Complex Systems 4 (1990) 401–414

The Effect of Learning on the Evolution of Asexual Populations

J.F. Fontanari
Instituto de Física e Química de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560 São Carlos SP Brazil

R. Meir
Division of Chemistry, California Institute of Technology, Pasadena, CA 91125 USA

Abstract. We study a model for the effect of learning on a population of asexually reproducing individuals. While current models are usually based on simulations of the actual hypothetical organisms, we adopt a classic population genetic approach, which deals directly with gene frequencies. This allows us to investigate the model in great detail, obtaining results which are difficult to obtain using other techniques. In particular, we demonstrate that learning has a drastic effect on evolution, as has already been noted by Hinton and Nowlan [3] using computer simulations. A detailed quantitative description of the temporal behavior of the model is also presented. In particular, an interesting interplay between mutation and learning is demonstrated.

1. Introduction

While most biologists today agree that acquired characteristics cannot be passed to future generations, it is much less clear how learning during an individual's lifetime can influence evolution. Since in an unpredictable environment it is more advantageous to leave decisions to learning than specifying them genetically, it is rather obvious, from a Darwinian viewpoint, that differences in the learning capabilities of the individuals will produce a selective pressure for the genes responsible for the learning advantage, thereby increasing their frequency. However, anything but obvious is the proposal put forward by Baldwin [1] that in a fixed environment properties that were behavioral goals in earlier generations could be genetically determined in the course of the evolution. Although the idea is simple enough, it has not always been clear how such a scheme would work.

In a recent paper, Hinton and Nowlan [2] have proposed a model that explains how such a facilitation of evolution through learning could come

about. The model considers a population of N haploid (see below) individuals, each containing a neural network with L connections. It is assumed that each connection is determined by a single gene, which comes in three basic types (alleles): 1, 0, and ?. A 1 allele specifies that a connection is present, a 0 allele that it is absent, and a ? allele specifies a plastic connection that can be determined through learning. To make the model as simple as possible (but not simpler), Hinton and Nowlan assume that there is a single combination of alleles, say all alleles 1, corresponding to a specific neural network structure, which confers a reproductive advantage. Clearly in such a situation, termed a "needle-in-a-haystack" problem, the evolutionary search for such a combination would be extremely difficult since there is no slope leading to the maximum ("needle") so the usual evolutionary hill-climbing process would be no better than a random search. What Hinton and Nowlan have shown is that learning can create an increased fitness zone around the "needle" by allowing individuals whose connections are near perfect to learn the correct setting.

To check Baldwin's proposal, Hinton and Nowlan simulated a population of 1000 individuals with 20 genes each. Learning consists of giving each individual 1000 random trials to guess the correct setting, the fitness being larger for individuals that learn more rapidly. Reproduction in their model was sexual, in the sense that in each generation 1000 matings were performed between parents chosen according to their fitness. Note however that the "organisms" are haploid, i.e., contain a single version of each allele for every locus as opposed to most real organisms, which are diploid, i.e., they contain two alleles for each locus. Crossover between two "chromosomes" takes place by picking the crossing-over position at random. They demonstrated that even for low initial values of the correct allele 1, the population evolved a very high proportion of correct alleles, due to the increased fitness conferred on organisms through learning. An exhaustive analysis of Hinton and Nowlan's model can be found in reference [3].

While sexual reproduction seems to be the rule rather than the exception in the natural world, it is not at all clear what the advantages are of sexual versus asexual reproduction. In fact, this is a hotly debated issue to this very day (see for example reference [4]). Two points are usually cited in favor of sexual reproduction [4]: (1) Sexual organisms evolve more quickly than asexual ones, due to the possibility of two (or more) favorable genes coming together, and (2) sexual organisms are less likely to accumulate harmful mutations as these can be reconstituted through recombination. In the asexual case, a reverse mutation would have to take place. These points should be weighed against the advantage of asexual reproduction, namely that favorable genes are transmitted directly to the offspring without the potential harm of mixing them with unfavorable genes of the mate. While it is not our wish to discuss these fascinating issues, and we are certainly not qualified to do so, we note that in a fixed environment, as is the one studied by Hinton and Nowlan, sexual reproduction does not have any advantage over asexual reproduction. Moreover, as pointed out by Maynard Smith [5], in the

absence of learning a sexual population would never evolve the correct setting in the above setup, while an asexual one would (although it would take a very long time). From our point of view, however, asexual reproduction has the added advantage of facilitating the mathematical analysis and enabling us to explore analytically the full dynamical behavior of the model.

The organization of the paper is as follows. In section 2 we define our model and derive the equations governing the evolution of the allele frequencies. We take a classic population genetics approach [6] in deriving these equations, as opposed to the "microscopic" one taken by Hinton and Nowlan. To better appreciate the effect of learning, we study in section 3 the evolution of a population without learning on the "needle-in-a-haystack" fitness landscape discussed above. We then go on to derive the full consequences of incorporating learning into the model in section 4. We find very similar results to those obtained by Hinton and Nowlan in their simulations using sexual reproduction. We are able, however, to derive much more, as will be demonstrated. In particular, an interesting interplay between learning and mutation is demonstrated. Section 5 then summarizes our results and points out how the model may be generalized.

2. The model

We consider an infinite population of haploid individuals, each one represented by its genotype, which we model by a string of L genes. The genes may occur in three different forms (alleles): 1, ?, and 0. At generation n, these alleles have frequencies p_n, q_n, and r_n, respectively. These frequencies are defined in terms of the total population's genetic distribution so that, at each generation, one has $p_n + q_n + r_n = 1$. We assume that a genotype is completely characterized by three parameters P, Q, and R standing for the number of alleles 1, ?, and 0, respectively, that it possesses. Thus, the genotype frequency in the population at generation n is

$$\Pi_n(P,Q,R) = C^L_{P,Q,R} p_n^P q_n^Q r_n^R \tag{2.1}$$

where $C^L_{P,Q,R} \equiv L!/P!\,Q!\,R!$ and $P + Q + R = L$.

Natural selection is taken into account by assigning to each genotype a fitness value or rate of reproduction $\mu(P,Q,R)$. Without loss of generality, we assume that the correct neural network is achieved by the genotype $P = L$ and $Q = R = 0$. Mutation is introduced into the model by allowing the alleles to change spontaneously with some probability. More specifically, we define the mutation rate ν as the fraction of alleles that underwent mutation in one generation and treat ν as a control parameter. We focus on the evolution of the allele frequencies, since the genotype frequencies can be obtained from them through equation (2.1).

In the following we derive the recursion relations for the allele frequencies. In the absence of mutations, the fraction of allele 1 that each genotype contributes to the next generation is proportional to the product of three factors: the number of alleles 1 it contains, its frequency in the population,

and its rate of reproduction. Hence the frequency of allele 1 in generation $n+1$ is

$$p_{n+1} = \frac{1}{\bar{w}_n} \sum_{P,Q,R} P\Pi_n(P,Q,R)\,\mu(P,Q,R) \tag{2.2}$$

where $\bar{w}_n$ is the average rate of reproduction of the entire population in generation n,

$$\bar{w}_n = L \sum_{P,Q,R} \Pi_n(P,Q,R)\,\mu(P,Q,R) \tag{2.3}$$

The significant parameter measuring the reproductive advantage of a genotype over its competitors is its relative rate of reproduction $\mu(P,Q,R)/\bar{w}_n$.

To introduce the effect of mutation, we notice that the probability of an allele 1 escaping mutation to ? or 0 is $1-2\nu$ and the probability of an allele ? or 0 changing to 1 is ν. Thus, in order to take mutation into account, equation (2.2) must be rewritten as

$$p_{n+1} = \frac{1}{\bar{w}_n} \sum_{P,Q,R} \Pi_n(P,Q,R)\,\mu(P,Q,R)[(1-2\nu)P + \nu(Q+R)] \tag{2.4}$$

which after some simple algebra, reduces to

$$p_{n+1} = \nu + \frac{1-3\nu}{\bar{w}_n} \sum_{P,Q,R} P\,\Pi_n(P,Q,R)\mu(P,Q,R) \tag{2.5}$$

The same procedure may be used to derive the recursion relations for q_n and r_n. It yields

$$q_{n+1} = \nu + \frac{1-3\nu}{\bar{w}_n} \sum_{P,Q,R} Q\,\Pi_n(P,Q,R)\mu(P,Q,R) \tag{2.6}$$

and

$$r_{n+1} = \nu + \frac{1-3\nu}{\bar{w}_n} \sum_{P,Q,R} R\,\Pi_n(P,Q,R)\mu(P,Q,R) \tag{2.7}$$

To examine the behavior of this system of coupled recursion relations one must specify the rate of reproduction assigned to the genotypes. This can be done rather straightforwardly in the case of a population without learning, and the next section is devoted to the study of the evolution of such a population. However, in the case of a population with learning the rate of reproduction *is not* specified genetically, i.e., it is not uniquely determined by P, Q, and R and equations (2.5), (2.6), and (2.7) are not readily applicable.

3. Evolution without learning

Within the "needle-in-a-haystack" scenario we ascribe to the correct genotype the rate of reproduction L and to the remaining genotypes the rate of reproduction 1. Moreover, since in this case there is no difference between alleles ? and 0, we consider only two types of alleles, 0 and 1. We choose as the relevant variable the frequency of correct alleles p_n, the fraction of incorrect alleles being obtained from it through the normalization constraint $r_n = 1 - p_n$. Following the reasoning of the last section and using $\mu(L, 0) = L$ and $\mu(P, R \neq 0) = 1$, it is straightforward to write down the recursion relation for p_n,

$$p_{n+1} = \nu + \frac{1 - 2\nu}{\bar{w}_n} \left[p_n + (L - 1) p_n^L \right] \tag{3.1}$$

where $\bar{w}_n = 1 + (L - 1) p_n^L$. In the following analysis we take $L = 20$ as in Hinton and Nowlan's original model.

Let us consider first the simpler case where mutations are not allowed, $\nu = 0$. There are two equilibria, namely $p^* = 0$ and $p^* = 1$. The former is unstable and the latter is globally stable, i.e., $p \to 1$ irrespective of its starting value $p_0 \neq 0$. Figure 1 shows the frequency of allele 1 when introduced at three different initial frequencies, $p_0 = 0.7$, 0.6, and 0.57. Since the frequency

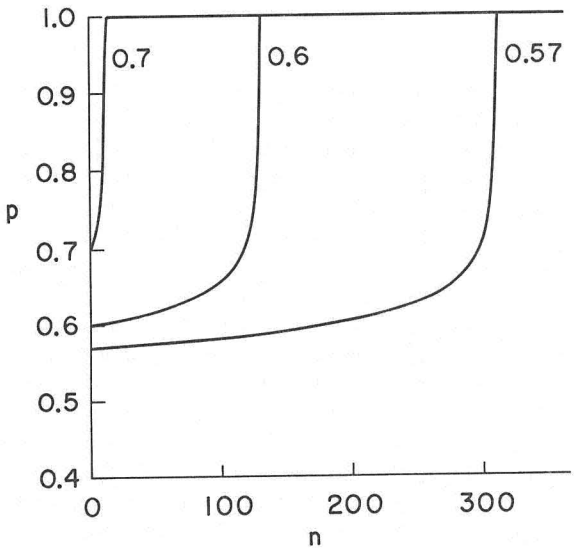

Figure 1: The frequency of allele 1 in a population without learning as a function of the generation number for $p_0 = 0.7$, 0.6, 0.57 with $L = 20$ and $\nu = 0$.

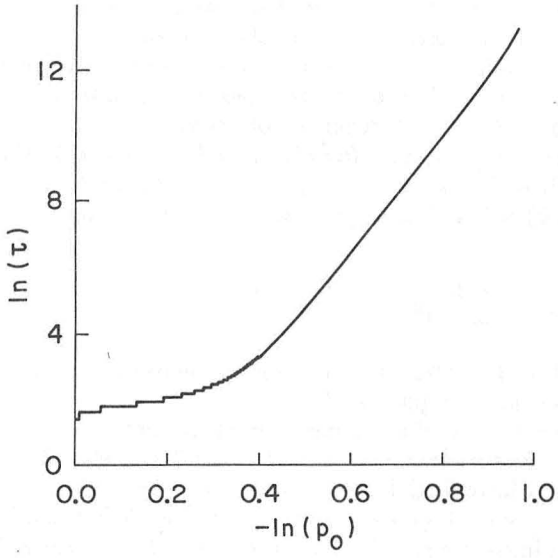

Figure 2: The dependence of the number of generations that a population without learning needs to reach fixation on the initial frequency of correct allele for $L = 20$ and $\nu = 0$.

of the correct genotype is p^L, small values of p do not affect the average rate of reproduction significantly, resulting in a slow increase of p in the earlier generations as observed in the lower curve of figure 1. However, a the correct genotype becomes common ($p > 0.6$) there is an explosive growth leading to fixation in a few generations. Thus, an asexual population can do quite well in the "needle-in-a-haystack" fitness landscape if it is given enough time to evolve. How well it does will depend on the number of generations it needs to reach equilibrium (τ), which, as can be seen in figure 1, depends on p_0. In figure 2 we show this dependence with τ defined as the generation for which $1 - p < \epsilon = 10^{-6}$. The specific value of ϵ is not important, any smaller than 0.2 will give very similar results. This and also the staircase shape of the curve in figure 2 are consequences of the steep gradient of the curves in figure 1 for p close to 1. The straight part of the curve in figure 2 indicates that τ grows as a power of p_0; the best fit being $\tau = 0.02 p_0^{-17.3}$. For instance, for $p_0 = 0.25$ approximately 5.2×10^8 generations will be needed until the population is composed only of individuals carrying the correct setting of alleles.

A noteworthy effect of incorporating mutation into the model is the elimination of the fixed point $p^* = 0$. In fact, it can be seen easily from equa-

tion (3.1) that p^* cannot be smaller than $1/2$. For small ν the equilibrium frequency of correct allele 1 decreased to

$$p^* = 1 - \frac{L}{L-1}\nu + \mathcal{O}(\nu^2) \tag{3.2}$$

although the convergence is slightly speeded up. Numerical analysis indicates that equation (3.2) gives the only fixed point of equation (3.1) for $\nu < \nu_1 \simeq 4.35 \times 10^{-4}$. A second stable fixed point, $p^* \simeq 1/2$, with a large basin of attraction appears for $\nu > \nu_1$. As we increase ν further on, the high p fixed point disappears discontinuously at $\nu = \nu_2 \simeq 0.08$, the driving of natural selection being completely eliminated by the random perturbations due to mutation. In figure 3 we show the basins of attraction of the two fixed points in a phase diagram in the plane (ν, p_0). The region below the curve belongs to the basin of attraction of $p^* \simeq 1/2$ and the region above it to the high p fixed point, equation (3.2). The curve starts at $\nu = \nu_1$ and ends at $\nu = \nu_2$; beyond this point only the low p fixed point exists. We notice that due to our probabilistic interpretation of the mutation rate, the maximal value it can have is $\nu = 1/2$, though evolution is disrupted for a much smaller value, $\nu \simeq 0.08$.

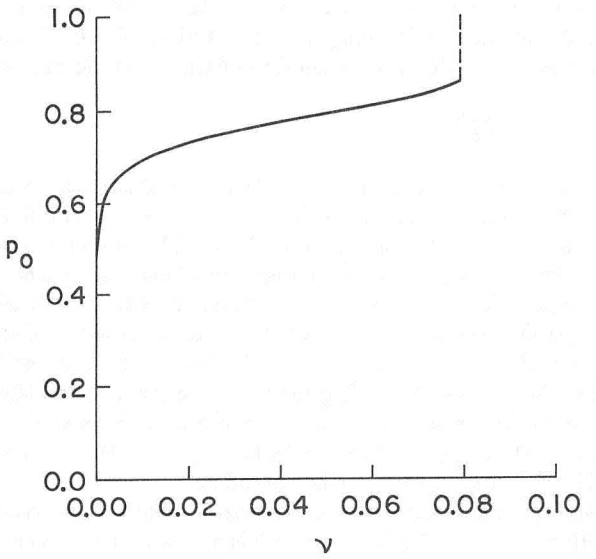

Figure 3: Phase diagram showing the basins of attraction of the high p fixed point (above the curve) and of the low p fixed point (below the curve) as a function of the mutation rate for a population without learning with $L = 20$. The solid curve begins at $\nu \simeq 4.35 \times 10^{-4}$ (before it only the high p fixed point exits) and ends at $\nu \simeq 0.08$ (beyond it only the low p fixed point exists).

4. Evolution with learning

Reverting to the three alleles model described in section 2, we proceed in the incorporation of learning into the model. Following Hinton and Nowlan we allow each individual a maximal number G of random guesses to find the correct set of neural connections. It is assumed that the individuals can sense when they have guessed correctly so the learning or guessing process is stopped at that moment. Since there are Q learnable connections, the probability that an individual sets them correctly in one trial is $c \equiv 2^{-Q}$ and during its lifetime (G trials) its odds are increased to $1-(1-c)^G$.

Reproductive advantage is conferred on individuals with the correct neural network, not necessarily with the correct genotype, as opposed to the case without learning discussed before. Thus, the rate of reproduction must be assigned to the phenotype rather than the genotype. The learning capability, however, is genetically determined in the sense that an individual whose genotype has at least one allele 0 will never be able to guess the correct neural network, while the rapidity with which individuals guess it depends on the number of undefined alleles in their genotypes. There are two situations where the rate of reproduction is genetically determined: when the individual carries the correct genotype, $P = L$, $R = Q = 0$, and when the individual has a least one allele 0 in its genotype, $R \neq 0$. The former is assigned rate of reproduction L and the latter, 1. It remains to define the rate of reproduction of individuals possessing only alleles ? and 1, $R = 0$. According to Hinton and Nowlan, the rate of reproduction of these individuals is

$$\rho(g) = L - (L-1)\frac{g}{G} \tag{4.1}$$

where $1 \leq g \leq G$ is the number of guesses that the individual made before finding the correct setting. Thus $\rho(g)$ depends not only on the individual's genotype but also on the stochastic valuable g. This dependence on g is the main obstacle to applying the formalism developed in section 2 to the evolution of a population with learning. However, rather than studying the evolution of a specific population depending on the particular realizations of the stochastic variable g, we focus on the evolution of an "average" population whose allele frequencies at each generation are regarded as the average of the allele frequencies of an infinite number of populations at that generation. Thus, $\rho(g)$ will be replaced by its average $\langle\rho(g)\rangle$. To accomplish that we must consider the statistical distribution of g.

The probability of an individual guessing correctly at the gth trial is equal to the probability it fails in the first $g-1$ trials and succeeds in the gth trial,

$$(1-c)^{g-1}c \tag{4.2}$$

However, we must account for the possibility that the individual never guesses correctly, the probability of this being $\delta_{g,G}(1-c)^G$. Note that if $g = G$ there are two possibilities: (1) with probability $(1-c)^{G-1}c$ the individual guesses correctly in its last trial or (2) with probability $(1-c)^G$ it fails in the G

trials; the rate of reproduction being 1 in both cases. Hence g is distributed according to

$$\Upsilon(g) = c(1-c)^{g-1} + \delta_{g,G}(1-c)^G \tag{4.3}$$

and, since g appears only linearly in equation (4.1), we must compute only the first moment

$$\langle g \rangle = \sum_{g=1}^{G} g\,\Upsilon(g) \tag{4.4}$$

which, after performing the summation, reduces to [3]

$$\langle g \rangle = \frac{1-(1-c)^G}{c} \tag{4.5}$$

Thus, averaging $\rho(g)$ makes explicit the dependence on the genotype parameters

$$\langle \rho(g) \rangle = \mu(P \neq L, Q, R = 0) = L - (L-1)\left[\frac{1-(1-2^{-Q})^G}{2^{-Q}G}\right] \tag{4.6}$$

At this stage the framework presented in section 2 can be readily applied. Inserting equation (4.6) together with $\mu(P = L, Q = 0, R = 0) = L$ and $\mu(P, Q, R \neq 0) = 1$ into equations (2.3), (2.5), and (2.7) yields

$$\frac{\bar{w}_n}{L} = 1 + (L-1)\left\{(1-r_n)^L + \frac{p_n^L}{G}\left[1 + \Sigma_n^0 - (1+\gamma_n)^L\right]\right\} \tag{4.7}$$

$$\begin{aligned} p_{n+1} &= \nu + \frac{1-3\nu}{\bar{w}_n/L}\left\{\frac{L-1}{G}p_n^L\left[1 + \Sigma_n^0 - \Sigma_n^1/L - (1+\gamma_n)^{L-1}\right]\right. \\ &\quad \left. + p_n\left[(L-1)(1-r_n)^{L-1} + 1\right]\right\} \end{aligned} \tag{4.8}$$

$$r_{n+1} = \nu + \frac{1-3\nu}{\bar{w}_n/L} r_n \tag{4.9}$$

respectively, where

$$\Sigma_n^0 = \sum_{Q=0}^{L} \frac{L!}{Q!(L-Q)!}\gamma_n^Q (1-2^{-Q})^G \tag{4.10}$$

$$\Sigma_n^1 = \sum_{Q=0}^{L} \frac{L!}{Q!(L-Q)!}\gamma_n^Q (1-2^{-Q})^G \tag{4.11}$$

and $\gamma_n = 2q_n/p_n$. The frequency of undefined alleles in generation $n+1$ is obtained through the normalization constraint,

$$q_{n+1} = 1 - p_{n+1} - r_{n+1} \tag{4.12}$$

These equations reduce to the ones of the evolution without learning by setting $q_n = 0$ provided $\nu = 0$. The nonlearning equations are not recovered for $\nu \neq 0$ because there is no equilibrium with $q_n = 0$. In the remainder of this section we concentrate on the numerical analysis of the above system of recursion relation with the parameters set as in Hinton and Nowlan's paper, $L = 20$ and $G = 1000$.

Figure 4 shows the evolution of the allele frequencies for $p_0 = r_0 = 0.25$ and $q_0 = 0.5$ with $\nu = 0$. These are the same initial frequencies chosen by Hinton and Nowlan in their simulation. There is a single stable equilibrium, $p^* = 1$, while the other two, $q^* = 1$ and $r^* = 1$, are unstable. Note that in only 200 generations the correct alleles amount to 90% of the total number of alleles present in the population, in contrast to the evolution without learning where 10^8 generations are needed to reach this proportion. A further increase of p becomes extremely slow because, as pointed out by Hinton and Nowlan, there is very little selective pressure against alleles ? since a few trials will suffice to guess the correct setting of neural connections. In figure 5 we use the same parameters as in figure 4 except for the mutation rate, which takes on the value $\nu = 0.01$. This figure illustrates the speedup of the relaxation to equilibrium when low mutation rates are allowed, as we have mentioned in the last section. Of course, the price to pay is the decrease in the average rate

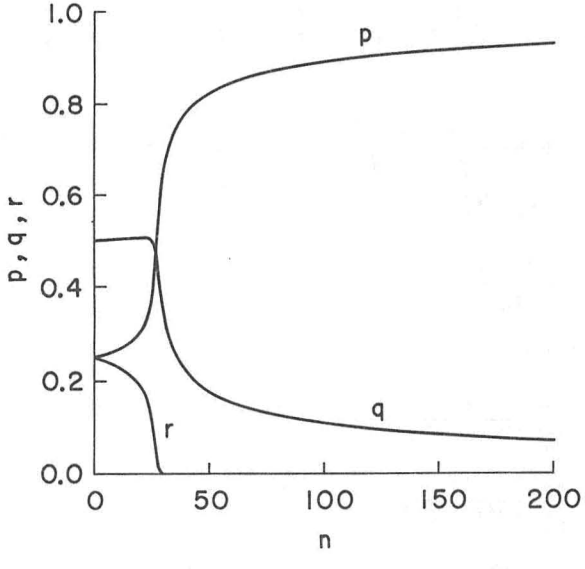

Figure 4: The frequency of alleles in a population with learning as a function of the generation number for $p_0 = r_0 = 0.25$, $q_0 = 0.5$, $L = 20$, $G = 1000$, and $\nu = 0$.

of reproduction shown in figure 6 for different mutation rates and the same initial frequencies used in figures 4 and 5. For small ν there is a single stable equilibrium. For $\nu > \nu_1 \simeq 0.019$, however, a second stable equilibrium with $p^* \simeq q^* \simeq r^* \simeq 1/3$ appears, which accounts for the lower curve (horizontal line) in figure 6. A similar behavior was observed in the evolution of the population without learning, though there the second equilibrium appeared for a much smaller value of ν. This interesting result indicates that learning strongly increases the robustness of the population to mutation. For $\nu > \nu_2 \simeq 0.074$ the high p equilibrium disappears. The interplay between the two equilibria as ν varies is depicted in the flux diagrams of figure 7. Several pieces of information can be obtained from these diagrams: the fixed-point locations, their basins of attraction, and the number of generations needed to reach fixation that is roughly measured by the density of dots. In figure 7a, $\nu = 0.01$, there is a single fixed point $p^* \simeq 0.77$, $q^* \simeq 0.22$. Notice the sticky region around $p \simeq q \simeq 1/3$ indicating that the landscape is rather flat there, and the turnpike-like topology around the fixed point with all paths converging to a single one that leads to the fixed point. In figure 7b, $\nu = 0.02$, the second stable fixed point $p^* \simeq 0.35$, $q^* \simeq 0.4$ has already shown up, draining almost all paths toward it. The high p fixed point is $p^* \simeq 0.72$, $q^* \simeq 0.25$. Figure 7c, $\nu = 0.07$, illustrates the shrinking of the basin of attraction of the high p fixed point, $p^* \simeq 0.56$, $q^* \simeq 0.34$, the turnpike-like topology having disappeared. The low p fixed point is $p^* \simeq 0.34$, $q^* \simeq 0.33$. Finally, in figure 7d, $\nu = 0.08$, there is again a single fixed point $p^* \simeq q^* \simeq 0.33$. As can be seen in figure 6, the low p fixed point has rate of reproduction $\simeq 1$ corresponding to a situation where natural election does not affect the distribution of alleles in the population.

5. Conclusion

We have introduced and investigated a model of how learning may affect a population of asexually reproducing organisms. One of the main advantages of considering asexual reproduction is the analytic tractability of the model and the ease of analyzing its behavior under a variety of conditions. Moreover, as stressed in the introduction, in a fixed environment like the one considered in this paper, sexual reproduction confers no advantage over asexual reproduction. We have shown that the recursion relations obtained, equations (4 7)–(4.9), give rise to behavior that is qualitatively very similar to that observed by Hinton and Nowlan in their simulations of sexually reproducing organisms.

The main conclusions we can draw from our study are the following. First, as has already been observed by Hinton and Nowlan, learning has a drastic effect on evolution. Our study corroborates their claims in a slightly different model using a classic population genetics approach. Second, we have been able to quantitatively analyze the model obtaining many interesting features that would be very difficult to detect solely through simulations. In particular, we have calculated the time it takes a population without learning

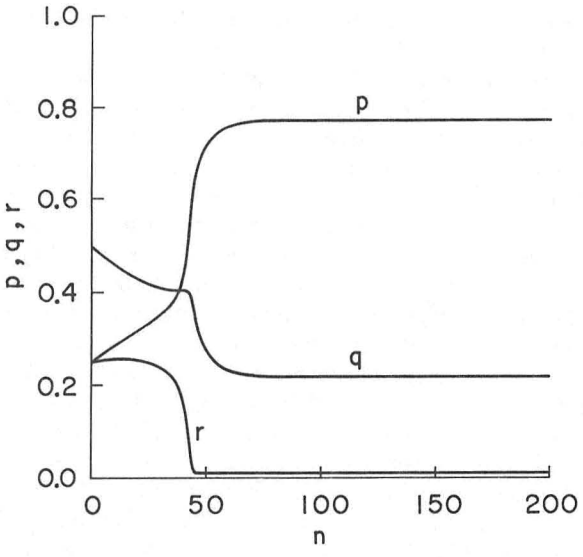

Figure 5: Same as figure 4 except that $\nu = 0.01$.

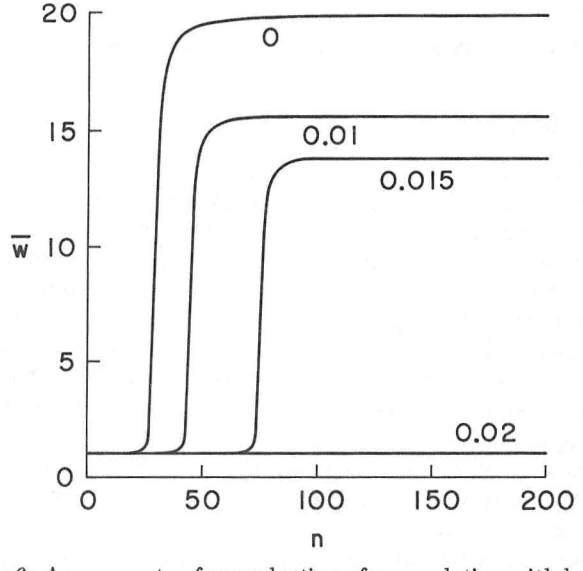

Figure 6: Average rate of reproduction of a population with learning as a function of the number of generations for different mutation rates $\nu = 0$, 0.01, 0.015, 0.02 with $L = 20$ and $G = 1000$.

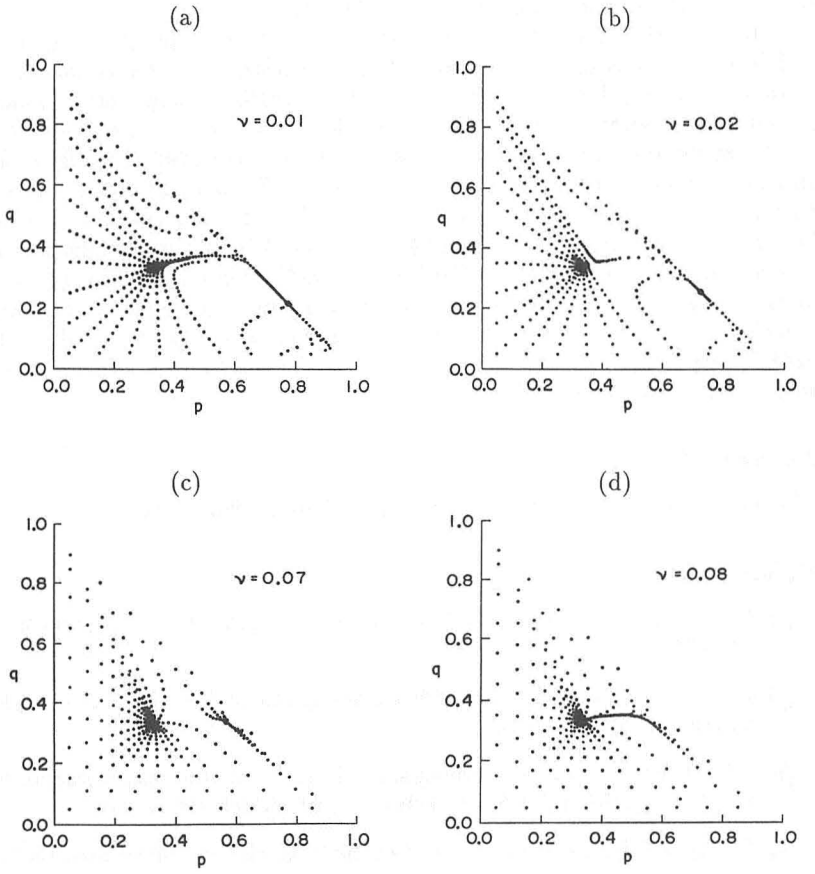

Figure 7: Flux diagrams for different mutation rates: (a) $\nu = 0.01$, (b) $\nu = 0.02$, (c) $\nu = 0.07$, (d) $\nu = 0.08$ with $L = 20$ and $G = 1000$.

to reach equilibrium (figure 2) as well as the basin of attraction as a function of the mutation rate (see figure 3 for the nonlearning case and figure 7 for the learning case). Third, the behavior of the model for various mutation rates has been fully analyzed. For small mutation rates, we find a single stable equilibrium with the correct setting of alleles, i.e., $p \simeq 1$, and high fitness value. Mutation has the beneficial effect of speeding up the evolution at the price of lowering the final fraction of correct alleles. Large mutation rates give rise to a second stable equilibrium with a very small fitness value that ends up destabilizing the high fitness equilibrium. We further find an interesting interdependence between learning and mutation. In particular, we find that learning populations are much more robust to mutation than

nonlearning ones, in the sense that they can tolerate a much higher mutation rate before natural selection becomes ineffective.

Although the approach we have taken has been very fruitful in gaining insight into various aspects of learning and evolution, it does have some limitations. One very interesting development of the Hinton and Nowlan model is that of incorporating culture into the model [3]. To incorporate culture one must assume some form of communication between organisms that allows the transmission of information between individuals. For instance, an individual with a good learning ability may wish to "teach" other organisms with similar genetic makeup, thus increasing their fitness. While it is straightforward to study the effect of culture in the Hinton and Nowlan model through simulations, it does not seem so easy to do so in the population genetic approach. Another point to note is that our approach can be easily extended to the case of sexual reproduction, though the equations obtained in this case are too complicated to be useful.

Acknowledgments

The work of R.M. was supported by the Weizmann Foundation.

References

[1] J.M. Baldwin, "A new factor in evolution," *American Naturalist*, **30** (1896) 441–451.

[2] G.E. Hinton and S.J. Nowlan, "How learning can guide evolution," *Complex Systems*, **1** (1987) 495–502.

[3] R.K. Belew, "Evolution, learning and culture: Computational metaphor for adaptive algorithms," CSE Technical Report #CS89-156 (1989).

[4] J. Maynard Smith, *Evolutionary Genetic* (Oxford University Press, 1989).

[5] J. Maynard Smith, "When learning guides evolution," *Nature*, **32** (1987) 761–762.

[6] D.L. Hartl and A.G. Clark, *Principles of Population Genetics* (Sinauer Associates, Inc., 1989).