural network. The *Turn* input is 1 when a turning point is encountered. M-E is Maze-End: it goes to 1 at the end of the maze. *Home* becomes 1 at the home location. The *Reward* input returns the amount of reward collected at the maze-end, it remains 0 during navigation. One output determines the actions of turning left (if less than -1/3), right (if greater than 1/3) or straight navigation otherwise. Inputs and internal neural transmission are affected by 1% noise.

The difficulty of the problem lies in the fact that the position of the reward changes across trials. When this happens, the agent has to forget the position of the reward that was learnt previously and explore the maze again. In our experiments, the position of the high reward is changed at least once during lifetime, resulting in an uncertain foraging environment where the pairing of actions and reward is not fixed: turning left might result in a high reward at a certain time but in a lower reward later on. The intent is to foster the emergence of learning behaviour.

The complexity of the problem can be increased, as shown in Figure 4, by enlarging the maze to include two sequential turning points and four possible endings. In this problem an optimal strategy is achieved when the agent explores sequentially the four possible maze-ends until the high reward is found. At this point, the sequence of turning actions that leads there should be learnt and memorised together with the return sequence to the home location.

An agent is exposed to 100 trials in the experiments with the single T-maze and to 200 trials in the double T-maze. Each trial consists of a number of steps during which the neural network is updated and the agent moved accordingly (Figure 5). The large reward is randomly positioned and relocated after 50 trials on average, with a random variability of ± 15. The high reward value is 1.0 whereas the low reward is 0.2. The agent that fails to return to the home position (within a trial) will be relocated automatically to the home position and will suffer a penalty of 0.3, which is subtracted from the total amount of reward collected. The agent is required to maintain a forward direction in corridors and perform a right or left turn at the turning points: failure to do so results in the agent crashing, a penalty of 0.4 and being relocated to the home position. Each corridor and turning point stretches for three steps of the agent. Higher or variable numbers of steps have been tested providing similar results.

The control systems of the agents are evolved using the agents' performance as a measure of fitness.

Evolutionary Search

An Evolution Strategy (ES) (Bäck et al., 1997) was used to search for network topologies. The genome was encoded as a matrix of real-valued weights that represent the strengths of the initial connections w_{ij}. The 5 parameters for the plasticity rule A, B, C, D and η of Equation 4 were separately encoded and evolved in the range [-1,1] for A-D, and [-100,100] for η. A set of special genetic operators was devised to perform the topology search: insertion, duplication and deletion of neurons were introduced respectively to insert a new neuron in the network (a new line and row are added to the weight matrix) with probability 0.04, to duplicate an existing neuron (a line and a row are duplicated in the weight matrix) with probability 0.02, and delete a neuron (a line and a row are deleted in the weight matrix) with probability 0.06. Inserted neurons have the same probability (0.5) of being standard or modulatory.

All real values in the genome (GeV_i) are in the range [-1,1], and the phenotypical values PhV_i (with the exception of η), are mapped as $PhV_i = R_i \cdot (GeV_i)^3$, where R is the range (10 for weights, 1 for A..D). The mapping with a cubic function was introduced to favor small weights and parameter initially, and allow for the evolutionary growth of larger values by selection pressure when those are needed. Weights below 0.1 were set to 0.

Mutation is applied to all individuals (except the best) at each generation by adding to each gene a positive or negative perturbation $d = W \cdot exp(-P \cdot u)$, where u is a random number drawn from a uniform distribution [0,1] and P is a precision parameter here set to 180. This probability distribution favours local search with occasional large jumps, see (Rowe and Hidovic, 2004) for details. The function, although differently shaped than the traditional Gaussian, does not introduce a conceptual difference in the evolutionary algorithm. One point crossover on the weight matrix was applied with probability 0.1. A selection mechanism to enhance diversity in the population was devised. All individuals were positioned sequentially on an array. At each generation, the array was divided into consecutive segments of size 5 (with random segmentation offset at each generation), and the best individual of each segment was copied over the neighboring four. In this way, a successful individual spreads its genes only linearly with the generations. A population size of 300 for the single T-maze and 1000 for the double T-maze were used with termination criterion of 600 and 1000 generations. Generation zero was initialised with networks with one neuron per type and random connections.

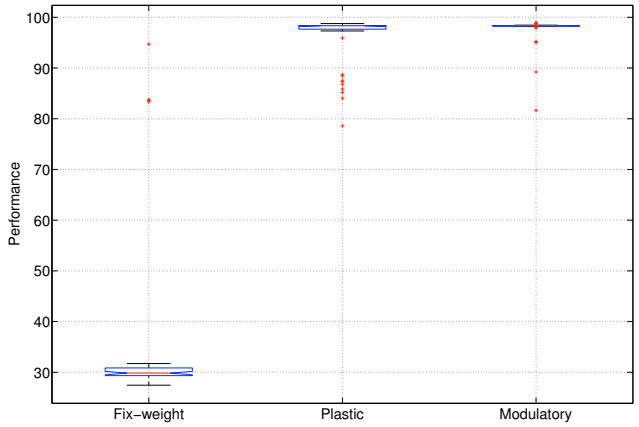

Figure 6: Box plots with performances of 50 runs on the single T-maze. The boxes are delimited by the first and third quartile, the line inside the boxes is the median value while the whiskers are the most extreme data samples from the box not exceeding 1.5 times the interquartile interval. Values outside this range are outliers and are marked with a cross. Boxes with non overlapping notches have significantly different median (95% confidence) (Matlab, 2007)

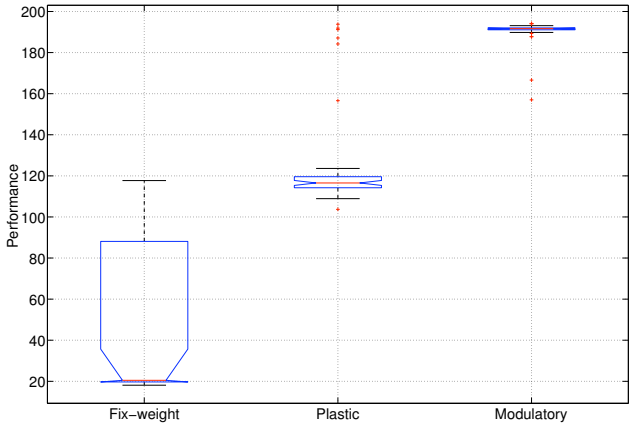

Figure 7: Box plots with performances of runs on the double T-maze.

Experimental Results

We conducted three types of evolutionary experiments, each characterised by different constraints on the properties of the neural networks: 1) fixed-weight, 2) plastic, and 3) plastic with neuromodulation. The fixed-weight networks were implemented imposing a value of zero on the modulatory activity, which resulted in a null update of weights (Equation 3). Plastic networks had a fixed modulatory activity of 1 so that all synapses are continuously updated (Equation 3 becomes $\Delta w = 0.462 \cdot \delta$). Finally, neuromodulatory plastic networks could take advantage of the full model described in Equations 1-4.

Fifty independent runs were executed for each of the three conditions. For each run, the individual that performed best at the last generation was tested 100 lifetimes with different initial conditions. The average reward collected over the 100 tests is the numerical value of the performance. The procedure was repeated for all the 50 independent runs. The distribution of performance is summarized by box plots in Figure 6 for the single T-maze, and in Figure 7 for the double T-maze.

For the single T-maze, the theoretical and measured maximum amount of reward that can be collected on average is 98.8, and not 100 due to the minimum amount of exploration that the agent needs to perform at the beginning of its lifetime and when the reward changes position. For the double T-maze, the theoretical and measured maximum amount of reward that can be collected is 195.2 when averaged on many experiments.

The experimental results indicate that plastic networks achieve far better performance than the fixed-weight networks. Fixed-weight networks could potentially display levels of learning-like behaviour by exploiting recurrent connections and storing state-values in the activation of neurons (Blynel and Floreano, 2002). However, our experiments show that such solutions are more difficult to evolve.

Among plastic networks, those that could exploit modulation displayed only a small advantage in the single T-maze. However, when memory and learning requirements increase for the double T-maze, modulated plasticity displayed a considerable advantage. Figure 7 shows that modulatory networks achieved nearly optimal performance in the double T-maze experiment. Simplified versions of the single and double T-maze can be obtained by removing the requirement for homing. Experiments not reported here on T-mazes without homing confirmed the results showing an advantage for modulatory networks.

It is important to note that the exact performance reported in Figures 6 and 7 depend on the specific design and settings of the evolutionary search. Higher or lower population numbers, available generations, different selection mechanisms and mutation rates affect the final fitness achieved in all cases of fix-weight, plastic and modulatory networks. However, a set of preliminary runs performed by varying the above settings confirmed that the differential in performance between modulatory networks and plastic or fix-weight networks is consistent, although not always the same in magnitude.

Analysis and Discussion

The agents achieving optimal fitness in the tests display an optimal control policy of actions. This consists in adopting an exploratory behaviour initially – until the location of the high reward is identified – followed by an exploitative

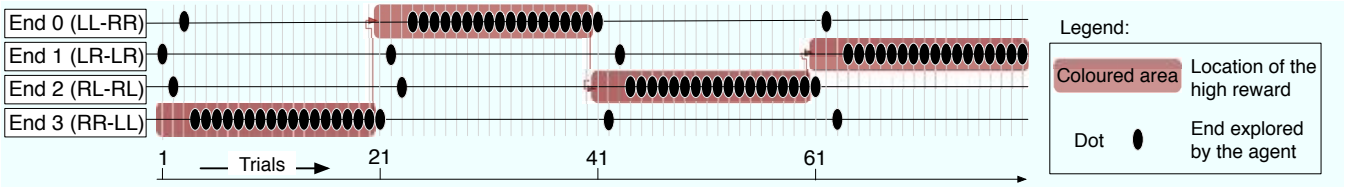

Figure 8: Behaviour of an agent exploring the double T-maze of Figure 4. A test of 80 trials is performed. The four horizontal lines track the events at each of the four maze-ends. The position of the reward is changed every 20 trials. The coloured area indicates where the high reward is located. The black dots show the maze-end explored by the agent at each trial. The agent adopts an explorative behaviour when it does not find the high reward, and settles on an exploitative behaviour after the high reward is found.

behaviour of returning continuously to the location of the high reward. Figure 8 shows an evolved behaviour, which is analogous to operant conditioning in animal learning. This policy involves the exploration of the 4 maze-ends. When the high reward is discovered, the sequence of turning actions that lead there, and the correspondent homing turning actions, are memorised. That sequence is repeated as long as the reward remains in the same location, but abandoned when its position changes. At this point the exploratory behaviour is resumed. This alternation of exploration and exploitation driven by search and discovery of the reward continues indefinitely across trials.

Although this strategy is a mandatory choice to maximise the total reward, from the performance indices presented in the previous section (Figures 6 and 7) we deduce that this behaviour can be more easily evolved when modulatory neurons are allowed into networks.

Functional Role of Neuromodulation

The experimental data on performance showed a clear advantage for networks with modulatory neurons. Yet, the link between performance and characteristics of networks is not easy to find due to the large variety of topologies and learning rules that evolved from independent runs. Figure 9 shows an example of a network that solves the double T-maze. The neural topology, number of neurons and learning rule may vary considerably across evolved networks that perform equally well.

Nonetheless, it is possible to check if the better performance in the double T-maze agents evolved with neuromodulated plasticity is correlated with a differential expression of modulatory and standard neurons. The architecture and composition of the network are modified by genetic operators that insert, duplicate and delete neurons. We measured the average number of the two types of neurons in evolving networks for the condition where plasticity is not affected by modulation (Figure 10, top left graph) and for the condition where plasticity is affected by modulatory inputs (Figure 10, bottom left graph). In both conditions, the number of modulatory neurons is higher than the number of standard neurons. However, the presence of modulatory neurons when those are not active (top left graph) depends only on insertion, duplication and deletion rates, whereas in the case when they are enabled (bottom left graph) their presence might be linked to a functional role. This fact is suggested by the higher value of the mean fitness.

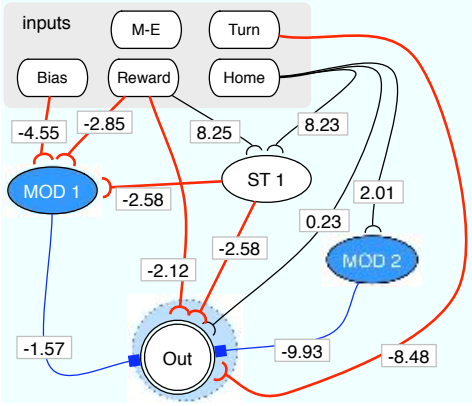

Figure 9: Example of an evolved network that solves the double T-maze. This network has two modulatory neurons and one standard neuron beside the output neuron. Arcs represent synaptic connections. The inputs (Bias, Turn, Home, M-E, Reward) and standard neurons (ST 1 and OUT) send standard excitatory/inhibitory signals to other neurons. Modulatory neurons (MOD 1 and MOD 2) send modulatory signals which affects only plasticity of postsynaptic neurons, but not their activation level. The evolved plasticity rule is $A = 0, B = 0, C = -0.38, D = 0, \eta = -94.6$.

In a second phase, we continued the evolutionary experiments for additional thousand generations, but we set to zero the probability of inserting and duplicating neurons, while the probability of deleting neurons was left unchanged. In both conditions all types of neurons slightly decreased in number. However, modulatory neurons completely disappeared in the condition where the modulatory input had no effect on plasticity (Figure 10, top right graph) while on average two modulatory neurons were observed in the condition where modulation could affect plasticity. This repre-

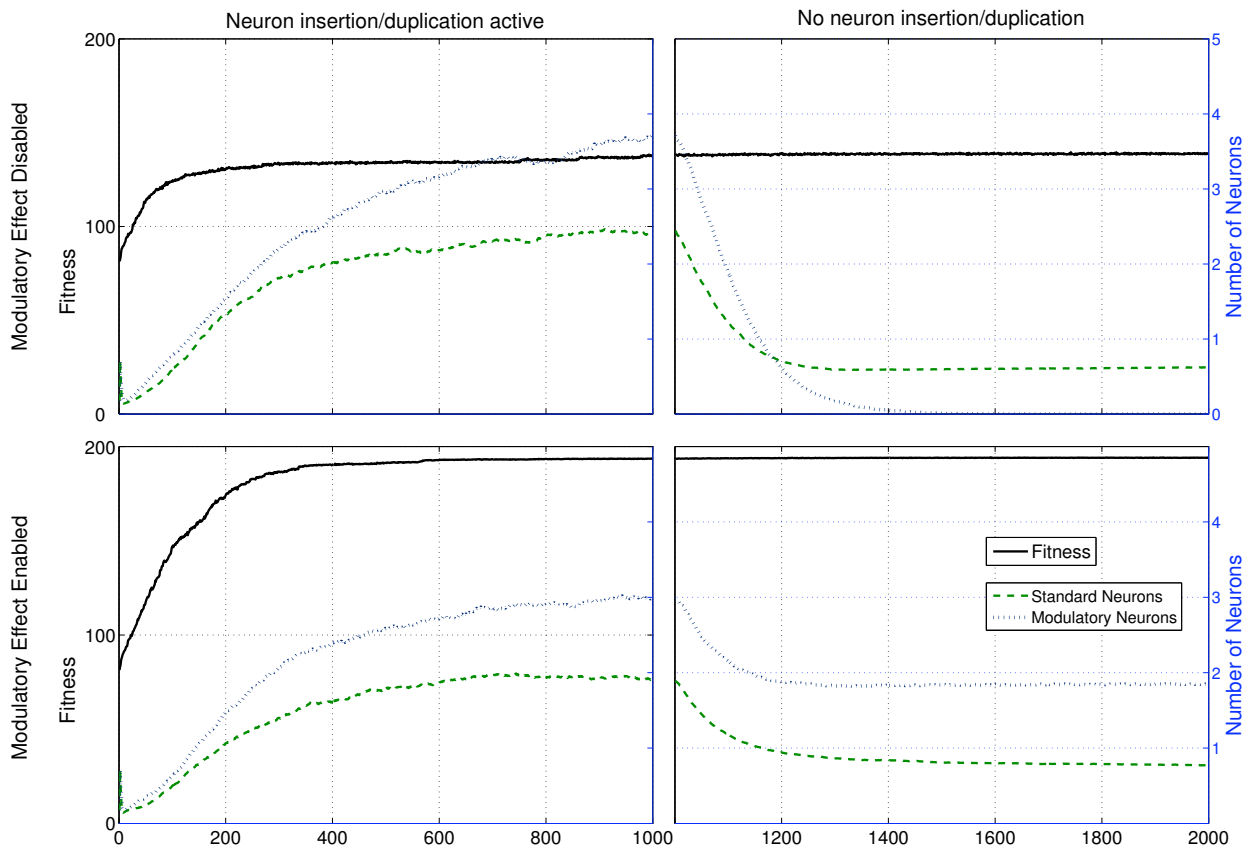

Figure 10: Fitness (continuous line) and number of neurons (dashed lines for standard and dotted lines for modulatory) in networks during evolution (average values of 50 independent runs).

sents a further indication that neuromodulation of synaptic plasticity is responsible for the higher performance of the agents in the double T-maze problem and that they play a functional role in guiding reward-based learning.

A further test was conducted on the evolved modulatory networks when the evolutionary process was completed. Networks with high fitness that evolved modulatory neurons were tested with modulation disabled. The test revealed that modulatory networks, once deprived of modulatory neurons, were still capable of navigation by turning at the required points and maintaining straight navigation along corridors. The low level navigation was preserved and the number of crashes did not increase. However, most of networks seemed capable of turning only in one direction (i.e. always right, or always left), therefore failing to perform homing behaviour. None of the networks appeared to be capable of learning the location of the high reward. Generally, networks that were evolved with modulation and that were downgraded to plastic networks (by disabling modulatory neurons) performed worse than evolved plastic networks. Hence, we can assume that modulatory neurons are not employed to implement a higher level of functionality, otherwise not achievable with simple plasticity. Rather, modulatory neurons are employed to design a completely different neural dynamics that, according to our experiments, are easier to evolve, and on average resulted in better performance at the end of the simulated evolution.

Conclusion

The model of neuromodulation described here applies a multiplicative effect on synaptic plasticity at target neurons, effectively enabling, disabling or modulating plasticity at specific locations and times in the network. The evolution of network architectures and the comparison with networks unable to exploit modulatory effects allowed us to show the advantages brought in by neuromodulation in environments characterised by distant reward and uncertainties. We did not observe an obvious correspondence between performance and architectural motifs: we assume that the unconstrained topology search combined with different evolved plasticity rules allow for a large variety of well performing structures. In this respect, the search space was explicitly unconstrained in order to assess modulatory advantages independently of particular or hand-designed neural structures. In this condition, the philogenetic analysis of evolving networks supports the hypothesis that modulated plasticity

is employed to increase performance in environments where sparse learning events demand memorisation of selected and timed signals.

Future work includes the analysis of working architectures to understand the relation between the requirements of the problems and the type and size of networks that solve them. This study does not address in detail the neural dynamics that allowed for the improved learning and memory capabilities. Further analysis could possibly clarify the properties of the internal neural pattern of activations and weight changes. The relation between reward signals and modulatory activations could unveil important properties of the neural dynamics and explain how information from global reinforcement signals is transferred to the synaptic level, and consequently modify behavioural responses.

Acknowledgements

This work was partly supported by the Swiss National Science Foundation.

References

Abbott, L. F. and Regehr, W. G. (2004). Synaptic computation. *Nature*, 431:796–803.

Bäck, T., Fogel, D. B., and Michalevicz, Z., editors (1997). *Handbook of Evolutionary Computation*. Oxford University Press, Oxford.

Bailey, C. H., Giustetto, M., Huang, Y.-Y., Hawkins, R. D., and Kandel, E. R. (2000a). Is heterosynaptic modulation essential for stabilizing hebbian plasticity and memory? *Nature Reviews Neuroscience*, 1(1):11–20.

Bailey, C. H., Giustetto, M., Zhu, H., Chen, M., and Kandel, E. R. (2000b). A novel function for serotonin-mediated short-term facilitation in aplysia: Conversion of a transient, cell-wide homosynaptic hebbian plasticity into a persistent, protein synthesis-independent synapse-specific enhancement. *PNAS*, 97(21):11581–11586.

Birmingham, J. T. and Tauck, D. L. (2003). Neuromodulation in invertebrate sensory systems: from biophysics to behavior. *The Journal of Experimental Biology*, 20:3541–3546.

Blynel, J. and Floreano, D. (2002). Levels of dynamics and adaptive behavior in evolutionary neural controllers. In *Proceedings of the seventh international conference on simulation of adaptive behavior on From animals to animats*, pages 272–281. MIT Press Cambridge, MA, USA.

Brembs, B., Lorenzetti, F. D., Reyes, F. D., Baxter, D. A., and Byrne, J. H. (2002). Operant Reward Learning in Aplysia: Neuronal Correlates and Mechanisms. *Science*, 296(5573):1706–1709.

Britannica (2007). Animal learning. Encyclopedia Britannica 2007 Ultimate Reference Suite.

Burrell, B. D. and Sahley, C. L. (2001). Learning in simple systems. *Current Opinion in Neurobiology*, 11:757–764.

Carew, T. J., Walters, E. T., and Kandel, E. R. (1981). Classical conditioning in a simple withdrawal reflex in aplysia californica. *The Journal of Neuroscience*, 1(12):1426–1437.

Cooper, S. J. (2005). Donald O. Hebb's synapse and learning rule: a history and commentary. *Neuroscience and Biobehavioral Reviews*, 28(8):851–874.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, 15(4-6):495–506.

Fellous, J.-M. and Linster, C. (1998). Computational models of neuromodulation. *Neural Computation*, 10:771–805.

Jay, M. T. (2003). Dopamine: a potential substrate for synaptic plasticity and memory mechanisms. *Progress in Neurobiology*, 69(6):375–390.

Kondo, T. (2007). Evolutionary design and behaviour analysis of neuromodulatory neural networks for mobile robots control. *Applied Soft Computing*, 7(1):189–202.

Marcus, E. A. and Carew, T. J. (1998). Developmental emergence of different forms of neuromodulation in Aplysia sensory neurons. *PNAS, Neurobiology*, 95:4726–4731.

Matlab (2007). *Box plots*. The Mathworks Documentation.

Niv, Y., Joel, D., Meilijson, I., and Ruppin, E. (2002). Evolution of Reinforcement Learning in Uncertain Environments: A Simple Explanation for Complex Foraging Behaviours. *Adaptive Behaviour*, 10(1):5–24.

Roberts, A. C. and Glanzman, D. L. (2003). Learning in aplysia: looking at synaptic plasticity from both sides. *Trends in Neuroscience*, 26(12):662–670.

Rowe, J. E. and Hidovic, D. (2004). An evolution strategy using a continuous version of the gray-code neighbourhood distribution. In *GECCO (1)*, pages 725–736.

Schultz, W., Apicella, P., and Ljungberg, T. (1993). Responses of Monkey Dopamine Neurons to Reward and Conditioned Stimuli during Successive Steps of Learning a Delayed Response Task. *The Journal of Neuroscience*, 13:900–913.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A Neural Substrate for Prediction and Reward. *Science*, 275:1593–1598.

Smith, T., Husbands, P., Philippides, A., and O'Shea, M. (2002). Neuronal Plasticity and Temporal Adaptivity: GasNet Robot Control Networks. *Adaptive Behaviour*, 10:161–183.

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

Sporns, O. and Alexander, W. H. (2002). Neuromodulation and plasticity in an autonomous robot. *Neural Networks*, 15:761–774.

Sun, Z.-Y. and Schacher, S. (1998). Binding of Serotonin to Receptors at Multiple Sites Is Required for Structural Plasticity Accompanying Long-Term Facilitation of Aplysia Sensorimotor Synapses. *The Journal of Neuroscience*, 18(11):3991–4000.

Ziemke, T. and Thieme, M. (2002). Neuromodulation of Reactive Sensorimotor Mappings as Short-Term Memory Mechanism in Delayed Response Tasks. *Adaptive Behavior*, 10:185–199.