

# Neuromodulation Increases Decision Speed in Dynamic Environments

Andrea Soltoggio

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

## Abstract

Neuromodulation is considered a key factor in learning and memory in biological neural networks. Recent computational models of modulated plasticity have shown increased learning capabilities also in artificial neural networks. In this study, a reward-based dynamic scenario is employed to test networks evolved with modulatory dynamics. The analysis on the networks shows that neuromodulation does not only allow for better learning, but accelerates the computation in decision processes. This appears to derive from topological features in modulatory networks displaying more direct sensory-motor connections, whereas non-modulatory networks require longer pathways for signal processing. This computational advantage in increased decision speed could contribute to unveil the fundamental role of neuromodulation in neural computation.

## 1. Introduction

Neuromodulation has become the focus of extensive studies with the realisation that a large variety of neural functions are regulated by modulatory chemicals, both in invertebrate and vertebrate organisms (Burrell and Sahley, 2001, Birmingham and Tauck, 2003). Neurotransmitters such as Acetylcholine (ACh), Norepinephrine (NE), Serotonin (5-HT) and Dopamine (DA) closely affect synaptic plasticity, neural wiring, Long Term Potentiation (LTP) and Long Term Depression (LTD). Short and long term configurations of brain structures are affected by modulatory chemicals, and consequently those have been linked to the formation of memory, the implementation of new neural functions, learning and adaptation (Jay, 2003, Abbott and Regehr, 2004).

The realisation that Hebbian plasticity does not account entirely for experimental evidence on synap-

tic modification has brought further focus on modulatory dynamics. Associative learning, like classical and operant conditioning, seems to be based on additional mechanisms besides the Hebbian synapse. Studies on the mollusc *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).

At a cellular level, modulatory neurons appear to release modulatory chemicals at target synapses, affecting plasticity mechanisms rather than postsynaptic activations. In (Bailey et al., 2000a), heterosynaptic modulation is shown to stabilise Hebbian plasticity by means of nonlinear effects of modulatory signals; when neuromodulation is coupled with presynaptic stimuli, it results in the activation of transcription factors and protein 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. A graphical representation is shown in Figure 1.

These cellular processes are believed to be the working mechanisms from which system and behavioural dynamics emerge. The release of dopamine, for example, has been linked to learning in (Schultz et al., 1993), where activation patterns in monkeys' brains followed a measure of prediction-error in classical conditioning.

Computational models of neuromodulation (Fellous and Linster, 1998, Doya, 2002) offer a valid tool to investigate the potential of neural dynamics both for understanding biological networks and for the design of artificial bio-inspired systems. The modulating action on synaptic plasticity, and consequently the role in the long term modification of neural pathways, place neuromodulation as

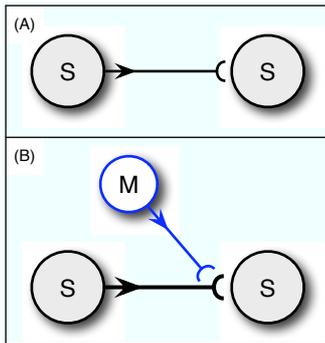


Figure 1: (A) Hebbian (homosynaptic) plasticity: the connection strength is updated as function of pre- and postsynaptic activity of standard neurons (S). (B) Heterosynaptic mechanism, or neuromodulation: a third neuron (M) mediates synaptic growth, i.e. the amount of modulatory signal determines the response to Hebbian plasticity.

ideal candidate to explain the acquisition of new input-output relations – achieving learning – and the retention of those – achieving memory.

For the reason above, studies on computational models often employ learning and memory tasks as benchmark. A modulatory neuron was implemented to evolve learning behaviour for a foraging task in uncertain environments (Niv et al., 2002), where a simulated bee learnt and remembered the colour of the best rewarding flowers. The same experimental setting was chosen also in (Soltoggio et al., 2007) to show that modulatory architectures could freely develop throughout evolution to achieve higher performance. Although various implementations and models of modulatory dynamics have been proposed so far, the concept of *modulatory neurons* as computational units for building networks was introduced and exploited only recently in (Soltoggio et al., 2007, Soltoggio et al., 2008). In (Soltoggio et al., 2008), plastic networks could be enriched by second-type modulatory neurons, resulting in evolutionary advantages in dynamic, reward-based scenarios. The experiments included a single and double T-maze explored by an agent where the location of a high reward was periodically changed during the agent’s lifetime.

The studies above describe a growing evidence that neuromodulation favours better performance in learning problems. However, many aspects regarding the computational advantages of neuromodulation have not been well described yet. To address this issue, this paper focus on the propagation of signals in modulatory networks. The reward-based scenario of the double T-maze described in (Soltoggio et al., 2008) is implemented here with the aim of reproducing and analysing modulatory neural networks. Surprisingly, the analysis reveals that

modulatory networks do not only achieve better learning, but allow for a faster computation at decision points which derives from shorter input-output connections. A further test with a stricter time constraint confirms the advantage in decision speed of networks with modulated plasticity.

The next sections 2, 3 and 4 introduce respectively the model of neuromodulation, the T-maze learning problem, and the evolutionary process that constitute the preliminary steps and experimental settings as in (Soltoggio et al., 2008). Following, the analysis of the networks outlines the learning performance, decision speed, topologies, and the test with enforced decision speed. The paper ends with final remarks in the conclusion.

## 2. A Model for Modulated Plasticity

Although different models of neurons have been proposed in the last decades, most theoretical and empirical studies on neural networks consider network graphs where nodes are instances of the same class. Biological neural networks, on the other hand, display a large variety of neural types, shapes and neurotransmitters that operate in close interaction. On this assumption, the idea of modelling two types of neuron, a *standard neuron* and a *modulatory neuron*, has been recently exploited in (Soltoggio et al., 2008). The underlying principle is that the two kinds of neuron transmit fundamentally different types of signal: standard neurons transmit signals that activate other connected neurons in cascade; on the contrary, modulatory neurons propagate modulatory signals that do not affect activations, but rather modulate synaptic growth of target neurons. This requires the existence of a traditional form of plasticity (for instance Hebbian plasticity) on which modulatory signals apply a gating operation. Therefore, as in standard homosynaptic networks, standard neurons are capable of increasing or decreasing incoming connection strengths according to models of homosynaptic plasticity. Modulatory neurons, on the other hand, intervene hierarchically on standard neurons by gating (or modulating) plasticity of specific target neurons. A graphical representation of the model is illustrated in Figure 2.

The model, inspired by the biological findings presented in the review paper (Bailey et al., 2000a), was devised to address also limitations in artificial neural networks and robotics: continuously learning networks might display instability of neural structures or catastrophic forgetting when new input-output signals update the network connectivity. Neuromodulatory signals are a possible solution to implement event-based learning, enabling plasticity at target neural areas and precise timing, and stabilising information into synaptic weights.

The model is expressed analytically by a multi-

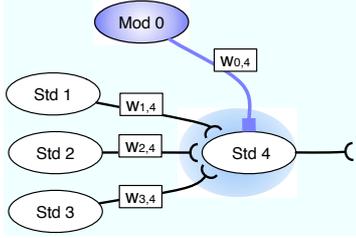


Figure 2: Ovals represent standard and modulatory neurons labelled with *Std* and *Mod*. A modulatory neuron transmits a modulatory signal – represented as a coloured shade – that diffuses around the synapses of the target neuron. Modulation affects the update of the afferent weights  $w_{1,4}$ ,  $w_{2,4}$  and  $w_{3,4}$ .

plicative operation on the weight updates

$$\Delta w_{ji} = \delta_{ji} \cdot \tanh(m_i/2) \quad (1)$$

where  $\delta_{ji}$  is a *plasticity term* (described later) between neurons  $j$  and  $i$ ,  $j$  is a standard neuron, and  $m_i$  is the modulatory signal perceived by the postsynaptic neuron  $i$  (standard or modulatory).

The activation  $a_i$  of a neuron and the value of modulation  $m_i$  are computed by summing the inputs from the two subsets (standard and modulatory) of neurons in the network

$$a_i = \sum_{j \in Std} w_{ji} \cdot o_j \quad , \quad (2)$$

$$m_i = \sum_{j \in Mod} w_{ji} \cdot o_j + m_d \quad , \quad (3)$$

where  $w_{ji}$  is the connection strength from neuron  $j$  to  $i$ ,  $o_j$  is the output of a presynaptic neuron computed as function of the standard activation  $o_j(a_j) = \tanh(a_j/2)$ ,  $m_d$  is a default modulation value that was set to 1.

The plasticity rule can effectively be implemented with any form of homosynaptic plasticity. Here  $\delta$  was described by the rule

$$\delta_{ji} = \eta \cdot [A o_j o_i + B o_j + C o_i + D] \quad (4)$$

where  $o_j$  and  $o_i$  are the pre- and postsynaptic neuron outputs,  $A, B, C, D$  and  $\eta$  are tuneable 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. The use and tuning of one or more of these terms allow for the implementation of a large variety of plasticity rules (Floreano and Urzelai, 2001).

### 3. Learning Sensory-Motor Relations to Maximise Reward Intake

Dynamic, reward based environments are frequent in the natural world. Even simple invertebrates have to deal with changing environments where the location, quantity and type of food vary according to many factors like the season, weather, competition among or within species, etc. In such conditions, the ability to adapt and learn is a fundamental skill to survive.

For these reasons, the learning capabilities of modulatory networks have been tested on dynamic scenarios where uncertain reward locations were to be learnt by simulated agents (Niv et al., 2002, Soltoggio et al., 2007).

Here, the task of an agent in a double T-maze, as in (Soltoggio et al., 2008), is to navigate the corridors, turn when it is required, collect the reward at the end of the maze and return home (see Figure 3). This is repeated many times (*trials*) during a lifetime. The task is to maximise the total amount of reward collected. To do so, the agent needs to learn where the high reward is located. During each trial, the agent encounters two turning points during the outgoing trip, and two turning points during the return trip. At each turning point, the agent will select a direction, either a left turn or a right turn, with the purpose of reaching the reward first, and then return home (the sensory-motor signals are illustrated in Figure 4). Therefore, each trial is composed of four critical decision points where the neural controller – given the current state and the history of collected rewards – has to take a decision. 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. This results in an uncertain foraging environment where the optimal pairing of actions and reward is not fixed, requiring online learning.

An agent is given 200 trials. Each trial consists of a number of steps during which the neural network is updated and the agent moved accordingly. The large reward is randomly positioned and relocated after 50 trials on average, with a random variability of  $\pm 15$ . The agent that fails to return to the home position collects a negative reward of 0.3 and is repositioned at home. 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, and a negative reward of 0.4 is given. Each corridor and turning point stretches for three steps of the agent. Higher or variable numbers of steps have been tested providing similar results.

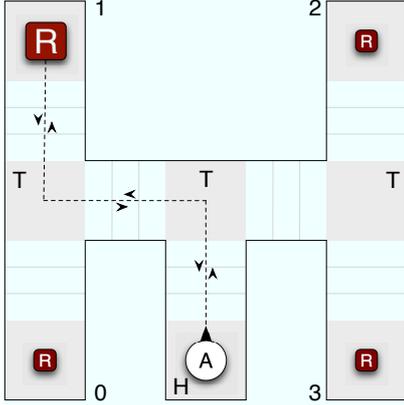


Figure 3: Double T-maze with homing. The agent navigates on a 1D track to one maze-end and returns home (H) after collecting the reward. The amount of reward is proportional to the size of the token (1.0 for the large reward, 0.2 for the small one). The agent performs many trials during a lifetime. After a number of trials, the high reward is moved to another location. The agent, unaware of the change, discovers a low reward upon returning to the usual location. When this happens, the agent will have to try another maze-end until the new location of the reward is discovered.

#### 4. Phylogenetic Outset of Neuromodulation

Adaptive behaviour was sought here by applying the Darwinian principle in a simulated evolutionary process. Standard and modulatory neurons were utilised as building blocks, and networks with such components were bred in the environment described in the previous section.

The process was implemented to reproduce the algorithm in (Soltoggio et al., 2008) and it is briefly outlined here. Network topologies were searched by means of an Evolution Strategy (ES) (Bäck et al., 1997). A matrix of real values encoded the network weights  $w_{ij}$ . The 5 parameters for the plasticity rule A, B, C, D and  $\eta$  of Equation 4 evolved in the range  $[-1,1]$  for A-D, and  $[-100,100]$  for  $\eta$ . Genes in the range  $[-1,1]$  were mapped into phenotypical values with a cubic function to introduce a bias towards small values. Phenotypical weights were in expressed in the range  $[-10,10]$ . Insertion, duplication and deletion of neurons were applied with probabilities 0.04, 0.02 and 0.06 respectively: those operators add, duplicate and remove respectively rows and columns in the weight matrix. Newly inserted neurons had the same probability (0.5) of being standard or modulatory.

Mutation acted on each gene with a positive or negative perturbation  $d = W \cdot \exp(-P \cdot u)$ , where  $u$  was a uniformly distributed  $[0,1]$  random number, and  $P$  a precision parameter here set to 180

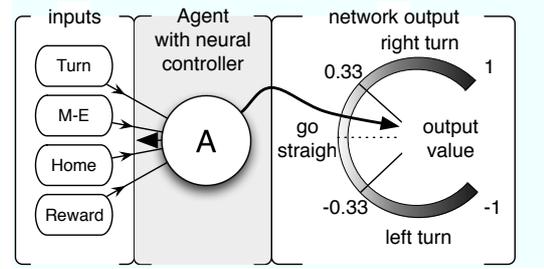


Figure 4: Inputs and output of the neural network. The *Turn* input is 1 when a turning point is encountered. *M-E* is Maze-End: it goes to 1 at the end of the maze. *Home* becomes 1 at the home location. The *Reward* input returns the amount of reward collected at the maze-end, it remains 0 during navigation. One output determines the actions of turning left (if less than  $-1/3$ ), right (if greater than  $1/3$ ) or straight navigation otherwise. Turning while in a corridor, or going straight at a turning point result in crashing, and the agent starting from the home location. Both inputs and internal neural transmission are affected by 1% uniform noise.

(Rowe and Hidovic, 2004). One point crossover on the weight matrix was applied with probability 0.1. The selection mechanism was implemented as a local tournament selection of size 5 on a spatially distributed population. A population of a 1000 individuals was employed with 2000 generations as termination criterion. After generation 1000, insertion and duplication of neurons were set to zero, while leaving deletion probability unchanged, to minimise the number of nodes in the networks.

With these settings, two parallel sets of experiments were executed. The evolutionary algorithm with the same initial conditions was run with modulatory neurons disabled in one case, and enabled in a second case. Fifty independent runs were performed for plastic and modulatory networks.

#### 5. Analysis of Results

Tests indicated that 47 out of 50 modulatory networks and 4 out of 50 standard plastic networks solved the task. The problem was considered solved when an agent scored on average at least 180 of total reward collected, out of 200 available<sup>1</sup>. This data confirms and reproduces the results in (Soltoggio et al., 2008), showing that modulatory networks evolved to achieve a higher level of learning. Because that previous study showed already the advantage in performance in terms of total reward collected, here the focus is on network features and topologies.

<sup>1</sup>Because the location of the reward is hidden to the agent, until it comes across it, the maximum fitness is 195.2 due to the exploratory trials that occur initially and when the reward changes location.

To compare network features, two fundamental points have to be considered: 1) different runs evolved considerably different topologies, number of neurons and plasticity rules, given the large search space and the neuron insertion/duplication/deletion operators; 2) plastic networks, achieving inferior performance, have a more limited functionality than modulatory networks: comparing modulatory networks that solved the problem with plastic networks that failed on average might not be significant. As a result of this last observation, it was decided to consider for analysis only the networks that achieved full functionality: in all, 47 modulatory networks, and 4 plastic networks. Unfortunately, this small number of plastic networks did not allow sufficient statistical analysis. Consequently, additional 100 runs were launched, resulting in 7 new successful standard plastic networks. In conclusion, the statistical analysis was carried out considering 11 plastic networks and 20 modulatory networks.

Even considering networks with similar performance, the evolutionary process designed a large variety of neural topologies, number of neurons and plasticity rules across different independent runs. However, this is true only for modulatory networks: a closer inspection revealed that all standard plastic networks evolved with the same topology. An example is reported in Figure 6. On the contrary modulatory networks had an average of 3.7 neurons and 17.4 connections with standard deviations of 0.9 and 9.2 respectively, resulting in high diversity of networks, all of them however achieving optimal behaviour. This first finding might provide an explanation for the considerable difference in successful rate of the evolutionary runs: while standard networks achieve full functionality with only one specific architecture, modulatory networks display a variety of optimally performing topologies. This suggests that the search space – when modulatory neurons are introduced – becomes richer of multiple global optimal solutions. It is also possible that modulatory neurons create neutral networks in the search space, allowing for a higher evolvability.

### 5.1 Decision Speed

Despite the number of neurons varied across different modulatory networks, the input and output, imposed by the environmental settings, were the same for all networks. On this basis, it was decided to compare input-output signal propagation considering the networks as a black box.

Surprisingly, the analysis revealed that the outputs of modulatory networks on average seem to react faster at turning points than the output of plastic networks. Figure 5 shows the absolute values of the output neurons (one for each network) when the network under test encounters a turning point. The

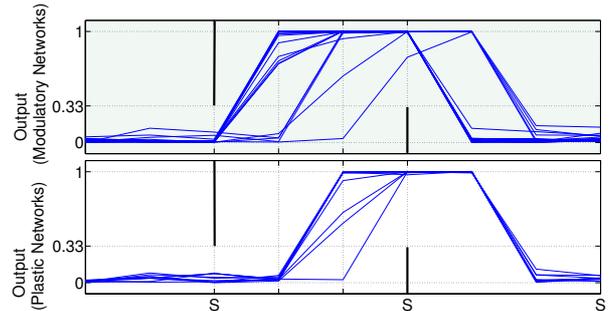


Figure 5: Absolute values of output signals at a turning point of modulatory and plastic networks with similar performance. Modulatory networks (upper graph) appear to react faster to the turning point and provide a quicker decision. Plastic networks show a longer reaction time. The thick vertical lines indicate the constraints at Sampling points (S): the first line from left indicates that the output is required to be less than 0.33 (to maintain a straight direction in the corridor). The second line shows that the output is required to be higher in absolute value than 0.33 (to perform a turning action).

number of computational steps required by modulatory networks to indicate a turning direction is 1.43 (average on 20 networks). Plastic networks, on the other hand, take 2.21 steps (average on 11 networks) to indicate the turning preference. Moreover, Figure 5 shows that whereas a substantial number of modulatory networks react in one step, none of the plastic networks had such a short reaction time.

The turning action expressed by the output is a required reaction at turning points: failure to turn – i.e. the output to indicate left or right turn – results in the agent crashing. Therefore, it was assumed here that the relevant part of the computation involved in the decision of which direction to take had to lie in the pathway between in the turn-input signal and the output. Accordingly, the network topologies were analysed to discovered relevant features in pathways from turn-input to output neuron.

The networks resulted to have on average a distance of 1.1 connections between input and output in the modulated case. Plastic networks have always 2 connections between turn-input and output, i.e. there is never a direct connection between turn-input and output. These data have a correspondence to the time – expressed in network steps – to trigger the output and indicate a turning direction. Therefore, one can assume that the number of connections through which the turn-input propagates correspond roughly to the time required to complete the computation at the turning point and provide a direction of navigation at the output neuron. For modulated networks, a direct connection between turn-input and

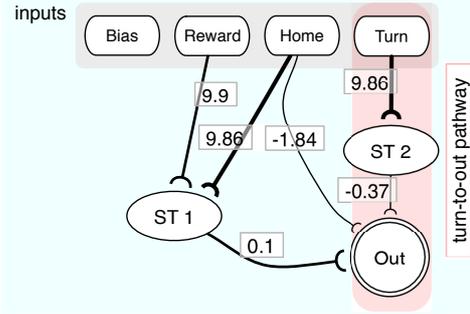


Figure 6: Example of a plastic network achieving near-optimal performance (plasticity rule A:-0.261,B:0,C:-1,D:0,eta:-31.8). All plastic networks that achieved optimal fitness had this topology with one inner neuron between the turn-input and the output.

output is frequently present; in plastic networks, the turn-input requires to be processed by at least one inner neuron. Examples of two representative networks are shown in Figure 6 for a plastic network and Figure 7 for a modulated network.

It is important to note that, according to the experimental settings, the networks are given three computational steps for each sensory-motor (input-output) update. The output of the network is sampled each three network steps, implying that no difference in behaviour or fitness can be detected if the output changes in 1, 2 or 3 computational steps. so long as the output reaches the required level before being sampled. Therefore plastic networks derive no disadvantage on performance<sup>2</sup> by having a path of two serial connections between input and output. Such configuration might have originated from implementation aspects of the evolutionary process.

Similarly, although modulatory networks display frequently a direct turn-input to output connection, it is not excluded that other parts of the network require longer processing time. In fact, the analysis of modulatory networks showed other longer pathways departing from input signals like the "reward" or "home" and innervating several other neurons.

Hence, although the analysis so far seems to indicate a faster computation for the decision process in modulatory networks, a further test is necessary.

## 5.2 Enforcing Speed

Reducing the available computational time at decision points is a way of compelling networks to react quickly. If modulatory networks do indeed implement a faster decision process, this will result in an advantage in performance whenever strict timing is imposed. Accordingly, a new evolutionary process

<sup>2</sup>Remember that modulatory and plastic networks with identical performance are compared here.

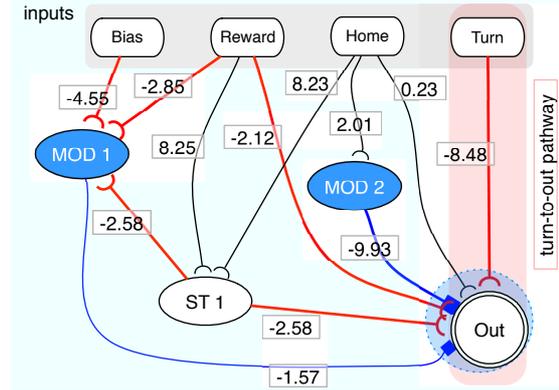


Figure 7: Example of a modulatory network achieving near-optimal performance (plasticity rule A:0, B:0, C: -0.38, D:0, eta: 94.6). Most of these networks, like in this case, show a direct connection between turn-input and output. None of the plastic networks had this feature.

was devised with identical settings as previously, but with only one computational step available at critical points in the maze. All the grey areas in the maze of Figure 3 were presented to the network for one computational step only. The new constraint requires networks to take decision at turning points in one computational step only, but also to acquire information on rewards and home in an equally short time.

The results of 50 independent evolutionary runs are illustrated in the box plots of Figures 8. What is relevant here, more than the distance between the boxes, is that none of the plastic networks scored fitness close to the optimal values above 180. This result allows for two considerations: 1) the constraint on the decision speed is determinant for plastic networks that without sufficient computational time fail to achieve learning and memory. 2) Modulatory networks are minimally affected by the reduced decision time, resulting in a high percentage of successful solutions.

Interestingly, these facts suggest that other longer pathways in modulated networks are not employed during the turning decision process, but are devoted to other functions. Therefore it can be concluded that decision speed is effectively enhanced in modulatory networks. What is the precise role of longer modulatory pathways is not entirely clear, however, it is evident that the direct connection between turn-input and output pre-encodes the next turning direction: a negative turn-input to output connection will result in a left turn, whilst a positive connection will result in a right turn. Plastic networks are unable to achieve this, implying that neuromodulation is responsible for a pre-computation, resulting in the encoding of the next turning direction into a connection weight before the turn occurs. In modu-

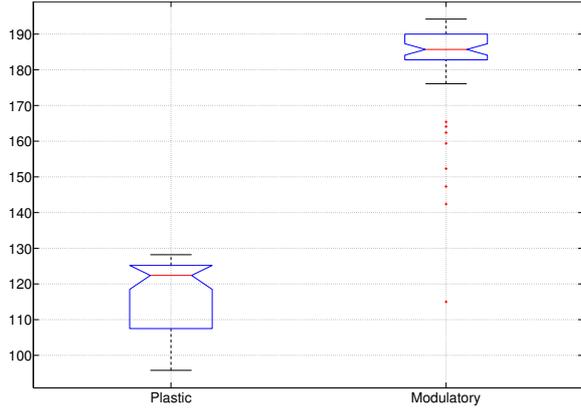


Figure 8: Box plots with performances of 50 runs with the additional constraint of one computational step at turning points. The boxes are delimited by the first and third quartile, the line inside the boxes is the median value while the whiskers are the most extreme data samples from the box not exceeding 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). Although modulatory networks register slightly decreased performance with respect the experiments with longer decision time, plastic networks were unable to evolve any optimal solution in this case.

latory networks, the learning process appears to correspond to the transfer of information into weights. Subsequently, the pre-computed information results in a faster signal processing at turning points. Figure 9 shows an example of network solving the learning problem with the strict time constraint.

### 5.3 The Dynamical System

Plots of activation and weight values in networks revealed that the neural dynamics are highly complex even for small networks. All the weights undergo plasticity of Equation 4, implying that the weight update is a linear combination of four terms depending of pre-, postsynaptic, correlated activities and decay<sup>3</sup>. In addition, some or all neurons can be targeted by modulatory signals, resulting in a modulation whose intensity is the hyperbolic tangent of the summation of all incoming modulatory signals. Considering a 5-input 4-neuron network, the neural dynamics result from a high complex interaction of 9 activation values, up to 4 modulated activities and up to 36 plastic weights. The system is also strongly affected by nonlinearity from saturation on weights and the squashing effect of the hyperbolic tangent on activations. Hence, this network is effectively a non-linear system with up to 44 states, and 14 inputs (5

<sup>3</sup>Evolution might evolve to zero the contribution of some of these terms.

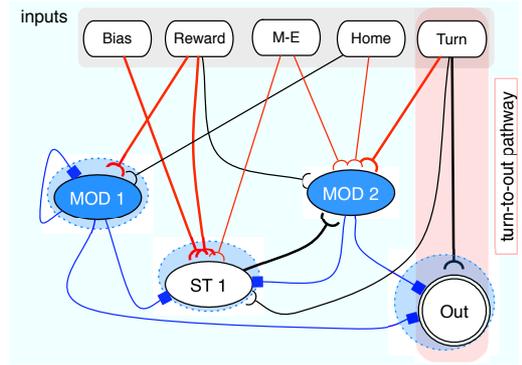


Figure 9: Example of a modulatory network achieving near-optimal performance with the strict time constraint of one step at decision points. The intricate network structure appears to pre-compute the turning direction and encode it into the direct turn-input to output pathway. The thickness of the lines are approximate indicators of connections strengths. This networks uses only a postsynaptic plasticity rule (term C in Equation 4).

external input and 9 noise inputs on neural transmission). Although a thorough analysis of such system could reveal important properties of the modulatory model given by Equations 1-4, the task falls outside the scope of this study.

## 6. Conclusion

This study considers performance and computational aspects of plastic and plastic-modulated networks evolved in a dynamic, reward based scenario where learning events and decision processes determine the fitness of an agent.

The learning capabilities of modulatory networks are reproduced here to gain a further insight on computational and topological aspects of networks with modulated plasticity. A fundamental difference between plastic and modulatory networks was shown in an increased sensory-motor propagation speed and quicker responses in decision making for modulatory networks with respect to standard plastic networks. At a further inspection, this property appeared to derive from more direct sensory-motor connections in modulatory networks. The magnitude and signs of those direct connections store a value that indicates the direction at the next turning point. This fact suggests that the decision at turning points is pre-computed and hierarchically encoded by neuromodulation onto the sensory-motor direct connection. In conclusion, this results in a faster signal processing during decision processes.

Modulated networks displayed a faster input-output response than plastic networks even without strict speed constraints. However, when the speed constraint was imposed in the second evolutionary

experiment, forcing control networks to take quick turning decisions at turning points, modulatory networks exhibited an even more considerable advantage in performance by evolving successful solutions where plastic networks failed.

The evolved modulatory networks have features that depend strongly on the environment in which the networks are evolved: other experiments on a simpler maze without homing did not result in a faster computation for modulatory networks. Therefore, this study on a single learning problem, although complex, does not allow to generalise the results on other learning problems. However, the interesting features displayed in this particular instance could emerge in a variety of similar or more complex learning problems.

Future work could extend this analysis to other learning problems and carry out a more extensive analysis of neural dynamics. The direct input-output pathways hierarchically modulated recall the concept of feed-forward models in control systems and biological networks: whether neuromodulation is a biological mechanism to implement feed-forward models in neural network is a challenging research question. However, a careful analogy can be drawn by comparing the pre-encoded information in the networks here with feed-forward architectures that bypass neural circuits to provide quicker motor responses. The remarkable performance and computational speed exhibited in the experimental results suggest the possible application of the model to a variety of learning and decision making problems.

## References

- Abbott, L. F. and Regehr, W. G. (2004). Synaptic computation. *Nature*, 431:796–803.
- Bäck, T., Fogel, D. B., and Michalewicz, Z., (Eds.) (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.
- 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.
- 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.
- 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.
- Floreano, D. and Urzelai, J. (2001). Evolution of plastic control networks. *Auton. Robots*, 11(3):311–317.
- Jay, M. T. (2003). Dopamine: a potential substrate for synaptic plasticity and memory mechanisms. *Progress in Neurobiology*, 69(6):375–390.
- 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.
- Soltoggio, A., Bullinaria, J. A., Mattiussi, C., Dürr, P., and Floreano, D. (2008). Evolutionary Advantages of Neuromodulated Plasticity in Dynamic, Reward-based Scenarios. In *Proceedings of the Artificial Life XI Conference 2008 (to appear)*.
- 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*.
- 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.