An Examination of Hypotheses Regarding Learning and Evolution

Marie Cosgrove-Davies and Mary Wootters

May 11, 2008

Abstract

The idea that learning aids evolution is an old one which has recently come under examination by computer scientists. We examined this concept by replicating the "light room"/"dark room" experiment performed by Nolfi and Parisi in simulation. In this experiment, both learning and nonlearning robots were evolved in environments with varying visibility. Several variations on the experiment were performed to test theories advanced by Nolfi and Floreano about how learning may aid evolution. Theories found to be supported were that learning allows individuals to adapt to fast changes in environment and that learning allows individuals to make more use of available information.

1 Introduction

1.1 Learning and Evolution: A Brief History

The idea that learning and evolution can aid each other dates back to Baldwin and others who, in the late 19^{th} century suggested that characteristics acquired during an indvidual's lifetime could influence the evolution of future generations [1]. The Baldwin theory is distinct from Lamarckian evolution in that it does not postulate direct passing on of traits, but instead suggests that an individual's survival rate is affected by traits developed during its life and contributes to the propagation of its genes. Since the individual's descendants are likely to be able to develop similar traits to those the individual developed, they will have an advantage if the traits are advantageous. In the later 20^{th} century computer scientists began to examine this idea with an eye towards evolving desired behaviors [2][3]. It has also been found that learning robots evolved with non-related learning and evolutionary tasks will still do better than robots which are simply evolved [4].

1.2 Self-Teaching Robots

In [5], Nolfi and Floreano make several arguments as to why a combination of learning and evolution may be beneficial to an individual. These arguments can be summarized as follows and will be referred to by number:

1. Learning allows individuals to adapt to fast changes in their environment.
2. Learning allows individuals to make more use of the information available to them, rather than restricting them to a single fitness number.
3. Learning guides evolution by smoothing the search space. If a robot is able to easily learn a good solution, its children are likely to be able to easily learn that solution, causing the population to tend towards robots able to find the solution. This is not an option with nonlearning robots, since the only change they can experience is mutation.
4. Learning allows for more complex phenotypes without adding much to the genotype.

In [6], Nolfi and Parisi put forth the idea of neural networks that generate their own teaching input. They note that, in nature, there is no external teacher to impose a goal state on an organism, so goals must originate internally. They also point out that teaching inputs for neural networks are static objects that do not adapt to changes in robot state or environment, whereas in biological systems goals emerge, develop, and are learned. Furthermore, they say that often a developer will not know the correct solution for a given task, and it would be advantageous for networks to be able to determine their own solutions, which can then be evaluated by experimenters. In their system, supervised learning is facilitated through evolution, with the best candidates being propagated but not supervised during the evolutionary process. The teaching units are a part of the network and their output units are the goal states of the non-teaching output units. They found that networks which were well-equipped to learn were the ones which survived. These networks had increased learning capacity which they explained as a co-evolution of the teaching and standard networks.

In [7], Nolfi and Parisi expand on the ideas put forth in [6] and by those put forth by Baldwin and others. They combine adaptation and evolution to create robots which, according to their results, benefit greatly from the combination of the two. In their experiment, the robot was assigned to find a target it could not see, leading to exploration of the environment. Every generation the environment toggled between "light" (high visibility) and "dark" (low visibility), meaning that the best robots would be those who had a robust search method suitable for both dark and light conditions. The thought was that robots which were able to adapt to their specific environment within their life span would have more robust methodologies.

Two sets of simulations were run, one with evolution and learning and one with evolution and no learning. The hypothesis was that evolution and learning, when combined, would produce individuals well-suited to quickly learn the task at hand, while non-learning individuals would display less of this quality. Their findings supported the hypothesis: learning robots were found to develop behaviors that were easily adaptable to light or dark environments, while non-learning robots were found to develop behaviors that worked moderately well for light and dark environments but did not work as well as the learned behaviors. They note that the non-learning robots adapted a strategy which was somewhat good for both light and dark environments, while learning robots were able to determine the type of their environment and implement an environment-based exploration strategy. Generally, characteristics of a good strategy included covering a large amount of ground and not running into a wall, since hitting a wall ended the trial.

2 Experiment One

2.1 Procedure

Our objective was to perform a variety of experiments designed to test the hypotheses put forth by Nolfi and Floreano in order to determine which were able to best explain our results. We modeled our experiments on the dark/light room experiment performed by Nolfi and Parisi.

2.1.1 Environment

Nolfi and Parisi used a real Khepera robot in an environment measuring 60x20 cm (Figure 1). The robot had eight sensor values, which were averaged into four neural network inputs (front, back, right, and left). Their target circle had a diameter of 2 cm and was positioned in a random place in the environment.

Trials ended when the robots hit the wall, since they got stuck.

We used a simulated Pioneer robot in an environment with an environment measuring 12x4 feet and a target spot with a diameter of 0.3 feet. Our light- and dark-sensing areas were initially the same, scaled up for the larger size of the Pioneer. We altered them when we found that, in dark conditions, the robot could not see the wall before crashing into it (Figure 2). We will refer to this task as the "easy task" and the Nolfi and Parisi task as the "original task". Our theory as regards this effect is that running Pyrobot at high speeds (rather than running a real robot at painfully slow speeds) altered the robot's perceptions and step sizes. This will be discussed in more detail later, as it pertains to our results as well. The Pioneer had sixteen sensor values, which we averaged into four neural network inputs as they did. Our trials ended when the robots hit the wall.

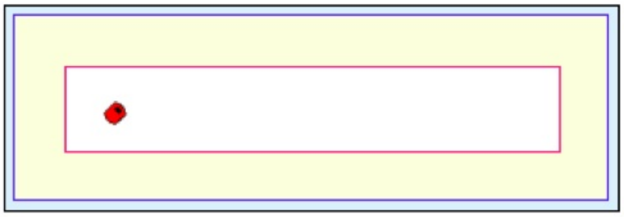


Figure 1: The world used by Nolfi and Parisi. When the walls are light-colored, the robot is able to see the walls while outside the center magenta box (in the cream and blue areas); when it's dark, it can only see the walls when it is outside the dark blue box (in the light blue area).

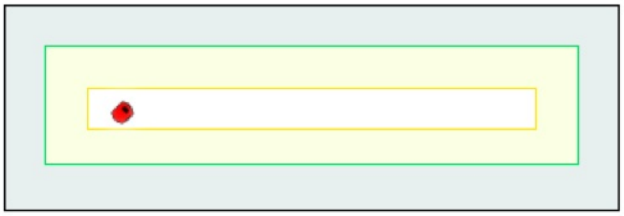


Figure 2: The world used in our experiment. When the walls are light-colored, the robot is able to see the walls while outside the center yellow box (in the cream and grey areas); when it's dark, it can only see the walls when it is outside the green box (in the grey area). Note that these dimensions are significantly different from those used by Nolfi and Parisi.

2.1.2 Brains

Their robots were controlled by a feed-forward self-teaching neural network consisting of an input layer and an output layer with no hidden units. There were four input units (one for each sensor group) and four output units. Two of the output units controlled the robot's motion – one the rotation motor and the other the translation motor – and the other two were 'teaching units' which encoded a teaching input for the other two output nodes, as described in [6]. This teaching was used for reinforcement and controlled how the network changed its connection weights during the life of the robot. The teaching network's output was used as training data for the standard network as part of its backpropagation training. The teaching network's connection weights were fixed for the life of the robot. The connection weights of the standard network were not fixed and changed according to the input from the teaching nodes. See Figure 3 for a diagram of the layout of the robot's brain.

2.1.3 Evolution

Nolfi and Parisi used genetic algorithm to evolve the robots. The population was initialized with 100 randomly generated genotypes, each of which was then tested for ten epochs, each epoch consisting of 500 learning input/output cycles. The robots were not reset between epochs. Every even generation, the robot was in a "light" room, and every odd generation, the robot was in a "dark" room, to encourage them to evolve exploration strategies suited to both environments.

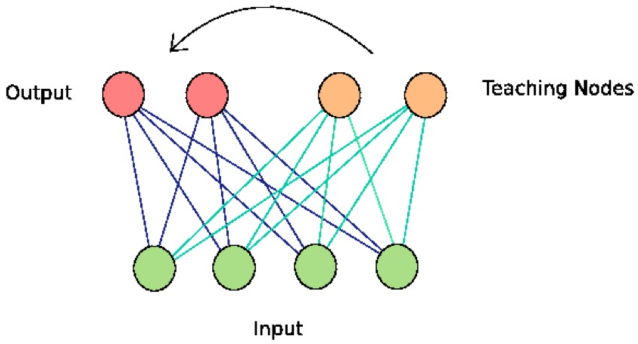


Figure 3: The robot's brain.

The fitness they assigned to each run was:

$$\text{fitness} = \sum_{\text{epochs}} \left(500 - \begin{array}{c}\text{number of timesteps} \\ \text{it took to reach target}\end{array} \right).$$

The total fitness of a genotype was the sum of its fitnesses for each run. The twenty genotypes with the highest fitness ratings were each allowed to reproduce by making five of themselves for the next generation. During reproduction, 10% of connection weights were mutated by adding a randomly selected number between -1 and 1 to their values. This was repeated for 1000 generations, alternating between "light" and "dark" environments, as pictured above. There was no crossover.

We replicated their experiment, with parameters slightly altered. Because of time constraints, we ran 1000 generations of 50 individuals, rather than 100 individuals. We ran six evolutionary runs rather than ten. Each individual did six epochs of training rather than ten. Because we observed that our robots did not seem to be moving as far as theirs and were not having a fair chance to cover ground, we allowed our robots 2000 input/output steps but only allowed them to learn every fourth step, resulting in 500 learning input/output steps. With this change made, the fitness we assigned to each run was:

$$\text{fitness} = \sum_{\text{epochs}} \left(2000 - \begin{array}{c}\text{number of timesteps} \\ \text{it took to reach target}\end{array} \right).$$

We flirted with resetting the robots between epochs but found that not resetting them produced more interesting results.

2.2 Results

Our first finding was that our robots were unable to do much of anything in the dark in the original environment, which replicated that of Nolfi and Parisi (Figure 4).

When sensor range was extended, as shown in Figure 2, there was some improvement in the dark robots' abilities to navigate (Figure 7). There was further improvement when we prevented the target from being placed around the very edge of the world, where the robots were consistently unable to find it. The behavior manifested by the best robots of the last generation is shown in Figure 5. As is expected, robots in the light and dark environments show similar behaviors, except that in light environments they are far from the walls (because they can see them from farther away) and in dark environments they are closer to the walls. It can also be noted that on this easier task, the learning and nonlearning robots had similar approaches to solving the problem: both develop some sort of wall-following.

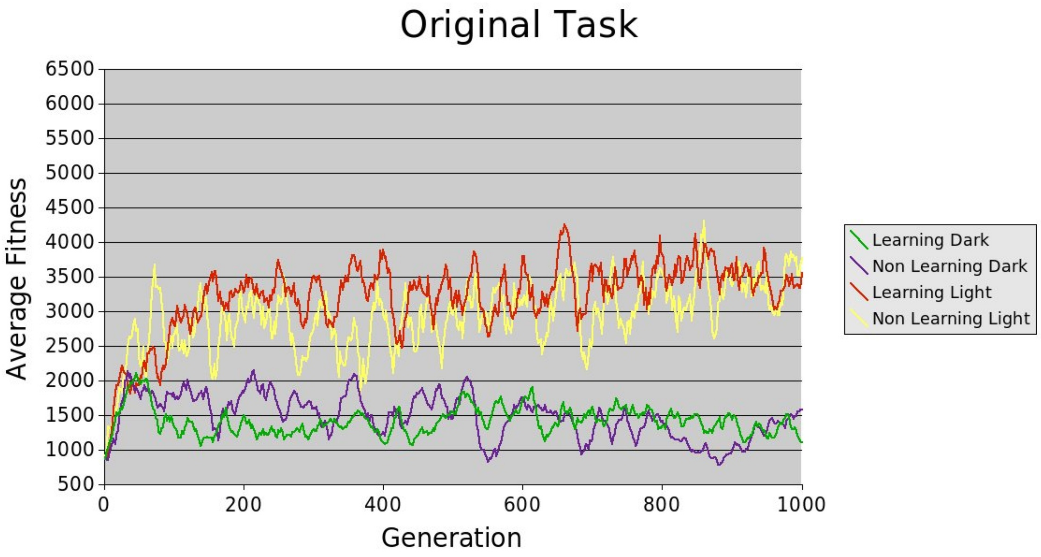
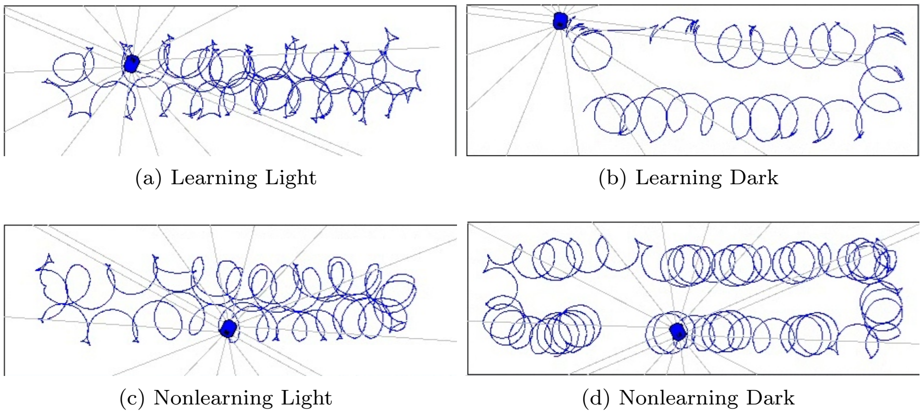


Figure 4: Our initial test of our robots' performance in an environment equivalent to Nolfi and Parisi's.


(a) Learning Light

(b) Learning Dark

(c) Nonlearning Light

(d) Nonlearning Dark

Figure 5: Robot behavior in the buffered environment. These behaviors are those manifested by the best robots of the final generations.

What we observed in the learning robots is that their behavior becomes increasingly erratic as time increases and they do not find the target (Figure 6). They begin to manifest behavior that would seem "stupid" at the outset, but since they have not been successful in finding their target with more conventional means it makes sense for them to exhibit riskier behavior that allows them to get closer to the wall. This does allow learning robots to find harder targets, such as those located close to the wall, by allowing for a change in actions from epoch to epoch. If the target is centrally located, the robot will not have to move much before finding it and will not need to learn. However, as can be seen in Figure 7, learning did not confer any significant benefit on robots who had it in either environment.

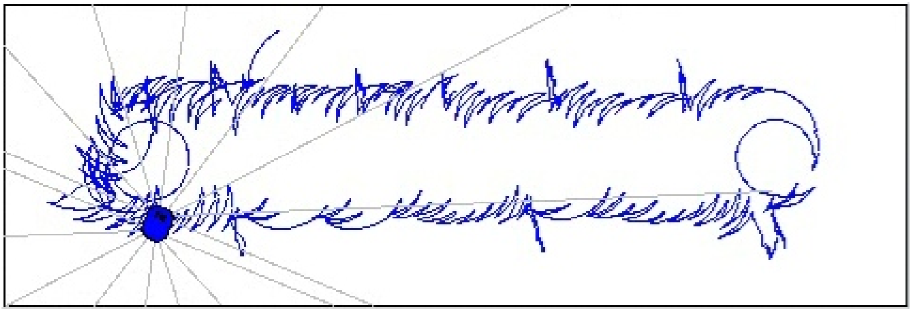


Figure 6: The behavior of a learning robot in a light environment after a few epochs.

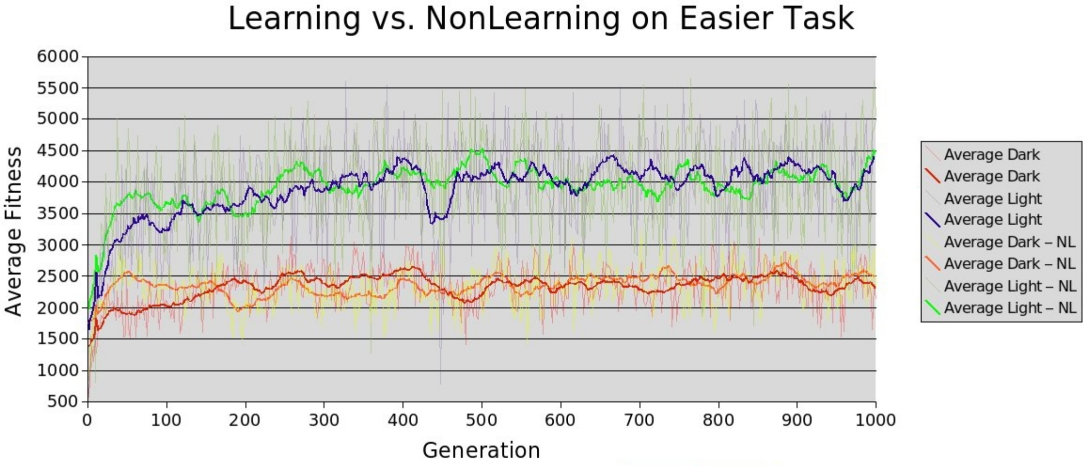


Figure 7: Our robots' performance in our additionally-buffered environment. These results are the average of six runs smoothed with a 20-point moving average. The light lines in the background are the unsmoothed data.

To see if learning would be helpful in a stable environment, we ran two trials giving the robots full sensor range within the room (Figure 8). Not only did learning robots converge more quickly but they also remained fairly consistently better at finding the target. However, looking at the unsmoothed curves in the background of the figure, it still seems that the large variance between generations prevents any firm conclusions from being made about the performance of the two.

2.3 Conclusions

Our results are quite different from the results found by Nolfi and Parisi, in which learning robots showed significantly better performance than nonlearning robots. However, the paths taken by all of our robots, learning and nonlearning, at least initially resemble the paths taken by their nonlearning robots. It is likely that the speed with which our trials were run is the problem, since we observed that the robots manifested different behavior when run at different speeds.

Nolfi and Parisi found that robot performance increased over multiple epochs in the same environment, indicating that the robot was somehow adapting to its environment. We found that this did not happen. Rather, our learning robots tended to develop stranger and stranger behavior over the course of several epochs. Our hypothesis is that drastic changes in behavior make it more likely for the robot to hit the target in at least one epoch, so this strategy was adopted. This would imply that perhaps the task of "adapt to

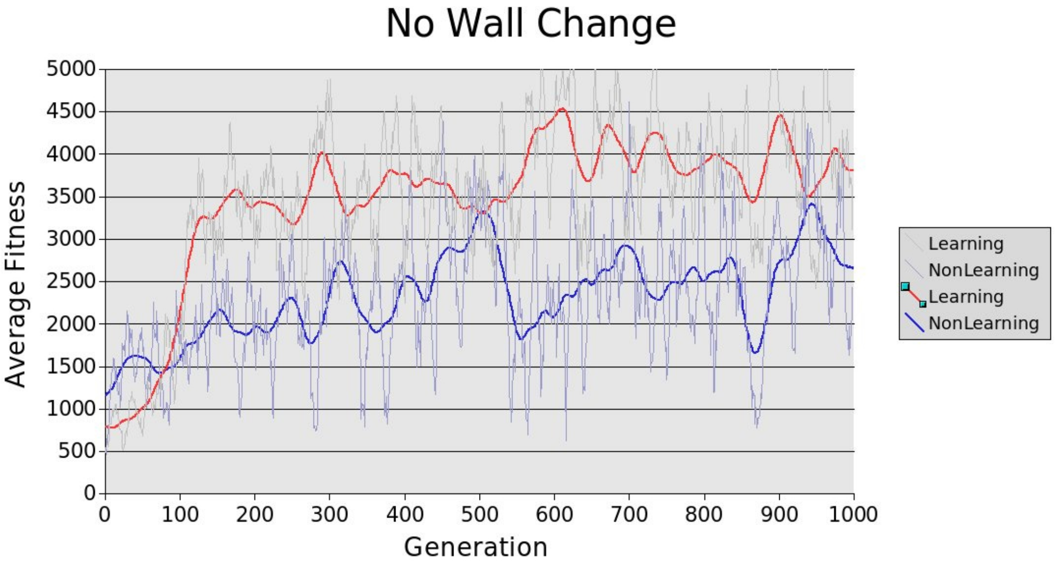


Figure 8: Our robots' performance with full sensor range in an unchanging environment. These results are the average of two runs smoothed with a 20-point moving average. The light lines in the background are the unsmoothed data.

the environment" (in a positive way) was too difficult with our setup and parameters.

If there is one thing we have learned this semester, it is that seemingly-insignificant changes in parameters can lead to completely different results, so that is a likely reason for the different results. As mentioned, the speed at which the simulation is run seems to affect the robots' behavior. We also could have had another parameter which was set strangely. Potential culprits are the vision distances and our method of shortening them (we bumped the sensor readings up to 10 at any point when the robot should not see the wall and allowed it normal sensors when it could see the wall, leading to sharp transitions). Although the increasingly strange behavior of learning robots when they could not find the target might be considered to support Hypothesis 4, we believe that we are not able to support or refute any of the enumerated hypotheses.

With regard to the trials run in the static environment, the data seem to support Hypothesis 2, 3, or 4, or some combination. It is clear that Hypothesis 1 cannot be having an effect, since there is nothing changing in the environment. Comparing Figure 8 to Figure 7 could lead one to conclude that the opposite of Hypothesis 1 is true, since it is in the stable environment that learning robots seem to do better than nonlearning robots, but as discussed above there may be other parameter factors in play. Since the conditions for these trials were not the same (even the light robots were not able to see their whole environment), it is difficult to draw firm conclusions.

3 Experiment Two

3.1 Procedure

We wanted to see how the addition of internal nodes would affect the performance of the robot. We were fairly certain that performance would improve, and we were curious to see whether the addition of internal nodes would benefit learning robots more, benefit nonlearning robots more, or bestow equal benefit.

3.1.1 Environment

The environment was our “easy” environment in terms of robot vision (Figure 2). We found that the addition of the internal nodes allowed for the unbuffered-target task to be solved (the robots could finally find targets around the edge of the area), so we reverted to unbuffered target placement.

3.1.2 Brains

Four internal nodes were added too the network, each fully connected to the input layer. Two were fully connected to the two output nodes and two were fully connected to the two teaching nodes. It should be noted that this is a substantial growth in the size of the network: it increased it by 50%.

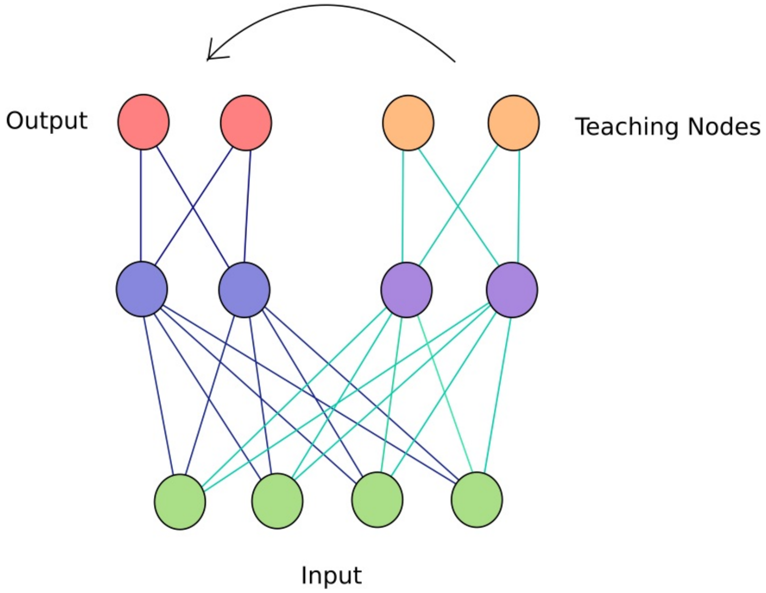


Figure 9: The robot’s brain.

3.1.3 Evolution

The robots were evolved with the parameters outlined in Experiment One, above. Nothing was changed, aside from the number of evolutionary runs.

3.2 Results

With internal nodes, the robots were finally able to navigate in the easy, un-buffered environment. In general, the learning robots converged faster than the nonlearning robots, and they tended to develop a coherent wall-following strategy, in which the robot circled farther from the wall in the light environment and closer in the dark environment. The nonlearning robots’ behaviors converged separately, developing two separate strategies. Two types of robot behavior evolved, one for light environments and one for dark environments. The light strategy consisted of wall-following, while the dark strategy was a seemingly random series of lines across the space, changing directions whenever a wall was visible (Figure 10).

From the data, it can be seen that the learning robots converged to an effective solution much more quickly than the nonlearning robots (and, as mentioned, the nonlearning robots never converged to a single effective solution!). However, it was not clear that either population dominated the other after generation 300 or so (Figure 11).

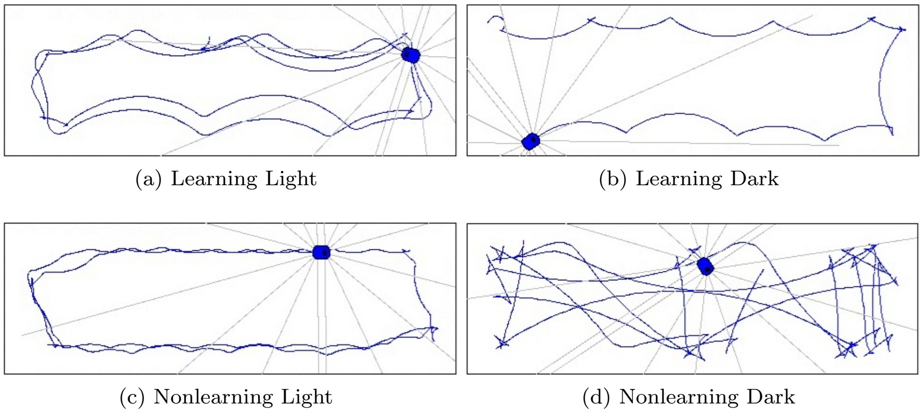


Figure 10: Internal-node-enhanced robot behavior in the original environment. These behaviors are those manifested by the best robots of the final generations.

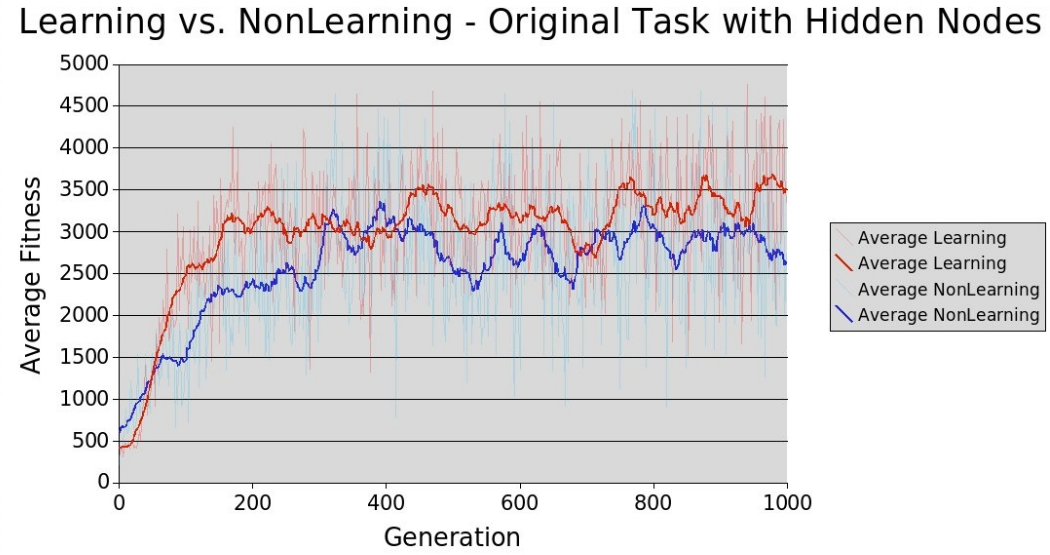


Figure 11: Our hidden-node-enhanced robots' performance on the original task. Light and dark results were averaged to produce a learning and non-learning curve. These results are the average of four runs smoothed with a 20-point moving average.

3.3 Conclusions

Observing the robots' behavior, we noticed that the learning robots generally discovered wall-following early and refined it to work with dark walls. In contrast, the non-learning robots actually developed *two* solutions, one for dark environments and one for light environments. These solutions never converged.

Both this result and the finding that learning robots converged more quickly than nonlearning robots seem to make a case for Hypothesis 3, that learning augments evolution by smoothing the search space. The addition of internal nodes renders the experiment so different from the original Nolfi and Parisi experiment that we cannot make any comparisons to their results.

4 Experiment Three

4.1 Procedure

4.1.1 Environment

For the second portion of our experiments, we examined the behavior of learning and nonlearning robots in a T-shaped environment in which the target was in either the left or right corridor at the top of the T (Figure 12). The robot obtained a small reward for each time step spent to the right of the yellow line and obtained a large reward when it found the target, located either at the green or the red spot. The reward for being to the right of the yellow line was added as an incentive to move when we noticed that the robots tended to just hover near the starting point in order to avoid accruing penalties for hitting walls.

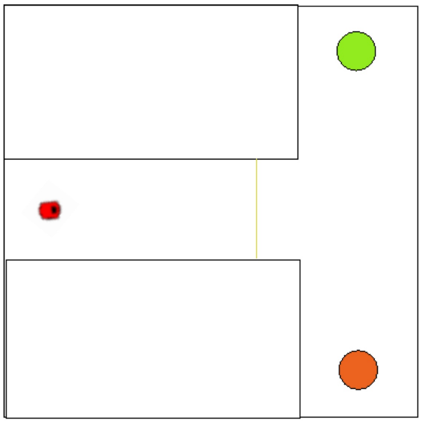


Figure 12: The world used in the second portion of our experiment.

4.1.2 Brains

The brain used in the robot is almost the same as described in Experiment Two, except for the addition of one more input node (Figure 13). The new input node was fully connected to all the internal nodes, with an input consisting of either location information or static, as described below.

4.1.3 Evolution

The robots were evolved with parameters similar to those outlined in Experiment One, above. The one difference is that, rather than changing wall color each generation, instead during odd generations the robot had an input bit telling it the location of the target ('0' for bottom, '1' for top) and during even generations it had a value of '0.5', representing no information.

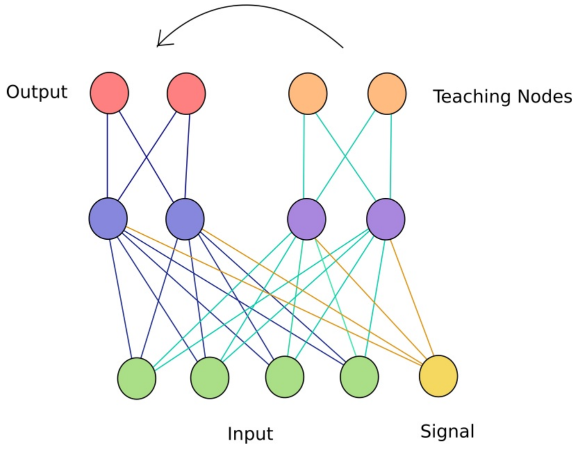


Figure 13: The robot's brain.

The fitness function was:

$$\text{fitness} = \sum_{\text{epoch}} .1 \times \left(\begin{array}{c} \text{Number of} \\ \text{timesteps} \\ \text{right of line} \end{array} \right) + 500 - \left(\begin{array}{c} \text{Number of} \\ \text{timesteps} \\ \text{to target} \end{array} \right).$$

4.2 Results

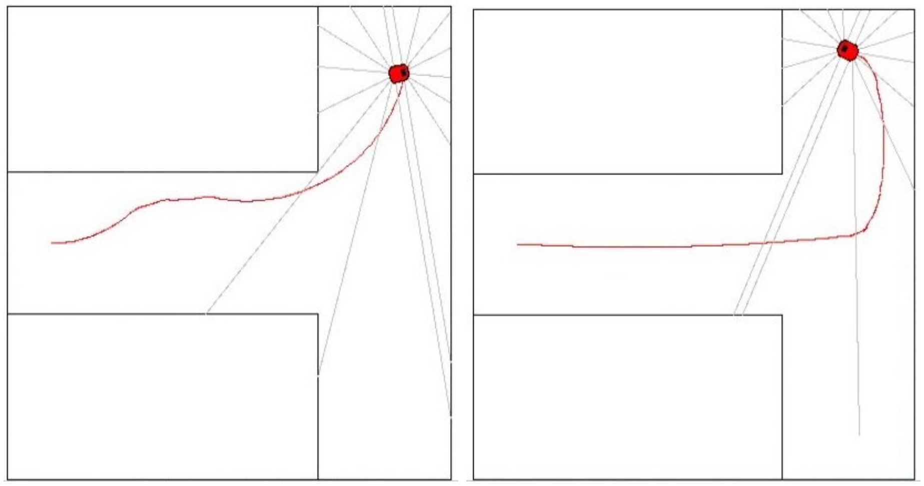
Looking at our initial results, we were excited because in the first two trials we ran, the learning robots did significantly better than the nonlearning robots. We thought that this meant that the learning robots were able to use the location bit supplied to them. However, upon observing robot behavior, we found that neither learning nor nonlearning robots were able to use the position bit: learning robots achieved their higher scores by cutting corners (Figure 14). After several trials, the advantage that we saw in learning robots has largely disappeared, although learning robots seem to do a little better (Figure 15).

It was then suggested to us that the task may have been too difficult for robots to learn even *without* the toggling on and off of the location bit, so we ran some trials to see if robots could learn to use a location bit always giving them good information. We found that they could not (Figure 16), which was rather disappointing.

It should be noted that in both the changing and stable environments there was one "dud" learning robot trial in which the robots converged to moving quickly to the right of the yellow line and then just staying there collecting small rewards. We believe that this is because the robots converged too quickly, preventing them from discovering the options available to them which would have led to more reward.

4.3 Conclusions

Although robots with internal nodes could solve the original problem, which Nolfi and Parisi solved with a perceptron, they were unable to develop a solution to a navigation problem which took advantage of a high-level location signal. The learning robots did develop a more sophisticated solution to the task of hitting both target points as efficiently as possible, indicating once again that Hypothesis 3, smoothing of search space, is the case. Our observations about hasty convergence also support Hypothesis 3, indicating a potential problem with learning over-smoothing the state space.



(a) Learning (b) Nonlearning

Figure 14: Internal-node-enhanced robot behavior in the T-Maze.

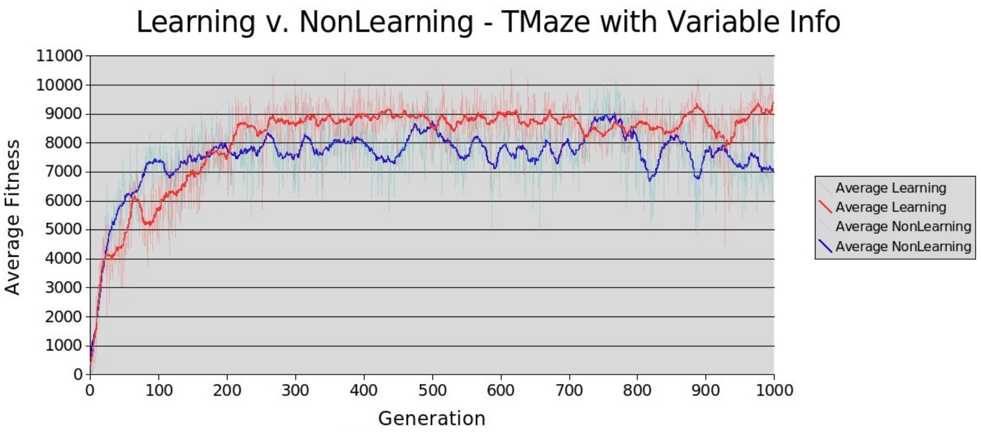


Figure 15: Our robots' performance on the T-maze task with location information 50% of the time. These results are the average of four runs smoothed with a 20-point moving average.

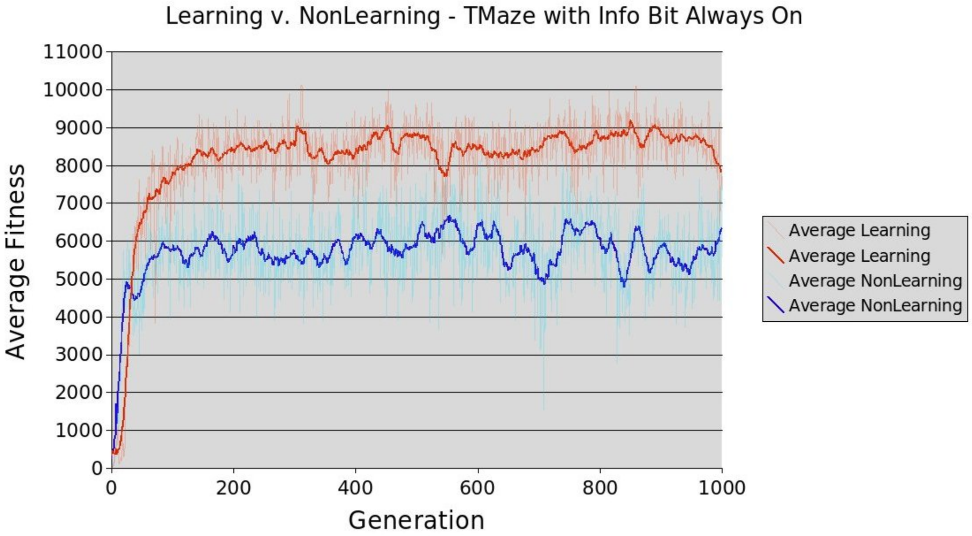


Figure 16: Our robots' performance on the T-maze task with location information all of the time. These results are the average of three learning and two nonlearning runs smoothed with a 20-point moving average.

5 Conclusions

We decided to replicate the light room/dark room experiment performed by Nolfi and Parisi, and produced decidedly different results. We believe that our results are different because of differences in test practices: whereas they ran their tests on an actual robot running in realtime, we ran ours on a simulated robot running at many times its realtime speed.

We did obtain evidence that learning during life is a valuable augmentation of evolution, and at the very least inspires different (and likely beneficial) behavior. We saw in the unchanging version of Experiment 1 and also in Experiment 3 that the learning population developed a better version of the solution developed by the nonlearning population. We saw in Experiment 2 that the learning population developed a different and more general solution than the nonlearning population, and converged on it faster. We explain these successes by referring to Nolfi and Floreano's Hypotheses 2 and 3, which provide mechanisms for learning to "guide" evolution to a more refined solution more quickly than solution could obtain the solution without help.

We did not see evidence of Hypothesis 1, as Nolfi and Parisi did. If Experiment 1 had produced the expected result of better performance by learning robots, we would have considered Hypothesis 1 to be supported. We also did not see evidence of Hypothesis 4. Only in Experiment 1 did the learning robots display behavior showing the ability to change their course over their lifetime – that is, behavior indicating a significantly more complex phenotype than their non-learning peers. This did not seem to help them. Further, if Hypothesis 4 had been active in our results, we would expect the gap between learning and nonlearning perceptrons to be larger than the gap between learning and nonlearning 3-layer networks, since the additional complexity in the 3-layer networks would have reduced the learners' advantage. This was not the case.

We did, however, see a lot of evidence that learning helped to guide evolution, as discussed above. Our conclusions therefore support Hypotheses 2 and 3. Our conclusions also supported the Baldwin theory that learning aids evolution.

References

[1] J.M. Baldwin. A new factor in evolution. *American Naturalist*, 30:536–553, 1896.

[2] Brian Carse and Johan Oreland. Evolution and learning in neural networks: dynamic correlation, relearning and thresholding. *Adapt. Behav.*, 8(3/4):297–311, 2000.

[3] Jürgen Branke. Evolutionary algorithms in neural network design and training – A review. In Jarmo T. Alander, editor, *Proc. of the First Nordic Workshop on Genetic Algorithms and their Applications (1NWGA)*, number 95-1, pages 145–163, Vaasa, Finnland, 1995.

[4] Stefano Nolfi. How learning and evolution interact: The case of a learning task which differs from the evolutionary task. *Adaptive Behavior*, 2:231–236, 1999.

[5] Stefano Nolfi and Dario Floreano. Learning and evolution. *Autonomous Robots*, 8(1), 1999.

[6] Stefano Nolfi and Domenico Parisi. Auto-teaching: Networks that develop their own teaching input. In J. L. Deneubourg, H. Bersini, S. Goss, G. Nicolis, and R. Dagonnier, editors, *Proceedings of the Second European Conference on Artificial Life*, pages 845–862, 1993.

[7] Stefano Nolfi and Domenico Parisi. Learning to adapt to changing environments in evolving neural networks, 1995.