Learning and Evolution

Stefano Nolfi* **Dario Floreano~**

*Institute of Psychology, National Research Council
Viale Marx 15, Roma, Italy
stefano@kant.irmkant.rm.cnr.it

~LAMI - Laboratory of Microcomputing
Swiss Federal Institute of Technology
EPFL, Lausanne, Switzerland
Dario.Floreano@epfl.ch

Abstract

In the last few years researchers have used artificial evolution techniques (e.g. genetic algorithms) and learning techniques (e.g. neural networks) to the study of the interaction between learning and evolution. These studies have been conducted with two different purposes: (a) looking at the advantages, in terms of performance, that the interaction gives to evolution; (b) understanding the role of the interaction between learning and evolution in natural organisms. In this paper we describe some of the most representative experiments conducted in this area and we will try to point out the implications from both these points of view. As we will show the problem of understanding the interaction between learning and evolution is probably one of the best examples in which computational studies have shed light on problems that are difficult to study with the research tools of evolutionary biology and biology in general. From and engineering point of view, the most relevant results are those showing that adaptation in dynamic environments is favored by a combination of evolution and learning. These studies also demonstrate how the interaction between learning and evolution deeply alters the evolutionary and the learning process themselves. The study of learning in an evolutionary perspective, is still in its infancy. We believe that the study of learning in interaction with evolution will produce in the next years an enormous impact on our understanding of what learning and evolution are.

1. Introduction

Evolution and learning are two forms of biological adaptation that differ in space and time. Evolution is a process of selective reproduction and substitution based on the existence of a geographically-distributed population of individuals displaying some variability. Learning, instead, is a set of modifications taking place within each single individual during its own life time. Evolution and learning operate on different time scales. Evolution is a form of adaptation capable of capturing relatively slow environmental changes that might encompass several generations, such as perceptual characteristics of food sources for a given bird species. Learning, instead, allows an individual to adapt to environmental changes that are unpredictable at the generational level. Learning might include a variety of mechanisms that produce adaptive changes in an individual during its lifetime, such as physical development, neural maturation, and synaptic plasticity. Finally, whereas evolution operates on the genotype, learning affects only the phenotype and phenotypic changes cannot directly modify the genotype.

In the last few years researchers have used artificial evolution techniques (e.g. genetic algorithms) and learning techniques (e.g. neural networks) to the study of the interaction between learning and evolution. These studies have been conducted with two different

purposes: (a) looking at the advantages, in terms of performance, that the interaction gives to evolution; (b) understanding the role of the interaction between learning and evolution in natural organisms. In this paper we will describe some of the most representative experiments conducted in this area and we will try to point out the implications from both these points of view.

We will use the term 'learning' to indicate modifications to the synaptic weights of a neural network during the lifetime of one individual. These modifications may produce behavioral changes that increase the adaptivity of the individual. However, it is clear that adaptive changes might be obtained by modifying other entities such as, for example, the activation states of internal neurons, the network architecture, and the sensory-motor characteristics of the organism. Connectionists usually refer to long-lasting changes in the synaptic weights as *learning* and to the activation states of neurons with recurrent connections as *memory*. Learning and memory are both forms of plasticity in that they may produce similar outcomes from a behavioral point of view. For example, Yamauchi and Beer (1995) have evolved and analyzed continuous-time recurrent neural networks that give the external appearance of performing reinforcement learning while in fact they have fixed connection weights and use only the dynamics of the neuron activations. A better understanding of the functional differences between different ways of realizing ontogenetic plasticity is an important issue that remains to be investigated.

2. The adaptive functions of learning in evolution

Within an evolutionary perspective, learning has at several different adaptive functions:

1) *It allows individuals to adapt to changes in the environment that occur in the lifespan of an individual or across few generations.* As mentioned in the previous section, learning has the same function attributed to evolution: adaptation to the environment. Learning supplements evolution in that it enables an organism to adapt to changes in the environment that happen too quickly to be tracked by evolution (Todd & Miller, 1991; Nolfi, Miglino & Parisi, 1994; Floreano & Nolfi; 1997, Nolfi & Parisi, 1997; Sasaki & Tokoro, 1997).

2) *It allows evolution to use information extracted from the environment thereby channelling evolutionary search.* Whereas ontogenetic adaptation can rely on a very rich, although not always explicit, amount of feedback from the environment, evolutionary adaptation relies on a single value which reflects how well an individual coped with its environment. This value is the number of offspring in the case of natural evolution and the fitness value in the case of artificial evolution. Instead, from the point of view of ontogenetic adaptation, individuals continuously receive feedback information from the environment through their sensors during the whole lifetime. However, this huge amount of information encodes only very indirectly how well an individual is doing in different moments of its life or how it should modify its own behavior in order to increase its fitness. However, ontogenetic and phylogenetic adaptation together might be capable of exploiting this information. Indeed evolution may be able to transform sensory information into self-generated reinforcement signals or teaching patterns (Ackley and Litmann, 1991; Nolfi & Parisi, 1993; Nolfi & Parisi, 1994; Floreano & Mondada, 1996; Nolfi & Parisi, 1997)

3) *It can help and guide evolution.* Although physical changes of the phenotype, such as strengthening of synapses during learning, cannot be written back into the genotype, Baldwin (1896) and Waddington (1942) suggested that learning might indeed affect the evolutionary course in subtle but effective ways. Baldwin's argument was that learning accelerates evolution because sub-optimal individuals can reproduce by acquiring during life necessary features for survival. However, since learning requires time (and might thus be a disadvantage), Baldwin suggested that evolution tends to select individuals who have already

at birth those useful features which would otherwise be learned. This latter aspect of Baldwin's effect, namely indirect genetic assimilation of learned traits, has been later supported by scientific evidence and defined by Waddington (Waddington 1942) as a *canalization effect*. Recently, Hinton & Nowlan (1987) have provided a clear computational model that demonstrates how learning may help and guide evolution. Nolfi, Elman & Parisi (1994) have further investigated this issue a the case in which the learning task differs from the evolutionary task.

4) *Other advantages*. Learning might allow to produce complex phenotype with short genotypes by extracting some of the information necessary to build the corresponding phenotypes from the environment (Todd & Miller, 1991; Mayley, 1997). Moreover learning can allow the maintenance of more genetic diversity. Different genes, in fact, have more chances to be preserved in the population if the individuals who incorporate those genes are able to learn the same fit behaviors (Whitley, Gordon & Mathias, 1994)

However, learning has also costs:

1) *A delay in the ability to acquire fitness*. Learning individuals will necessarily have a sub-optimal behavior during the learning phase. As a consequence they will collect less fitness than individuals who have the same behavior genetically specified. The longer the learning period, the more accumulated costs have to be paid (Mayley, 1997).

2) *Increased unreliability*. "Since learned behavior is determined , at least partly, by the environment, if a vital behavior-defining stimulus is non encountered by a particular individual, then it will suffer as a consequence. The plasticity of learned behaviors provides the possibility that an individual may simply learn the wrong thing, causing it to incur an incorrect behavior cost. Learning thus has a stochastic element that it is not present in instinctive behaviors" (Mayley, 1997, pp.216).

3) *Other costs*. In natural organisms or in biologically inspired artificial organisms learning might implies additional costs. If individuals are considered juvenile during the learning period, learning also implies a delayed reproduction time (Cecconi, Menczer & Belew 1996). Moreover, learning might implies the waste of energy resources for the accomplishment if learning process itself (Mayley, 1997) or for parental investment (Cecconi, Menczer & Belew 1996). Finally, while learning, individuals without a fully formed behavior may irrevocably damage themselves (Mayley, 1997).

In the next sections we will present a set of models and experiments devised to study the interaction between learning and evolution and we will discuss the implications from the different perspectives described above.

3. How learning can 'help and guide' evolution

Hinton and Nowlan (1987) have described a simple computational model that shows how learning might help and guide evolution in some circumstances. The authors considered the case in which "a neural network confers added reproductive fitness on an organism only if it is connected in exactly the right way. In this worst case, there is no reasonable path toward the good net and a pure evolutionary search can only discover which of the potential connections should be present by trying possibilities at random. The good net is like a needle in a haystack" (p. 495). In their experiment individuals have genotypes with 20 genes which encode a neural network with 20 potentials connections. Genes can have three alternative value: 0, 1, and ? that represent, respectively, the presence of the connection, the absence of the connection, and a modifiable state (presence or absence of the connection) that can change its value according to a learning mechanism. The learning mechanism is a simple random process that keeps changing modifiable connection weights until a good combination (if any) is found during the limited life time of the individual.

In the absence of learning (i.e., when genes can only have 0 and 1 values), the probability of finding a good combination of weights would be very small given that the fitness surface (see box) would look like a flat area with a spike in correspondence of the good combinations (see Figure 1, thick line). On such a surface genetic algorithms do not perform better than any random search algorithm. However, if learning is enabled, it will be more probable that some individuals will achieve the good combinations of connection values at some point during life and start to collect fitness points. The addition of learning, in fact, produces an enlargement and a smoothing of the fitness surface area around the good combination which can be discovered and easily climbed by the genetic algorithm. This is due to the fact that not only the right combination of alleles but also combinations which in part have the right alleles and in part have unspecified (learnable) alleles have an average fitness greater than 0 (fitness is proportional to the number of fixed right values because the time needed to find the right combination is inversely proportional, on the average, to the number of learnable alleles).

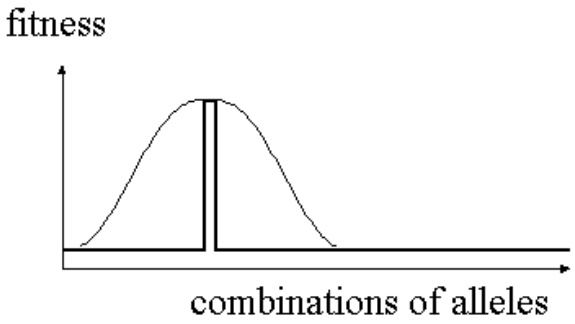


Figure 1. Fitness surface with and without learning. In absence of learning, the fitness surface is flat, with a thin spike in correspondence of the good combinations of alleles (thick line). When learning is enabled, the fitness surface has a nice hill around the spike which includes the alleles combination which have in part right fixed values and in part unspecified (learnable) values (thin line). The thick line represents the fitness for each possible combination of two alleles ([0, 1]) while the thin line represents the fitness for each possible combination of three alleles [0, 1, ?]). Redrawn from Hinton and Nowlan (1987).

In other words, learning makes the fitness surface smoother, and this, in turn, simplifies the search which should be performed by evolution. As claimed by Hinton and Nowlan, with learning "it is like searching for a needle in a haystack when someone tells you when you are getting close" (p. 496). This simple model also accounts for the Baldwin effect that postulates that characters that are initially acquired through learning may later be fixated in the genotype. Once individuals which have part of their genes fixed on the right values and part of their genes unspecified (learnable) are selected, individuals with less and less learnable genes tend to be selected given that fitness is inversely proportional to the number of learnable genes (although an equilibrium point is reached, see Hinton and Nowlan, 1987). In other words, characters that were first acquired through learning tend to become genetically specified later on[1].

[1] One might wonder if Lamarckian evolution (i.e. an evolutionary process in which characters acquired through learning are directly coded back into the genotype and transmitted to the offspring) could be more effective that Darwinian evolution (i.e. an evolutionary process in which characters acquired through learning are discarded). Ackley & Littman (1994) for instance claimed that in artificial evolution, in which inherited characters can be easily coded into the genotype given that the mapping between the genotype and the phenotype is generally quite simple, there is no reason to not use Lamarckian evolution. Indeed the authors showed how Lamarckian evolution is far more effective than Darwinian evolution in a stationary environment. On the other hand, as

In the representation adopted in Figure 1 each individual is represented as a point in the fitness surface with a height corresponding to the average fitness of the individual during its lifetime. This is a static representation in which changes in performance during lifetime cannot be visualized. Another way of representing the individuals in the search space is to imagine that each individual network corresponds to a point in a phenotype space. In this case the changes in the connection weights introduced by learning correspond to a movement of the learning individual in the phenotype space and on the fitness surface. As a consequence, changes in performance correspond to movements toward higher or lower areas of the fitness surface. By using this type of representation the interaction between learning and evolution in the Hinton and Nowlan's model can be explained in the following way. The fitness surface (i.e. the distribution of fitness value that are assigned to each possible combination of 0 and 1 alleles) is flat with the exception of one spike of high fitness both in the case or learning and non-learning individuals both in the case of learning and non-learning individuals. The initial population will be represented as a set of fixed points randomly distributed on the fitness surface (see the cross-marks in Figure 2). Non-learning individuals do not move in the phenotype space during their lifetime while learning individuals do (given that part of their connections continually changes value during lifetime). Clearly, the probability that at least one will step on the spike will be much higher in the case of learning individuals. In other words we can say that learning allow the evolutionary process to explore the landscape surrounding each candidate for reproduction (Nolfi, Elman, and Parisi, 1990).

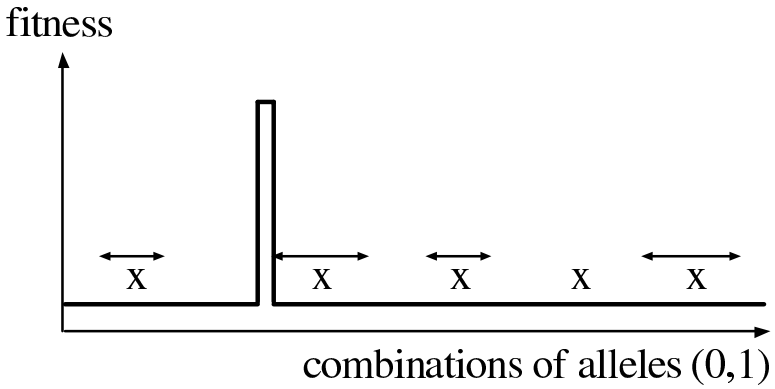


Figure 2. Representation in the phenotype space of the same fitness surface described in Figure 1. The full line represents the fitness for each possible combination of two alleles (0, 1). The cross-marks represent the positions of the individuals at birth. The arrows represent the movements of the individual in the search space corresponding to the changes modifiable alleles. Notice however that only learning individuals move in the search space during lifetime.

If learning is represented as a random process (such as in the case of Hinton and Nowlan's model), the representations shown in Figure 1 and 2 are functionally equivalent. However the latter representation is more appropriate in the cases in which learning is modeled as a form of change that has a directionality.

Despite its explicative power, Hinton and Nowlan's model has several limitations: (1) learning is modeled as a random process; (2) there is no distinction between the learning task and the evolutionary task; (3) The environment does not change; (4) The learning space and the evolutionary space are completely correlated. The two spaces are correlated if genotypes which are close in the evolutionary space correspond to phenotype which are close in the phenotype spaces (Mayley, 1997).

---

shown by Sasaki & Tokoro (1997), Darwinian evolution largely outperform Lamarckian evolution when the environment is not stationary or when different individuals are exposed to different learning experiences.

The results obtained by Hinton and Nowlan may not generalize completely to other circumstances in which these limitations are released. In particular, they may not generalize to cases in which the learning and the evolutionary spaces are less correlated. In the case of the Hinton and Nowlan model, learning and evolution operate on the same entities (i.e. the connection weights) with the same operators (i.e. both changes produced by mutations and changes produced by learning correspond to substitutions of genes with new values which are randomly selected). Therefore the two spaces are completely correlated. By systematically varying the cost of learning and the correlation between the learning space and the evolutionary space in another experimental set Mayley (1997) showed how: (1) the adaptive advantage of learning is proportional to the correlation between the two search spaces; (2) the assimilation of characters first acquired through learning is proportional to the correlation between the two search spaces and to the cost of learning (i.e. to the fitness lost during the first part of the lifetime in which individual have sub-optimal performance); (3) in certain conditions learning costs may exceed learning benefits.

4. Evolving individuals which learn a task different from what they are selected for

As we claimed in the previous section, one of the limitations of the Hinton and Nowlan's model is that there is no distinction between the learning task and the evolutionary task. This is possible because the experimenter provide the supervision signals both for the evolutionary and the learning task. In natural evolution instead, the environment does not usually provide cues that directly indicate to the individual how it should change in order to produce more adapted behavior. Natural selection is the only source of supervision for many living systems. However, natural organisms can use environmental information, made available to them through their sensors, to acquire competencies (such as the ability to predict the next sensory states see Nolfi & Tani, in press) that may indirectly increase their ability to reproduce.

Nolfi, Elman & Parisi (1994) have studied the case of artificial agents (also known as animats, see Wilson, 1987) that evolve (to become fitter at one task) at the population level and learn (a different task) at the individual level. In particular, individuals which were selected for their ability to find food in their environment were also asked to learn to predict the sensory consequences of their motor actions during their lifetime. Notice how the supervision necessary for learning this task is directly available from the environment (i.e. the correct prediction correspond to the state of the sensor at time t_{+1}).

Each individual animat lives in a two-dimensional grid world where a number of food tokens are randomly distributed. Each food token occupies one cell; if the animat happens to step on one of these cells, the food token is automatically "eaten" and the animat's fitness is increased. Individuals are equipped with a neural network interfaced to a sensorimotor system that provides input information on the distance and angle (with respect to the facing direction of the animat) of the nearest food token, and on the planned motor action. Two input units encode the angle and the distance of the nearest food token and two other units (thresholded to the nearest binary value) encode one of four possible actions: turn 90° right, turn 90° left, move one cell forward, and remain still. At each time step, the neural network receives as input the sensory information on the nearest food token and the current planned motor action and produces as output the next planned action and a prediction of the sensory state after the execution of the current planned action. At this point: (a) the planned action that was used as input is executed and the next planned action is passed as new input; (b) the freshly-gathered sensory information is used both as input and as teaching input for the output units encoding the predicted state of the sensors (the new sensory state is compared

with the predicted state and the difference (error) is used to adjust by back-propagation the connection weights between the four input, the seven hidden, and the two prediction units).

The initial population is composed of 100 individuals. At the end of life the 100 individuals are ranked in terms of their fitness (total number of food elements eaten during life) and the best 20 individuals are allowed to reproduce by generating 5 copies each of their connection weights. The inherited original weight matrices (changes due to learning during life are discarded) are mutated by selecting 5 weights at random and perturbing the weights value by adding a quantity randomly selected.

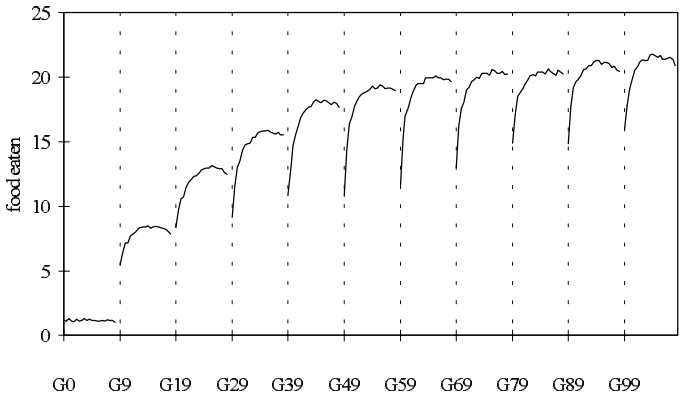


Figure 3. Average number of food elements eaten by populations of successive generations which learn to predict. Each curve represents performance prior to learning and then for each of the 20 epochs of life (performance prior to learning are obtained by measuring the number of food tokens eaten by individuals during one epoch of life without updating the weights). For reasons of space, performance are displayed only each 10 generations. Average results over 10 replications of the experiment.

The results showed that, after a few generations, individuals learning to predict also increased their ability to find food during life (Figure 3).

Moreover, by comparing the results of the experiments described above with another set of experiments in which individuals were not allowed to learn to predict during their lifetime, it was shown that learning populations displayed faster and higher fitness values across generations than populations without learning (Figure 4). The same type of results were obtained in other cases and in particular in cases in which the learning task and the evolutionary task were clearly "uncorrelated" (see Parisi, Nolfi & Cecconi, 1992; Harvey, 1997).

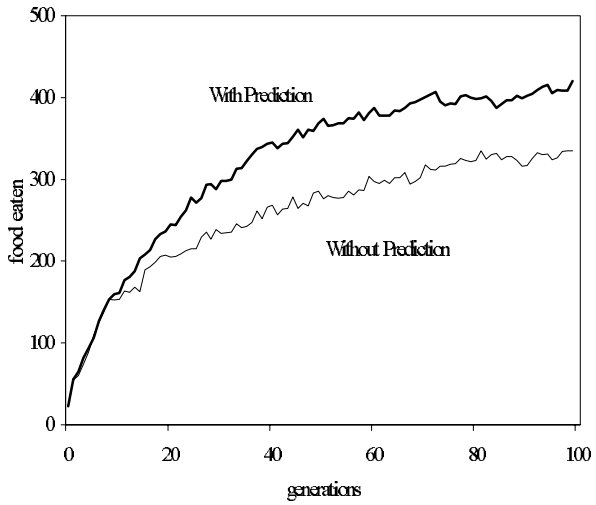


Figure 4. Average of food tokens eaten by populations of animats throughout generations for experiments with and without learning. Each curve is the average result of 10 replications. The difference between the two curves is statistically significant from generation 25 (see Nolfi, Elman, and Parisi, 1994).

Since here the learning criterion is different from the evolutionary goal and learning has a directionality (i.e., the weights are not changed in a random fashion), the explanation by Hinton and Nowlan depicted in Figure 1 and 2 is not sufficient for explaining these results. Therefore, a new explanation of the interaction between learning and evolution has been proposed (Nolfi, Elman & Parisi, 1994; Parisi & Nolfi, 1996). Imagine two different search surfaces, an evolutionary surface and a learning surface (Figure 5). Changes due to learning produce a movement of the individual phenotype both on the learning and the evolutionary surfaces. However, because learning tries to maximize performance on the learning task, individuals will move toward the higher area of the learning surface. Given that the way in which individuals move in weight space affects their fitness (the total fitness of the individual is the sum of the fitness values received during such displacements on the weight space) evolution will tend to select individuals located in areas in which, by increasing their performance on the learning task, they also increase their performance on the evolutionary task.

Consider for example two individuals, *a* and *b*, which are located in two distant locations in weight space but have the same fitness at birth; i.e., the two locations correspond to the same height on the fitness surface (cf. Figure 5). However, individual *a* is located in a region in which the fitness surface and the learning surface are dynamically correlated; i.e., a region in which movements that result in an increase in height with respect to the learning surface tend to cause an increase also with respect to the fitness surface. Individual *b*, on the other hand, is located in a region in which the two surfaces are not dynamically correlated. If individual *b* moves in weight space it will go up in the learning surface but not necessarily in the fitness surface. Because of learning, the two individuals will move during their lifetime in a direction that improves their learning performance, i.e., in a direction in which their height on the learning surface tends to increase. This implies that individual *a*, which is located in a dynamically correlated region, will end up with a higher fitness than individual *b* and, therefore, will have a better chance to be selected. The final result is that evolution will have a tendency to progressively select individuals which are located in dynamically correlated regions. In other words, learning forces evolution to select individuals which improve their performance with respect to both the learning and the evolutionary task.

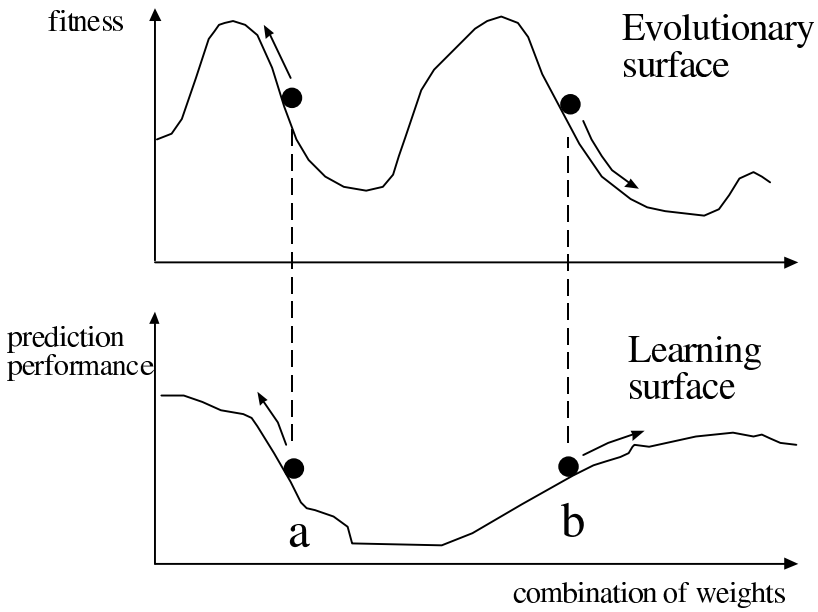


Figure 5. Fitness surface for the evolutionary task and performance surface for the learning task (sensory prediction) for all possible weight matrices. Movements due to learning are represented as arrows. Point *a* is in a region in which the two surfaces are dynamically correlated. Even if *a* and *b* have the same fitness on the evolutionary surface at birth, *a* has more probability to be selected than *b* since it is more likely to increase its fitness during life than *b*.

These results show that although evolution and learning are two distinct types of change occurring in two distinct entities (populations and individual organisms), they strongly influence each other. The influence of evolution on learning is not surprising. Evolutionary change leaves its trace in the genotype. Hence, each individual inherits a genome which is the cumulative result at the level of evolutionary changes that have occurred at the level of the population. Since an individual's genome partially specifies the resulting phenotypic individual and it constrains how the individual will behave and what it will learn, the way is open for an influence of evolution on learning. However, the experiments described in this and in the previous section clearly show that the reverse is also true: learning affects evolution. The interaction between the two processes is so profound that learning tends to produce a positive effect on evolution even if the learning task and the evolutionary task are different (and, apparently, independently from what the learning task is, see Parisi, Nolfi, and Cecconi, 1992).

In two recent articles, Harvey (1996, 1997) proposed a different explanation of the interaction between evolution and learning than the one described in this section. He claimed that the improvement in average performance observed in the learning individuals described in Figure 4 and 5 is not due to an interaction between learning and evolution but "rather from a relearning effect on weights perturbed by mutation" (Harvey, 1997, p.328). Harvey's hypothesis is based on evidence that by perturbing the weights of a neural network previously trained with back-propagation on a set of input-output pairs and then retraining the network on a subset of the original training set, performance also improves on the other input-output pairs of the original training set not included in the retraining (Hinton & Plaut, 1987). The same effect is observed also when the second training set is uncorrelated with the original training set (Harvey & Stone, 1996).

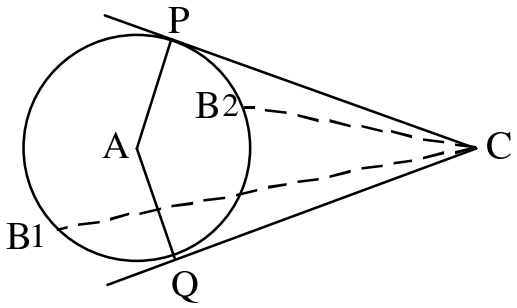


Figure 6. A two-dimensional representation of the search space. Redraw from Harvey (1996).

Harvey's explanation is based on a geometrical argument. Briefly stated, the trajectory of the network in the weight space during training on the second set is very likely to transit by the original point where it was before being perturbed by noise. Assume that A represents the weights of the network trained on the original training set, B1 and B2 are two possible positions of the network after perturbation, and C is the position of the network after being trained on the second training set. Finally, assume that performance on the original set is inversely proportional to distance from point A. Therefore, whenever B lies outside the inner arc PQ (e.g., B1), its trajectory gets closer to A for some time; instead, whenever B lies inside the inner arc PQ (e.g., B2), its trajectory always goes away from A. Regardless of the position of C, the former situation happens at least 50% of the times for a 2-dimensional weight space and much more often in a high-dimensional weight space (Harvey, 1997); furthermore, it happens 100% of the times when C lies within the circle.

On the basis of this new explanation, Harvey claims that the beneficial effects of learning a different task than the evolutionary task can be explained by considering a highly converged evolved population sitting on point A (food finding) being pulled away by mutations to point B, and then transiting to point C with prediction learning. As a consequence he hypothesized that "if one substituted for the elite member of a population evolved on the food-finding task one individual trained by back-propagation using an external teacher (or any other learning mechanism) on the same task, then one should expect similar responses after weight perturbations" (Harvey, 1986, p. 83). In order to test this prediction, we measured the performance of individuals of successive generations which were allowed to learn for the first time (i.e. individuals which had the same architecture of learning individuals but which evolved without being exposed to learning during lifetime, see Nolfi, 1998). In contrast with Harvey's expectation, learning to predict produces a significant decrease in performance of these individuals even though their weights have been perturbed by mutations (exactly like the weights of individuals which were exposed to learning in previous generations). This and other results described in Nolfi (1998) suggest that: (a) the advantages produced by lifetime learning are due to the interaction between learning and evolution; (b) in the case of learning individuals, the population does not converge on A but on some point on the left side of A which ensures that by learning (i.e. by moving toward C) individuals will spend most of their lifetime close to A. This explanation fits nicely the suggestion given above that evolution tends to select individuals which are located in dynamically correlated regions of the fitness and learning surfaces.

It should be noted that in these experiments characters initially acquired through learning are not assimilated in the individuals' genotype later on, at least completely, even if learning has a cost (i.e. individuals which acquire fit behaviors through learning start with lower performance and therefore collect less fitness in the first part of their lifetime). This may be explained by considering that: (a) the cost of learning can be reduced by increasing the learning speed (as can be observed in Figure 3, evolved individuals reach optimal

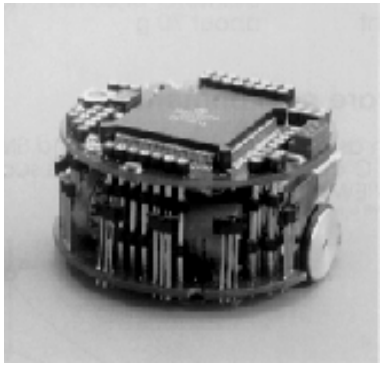
performance in the very first epochs of their lifetime); (2) the learning space and the evolutionary space are not completely correlated (on the correlation between the learning space and the evolutionary space see Mayley, 1997). In fact, although learning and evolution operate on the same entities (i.e. the connection weights), the genetic operators and the learning operators are quite different (mutations are accomplished by adding a random selected values to a set of randomly selected connection weights while learning is accomplished by back-propagation). This implies that points which are close in the learning space may be far apart in the evolutionary space. In other words, points, which can be reached with few learning cycles, may require a prohibitively long list of mutations. As we claimed in the previous section the probability to genetically assimilate characters acquired through learning is inversely proportional to the correlation between the learning and the evolutionary space.

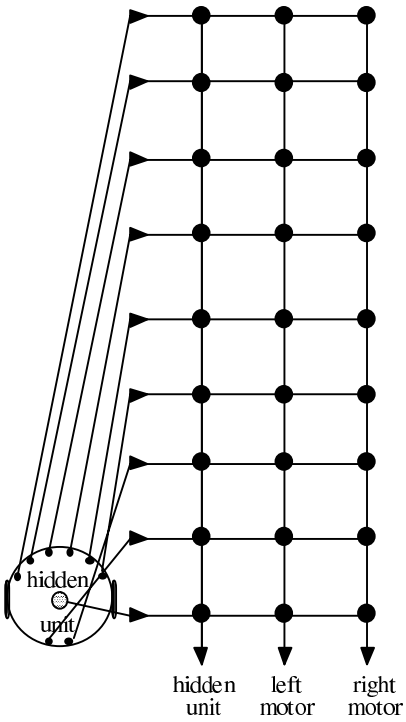
5. Exploiting the regularities of the environment through learning

In all the experiments described so far the weights were genetically inherited although, in the models described in the previous two sections, the inherited weights were also subjected to changes during the lifetime of individuals. However in most of the species the genome does not contain all the information necessary to build the corresponding phenotype. It is therefore unlikely that fine details such as initial weight values could be precisely encoded on the genotype.

In this section we will describe an experiment (Floreano & Mondada, 1996) in which the neural network is used to control a mobile robot and in which the synaptic weights are not precisely encoded into the genotype, but are continuously modified during lifetime through a learning process in which genetically-inherited instructions interact with information coming from the external environment. In other words, the genotype encodes only the architecture and "meta-properties" of the synapses, but not their precise strengths. Every time that a chromosome is decoded into the corresponding neural controller, all the synapses are intialized to small random values. As soon as the robot begins to move and sensory signals flow through the network, synaptic values can change according to the genetically encoded meta-properties. These meta-properties include four possible learning rules, a learning rate, and chemical properties of the synapse, such as whether it is excitatory or inhibitory and whether it drives or modulates the postsynaptic neuron[2]. Therefore, the accurate balance between weighted signals necessary to drive the motor neurons in a coordinated fashion must be learned during lifetime according to the genetically specified instructions.

[2] It should be noticed that such encoding is rather irrealistic from a biological point of view and one might correctly argue that biological genotypes do not encode characteristics of individual synapses. However, here the point is that of studying the interaction between learning and evolution by preventing evolution alone to find a precise behavioral solution and giving it the possibility to exploit lifetime learning.





Inserire foto del mondo qui orientata in verticale invece che in orizzontale

Figure 7. **Left**: The Khepera robot. **Center**: the architecture of the neural network employed. Black disks are synapses; the circle in the middle of the robot body represents the hidden neuron. The activation of the three units correspond, respectively, to the hidden unit, the left motor, and the right motor. **Right**: Bird-view of the environment with the robot.

The agent in this case was modeled by using the mobile robot shown on the left side of Figure 7 (Mondada, Franzi & Ienne, 1993). The robot is supported by two wheels that are controlled by two DC motors with an incremental encoder and that can move both forward and backward and has 8 infrared proximity sensors which can detect obstacles up to a distance of about 4 cms. The robot is controlled by a neural network with three units: one hidden neuron and two motor neurons, each one receiving input/activation via synaptic connections from all eight infrared sensors and from the hidden neuron itself (Figure 7, center). Signals going through the synaptic connections, which could have a driving or a modulatory effect on the postsynaptic neuron, were combined in a two-component activation function (Phillips, Kay & Smyth 1995) that generated an output between 0 and 1. For the motor neurons, the sum of the driving signals determined the direction of rotation of the wheel, whereas the sum of modulatory signals could enhance or reduce rotation speed, but could not change the direction of rotation. The genotype of each individual contains 6 bits for each synapse which encode the meta-properties of the corresponding synapse: driving or modulatory (1 bit), excitatory or inhibitory (1 bit), four learning rules (2 bits), and four learning rates (2 bits). Each individual synapse could change its strength according to one of four Hebbian learning rules (Willshaw & Dayan, 1990): pure Hebbian, postsynaptic, presynaptic, and covariance. These learning rules included a decay factor so that synaptic strengths were always bound within the interval [0.0, 1.0] and their signs were genetically specified (second bit of each gene). The final weight values were not coded back into the genotype. Individuals were selected for their ability to navigate in the environment shown in the lower left of Figure 7 as fast as possible while keeping far from obstacles. For all other details see Floreano and Mondada (1996).

All the best neural networks of the last generation could control the robot in order to keep a straight trajectory while avoiding obstacles. The evolved behaviors resulted in smooth paths around the arena. The neural networks learned to navigate starting from random initial

values assigned to the synapses. The acquisition of the ability to navigate in the environment is very fast: in less than 10 sensory motor loops the best individuals were already capable of moving forward without getting stuck into walls.

Figure 8 shows the trajectory of one of the best evolved controllers in two successive laps of the looping maze. Initially, the synapses were randomly initialized and the robot was positioned facing a corner of the inner wall (Figure 8, left; initial position corresponds to the set of superimposed bars in the lower portion of the environment). During the first 2 seconds (6-7 synaptic updates), the robot adjusted its position alternating back-and-forth movements until it found a wall on its right side. This initial behavior was quite stereotypical: it was displayed for any starting position. Once a wall had been found, the robot moved forward keeping it at a distance of 2 cm from its own right side. Every second or third action, it slightly turned toward the wall and then continued forward. This sort of jerky behavior was gradually reduced when coasting a straight long wall (e.g., the north and east walls). If the wall was slightly bent, the robot could still follow it without reducing speed, but when it encountered a convex angle smaller than 90 degrees (which means that most of the front infrared sensors were active) the robot stopped, backed rotating to the right, and then resumed forward motion in the new direction. After one lap around the maze, the path became smoother with less trajectory adjustments and more tuned to the geometric outline of the environment (figure 8: right).

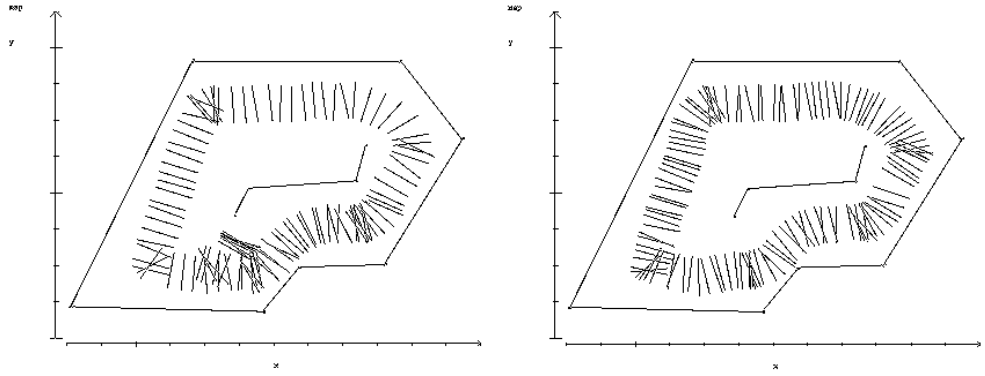


Figure 8. Trajectory of an evolved robot learning to navigate during its lifetime. Position data, visualized as bars representing the axis connecting the two wheels, were acquired with an external laser positioning device every 100 ms. **Left**: trajectory during the first lap (the robot starts in the lower portion of the environment and turns anti-clockwise). **Right**: trajectory at the second lap.

Interestingly, by looking at the internal dynamics of the evolved network, the authors observed that synapses keep changing throughout all individuals lifetime. In particular, synapses continue to change even if the behavior of the robot became rather stable after a few seconds (see Floreano and Mondada, 1996).

In the conventional view, synapses are relatively slow-changing and stable components of the nervous system. Synaptic change is identified with learning of new skills or acquisition of new knowledge, while neural activation is identified with behavior (or short term memory). Typically, it is assumed that acquisition of a stable behavior in a static environment (for example, learning to distinguish faces) corresponds to stability--no further change--of synapses in the network (e.g., see Hertz, Krogh & Palmer 1989). This assumption is explicitly included into the objective functions from which--both supervised and unsupervised--conventional learning algorithms are analytically derived: least-mean-square error minimization, energy reduction, maximization of node mutual information, etc. Since synaptic stability was not included in the fitness function employed in this experiment, which

was defined solely in behavioral terms, the evolved neurocontrollers were free to explore and exploit different ways of using synaptic change.

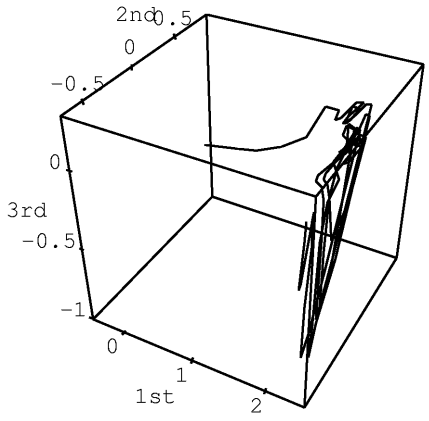


Figure 9. State-space representation of synaptic dynamics during the first 100 actions plotted as trajectory within the space of the first three principal components. Arrows indicate starting position and range of oscillation between action sequences 20-80 and 80-100. Oscillations within the subspace of the third (smallest) component correspond to fine trajectory adjustments. Method: Sanger's network (Sanger, 1989) for extracting the first three principal components of the input correlation matrix was trained to stability on the 27-component vectors corresponding to the synaptic activity recorded during the first 100 actions of the robot visualized in Figure 8. After training, input vectors were presented again to the network and output unit activations were plotted in the resulting 3-dimensional space.

One way of looking at a dynamical system such as this evolved neural network is to analyze its state-space behavior, that is the development of the synaptic vector in its 27-dimensional space over 100 updates (which correspond to two laps around the looping maze). Since it is impossible to draw a 27-dimensional space, a more convenient representation could be that of displaying how the 27-dimensional vector varies over 100 time steps. Principal Component Analysis does just that by extracting the directions of maximum variance of the distribution of data (our 100 synaptic vectors). Once we have extracted the first, say, three directions of maximum variance, we have a 3-dimensional space on which we can sequentially plot each of the 100 vectors. Figure 9 shows the trajectory of synaptic change in the reduced state-space of the first three principal components of the recorded synaptic vectors during the first 100 actions of the individual displayed in Figure 8.

During the first six actions the neural network moves toward a subregion of the space for which there is no change in the first two principal components; residual variation along the slice of space corresponding to the third principal component corresponds to fine trajectory adjustments and is further reduced as the robot gradually tunes its path to the geometry of the environment. This means that, after an initial phase of strong variation, the synapses as a whole change in a systematic and co-ordinated fashion. In other words, the stable behavior acquired during life is regulated by continuously changing synapses which are dynamically stable. Roughly speaking, this means that when one synapse goes up, there will be another synapse going down.

The synapses evolved in this experiment are responsible for both learning and behavior regulation. Knowledge in the network is not expressed by a final stable state of the synaptic configuration, but rather by a *dynamic equilibrium point* in an n-dimensional state-space (where n is the number of synapses). Learning can be seen as a displacement of the entire

system from a dynamically unstable state to a new dynamically stable state. Whether biological synapses can play a similar role or not, is an issue that remains to be investigated.

Learning of the evolved controller relies on simple genetically inherited abilities. For example, the controller analyzed above always starts by moving backward until it finds some object; then it rotates to the right until its rightmost sensors become active and synapses begin to change. These two simple motor programs result from weak sensory signals (mostly noise) filtered and amplified through the synaptic matrix of excitatory and inhibitory weights. They represent the basis from which learning can start to operate and are similar to instincts in that they are both genetically inherited and represent primitives upon which the organism can start to explore its environment. In other words, evolution not only shapes the learning modality, but also bootstrap learning by selecting basic behaviors useful for learning.

The analysis of the evolved behavior described above clearly indicates that the environment plays a great role in shaping the ontogenetically developed behavior. Behavior is an emergent property of the interaction between inherited instructions and the environment not only because evolution exploits the complexity of the environment and of the interaction between the robot and the environment, but also because inherited instructions only indirectly constraint how the robot react in the environment. The way in which the robot reacts to different sensory states itself is affected by the previous interactions of the robot with the environment. In this way the amount of information encoded in the genotype may be reduced given that part of the information will be filled up by the interaction between inherited instructions and the environment.

This model is also interesting from the point of view of the issues described in the previous two sections. In the Hinton and Nowlan's model described in section 3, learning is modeled as a random search process without any directionality. Instead, in the model described in the previous section, learning has a directionality but the learning task is fixed and pre-determined by the experimenter. In the model described in this section, learning has a directionality and the learning task itself (i.e. the learning constraints) is evolved. Interestingly some constraints on what can be learned are determined by the interaction between the robot and the environment. For example, only the weights departing from the sensors which are activated in a given environment can be affected by learning (for example, some weights from the left sensors will never learn because the robot follows walls on its right side). Instead, other constraints (e.g., the learning rates and the learning rules) are genetically inherited and therefore are subjected to the evolutionary process. This implies that in this model the exploration of the phenotypic space around the point corresponding to the individual at birth has a directionality which is determined by both environmental and genetic constraints.

One last point that should be stressed is the fact that in this model evolution and learning operate on two different synaptic entities (the meta-properties and the weight strengths, respectively) while in the experiments described in the previous two sections they operate on the same entity (i.e. the weight strengths).

6. Adaptation to fast changing environments

One of the adaptive functions of learning is the possibility to adapt to changes in the environment that are too fast for evolution to be able to track them. In the previous experiment this aspect was not taken into consideration because the environment did not change across generations.

Consider the case of a Khepera robot that should find a target in an arena in which walls change color from black to white (Nolfi and Parisi, 1997). The color of the walls significantly affects the response of the infrared sensors of the robot. Since the target is invisible, the robot

should explore the environment as much as possible. In order to do so it should adapt during lifetime to the different color of the walls.

The environment used for the experiments is a 60x20 cm arena surrounded by walls (Figure 10). The target area is a circle of 2 cm of diameter and is positioned at randomly chosen locations. Although the robots cannot directly perceive the target area, the fitness function selects individuals that can reach the target area in the shortest amount of time. This selection criterion indirectly encourages robots to explore the arena efficiently in order to increase their chance to end up on the target area.

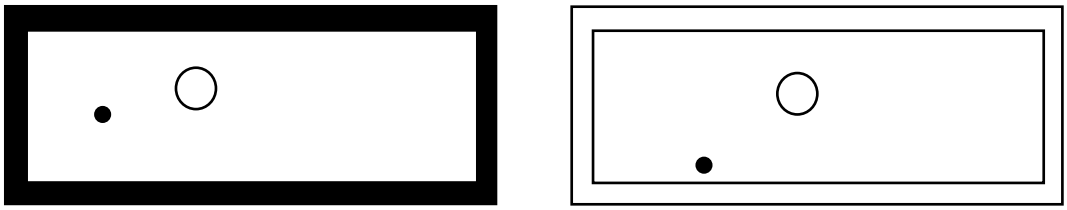


Figure 10. Two environments with a target area (small black circle) and the Khepera robot (large white circle). The target area is painted on the floor and therefore it cannot be detected by the robot's sensors. The two environments differ in the color of the walls. The wall of the environment on the right reflect six times more light than those of the environment on the left. The environments measure 60x20 cm, the target area has a diameter of 2 cm.

Robots can live in two different types of environments: (a) an environment with dark walls, and (b) an environment with bright walls, i.e. walls that reflect six times more light than dark walls. In the dark environment infrared sensors are activated within a distance of about 1 cm from the wall whereas in the light environment this distance is 6 cm. The robot should behave differently in the two environments in order to explore as much as possible the arena. If it lives in environment (a) the robot should move very carefully when sensors are activated because dark walls are detected only when they are very close. In contrast, if the robot lives in environment (b) the walls can be detected from farther away; therefore, if the robot wants to explore the portion of the arena which is close to the walls, it should begin to avoid them only when the sensors are strongly activated. Consider however that individual robots do not know in which type of environment they are going to live. Hence they should be capable of detecting the type of environment in which they are currently placed and should adapt to it through lifetime accordingly.

Robots are controlled by a feedforward neural network consisting of just an input and an output layer (Figure 11). The input layer includes four units that encode the activation level of robot's sensors. The first input unit encodes the average activation level of sensors 1 and 2, the second unit the average activation of sensors 3 and 4, etc. Hence, the network has four receptors: front, back, left, and right. These four input units are connected to four output units. The first two output units encode the speeds of the two wheels of the Khepera robot. The remaining two output units represent two 'teaching units' that encode a teaching input for the first two output units. (A more detailed description of this type of architecture is given in Nolfi and Parisi, 1993, 1994). This self-generated teaching input is used to change the weights from the input units to the two motor units with the Delta Rule (Widrow and Hoff, 1960). In other words, the neural architecture includes two distinct sub-networks that share the same input units but have separate output units. The first sub-network ("standard network"; thick connections in Figure 11) determines the robot's motor actions. The second sub-network ("teaching network"; thin connections in Figure 11) determines how the information coming from the environment is used to change the connection weights of the standard network. All connection weights are genetically encoded and evolved, but the connection weights of the teaching network (teaching weights) do not change during the

robot's lifetime while the connection weights of the standard network (standard weights) do change. This special architecture allows evolution to determine, by selecting the weights of the teaching network, the way in which environmental information modifies the nervous system of our creatures and, as consequence, their behavior during life (for a similar architecture in which the teaching network produce a self-generated reinforcement signal see Ackley and Littman, 1991).

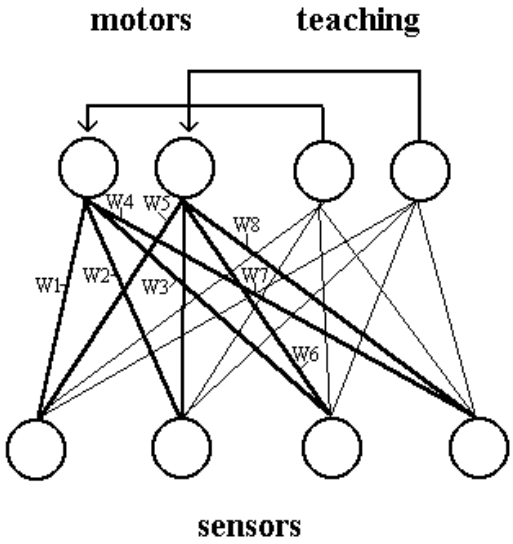


Figure 11. Self-teaching network. The output of the two teaching units is used as teaching input for the two motor units. The Delta Rule is used to change the weights from the input units to the motor units. The weights from the input units to the teaching units do not change during the lifetime of an individual.

The way in which our robots may adapt to different environments during their life becomes clear if one considers that the output of the teaching network, which functions as teaching input for the standard network, depends on two factors: the connection weights of the teaching network and the activation value of the four sensory units. While the connection weights of the teaching network are genetically inherited and are not directly influenced by the current environment, the sensory input does reflect the external environment. As a consequence, the teaching input generated by the teaching network may be influenced by the external environment and it can teach different things in different environments. Evolution has the possibility to select robots that are able to adapt to changing environments by selecting teaching weights that produce teaching inputs that are different in different environments and that produce changes that are appropriate to the current environment (for other details see Nolfi & Parisi, 1997).

The obtained results show that: (a) learning has an adaptive function. Individuals which were subjected to lifetime learning, in fact, outperformed non-learning individuals (i.e. individuals obtained by running another set of simulations in which learning was inhibited); (b) characters acquired through learning are adapted to the particular environment in which the learning takes place. It was observed, in fact, that individuals collect more fitness in the environment in which they were trained than in the other environment.

How is such adaptation to the current environment actually accomplished? How can robots 'recognize' the type of environment they happen to be born in and how can they modify themselves to adapt to that environment?

If we examine the type of stimuli that the two identical copies of the best individual of each generation experience in the dark and in the bright environment, we see that these

stimuli differ both quantitatively and qualitatively depending on the environment where the individual lives. We measured the activation level of the sensors during the entire lifetime of the best individuals of each generation and we discovered that the average activation level was 0.11 for the copy living in the dark environment and 0.24 for the copy living in the bright environment. In addition, we found that the percentage of times each of the four input units (corresponding to the left, right, front, and back pairs of sensors) is the most active one significantly varies at birth, i.e., prior to learning, between the two environments (Figure 12). This measure is obtained by allowing an individual to live for one epoch prior to learning in the two environments while measuring the percentage of times each of the four input units is the most active one.

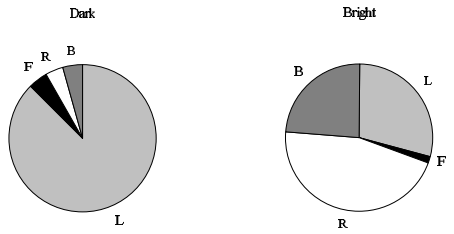


Figure 12. Percentage of time each of the four input units is the most active one during one epoch without learning (i.e., using the weight values inherited at birth) in both the dark and bright environment. The measures are carried out on an individual evolved to learn during lifetime. (F=front sensor (black); B=back sensor (dark-gray); L=left sensor (light-gray); R=right sensor (white)).

The different types of stimuli the robots experience in the two environments affect the type of teaching input computed by the teaching network and allow the robots to modify their standard weights (i.e., the weights that determine their motor behavior) differently in the two environments.

At this point we may ask ourselves what is the role of the inherited standard weights in the case of individuals that are allowed to learn during their life. For example, one might think that the standard weights incorporate the same general solution adopted by non-learning individuals and that learning is used to refine the inherited strategy by taking into consideration the specificity of the current environment. If we compare the performance exhibited prior to learning by evolved individuals belonging to the learning population with the performance of individuals belonging to the non-learning population, we discover that this is not the case. Individuals belonging to the learning population perform on the basis of their inherited standard weights less well than individuals of the non-learning population (see Nolfi & Parisi, 1997)[3]. This result contrasts with the comparison between the two populations when performance is assessed after learning. In these circumstances, as we said above, the individuals of the learning population outperform those of the non-learning population.

This result implies that inherited standard weights of learning individuals are selected not only for their ability to solve the task (as shown by their performance at birth prior to learning), but also to allow learning to produce a good performance. In other words, the genes (i.e., the inherited standard weights plus the inherited teaching weights) of evolved individuals that are allowed to learn do not incorporate a predisposition to behave efficiently but a *predisposition to learn to behave* efficiently.

To understand what a predisposition to learn can mean in the case of our robots we should consider two facts: a) initial conditions (e.g., initial weights) can determine the course

[3] This result is also obtained with evolved self-teaching networks living in a stationary environment (see Nolfi and Parisi, 1993, 1994).

of learning by error minimization (Kolen & Pollack, 1990), and b) evolution can select appropriate initial weights for learning (Belew, McInerney & Schraudolph, 1991). This implies that the standard weights are selected in order to enhance the adaptive advantage of changes produced by learning. If we allow our individuals to learn starting from random initial weights instead than from their inherited standard weights, their performance will remain constantly low throughout their life (see Nolfi & Parisi, 1997). Although the learning error will progressively decrease, the weight change does not improve the efficiency of their exploration of the environment even if the inherited teaching weights are left intact.

A predisposition to learn to explore the environment more efficiently, therefore, is at least in part incorporated in the inherited standard weights. However, the inherited teaching weights also incorporate a predisposition to learn (or, more precisely, to produce adaptive changes in the standard weights). If we allow our individuals to modify the genetically inherited standard weights but we randomize the teaching weights, in this case too learning will destroy whatever ability to explore is present at birth rather than increasing that ability (see Nolfi & Parisi, 1997). Moreover, if we let our robots to move using the output of the teaching units instead of the output of the standard units, once again we obtain a significant decrease in performance (see Nolfi & Parisi, 1997) with respect the case in which self-generated teaching are use to modify the standard weights that determine how the robots behave.

Therefore both the standard weights and the teaching weights incorporate a genetically inherited predisposition to learn rather than a predisposition to behave. The behavior of evolved robots emerge from the interaction between the two set of weights and cannot be traced back in part to one set and in part to the other set. More precisely, behavior is the emergent result of the interaction between standard weights, teaching weights, and the environment.

Interestingly, the predisposition to learn do not only consists in an ability to use the sensory patterns coming from the sensors to adapt to the environment but also in an ability to modify the patterns received from the environment in order to select patterns which produce adaptive changes. In other words evolved individuals have *a predisposition to select useful learning experiences*. It was found, in fact, that evolved robots which learn during lifetime behave at birth in a way that enhance the perceived difference between the two environment with respect to non-learning robots (see Figure 13).

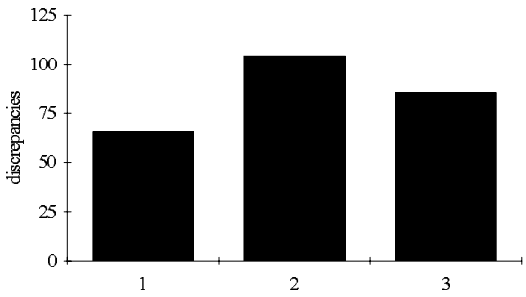


Figure 13. Difference in the percentage of time each of the four input units is the most activated one in the two environments for (1) nonlearning individuals at birth, (2) learning individuals at birth, and (3) learning individuals at the end of their life. Individuals evolved for learning behave so that input units are activated more differently in the two environments. This means that their behavior enhances perceptual differences.

To determine how the two environments differ in the inputs that they make available to the learning and nonlearning robots, we computed the percentage of cycles in which each of the four input units was the most active and we compared these percentages in the two

environments for both learning and non-learning individuals (details of how the measure has been conducted can be found in Nolfi and Parisi, 1997). The differences in the activation level among the four input units in the two environments reflect the different behaviors of an organism in the two environments. The first column of Figure 13 shows the average difference between the stimuli perceived at birth in the two environments by non-learning individuals. The second column shows the same average difference for the learning individuals at birth, i.e., before any learning. The third column shows the average difference for the learning individuals at the end of life, that is, after learning has had its effect. These data indicate that (a) learning individuals perceive at birth the two environments as more different than non-learning individuals, and (b) that there may be a tendency for the differences to decrease after learning (although the difference between the second and third column is not statistically significant). In other words, learning individuals behave at birth in a way that enhances perceived differences between the two environments which in turn allow them to learn to produce two different behaviors in the two environments.

7. Co-evolution of plastic neurocontrollers for competing robots

In the previous section we examined a case in which the environment included only abiotic elements. However, the external environment may also include other organisms. Thus changes in the subjective environment of one individual might be caused also by changes occurring in other organisms. One case which is interesting from the point of view of the adaptation to changing environment is the case in which the environment of one individual includes other organisms which compete with that individual. In this case, in fact, the environment will tend to change so to make the traits of evolving individuals no longer useful for reproductive success. It might thus happen that progress achieved by one population is reduced or eliminated by the other competing population. This phenomenon is sometimes referred to as the "Red Queen Effect" (van Valen, 1973) (from the imaginary chess figure, invented by novelist Lewis Carroll, who was always running without making any advancement because the landscape was moving with her).

We studied the case of two competing populations of predator and prey robots (Floreano & Nolfi, 1997; Floreano & Nolfi, 1997b; Nolfi & Floreano, in press). Two Khepera robots were used in these experiments, one of which (the Predator) was equipped with a vision module while the other (the Prey) had a maximum available speed set to twice that of the predator. The prey has a black protuberance, which can be detected by the predator everywhere in the environment (see Figure 14). The two species could evolve in a square arena 47 x 47 cm in size with high white walls so that predator could always see the prey (within the visual angle) as a black spot on a white background.

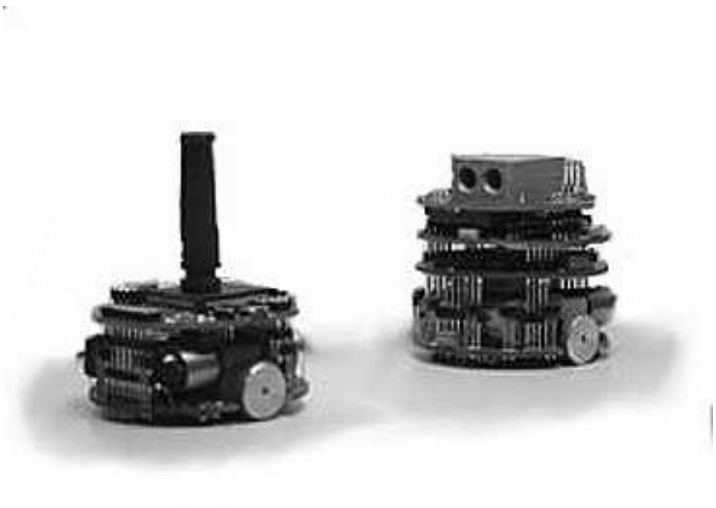


Figure 14. **Right**: The Predator is equipped with the vision module (1D-array of photoreceptors, visual angle of 36°). **Left**: The Prey has a black protuberance which can be clearly detected by the predator at a distance up to 70 cm, but its maximum speed is twice that of the predator. Both Predator and Prey are equipped with 8 infrared proximity sensors.

Both individuals were provided with eight infrared proximity sensors (six on the front side and two on the back) which had a maximum detection range of 3-4 cm in the environment. For the predator we considered the K213 module of Khepera which is an additional turret that can be plugged in directly on top of the basic platform. It consists of a 1D-array of 64 photoreceptors which provide a linear image composed of 64 pixels of 256 gray-levels each, subtending a view-angle of 36°. However the K213 module also allows detection of the position in the image corresponding to the pixel with minimal intensity. We used this facility by dividing the visual field into five sectors of about 7° each corresponding to five simulated photoreceptors. If the pixel with minimal intensity lay inside the first sector, then the first simulated photoreceptor would become active; if the pixel lay inside the second sector, then the second photoreceptor would become active, etc. From the motor point of view, we set the maximum wheel speed in each direction to 80mm/s for the predator and 160mm/s for the prey.

The robot controller was a simple perceptron comprising two sigmoid units with recurrent connection at the output layer. In the case of the predator, each output unit received connections from five photoreceptors and from eight infrared proximity sensors. In the case of the prey, each output unit received input only from 8 infrared proximity sensors, but its activation value was multiplied by 2 before setting the wheel speed.

Two populations of 100 individuals each were co-evolved for 100 generations. Each individual was tested against the best competitors of the ten previous generations (a similar procedure was used in (Sims, 1995; Cliff & Miller, 1996). At generation 0, competitors were randomly chosen within the same generation, whereas in the other 9 initial generations they were randomly chosen from the pool of available best individuals of previous generations. The competition ended either when the predator touched the prey or after 500 motor updates (corresponding to 50 seconds at maximum on the physical robot). The number of lifecycle (ranging between 0 and 499) was used as fitness value for both prey and predator. High values corresponded to high fitness for the prey and to low fitness for the predator[4].

[4] In Nolfi and Floreano (in press) we used a slightly different fitness function. In this case the fitness in each competition was simply 1 for the predator and 0 for the prey if the predator was able to catch the prey and, conversely 0 for the predator and 1 for the prey if the latter was able to escape the predator.

Two set of experiments were conducted. In one set the weights of the neural controllers were directly encoded into individuals' genotype (i.e. individuals were not able to adapt to their environment during lifetime). In the other set the genotype encoded some "meta-properties" of the synapses (as in the experiments described in section 5 the genotype encoded the sign of the synapse, the learning rule, and the learning rate). The weights of the synapses were assigned randomly and were allowed to change during lifetime.

Figure 15 shows the average population fitness in the case of the experiments conducted with non-plastic individuals.

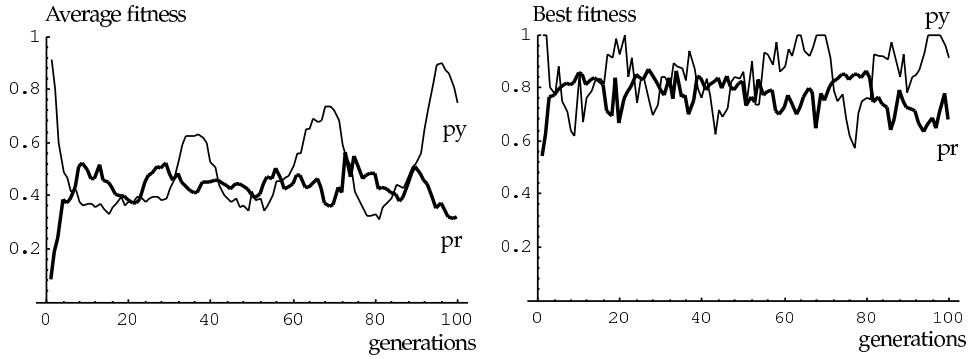


Figure 15. Co-evolutionary fitness measured in simulation in a typical experiments with non-plastic individuals. Average population fitness of the two populations (pr=predator; py=prey) At each generation, the fitness values of the two species do not sum to one because each individual is tested against the best opponent recorded from the previous 10 generations.

As expected, initially the prey score very high, whatever they might do, because the predators are not good at catching them; for the same reason, initially the predators score very low. Very quickly a set of counter-phase oscillations emerge in the two populations, as also reported by other authors (Sims, 1994, p.36), but we never observed dominance of one population on the other in any of our evolutionary runs (even when continued for 500 generations). However, the fitness for the prey always tended to generate higher peaks due to position advantage (even in the case of the worst prey and best predator, the latter will always need some time to reach the prey).

A remarkable aspect of these co-evolutionary experiments is the variety and complexity of behavioral strategies displayed by the two species. Figure 16 shows some typical tournaments recorded at different generations. In this experiment after few generations the prey moves quickly around the environment and the predator attacks only when the prey is at a certain distance (Figure 16, left). The strategy of the predator can be explained by considering if the prey move at an high speed, as in this case, the simple strategy to try to reach the prey as fast as possible will not pay off given that the prey is faster than the predator. Later on (Figure 16, center) the prey spins in place and, when the predator gets closer, it rapidly avoids it. Prey that move too fast around the environment sometimes cannot avoid an approaching predator because they detect it too late (remember that the other robot is more difficult to detect by infrared sensors than a large white wall). Therefore, it pays off for the prey to wait for the slower predator and accurately avoid it. However, the predator is smart enough to perform a small circle after having missed the target and re-attack until, by chance, the prey is caught on one of the two sides (where wheels and motors do not leave space for sensors). The predator of the following generations (Figure 16, right) temporary loose their ability to avoid walls (which was not required in the few previous generations because the predator very quickly localized and approached the prey). At the same time the

prey resumes a rapid wall following and obstacle avoidance which forces the predator to get closer to walls and collide if the prey is missed.

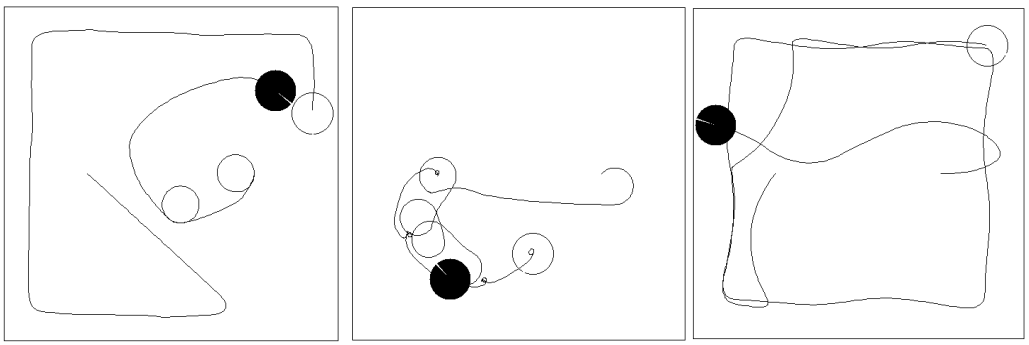


Figure 16. Behaviors recorded at interesting points of co-evolution representing typical strategies. Black disk is the predator, white disk is the prey.

By analyzing the behavior obtained in simulations throughout a longer evolutionary time, however, it can be shown that the same type of strategies are re-discovered over and over again (Nolfi and Floreano, in press). This does not imply that the co-evolutionary process is unable to find interesting solutions, as we just saw. It only means that effective strategies may be lost instead of being retained and refined. Such good strategies, in fact, are often replaced by other strategies that, although providing an advantage over the current opponents, may be ineffective against the previous strategies. When, as in this case, newly generated strategies are specialized to defeat the current competitors but not the old ones co-evolution ends in cycle in which the same strategies are adopted over and over again.

At this point we may ask if general strategies (i.e. strategies which can defeat several competitors adopting different strategies) can be found. One possible way to force co-evolution to produce general strategies is to save and use as competitors all the best individuals of previous generations (see Rosin and Belew, 1997). In this way individuals able to defeat a larger number of competitors, on the average, can be selected (see Nolfi & Floreano, in press). However, individuals selected against all the competitors of the previous generation are able to defeat most but not all of them. There is always some strategy, often adopted by competitors of several generation before, that they are unable to defeat and that can defeat them (see Nolfi & Floreano, in press).

These results point to the conclusion that in certain tasks it is always possible to find a simple strategy capable of defeating another single, albeit complex and general, strategy (although such simple strategy is a specialized strategy, i.e. it is able to defeat only that particular strategy and, of course, other similar strategies). If this is really true, in other words, if completely general solutions do not exist in some cases, we should re-consider the 'cycling problem'. From the point of view of non-plastic individuals, the fact that co-evolutionary dynamics lead to a limit cycle in which the same type of solutions are adopted over and over again should not be considered as a failure but as an optimal solution. We cannot complain that co-evolution does not find a more general strategy capable of coping with all the strategies adopted by the co-evolving population during a cycle if such general strategies do not exist, given the existing conditions (environment, sensory motor system, architecture of the neurocontroller, etc.). The best that can be done is to select the appropriate strategy for the current counter-strategy, which is what actually happens when co-evolutionary dynamics end in a limit cycle.

On the other hand, plastic individuals (i.e. individuals that can change their strategy during lifetime) have available another option. If these individuals posses a set of different

strategies which are effective against a variety of counter-strategies and are able to select the strategy which is most appropriate to the current competitor, they may be able to defeat a larger number of competitors than non-plastic individuals. In other words, plastic individuals can adapt to the changes occurring in the other populations during lifetime while non-plastic individuals can only adapt to the same changes after a certain number of generations. The ability to adapt during lifetime to the opponent's strategy would produce a significant increment in the adaptation power of a single individual because ontogenetic adaptations are much faster than phylogenetic one. Once individuals that are able to adapt to different counter-strategies during lifetime are selected, we may also expect that the co-evolutionary process will less likely fall into limit cycles. In this case, in fact, sudden changes of strategy will be less likely to produce advantage even in the short range.

The experiments conducted with plastic individuals seem to confirm these hypotheses in the case of predators (see Floreano & Nolfi, 1997b). Predators, in fact, reported higher average than prey in all six runs, except for short temporary oscillations (Figure 17). Furthermore, in all runs, the average fitness of the predator population was more stable than that of the prey. Relative performance of the two species (i.e., how many times one species wins over the other) in this condition significantly differed from the experiments described in Figure 15 in which individuals were not allowed to change during lifetime. Although also in this case behavioral strategies specifically tuned to the behavior of the competitor can be found, this pattern was less marked than in the experiments with non-plastic controllers.

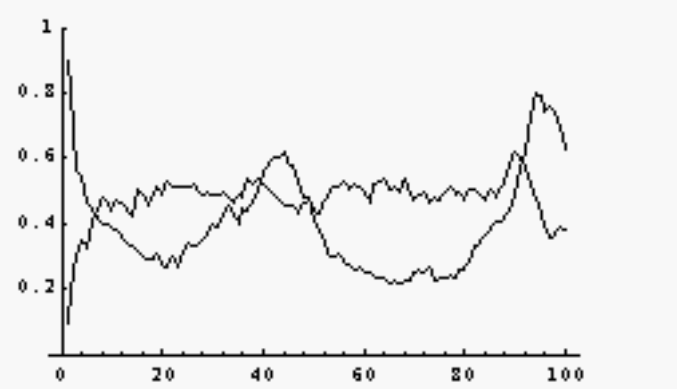


Figure 17. Average fitness across generations. Thick line = predator; thin line = prey.

More information can be gained by observing behavioral patterns of the two competitors during individual tournaments (Figure 18). There is not much variation in the behavior of the predator. It always displays a very good tracking ability across generations: once the prey has been locked in its visual field, it quickly accelerates to maximum speed until contact. As compared to the experiments with non-plastic individuals described above where the predator tended to efficiently track in only one direction, here it can turn in both directions at equal speed. For non-learning controllers proper tracking in both directions would have required accurate settings of all synaptic strengths from visual inputs (a rare solution that might be difficult to find on the genotype space). Here, instead, since synapses are temporarily increased depending on active visual units (Floreano & Mondada, 1996; Flotzinger, 1996), individual adjustments of synapses take place when and where required depending on current sensory input. The trajectory in the center image of Figure 18 shows an evident example of synaptic adjustment. Here, while the prey rotates always around the same circle, the predator performs three turns on itself during which synaptic values from the visual units are gradually increased; at the fourth turn, the synaptic values will be sufficiently high to cause a straight pursuit (eventually, the prey will try to avoid the predator without success). Finally, the

temporary drop in performance of the predator after generation 90 is due a more precise tracking combined with a slower motion (right image of Figure 18). Such behavior was probably developed because the prey were also slower and more careful in avoiding obstacles (including the predator).

Although activity-dependent synaptic change is exploited by the far-sighted predator, not the same happens for the prey. Prey are faster with respect to the experiment with non-plastic individuals, especially when turning near walls (where IR sensors become active and synapses temporarily strengthen), but plasticity does not allow an increment of their behavioral repertoire. Not even can they improve it because volatile changes of the synaptic values imply that most of the time they must re-develop on-the-fly appropriate strengths. Although this can be well-suited for avoidance of static obstacles, it does not represent an advantage when facing another fast-moving object such as the predator.

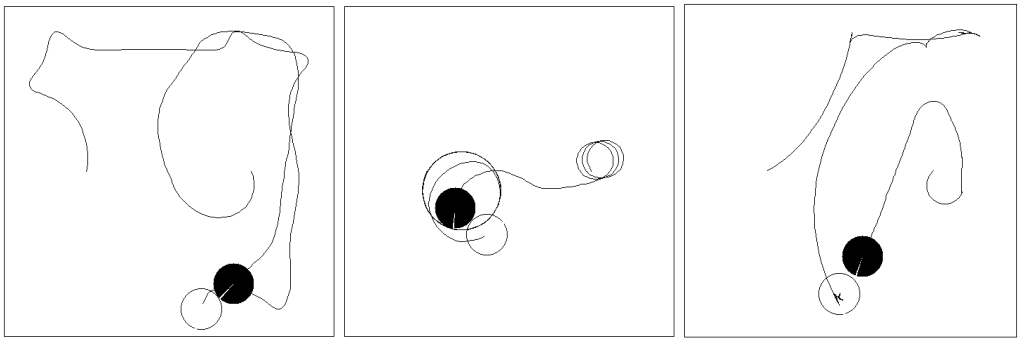


Figure 18. Behaviors of co-evolved individuals with ontogenetic learning. Black disk is predator, white is prey. **Left**: Generation 20; **Center**: Generation 70; **Right**: Generation 95.

These results indicate that plastic predators are capable of adapting their strategies to the strategy adopted by the current competitor during lifetime. Almost all predators are able to adapt to the two different classes of strategies adopted by the prey: (a) stay still or hidden close to a wall waiting for the predator and eventually trying to escape when the IR sensors detect the predator; (b) move fast in the environment, avoiding both the predator and the walls) by selecting the appropriate counter-strategy during lifetime. This explain why in these experiments predators are not compelled to abandon their strategy when prey suddenly change their counter-strategy. This is reflected by the fact that their fitness is more stable than that of the prey across generations. Instead, since the prey cannot develop more effective strategies because of their limited sensory ability, they display cyclic behaviors (as revealed by both fitness values and behavioral analysis)[5].

8. Discussion

In this paper we have investigated how learning can enhance the adaptive power of evolution. In section 3 we saw that learning can help and guide the evolutionary search even if characters acquired through lifetime learning are not inherited. In particular we saw that learning and evolution might solve tasks that evolution alone is unable to solve. Moreover, in section 4, we showed that learning can produce an increase in performance both ontogenetically and phylogenetically even if the learning task differs from the task for which individuals are selected.

[5] Prey can develop more interesting behavior however if they sensory system is enriched (see Nolfi and Floreano, in press).

We also showed that learning individuals outperform non-learning individuals in non-stationary environments by adapting during lifetime to their current environment (section 6 and 7). Evolved individuals are capable of detecting the type of environment in which they are placed and of modifying their behavior during lifetime accordingly in order to maximize their fitness.

In section 5 we showed that some characters (i.e. the connection weights in this case) can be extracted from the regularities present in the environment instead of being specified into the genotype. In the model described in that section, the weight values emerge from the interaction between genetically specified instructions and the environment.

We saw that information extracted from the environment can channel evolutionary search into promising directions. In section 5 and 6, for example, we saw that only weights departing from sensors which are stimulated in a given environment (i.e. only weights which have an effect on the corresponding behavior) change during lifetime. In other words, the information coming from the environments allows learning to exploit the most interesting dimensions of the search space.

Finally, we saw that evolution may channel learning into promising direction. In the case of the experiments described in section 5 and 6, the directionality of learning is not fixed but is determined by the inherited constraints (the combination of learning parameters and the teaching weights respectively) which are themselves under evolution. Therefore the directionality of learning is selected by evolution (i.e. evolution selects the learning task). Since lifetime learning affects the fitness of the individuals and consequently affects also the choice of individuals selected for reproduction, evolution will tend to select individuals that display good learning directions. In other words evolution will tend to select inherited constraints that produce ontogenetic changes which are adaptive on the average.

We want to conclude this paper by discussing in the next sections three general issues that can be raised on the basis of the results of the experiments described above.

8.1 The role of the interaction between learning and evolution

By exploring the adaptive functions of learning we discovered that the interaction between learning and evolution deeply alters both the evolutionary and the learning process themselves. Evolution in interaction with learning displays dynamics very different from those which are observed in evolution alone. While in non-plastic individuals the inherited characters are directly selected for their ability to produce successful behaviors, in the case of individuals that learn, the characters are selected for their ability to incorporate a *predisposition to learn*. This genetically inherited predisposition to learn may consist of different things:

1) the presence of starting conditions at birth (e.g., initial weights for learning) that canalize learning in the right direction. Evolution may select initial weight matrices or network architectures that cause a better and/or a faster learning. This has been shown to happen in the case where the learning task and the evolutionary task are the same (for weight matrices, see Belew et al., 1991; for network architectures, see Miller et al., 1989) and in the case where they are different (see Nolfi & Parisi, 1994). In the latter case, evolution does not only select individuals that have a predisposition to better learn, but also individuals that, by learning a certain task, improve their performance with respect to the evolutionary task (see section 4).

2) *an inherited tendency to behave in such a way that the individual is exposed to the appropriate learning experiences.* Evolution tends to select characters that produce initial behaviors that enhance the possibility to learn and/or that increase the probability to acquire adaptive characters through learning. In other words evolution tends to select individuals

which have an initial behavior suitable for learning and not necessarily for solving the evolutionary task (see initial back-and-rotate behavior of evolved individuals described in section 5, the behavior of individuals described in section 6 that enhance the perceptual differences between the two different environment, and the behavior of evolved predators which turn in circle and then attack directly prey that move in circle).

Similarly, learning within an evolutionary perspective has quite different characteristics from learning studied in isolation, as in "traditional" connectionist research (Rumelhart and McClelland, 1986). While in individuals that learn but are not subjected to an evolutionary process (e.g., neural networks trained with supervised methods) learning is usually accomplished by ignoring the characters of the individual prior to learning (which are typically generated at random), in evolving plastic individuals learning exploits such starting conditions. In other words, when the learning process itself is subjected to an evolutionary process, learning does not necessarily tend to incorporate the right solution to the problem; rather it tends to pull the learning individual in a direction that maximizes the chances of acquiring adaptive characters by taking into consideration its initial state. This explains the surprising result described in section 6 that self-generated teaching inputs do not correspond to the desired motor actions although they are capable of producing changes that generate suitable motor actions (on this point see also Nolfi & parisi, 1993).

8.2 Extracting supervision from the environment through learning

From the point of view of a natural or artificial organism the external environment does not provide any direct cue on how the agent should act to attain a given goal. However agents receive a large amount of information from the environment through the sensors. Such information (which is a function of both of the environmental structure and of the motor actions of the agent) may be used not only to determine how to react in different environmental circumstances but also to adapt to the current environment through lifetime learning. For example, a robot may learn the consequences of different actions in different environmental contexts or it may learn to classify sensory states not only on the basis of the current perceived sensory pattern but also on the basis of the preceding and following sensory patterns.

In principle, in an evolving population, any ability which can be acquired through lifetime learning can also be genetically acquired through evolution. However these two ways of adapting to the environment differ in one important respect: ontogenetic adaptation can rely on a very rich, although less explicit, amount of supervision. From the point of view of phylogenetic adaptation, individuals are evaluated only once on the basis of a single value which codifies how well they were adapted to their environment throughout all their lifetime (i.e., the number of offspring in the case of natural evolution and the fitness value in the case of Artificial Evolution). Instead, from the point of view of ontogenetic adaptation, individuals receive information from the environment through their sensors throughout their whole lifetime. However, this huge amount of information encodes only very indirectly how well an individual did in different moments of its own lifetime or how it should modify behavior in order to increase its own fitness. The problem is how such information can be transformed into an indication of what the agent should do or how well it is doing.

As we shown in sections 5, 6 and 7, evolution can solve this type of problem by producing subsystems capable of autonomously extracting supervision information that can be used for fast lifetime learning. In the case of the experiments described in section 6, the control system was divided into two sub-modules of which the former has the function of determining how to react to the current sensory state and the latter has the function of generating a teaching signal for the former. By subjecting the weights of the two sub-

networks to an evolutionary process, we showed how individuals which learn during their lifetime to adapt to the environment through self-generated teaching signals can be obtained. These individuals are able to transform the information which they receive from the environment into useful teaching inputs. Similarly, in the experiments described in section 5 and 7, evolution select the meta-properties of the synaptic weights which are able to channel the changes driven by the sensory states in the right directions.

8.3 Intelligence and generality

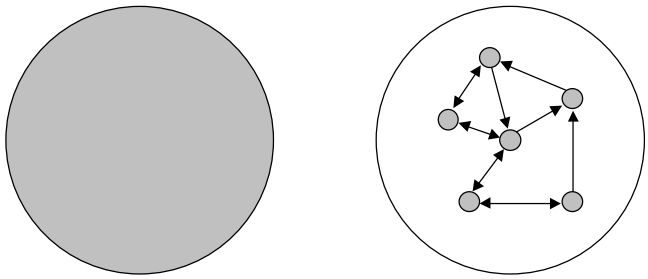
A key feature of intelligent systems is generality, i.e., the ability to carry out a certain task in different environmental conditions or the ability to carry out different tasks. In the context of predators and prey, for example, predators should be able to catch different types of prey.

By evolving simple neural controllers we can obtain interesting solutions. But these solutions often lack generality. Consider the case of non-plastic predator and prey described in section 7. These systems are interesting because can solve non-trivial tasks in simple ways. However, they are strongly dependent on the current state of the environment. If the environment changes (e.g., the strategy of the competitor changes) they may become unable to solve their task.

These two aspects (i.e., simplicity and lack of generality) are the two side of the same coin. These systems are able to solve non-trivial task with simple strategies because they exploit all the regularities available in the environment. One might conclude that, in order to be general, systems should be more autonomous from the environment. In other words, intelligent systems should rely less on the regularities available in the environment and more on their internal "nervous mechanisms". Such systems will probably require more internal complexity than the simple non-general systems which we described. The attempt of (Good Old Fashioned) Artificial Intelligence to build general purpose systems (e.g., universal planners) mostly ignoring the characteristics of the external and of the internal environment is a natural consequence of this line of thought.

Fortunately, this is not the only available option. Generality may also be achieved by systems that, instead of incorporating a single general strategy, posses a collection of simple strategies that are appropriate in different environmental circumstances and a mechanism which is able to select the strategy which is appropriate to the current environment.

To clarify this point, let us distinguish between 'full-general' and 'plastic-general' individuals (see Figure 19). Full-general individuals have a single strategy which is effective in different environmental. Plastic-general individuals, on the contrary, possess a set of different strategies which are effective in different environmental circumstances. Although these two types of individuals seem equivalent, there are some subtle differences. Full-general individual do not need to adapt during lifetime to changes occurring in the environments because their strategy ca face any environmental circumstance. Plastic-general individuals on the other hand should be able to select the appropriate strategy for their current competitor. In other words, they should be able to adapt through ontogenetic adaptation. From this point of view full-general individuals will be more effective because they can provide immediately the correct answer to the current environmental state. On the other hand, as we said above, it may be that in certain conditions a full-general individual cannot be selected because a full-general strategy does not exist or because it is too difficult to find for the evolutionary process. In this case, the only option left is that of plastic-general solutions.



full-general plastic-general

Figure 19. **Left**: Full-general individuals. The large gray circle represent a single general strategy which require a quite complex control system. **Right**: Plastic-general individuals. The small gray circles represents a collection of simple non-general strategies. The arrows indicate a mechanism which is able to select the strategy which is appropriate to the current environment state.

What is important to notice is that full-general and plastic-general individuals significantly differ in their internal organization. In general, full-general systems, by not relying on the regularities present in the internal and external environment, will require quite complex control systems. Plastic-general individuals on the other hand, by relying on the regularities present in the internal and external environment, will require a collection of simple strategies which may be produced by a much simpler control system. These strategies will be of the same type of those adopted by specialized individuals which adopt solutions which are non-general but extremely simple.

References

Ackley, D.E. & Littman, M.L. (1991). Interaction between learning and evolution. In C.G. Langton et. al (Eds.), *Proceedings of the Second Conference on Artificial Life*. Reading, MA: Addison-Wesley.

Baldwin, J.M. (1896). A new factor in evolution. *American Naturalist*, 30, 441-451.

Belew, R.K., McInerney, J. & Schraudolph, N.N. (1991). Evolving networks: using the genetic algorithm with connectionistic learning. In C.G. Langton et. al (Eds.) *Proceedings of the Second Conference on Artificial Life*. Reading, MA: Addison-Wesley.

Cecconi, F., Menczer F. & Belew R. K. (1996). Maturation and the evolution of imitative learning in artificial organisms. *Adaptive Behavior*, 4, 29-50.

Cliff, D., & Miller, G. F. (1996). Co-evolution of pursuit and evasion II: Simulation methods and results. In P. Maes, M. Mataric, J-A Meyer, J. Pollack, H. Roitblat & S. Wilson (Eds.), *From Animals to Animats IV: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, Cambridge, MA: MIT Press-Bradford Books.

Floreano, D. & Mondada, F. (1996) Evolution of plastic neurocontrollers for situated agents. In P. Maes, M. Mataric, J-A. Meyer, J. Pollack & S. Wilson. (Eds.), *From Animals to Animats IV*. Cambridge, MA: MIT Press.

Floreano, D., & Nolfi, S. (1997a). God save the red queen! Competition in co-evolutionary robotics. In J. R. Koza, D. Kalyanmoy, M. Dorigo, D. B. Fogel, M. Garzon, H. Iba & R. L. Riolo (Eds.), *Genetic Programming 1997: Proceedings of the Second Annual Conference*, pp. 398-406, San Francisco, CA: Morgan Kaufmann.

Floreano D. and Nolfi S. (1997b). Adaptive behavior in competing co-evolving species. In P. Husband, and I. Harvey (Eds). *Proceedings of the Fourth European Conference on Artificial Life*, MIT Press, Cambridge, MA, 378-387.

Flotzinger D. (1996). Evolving plastic neural controllers for autonomous robots. *Msc dissertation* 95800131, COGS, University of Sussex at Brighton.

Harvey, I. (1996). Relearning and evolution in neural networks, *Adaptive Behavior*, 4 (1): 81-84.

Harvey, I. (1997). Is there another new factor in evolution?, *Evolutionary Computation*, 4 (3): 313-329.

Harvey, I. & Stone, J. V. (1996). Unicycling helps your French: spontaneous recovery of associations by learning unrelated tasks. *Neural Computation*, 8: 697-704.

Hertz, J., Krogh, A., and Palmer, R.G. (1991). *Introduction to the theory of neural computation*. Redwood City, CA: Addison-Wesley.

Hinton, G.E. & Nowlan, S.J. (1987). How learning guides evolution. *Complex System*, 1, 495-502.

Hinton G. & Plaut D. (1987). Using fast weights to deblur old memories. In: *Proceedings of the 7th Annual Conferences of the Cognitive Science Society*, Seattle, U.S.A

Kolen, J.F. & Pollack, J.B. (1990). Back-propagation is sensitive to the initial conditions. *Complex Systems*, 4, 269-280.

de Lamarck, J.B.P.A. (1990). *Zoological Philosophy*. London: MacMillan, 1914. Relevant excerpts reprinted in Belew and Mitchell, 1996.

Mayley G. (1997). Landscapes, learning costs, and genetic assimilation, *Evolutionary Computation*, 4 (3): 213-234.

Miller, G.F., Todd, P.M. & Hedge, S.U., 1989, Designing neural networks using genetic algorithms, In L. Nadel and D. Stein (Eds.), *Proceedings Third International Conference on Genetic Algorithms*, San Mateo, CA: Morgan Kaufmann, pp. 379-384.

Mondada, R., Franzi, E. & Ienne, P. (1993). Mobile robot miniaturization: A tool for investigation in control algorithms, In T. Y. Yoshikawa & F. Miyazaki (Eds.) *Proceedings of the Third International Symposium on Experimental Robots*, Berlin: Springer-Verlag.

Nolfi, S., Elman, J.L. & Parisi, D. (1994). Learning and Evolution in Neural Networks. *Technical Report* CRL-9019, University of California, San Diego.

Nolfi, S., Elman, J.L. & Parisi, D. (1994). Learning and Evolution in Neural Networks. *Adaptive Behavior*, (3) 1, 5-28

Nolfi S., and Floreano D. (in press). Co-evolving predator and prey robots: Do 'arm races' arise in artificial evolution?, *Artificial Life*

Nolfi, S., Miglino, O., Parisi, (1994). Phenotypic Plasticity in Evolving Neural Networks. In D. P. Gaussier and J-D. Nicoud (Eds.) *Proceedings of the International Conference From Perception to Action*, Los Alamitos, CA: IEEE Press.

Nolfi, S. & Parisi, D. (1993). Auto-teaching: networks that develop their own teaching input. In. J.L. Deneubourg, H. Bersini, S. Goss, G. Nicolis, R. Dagonnier (Eds). *Proceedings of the Second European Conference on Artificial Life*, Brussels. pp.845-862

Nolfi, S. & Parisi, D. (1994). Desired responses do not correspond to good teaching input in ecological neural networks. *Neural Processing Letters*, (1) 2:1-4

Nolfi, S. & Parisi, D. (1997). Learning to adapt to changing environments in evolving neural networks. *Adaptive Behavior*, 1:99-105

Nolfi, S. & Tani, J. (in press). Extracting Regularities in Space and Time Through a Cascade of Prediction Networks: The Case of a Mobile Robot Navigating in a Structured Environment. *Connection Science*

Nolfi, S. (1998). How Learning and Evolution Interact: The Case of a Learning Task which Differs from the Evolutionary Task. *Technical Report*, Rome: Institute of Psychology, C.N.R.

Osborn, H. F.. (1896). Ontogenetic and phylogenetic variation. *Science*, 4:786-789.

Parisi, D., Nolfi, S. & Cecconi, F. (1992). Learning, Behavior and Evolution. In: Varela, F, Bourgine, P. (Eds.), *Toward a Practice of Autonomous Systems*. Cambridge, MA: MIT Press, pp.207-216

Parisi, D. & Nolfi, S. (1996). How learning can influence evolution within a non-Lamarckian framework. In Belew, R.K, Mitchell, M. (eds.) *Adaptive Individuals in Evolving Populations*. SFI Studies in the Science of Complexity, Vol. XXVI, Reading, MA: Addison-Wesley

Phillips, W. A. and Kay, J. and Smyth, D. (1995). The discovery of structure by multi-stream networks of local processors with contextual guidance. *Network*, 6: 225-246.

Rosin, C. D. & Belew, R. D. (1997). New methods for competitive coevolution, *Evolutionary Computation*, 5 (1), 1-29.

Rumelhart, D.E. & McClelland, J.L. (1986). *Parallel Distributed Processing*. Cambridge, MA: MIT Press.

Sanger, T.D. (1989). Optimal Unsupervised Learning in a Single-Layer Feedforward neural network. *Neural Network*, 2: 459-473.

Sasaky T. & Tokoro M. (1997). Adaptation toward changing environments: Why darwinian in nature?, In P. Husband, and I. Harvey (Eds*). Proceedings of the Fourth European Conference on Artificial Life*, MIT Press, Cambridge, MA, 378-387.

Sims, K. (1995). Evolving 3D morphology and behavior by competition. *Artificial Life*, 1, 353-372.

Todd, P.M. & Miller, G.F. (1991). Exploring adaptive agency II: simulating the evolution of associative learning. In J. A. Meyer and S.W. Wilson (Eds). *From Animals to Animats. Proceedings of the First International Conference on Simulation of Adaptive Behavior*. Cambridge, MA: MIT Press.

van Valen, L. (1973). A new evolutionary law. *Evolution Theory*, 1, 1-30.

Waddington, C.H. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, 150: 563-565.

Widrow B. & Hoff M.E. (1960). Adaptive switching circuits, In: *IRE WESCON Convention Record*, vol IV, 96-104.

Willshaw, D. and Dayan, P. (1990). Optimal plasticity from matrix memories: What goes up must come down. *Neural Computation*, 2: 85-93.

Wilson, S.W. (1987). Classifier systems and the animat problem. *Machine Learning*, 2:199-228.

Withley D., Gordon S. & Mathias, K. (1994). Lamarckian evolution, the Baldwin effect and function optimization. In Y. Davidor, H.P. Scwefel, & R. Manner (Eds.), *Parallel Problem Solving from Nature (PPSNIII)*, Berlin: Springer-Verlag.

Yamauchi, B. and Beer, R.D. (1995). Sequential behavior and learning in evolved dynamical neural networks. *Adaptive Behavior*, 2(3): 219-246.