Evolvability of Neuromodulated Learning for Robots

Peter Dürr
Ecole Polytechnique Fédérale de Lausanne
Laboratory of Intelligent Systems
1015 Lausanne, Switzerland
peter.duerr@epfl.ch

Claudio Mattiussi
Ecole Polytechnique Fédérale de Lausanne
Laboratory of Intelligent Systems
1015 Lausanne, Switzerland
claudio.mattiussi@epfl.ch

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

Dario Floreano
Ecole Polytechnique Fédérale de Lausanne
Laboratory of Intelligent Systems
1015 Lausanne, Switzerland
dario.floreano@epfl.ch

Abstract

Neuromodulation is thought to be one of the underlying principles of learning and memory in biological neural networks. Recent experiments have shown that neuroevolutionary methods benefit from neuromodulation in simple grid-world problems. In this paper we investigate the performance of a neuroevolutionary method applied to a more realistic robotic task. While confirming the favorable effect of neuromodulatory structures, our results indicate that the evolution of such architectures requires a mechanism which allows for selective modular targetting of the neuromodulatory connections.

1. Introduction

An autonomous agent – be it a biological organism or an artifact like a robot – can adapt to the contingencies of its environment by adjusting its behavioral strategy according to the consequences of its present and past behavior. If the consequences of the current behavioral strategy are satisfactory according to a suitable measure that depends on the agent's goal or purpose, the strategy can be maintained. Otherwise, a learning process must be activated to alter it. This implies in particular that the agent must be able to activate, to deactivate, and, more generally, to link the extent of the learning to the above-mentioned measure of satisfaction.

When the agent's control system is realized by a neural network (NN), the learning can be implemented by modifying the synaptic weights w according to the following generalized Hebbian plasticity rule [7]

$$\Delta w = \eta \, [Axy + Bx + Cy + D]$$

where η is a fixed learning rate, x and y are the activation of the presynaptic and postsynaptic neuron, respectively, and A, B, C, and D are parameters that determine the relative importance of the different types of learning (correlated, presynaptic, etc). When a neuron receives at least two inputs this plasticity rule permits the heterosynaptic control of learning. For example, in Figure 1 the activity x_2 of one presynaptic neuron can influence the activity y of the postsynaptic neuron and thus it can also influence the synaptic plasticity of the other presynaptic neuron. This mechanism permits the implementation of the link between learning and behavioral outcomes described above, in terms of the control of the learning process of one part of the network by another part of the network which assesses the value of the behavioral outcomes. This type of heterosynaptic plasticity, however, has a potential problem. The control of learning can interfere with the processing of information in the postsynaptic neuron, since both depend on the activation of the postsynaptic neuron.

A mechanism of plasticity control that avoids this difficulty is based on *neuromodulation*. The idea is to separate the control of plasticity from the signal processing by using the following modified plasticity rule

$$\Delta w = m \, \eta \, [Axy + Bx + Cy + D] \qquad (1)$$

where the new multiplicative term m represents a modulatory signal produced by a specialized modulatory neuron (Figure 2). There is abundant evidence that biological organisms use, among other things, neuromodulatory control

978-0-7695-3272-1/08 $25.00 © 2008 IEEE
DOI 10.1109/LAB-RS.2008.22

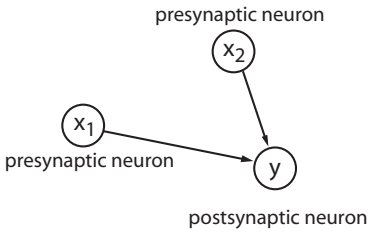

Figure 1. Using Hebbian learning the plasticity of the synapse between a pre- and a postsynaptic neuron can be affected by a signal produced by another presynaptic neuron.

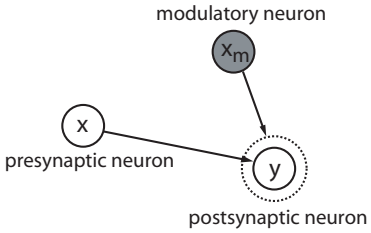

Figure 2. Using neuromodulation the plasticity of the synapse between a pre- and a postsynaptic standard neuron (white circles) can be affected by a modulatory signal produced by a modulatory neuron (gray circle) without interfering with the information processing of the standard neurons.

of synaptic plasticity based on similar principles [1]. One possible explanation for the existence of this mechanism in nature is that the use of neuromodulatory control of synaptic plasticity in place of plain Hebbian plasticity facilitates the evolutionary synthesis of complex adaptive neural structures. In [11] we presented results that corroborate this conjecture in the context of the artificial evolution of adaptive neural structures, showing that in some cases, the use of neuromodulation allows to evolve high-performing neural controllers, whereas plain Hebbian plasticity does not. This is in line with results from other experiments with related approaches (e.g., [10, 3]).

The recourse in [11] to artificial evolution for the synthesis of the neuromodulatory architectures is justified by the fact that it is difficult to hand-design both the underlying standard neural network and the additional neuromodulatory network that controls the learning of the former. In [12] we showed how an evolved neural architecture is typically more compact, while outperforming architectures designed by hand using the best neural networks and reinforcement learning practices. Recent results [5, 8] suggest that the use

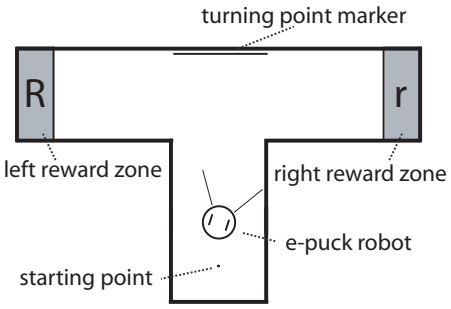

Figure 3. The T-maze experimental setup. Starting from the bottom of the maze, the e-puck robot must navigate to the left or the right reward zone. The robot can use infrared distance sensors to avoid colliding with the walls of the maze and a camera to detect when it arrives at the turning point.

of an implicit genetic encoding in place of a more conventional direct encoding endows the evolutionary process with several advantages, such as increased performance of the evolved structures, a more compact genome, and the possibility of complexification and simplification of the architecture via evolutionary duplication and mutation.

From a robotics point of view the limitation of the results presented in previous experiments of neuromodulatory evolution such as [7, 12, 11] is the use of simplified tasks based on grid-like worlds and a choice between a finite, small set of actions. It is therefore important to extend those results and the insights gained through them, to more realistic robotic scenarios. To this end, in this paper we investigate a more realistic neuromodulatory evolutionary scenario where a simulated robot is required to navigate and collect rewards in a T-maze, using the information provided by infrared sensors for obstacle avoidance. Our results show that the unrestricted application of neuromodulation in evolution can create difficulties due to the interference between the easily evolved basic navigation strategies and the more sophisticated reward-collecting strategies. We show how a hand-designed modularization of the effects of neuromodulation permits the preservation of the advantages of neuromodulation without interfering with evolution, thus suggesting to design neuroevolutionary systems that can reap both the architectural benefits of evolution and the adaptive benefits of neuromodulation.

2. Experimental setup

We consider an e-puck robot [2] which is placed in a simulated T-maze (Figure 3). At the end of each arm of the T-maze (left and right) there is either a high or a low reward

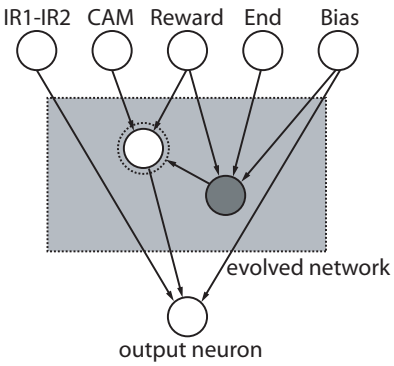

Figure 4. An example of an evolved neural network with a modulatory neuron (gray circle) and a standard neuron (white circle). The neural networks can use an infrared distance sensor input (*IR1-IR2*), a camera input (*CAM*), a reward signal (*Reward*), and an end zone detector (*End*) in addition to a bias unit (*Bias*) to control the behavior of the robot in the maze.

(*R* and *r*, with a value of 10 or 1 correspondingly). Starting from the bottom of the T-maze and facing in the direction of the turning point with a random angle $\gamma \in [-\pi/4, \pi/4]$, the task of the robot is to navigate in the maze and to collect one of the rewards by driving into the end-zone of either the left or the right arm. When the robot reaches either end-zone, it is awarded the respective reward and repositioned at the starting point.

The robot is controlled by an evolved artificial neural network with neuromodulation (see Figure 4). The modulatory signal m (see equation (1)) for all incoming synapses of each postsynaptic neuron is computed for each modulatory neuron as the product of the output of the respective modulatory neuron and the connection weight between the modulatory neuron and the postsynaptic neuron (see [12] for more details). Synapses leading to postsynaptic neurons which are not connected to a modulatory neuron do not undergo synaptic plasticity.

The network is connected to two continuous infrared distance sensors (which are merged into one sensory input $IR1 - IR2 \in [-1, 1]$). A turning point marker is placed in the middle of the maze. A camera sensor connected to the neural network ($CAM \in \{0, 1\}$) indicates that the turning point marker is in the field of view of the robot. Additionally, the robot can sense when it reaches the end-zone ($END \in \{0, 1\}$) and how big the obtained reward is ($Reward \in \{1, 10\}$).

The output o of the evolved neural network is used to control the behavior of the robot as follows: if the absolute value of the output is smaller than a threshold value ($|o| < o_t = 0.3$) the robot drives straight. If the output is smaller than the threshold value ($o < -o_t$) the robot rotates counterclockwise, and likewise, if the output is larger than the threshold ($o > o_t$), the robot rotates clockwise.

The fitness of the robot is calculated as the average of the total rewards the robot collects in two trials of 300s each. At the beginning of the first trial, the high reward is in the right arm of the maze. At a random time τ, uniformly selected from the interval $[125s, 175s]$, the reward position is flipped to the left arm. In the second trial, the reward is initially positioned in the left arm of the maze and is flipped to the other side at the same time τ.

Thus, as in the grid world experiments presented in [11], the artificial neural network must adapt to the changing position of the higher reward and change its strategy in order to gain maximal fitness.

However, unlike the grid world experiments described in [11], in this setup the evolved neural networks must control both the collision avoidance and the navigation of the robot in the maze. The more efficiently the robot avoids colliding with the walls, the more rewards, high or low, can be collected. The more efficiently the neural network adapts its behavior when the reward position changes, the higher the ratio of high rewards to low rewards that can be collected. While the collision avoidance behavior can be obtained using only the infrared distance sensors as an input, an adaptive navigation strategy requires connections from the other inputs. Under these circumstances, starting from an initial population of random networks, it is likely that simple collision avoidance networks, which use the infrared distance sensors and do not rely on synaptic plasticity, will appear early in evolution. This means that the neuromodulatory circuits which provide the reinforcement learning-like properties needed for adaptively switching navigation strategies, have a good chance of being forced to integrate themselves with the existing collision avoidance circuit.

In order to study the influence of modulated synaptic plasticity on the collision avoidance behavior, we introduced a factor α which adjusts the influence of synaptic plasticity on synapses connecting only from the infrared distance sensors. The synaptic weights of synapses w^* from the infrared distance sensors are updated by the modified update rule

$$\Delta w^* = \alpha\, m\, \eta\, [Axy + Bx + Cy + D]$$

with $\alpha \in [0, 1]$ where $\alpha = 1.0$ corresponds to the standard neuromodulatory update rule of equation (1) where all synapses leading to the corresponding post-synaptic neuron are affected by the synaptic plasticity, and $\alpha = 0.0$ corresponds to the case where the evolved weights of synapses coming from the infrared sensors are not plastic, while the other synapses remain plastic.

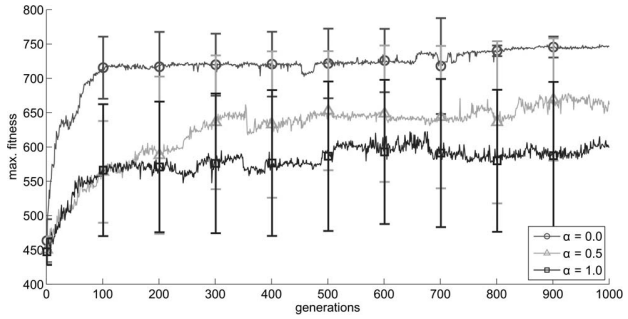

Figure 5. The average of the maximum population fitness of 10 evolutionary runs performed in the three conditions $\alpha = 0.0$, $\alpha = 0.5$ and $\alpha = 1.0$ over 1000 generations. The error bars every 100 generations indicate the standard deviations.

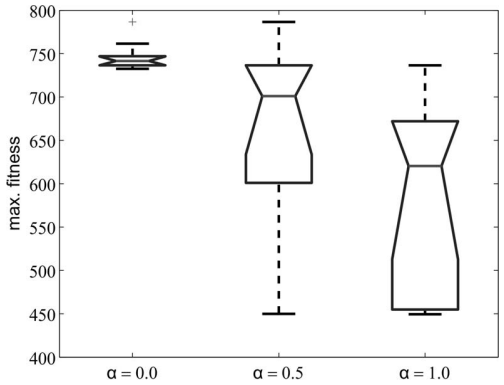

Figure 6. Maximum population finesses after 1000 generations of 10 evolutionary runs performed in the three conditions $\alpha = 0.0$, $\alpha = 0.5$ and $\alpha = 1.0$. The midline in each box is the median, the borders of the box represent the upper and the lower quartile. The whiskers outside the box represent the minimum and maximum values obtained, except when there are outliers which are shown as small crosses. We define outliers as data points which differ more than 1.5 times the interquartile range from the border of the box. The notches permit the assessment of the significance of the differences of the medians. When the notches of two boxes do not overlap, the corresponding medians are significantly different at (approximately) the 95% confidence level [6].

3. Evolutionary algorithm

As in [12], we used analog genetic encoding (AGE) to represent the neural network topology and the synaptic weights. The numerical parameters of both neurons and modulatory neurons were locally encoded using the center of mass encoding (CoME) [4] with search space intervals of $m \in [0.5, 5]$, $\eta \in [-10, 10]$, $\{A, B, C, D\} \in [-1, 1]^4$. We used a simple, generational genetic algorithm with a population size of 1000 individuals, tournament selection (with a tournament size of 2) and elitism (with an elite of size 1). We initialized the population with the best of 100000 random networks. The mutation probabilities used were 0.001 for nucleotide substitution, insertion, and deletion; 0.01 for chromosome fragment duplication, deletion, and transposition; and the probability of inserting a random device (e.g., a neuron or a modulatory neuron) was 0.2. The probability of recombination was 0.1. For more details on the algorithm see [5].

4. Results and discussion

We carried out 10 runs of the evolutionary experiment for the three different cases $\alpha = 1.0$, $\alpha = 0.5$ and $\alpha = 0.0$, that is, for unrestricted neuromodulation, reduced neuromodulation to the synapses coming from the infrared sensors, and absence of neuromodulatory plasticity for the synapses coming from the infrared sensors, respectively. Figure 5 shows that in the case of the unrestricted modulated plasticity ($\alpha = 1.0$) the task was not consistently solved within 1000 generations. We observed that while some of the evolved networks display near optimal performance, other networks fail to change strategy and end up collecting many small rewards. On the contrary, in the runs where synaptic plasticity is restricted to the synapses which are not connected to the infrared distance sensors ($\alpha = 0.0$), the evolved networks consistently display near optimal performance. In general, the fitness values obtained with $\alpha = 1.0$ are significantly smaller than the fitness values obtained with $\alpha = 0.0$, whereas the fitness values obtained with $\alpha = 0.5$ lie in between those produced by the other two cases. This difference is reflected in the significantly higher maximum population fitness obtained after 1000 generations with $\alpha = 0.0$, as shown in Figure 6.

In order to analyze the influence of modulated synaptic plasticity on the performance of the networks, we carried out a series of tests using the best networks resulting from the 10 runs under the condition $\alpha = 1.0$. We evaluated the performance of each network 100 times with unrestricted neuromodulatory synaptic plasticity ($\alpha = 1.0$) with restricted neuromodulatory plasticity ($\alpha = 0.0$) and without neuromodulatory plasticity on the whole network. Figure 7 shows the statistics of the amount of reward collected in the three conditions for the ten runs. In 9 out of

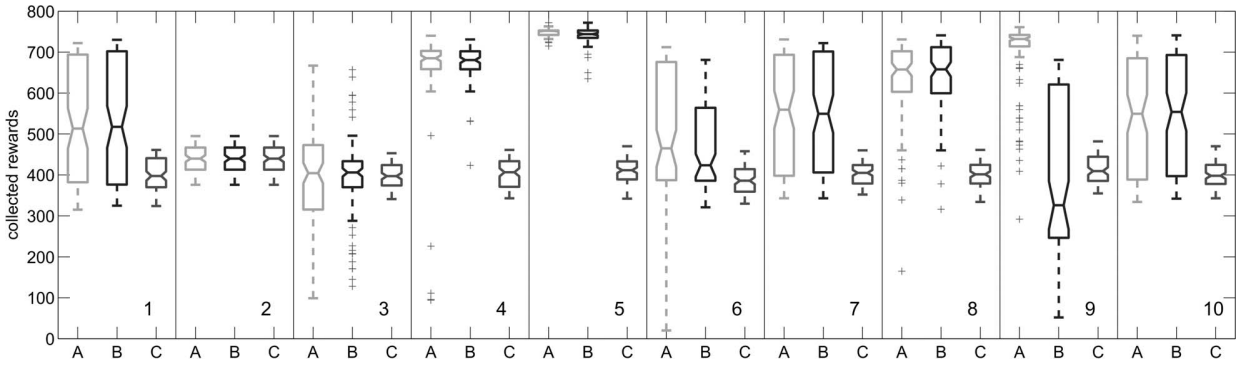

Figure 7. The rewards collected by the best individuals of 10 evolutionary runs after 1000 generations with unrestricted neuromodulatory plasticity, i.e., with $\alpha = 1.0$. The networks were tested in 100 trials each, with unrestricted neuromodulatory plasticity ($\alpha = 1.0$, columns A), with restricted neuromodulatory plasticity ($\alpha = 0.0$, columns B), and without neuromodulatory synaptic plasticity (columns C). For the details of the boxplot format, see the caption of Figure 6.

the 10 cases, disabling the plasticity of the synapses coming from the infrared sensors (i.e., setting $\alpha = 0.0$, box in columns B) on the networks evolved with $\alpha = 1.0$ does not significantly degrade the performance of the networks with respect to their evolved performance (box in columns A). When the neuromodulatory synaptic plasticity is disabled completely (box in columns C), the two worst networks in terms of evolved performance (runs 2 and 3) do not display a degradation in performance revealing that in these cases the neuromodulatory synaptic plasticity does not contribute to the performance. On the other hand, without neuromodulatory plasticity the networks produced by the other eight runs show a substantial degradation of performance with respect to their plastic counterpart. Of the networks evolved in these eight runs, all but that produced by run 9 do not show a significantly different performance when the synaptic plasticity is restricted, i.e. these networks seem to have been evolved with a built-in restriction of the influence of the synaptic plasticity on the synapses from the infrared distance sensors. These results suggest a potential reason for the improved performance of the algorithm when operated with restricted neuromodulatory plasticity (i.e., with $\alpha < 1.0$). We can conjecture that in order to gain high performance, the neuromodulatory synaptic plasticity needs to be restricted to certain parts of the network. As random mutations are very likely to disturb this restriction, lower values of α strengthen the restriction, resulting in increased evolvability.

In order to confirm that neuromodulatory synaptic plasticity is instrumental to the performance of the networks evolved with restricted plasticity, we further tested the best networks resulting from the 10 evolutionary runs using the condition $\alpha = 0.0$. We evaluated the performance of each network 100 times with the same restricted neuromodulatory plasticity used during evolution ($\alpha = 0.0$), and without neuromodulatory synaptic plasticity. Figure 8 shows the statistics of the amount of reward collected in the two conditions for the ten runs. The figure reveals that all networks display near optimal performance when neuromodulatory synaptic plasticity is active (box in columns B), while suffering a significant degradation of performance when neuromodulatory synaptic plasticity is deactivated (box in columns C). Further analysis of these results revealed that the degradation in performance is caused by the inability of the robots controlled by the networks with deactivated neuromodulatory plasticity to switch strategies when the position of the reward changed. This reveals that the neuromodulatory plasticity is actually instrumental to the performance of the evolved networks, and has not simply been ignored by the evolutionary process which synthesized the networks.

5. Conclusions

In this paper we have investigated the performance of a neuroevolutionary method based on an implicit genetic encoding when applied to the synthesis of neuromodulatory architectures for a realistic robotic task. The results of our experiments confirm the favorable effects previously observed in the experiments reported in [11] of the availability of neuromodulation on the performance of the evolved networks, with respect to the case where only plain Hebbian plasticity is available. However, these experiments, conducted in a more realistic simulated robotic scenario than those reported in [11], reveal that the presence of neuromodulation can hamper evolution by interfering with already

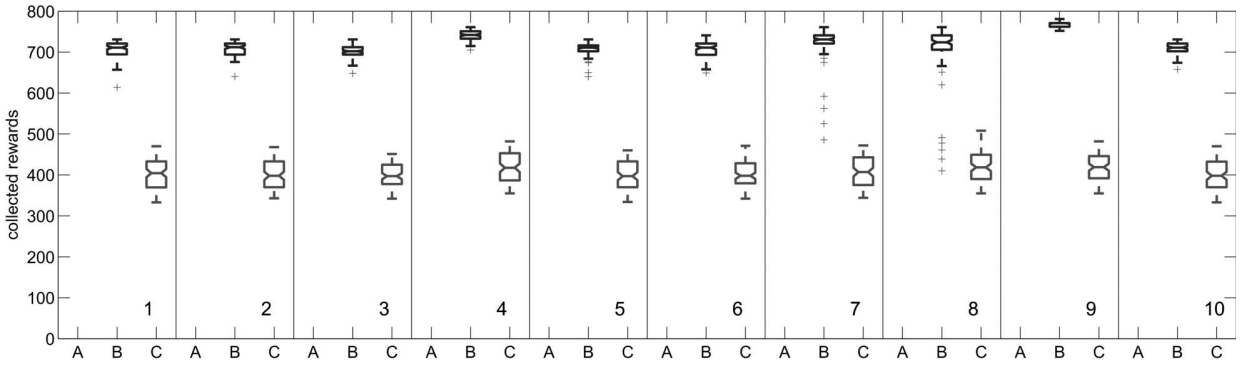

Figure 8. The rewards collected by the best individuals of 10 evolutionary runs after 1000 generations with restricted neuromodulatory plasticity, i.e., with $\alpha = 0.0$. The networks were tested in 100 trials each, with restricted neuromodulatory plasticity ($\alpha = 0.0$, columns B), and without neuromodulatory synaptic plasticity (columns C). For the details of the boxplot format, see the caption of Figure 6.

evolved behaviors that do not require synaptic plasticity. We have shown that by restricting the effects of neuromodulation it is possible to avoid this problem and reinstate a satisfying evolvability into the system.

In more complex robotic problems it is unlikely that the amount and topography of the required restriction can be estimated by human inspection. The results presented in this paper must be interpreted as an indication that the evolution of neuromodulatory architectures requires a framework capable of automatically evolving the required restriction of the neuromodulatory effects, for example, by permitting the modularization of the network and the selective modular targeting of neuromodulatory connections. This result is fully compatible with and can provide a rationale for the modular and targeted structure of neuromodulatory connections observed in biological organisms [9].

6. Acknowledgments

Thanks to Steffen Wischmann, Daniel Marbach and Sara Mitri for discussions and comments on the manuscript. This work was supported by the Swiss National Science Foundation, grant no. 200021-112060.

References

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

[2] M. Bonani. e-puck education robot. http://www.e-puck.org/ as online, the 10.05.08.

[3] T. Kondo. Evolutionary design and behavior analysis of neuromodulatory neural networks for mobile robots control. *Appl. Soft Comput.*, 7(1):189–202, 2007.

[4] C. Mattiussi, P. Dürr, and D. Floreano. Center of Mass Encoding: A self-adaptive representation with adjustable redundancy for real-valued parameters. In *Proceedings of the 2007 conference on genetic and evolutionary computation (GECCO 2007), University College, London.*, pages 1304–1311, New York, 2007. ACM Press.

[5] C. Mattiussi and D. Floreano. Analog genetic encoding for the evolution of circuits and networks. *IEEE Transaction on Evolutionary Computation*, 11(5):596–607, Oct. 2007.

[6] R. McGill, J. W. Tukey, and W. A. Larsen. Variations of box plots. *The American Statistician*, 32(1):12–16, Feb. 1978.

[7] Y. Niv, D. Joel, I. Meilijson, and E. Ruppin. Evolution of reinforcement learning in uncertain environments: A simple explanation for complex foraging behaviors. *Adaptive Behavior*, 10(1):5–24, Jan. 2002.

[8] J. Reisinger and R. Miikkulainen. Acquiring evolvability through adaptive representations. In *Proceedings of the 2007 conference on genetic and evolutionary computation (GECCO 2007), University College, London.*, pages 1045–1052, New York, 2007. ACM Press.

[9] N. Schweighofer, K. Doya, and S. Kuroda. Cerebellar aminergic neuromodulation: towards a functional understanding. *Brain Research Reviews*, 44(2-3):103–116, March 2004.

[10] T. Smith, P. Husbands, A. Philippides, and M. O'Shea. Neuronal Plasticity and Temporal Adaptivity: GasNet Robot Control Networks. *Adaptive Behavior*, 10(3-4):161–183, 2002.

[11] A. Soltoggio, J. A. Bullinaria, C. Mattiussi, P. Dürr, and D. Floreano. Evolutionary Advantages of Neuromodulated Plasticity in Dynamic, Reward-based Scenarios. In *Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*, 2008.

[12] A. Soltoggio, P. Dürr, C. Mattiussi, and D. Floreano. Evolving neuromodulatory topologies for reinforcement learning-like problems. In *Proceedings of the 2007 Congress on Evolutionary Computation*. IEEE Press, 2007.