Journal of Statistical Physics, Vol. 69, Nos. 5/6, 1992

Culture and Evolution on Populations of Neural Networks

T. J. P. Penna[1] **and P. M. C. de Oliveira**[1]

Received June 24, 1992

Although acquired characteristics are not incorporated into the genotype, some works have pointed to the influence of learning in evolution. We present a dynamic model of neural networks presenting evolutive features, even without modification in genotype, due to the introduction of culture. Our model presents other features that seem to reproduce some aspects of real world populations.

KEY WORDS: Evolution; neural networks; self-organizing criticality.

1. INTRODUCTION

It is well accepted that Lamarck's hypothesis, which suggest that adaptations acquired by learning are incorporated into the genotype, is not verified. Nevertheless, this fact does not invalidate the argument that learning can guide evolution. Clearly, for specimens living in an unpredictable environment, it is more advantageous to leave some decisions to learning rather than determine them genetically.(1) A situation where this occurs is described by Crowl and Covich.(2) They studied the freshwater snail *Physella virgata virgata*, which is able to change its life history due to the presence of the crayfish *Orconectes virilis*, one of its predators. In the absence of crayfish, the snails grow quickly up to 4 mm, when reproduction begins. In this case, the snail survives from 3 to 5 months. When crayfish are present, the snails grow up to 10 mm and survive up to 14 months. There is a compromise between reproduction and growth. An interesting and important fact is that crayfish feed selectively on the smallest snails. Two hypotheses might explain the two different life histories. First, it

---

[1] Instituto de Física, Universidade Federal Fluminense, C.P. 100296 24020, Niterói RJ, Brazil.

is result of genetic differentiation, i.e., long-term exposure to crayfish predation leads to natural selection of large snails. Second, it is only a phenotypically plastic response. A series of experiments was performed and some important conclusions determined the nature of life strategy. When snails of different sizes (some extracted from streams with crayfish and others from streams without them) are put in reservoirs without crayfish, the specimens of the first generation are small, which contradicts the first hypothesis. The snails decide for growth only in reservoirs where there are crayfish preying on them. In reservoirs where there are crayfish and snails, but without predation, the snails decide for rapid maturation. The information from predation, which means the warning to increase growth, might be a chemical substance released in this process. In this case, the evolution is clearly dictated by learning. In summary, snails can modify their life history due to learning in an unpredictable environment (a sudden addition of crayfish in streams where they were not present until this moment).

This behavior is different from the Baldwin effect,(3) which concerns fixed environment properties. This last is hypothesized because if the evolution takes place only with the production of whole organisms (mutations),(4) it would occur at very high cost, by discarding the learning of useful adaptations acquired during an organism's lifetime. In this case, the search for good genotypes would be simply like the "needle-in-a-haystack"(5) problem, i.e., like a random search. For the sake of clarity, imagine that individuals which are able to acquire a given ability during their lifetime have enhanced survival probability and consequently they will have a larger probability of reproduction than others that do not present it. The information that is determined in the genome is not the ability itself, but something like the capability of developing this ability during the lifetime. It is a less restrictive condition than the inclusion of this ability itself in the genome. Hence, the time and cost for adaptation are lowered. In summary, the purpose of learning is to enhance the efficiency of evolution by minimizing the number of possibilities to be tested for adaptation.

Recently, neural networks have been used as tools to study the effect of learning on evolution.(5,6) We will present some aspects of these references here. In these works, each individual in a population is a primitive neural network with many potential connections. For the sake of simplicity, it is supposed that each potential connection has only one corresponding gene in the genome. Each gene can assume three different forms (or *alleles*) 1, 0, and ?, specifying if its correspondent connection should be present (1), absent (0), or left to be determined through learning (?), i.e., a plastic connection that contains a switch which can be open or closed during a lifetime. The neural network structure determines the learning of a given set of informations.(7)

In ref. 5, the authors considered a population with 1000 individuals with 20 potential connections each. At the beginning, the alleles are chosen at random, with probability 0.5 for the ? allele and 0.25 for the 0 and 1 alleles. In average, 10 connections will be set during a lifetime and the other 10 will not be changed. Without loss of generality, it is supposed that there is a given combination of alleles (all of them 1 and called the *correct combination* hereafter) leading to a neural network structure that, in turn, corresponds to the learning of a specific information which gives a reproductive advantage to individuals that recognize it. During the lifetime, each individual performs up to 1000 learning trials. A learning trial means a proposal for determination of the plastic connections, i.e., set them to 0 or 1. When one of these trials leads to the correct combination, the individual no longer performs learning trials. Note that individuals with some connection absent will never learn the correct combination (all connections set to one). The probability of one individual having all its alleles set to one is $1/4^{20} \approx 10^{-12}$ in the first generation, an extremely small value. That is the probability of the correct combination to be reached by a random search, i.e., if learning does not play a role in evolution. On the other hand, the probability of an individual having no alleles set to 0 is $(3/4)^{20} \approx 10^{-3}$. Considering 1000 individuals, there is a macroscopic probability that at least one of them can have none of its connections set to 0, and consequently this one will be able to learn the correct combination. After 1000 learning trials (one generation), new individuals are generated from two others (sexual reproduction). The probability of a given individual to be selected as a parent is inversely proportional to the time it spends to learn the correct combination, such that individuals that learn rapidly have up to 20 times more chances to be a parent than individuals that never learn. The new individual's genome will be built choosing a fraction of the father's genome (an exact copy of its situation in $t=0$) and the remainder will be copied from the mother. As the information that is passed to the young must come from the genotype rather than the phenotype, the learned connections are passed as ?. After a few generations (typically 20, in this case) the relative frequency of alleles converges to a situation where only alleles 1 and ? are present in all individuals of the population. Considering the typical values in this simulation, virtually all individuals in this population will learn the correct combination during their lifetimes (the Baldwin effect). In ref. 6, the individuals are asexual ones, and mutations must be introduced to allow evolution. Fontanari and Meir(6) used analytical techniques to extract results similar to those of Hinton and Nowlan(5) in some aspects. It is important to note that, in both models, the interaction between individuals occurs only in the reproduction stage.

In this work we present a more complex form of populations and each individual is a more complex sort of neural network. In our model, the interactions between individuals do not occur only through reproduction, but throughout life. We incorporate "culture" in our model, in the sense that information can be exchanged by the individuals and they may stand in the population for several generations. Another difference presented in this work is that the convergence state of the population is determined by the dynamic evolution itself, whereas in the other quoted models this state is fixed *a priori*. In Section 2 we present the Hopfield model for neural networks, in Section 3 we define the dynamics of our model; in Section 4 we present the results and in Section 5 the conclusions.

2. THE HOPFIELD MODEL

A neural network is a complex system that presents associative memory features. In a simplified version, the state of each neuron is a binary variable $(S_k = \pm 1)^{(8)}$ and it is connected by synapses to other neurons. The synapse intensities are represented by J_{jk}, the so-called synaptic matrix elements. The state of a given neuron S_k is determined by the states of the others and by the synapse intensities, which can lead to excitation or inhibition of S_k. Hence, we can express the dynamics of a neural network as

$$S_k(t+1) = \operatorname{sign}\left(\sum_j J_{jk} S_j(t)\right) \tag{1}$$

i.e., the neuron S_k will be active if the excitatory signals received surpass the inhibitory ones. This dynamics minimizes the following Hamiltonian:

$$H = -\frac{1}{2} \sum_{j=1}^{\mathcal{N}} \sum_{k \neq j}^{\mathcal{N}} J_{jk} S_j S_k \tag{2}$$

considering a network with $\mathcal{N}$ neurons.

Hopfield$^{(9)}$ proposed that information $\{\xi_j = \pm 1, j = 1,..., \mathcal{N}\}$ is considered *learned* by the neural network if it corresponds to a minimum of the Hamiltonian (2) [or an attractor fixed point of (1)]. In this way, the recognition of a given information corresponds to a minimization process. Learning means the modification of the synaptic matrix in order to impose ξ as a minimum of (2) (in this paper, we will suppress the lower index, which refers to a given neuron, when considering the complete set of neurons). The original proposal uses Hebb's rule.$^{(7)}$ Starting from *tabula* *rasa,* i.e., $J_{jk}=0$, new information ξ^{μ} could be learned by modifying the couplings J_{jk} according to

$$\Delta J_{jk}=\frac{1}{\mathcal{N}}\,\xi_j^{\mu}\xi_k^{\mu} \tag{3}$$

However, this rule guarantees that ξ^{μ} $(\mu=1, 2,..., \mathcal{M})$ are fixed points of (1) only if two conditions are fulfilled. First, the information must be uncorrelated, i.e., the overlap between two patterns, defined as

$$m^{\mu\nu}=\frac{1}{\mathcal{N}}\sum_{j=1}^{\mathcal{N}}\xi_j^{\mu}\xi_j^{\nu} \tag{4}$$

must be statistically null for all pairs μ, ν of informations. Second, the number $\mathcal{M}$ of patterns stored in an $\mathcal{N}$-neuron network must be smaller than $\alpha_c\mathcal{N}$, where $\alpha_c\approx 0.14$.(10) If the patterns are correlated, this number is even smaller, depending on the overlaps between information. The larger the overlaps, the smaller the α_c. It is worth mentioning that, even with the above conditions satisfied, rule (3) does not guarantee that only ξ^{μ} are the equilibrium states of (1). In reality, there are other states that can be local minima of (2) and we will refer to them as "spurious" states.(9, 10) The true informations, i.e., those stored in J_{jk} by (3), will be referred as "informations" in this paper.

3. DYNAMICS OF POPULATION

In this work, we consider a population with $\mathcal{P}$ individuals, each one being a Hopfield neural network with $\mathcal{N}$ neurons. We generate at random $\mathcal{V}$ informations, i.e., we choose at random the state (± 1) of each neuron ξ_j^{ν}, $j=1,..., \mathcal{N}$ $(\nu=1,..., \mathcal{V})$. These $\mathcal{V}$ informations will be stored in all individuals using (3). Hereafter, we will refer to these informations as *vital* ones. Furthermore, each individual may have also other informations, called *individual* ones, that are not necessarily stored in other individuals. In most cases we have tested, we begin with only one random individual information for each individual—the initial number $\mathcal{I}_i(t=0)$ of individual informations in the individual I_i is equal to one, for all i. These informations lead to an initial diversification of the population. Thereby, each individual will have a different synaptic matrix. Informations will be acquired by each individual during its lifetime by receiving individual informations from others according to the dynamic rules we describe below.

In our simulations, each vital information and the initial individual ones are uncorrelated patterns, generated at random. At each time step, an individual I_1 passes one of its individual informations (ξ^{ν}, for instance), selected at random, to another individual I_2 also randomly chosen. To pass an information means to give it as input (*stimulus*) to the neural network I_2, i.e., $S_j(t=0)=\xi_j^\nu$, $j=1,\ldots,\mathcal{N}$, where S_j refers to the jth neuron of I_2. This information will be processed by I_2, following the dynamics (1), until an equilibrium (stationary) state is reached. Here "to process" means to apply the dynamic evolution (1) to each spin k, spanning the whole network, and corresponds to the recognition process. The equilibrium state can be either one of the informations introduced in (3) or a spurious state. In order to decide whether the information can be learned or not, the equilibrium state will be compared to all informations from I_2, vital and individual ones. If this final state is equal to any information from I_2, the stimulus will be taken as *known* and, in the next time step, we will choose another pair of individuals and repeat the process. Therefore a stimulus is taken as known if it belongs to some basin of attraction of the informations. If the final state is a spurious one, the stimulus will be taken as an *unknown* information by I_2. In this case, it may be learned by I_2, by (tentatively, at the beginning) modifying the couplings of I_2 ($J_{jk}\to\overline{J_{jk}}$) according to rule (3). Note that is the *stimulus* ξ^ν initially passed from I_1 to I_2, and not the stationary state reached after processing, that is included into $\overline{J_{jk}}$ through rule (3). The information will be said to be *learned*, however, only if after its inclusion in the I_2 synaptic matrix by (3) it remains as a minimum of the Hamiltonian (2) with $\overline{J_{jk}}$. In order to check whether the information is a minimum, it is given again as input to the neural network I_2 but this time with couplings $\overline{J_{jk}}$. If any neuron is flipped in the recognition process (1), we can conclude that this information is not a minimum of (2) and therefore it will not be learned. In this case the modification $J_{jk}\to\overline{J_{jk}}$ will not be performed, and we repeat the process for another pair of individuals. Furthermore, when some information is learned, we also perform the check of being minimum for all other individual informations from I_2, because the minima can be changed by the inclusion of a new information. Hereafter we will refer to an information that, due to the learning of new ones, no longer corresponds to a minimum as an *unlearned* information. Unlearned informations may still be modified or forgotten. In this case, given the unlearned information itself as initial stimulus, we take the equilibrium state for which the system converges and store it explicitly in the network using the prescription (3) (also tentatively), and simultaneously we discard the older information also using (3) with the negative sign, i.e.,

$$\Delta J_{jk} = -\frac{1}{\mathcal{N}}\xi_j^\phi\xi_k^\phi \tag{5}$$

where ξ^ϕ is the information to be discarded. We do it for all the patterns that are unlearned. Those patterns that stand as minima will be kept (this can be understood as a personal interpretation of a given information), but the other ones that are not learned, even with the modification, will be forgotten, at last, through (5).

We also performed the same checks on all $\mathcal{V}$ vital informations. The difference is that vital informations cannot be forgotten or modified. When any vital information no longer corresponds to a minimum of (2), the individual must be shut out of the population. We will refer to this process as *death.* In our simulations we performed a "canonical" version of the population, i.e., we kept the number of individuals as constant. Thus, when an individual dies another is born. Let us describe the birth step now. The synaptic matrix of the new individual (young) will be constructed as a birth based on two other individuals, termed parents (sexual reproduction). The young will have, at beginning, the $\mathcal{V}$ vital informations. Accordingly, our model does not allow *genetic* evolution, but only *learned informations* can be modified during the dynamic process, once we do not allow mutations. The choice of the parents is performed such that the greater is the number of individual informations, the greater is the probability of this individual to be chosen. The number of informations is a more (but is not *the most*) realistic time scale than the number of time steps, because, in the real world, there is no direct relation between the number of informations passed and time steps, whereas we need time to learn a given amount of informations.

The probability of a given individual I_i to be chosen as one of the parents, the father for instance, is

$$p_i = \frac{\mathcal{I}_i}{\sum_{j=1}^{\mathcal{P}} \mathcal{I}_j} \tag{6}$$

The mother will be chosen by the same process, except that she must be different from the father. All the individual informations that are present in both parents, simultaneously, will be passed to the young. If there are no common informations in the parents, a new one can be constructed from the crossover of the pair of informations (one from the father and the other from the mother) which has the largest overlap. In this new information, the state of a given neuron i will be chosen at random from the father (or from the mother) with equal probabilities. It is clear that this information will have a macroscopic overlap with the original pair. We also ran some cases where individuals are born with one individual random information, i.e., an information uncorrelated with all others. We will comment on the consequences of this modification in the next section.

In summary, our model presents the following stages: flux of informations, learning, forgetting, interpretation, death, and birth. The dynamics is sequential, i.e., at each time step, one information is passed from each individual, spanning the whole population. We consider a "mean-field" version of population, i.e., each individual is connected and can exchange information with all others.

4. SIMULATIONS AND RESULTS

We tested our model by carrying out simulations on IBM-PC 80286 and 80386 microcomputers, using the "multineuron coding."(11) In our simulations we varied the number of initial informations ($\mathcal{V}$ and $\mathcal{I}$), the number of individuals $\mathcal{P}$, and the number of neurons $\mathcal{N}$ in each individual. Also, for each run, we varied the random numbers that determine what informations and to whom they will be passed and also the parents in the birth step. We extract the number of steps to reach convergence, the final number of individual informations, the number of generations, the self-correlation (correlation between individuals at the end of the process), and its dispersion.

The first result is that we *always* found a stationary state for this dynamics. The population converges to one configuration of informations where all the individual informations available at this time are considered known by all individuals, i.e., *any* information from *any* individual that is passed to *any* other belongs to the basin of attraction of some information of this last individual. In this stage, the synaptic matrices of all individuals will remain unchanged. When we varied only the random numbers, keeping all other conditions unchanged, we obtained different stationary states starting from the same initial population. The difference between them is that some informations present in some stationary states may not be present in others. This result shows us that there are several stable configurations of stored informations, as in a spin-glass phase space. This multiple-equilibrium behavior is indeed observed in real evolution processes (see, for instance, ref. 12). Moreover, multiequilibria seem to be a common characteristic among distinct dynamically evolved complex systems (see, for instance, ref. 13). This is a remarkable difference between our model and the one proposed in refs. 5 and 6, where only one stable configuration is reached. In the stationary state, we can find informations that were present in the initial stage of the population and others that are generated during the dynamic process (by interpretation or births).

Different sequences of random numbers lead to different stationary states because the instant at which an information is passed plays an important role in the process. An information which is considered known by an individual at a given time and therefore will not be learned may no longer be recognized at a later time. That is why when an individual has few stored informations the basins of attraction are large, and therefore a stimulus even far away from these may be recognized. For example, rule (3) guarantees that any two informations are minima of (2) if only these two are stored in the synaptic matrix. As in this case there are no spurious states, any stimulus will be recognized as one of the two informations; it does not matter how far away it is from the two minima. As soon as more informations are stored, the phase space is crowded of local minima and the basins of attraction become smaller, and the probability that the stimulus is attracted to a spurious state (which number also increases with the number of learned informations) grows. As a consequence, in this work we have shown that the more individuals learn, the larger is the extension of their ignorance. This feature already had been pointed out by Socrates. As in the real world, learning ability is larger when the individual has more stored informations. As the instant at which an information is passed is determined by the random number sequence, it will lead to different configurations. We have interpreted the spurious states as indicative of ignorance about a given information.

In order to quantify our results, we defined some quantities measuring the correlation between a pair of individuals, by analyzing the individual informations of each one. We define the correlation matrix element $C_{i \to j}$ as the mean overlap (4) between all informations from I_i with the ones with largest overlaps from I_j. This matrix is *nonsymmetric*, i.e., $C_{i \to j} \neq C_{j \to i}$. Let us suppose, for instance, that a given individual I_i has the same four informations of I_j and one fifth additional one, different from the four others. In this case, $C_{j \to i} = 1$. However, due to the presence of an additional information in I_i, $C_{i \to j} < 1$. Explicitly, the mean correlation will be defined as

$$C_{ij} = \tfrac{1}{2}(C_{i \to j} + C_{j \to i}) \tag{7}$$

Thus, $C_{ij} = 1$ if and only if I_i and I_j have the same informations and no more beyond these. We also define $\mathbf{C}$ as the mean correlation of the population

$$\mathbf{C} = \frac{2}{\mathscr{P}(\mathscr{P} - 1)} \sum_{i=1}^{\mathscr{P}} \sum_{j>i}^{\mathscr{P}} C_{ij} \tag{8}$$

The case $\mathbf{C} = 1$ means that the individuals are identical, whereas $\mathbf{C} \approx 0$ means that the individuals are very different from each other. The dispersion ΔC_{ij}, presented in Table I, refers to the dispersion on C_{ij} for a given population, and it is related to the diversity of the individuals. If $\mathbf{C} = 1$, then $\Delta C_{ij} = 0$.

The occurrence (or not) of deaths and births folds the process in different regimes. Deaths do not occur when the total number of individual plus vital patterns in the population is lower than the limit $\alpha_c \mathcal{N}$. In this case, the convergence is reached in a few steps, but in the final state, the individuals are not exactly the same as one would expect by reasoning that all initial informations are passed to and learned by all individuals. Let us suppose that, for example, an individual had been generated with an information which is recognized as a vital one by the others. As the number of informations stored is low ($\alpha < \alpha_c$), there is no way by which this individual can lose this pattern, since that forgetting, interpretation, or death will not occur. In this case, this individual will be distinct from the others, presenting one more memorized information.

The most interesting results appear when deaths occur in the process. At the onset of death, the processes of forgetting and interpretation appear and new informations are passed to the population. We also ran the case where young were born with informations generated at random. In this case, we did not find stationary states, at least in a reasonable computational time that was several times larger then the one spent when we used our birth algorithm. After this large number of steps, we measured the mean correlation of the population and we found statistically insignificant values for this quantity. Let us stress that when we used the algorithm for births described before we always found stationary states. This shows the decisive influence of learning from parents in the evolution of our population. A summary of our quantitative results is shown in Table I, and will be discussed in the following paragraphs.

As can be seen from Table I, the final correlation is approximately the same ($\approx 70\%$) if deaths occur ($\mathcal{G}$ = number of deaths/$\mathcal{P} \neq 0$). We also can see that the larger the number of informations at $t = 0$, the larger the number of steps needed to reach an equilibrium state. We believe that the most realistic cases correspond to those with a large number $\mathcal{P}$ of individuals and a large number of initial individual informations $\mathcal{I}_i(0)$. This situation would lead to a large number of generations, as can be seen comparing cases 1 and 10, for example.

The convergence time, i.e., the number of steps needed to reach convergence, is also a relevant quantity. As in neural network studies,(14) this quantity has an important meaning in our model. The larger the convergence time, the larger the number of states similar to the equilibrium (equivalent to the metastable states of neural networks). Situations where few informations are learned reach the convergence with a smaller number of time steps (cases 3 and 7). On the other hand, when a large number of individual informations is available (cases 4, 9, and 10) the path to equilibrium can be retarded near some of these states, increasing the convergence times.

Another result that we extracted from simulations and that is not present in Table I concerns the introduction of reproductive advantage for the individuals with larger numbers of informations. One consequence of the introduction of this factor is the lowering of the number of steps needed to reach equilibrium, although even without it we always found stationary states for the dynamics of our model. As in the real-world situation described in ref. 2, the time scale for convergence (number of generations) is small compared to the time scale for evolutive processes in natural selection through mutations.

In order to test the stability of the equilibrium state of the population, we introduce, after convergence, a new individual with new individual informations, uncorrelated from all others. These informations may be passed to other individuals. As the original individuals of the population have a number of informations close to the limit α_c, deaths can occur frequently if the information is not recognized by some individuals. In this case, changes in different scales occur in the population: some individuals are not modified (those that recognize the new informations as older ones); others learn the new informations and stay alive (the inclusion of this new information does not lead them to discard the vital ones although it may lead to a reinterpretation of the individual informations), whereas other individuals can die, i.e., those for which the new informations are incompatible with the vital ones. This effect occurs because there are different clusters (or *classes*) of individuals that are similar, i.e., individuals that

Table I. Results of Simulations for Ten Situations[a]

#	$\mathcal{N}$	$\mathcal{P}$	$\mathcal{V}$	$\mathcal{I}(0)$	t	$\mathcal{I}(\infty)$	$\mathcal{G}$	C (%)	ΔC_{ij}
1	96	10	3	1	715	6.3	0	55	27
2	96	10	9	1	384	2.4	2.2	73	22
3	96	10	14	1	180	1.0	4.4	71	12
4	96	25	3	1	10029	8.8	2.5	84	20
5	96	25	9	1	6844	3.2	12.8	69	13
6	192	10	9	1	774	6.2	0.5	62	21
7	192	10	18	1	174	1.0	3.9	68	11
8	96	20	3	3	7740	7.9	3.2	67	18
9	96	20	3	6	9803	7.1	3.8	75	22
10	96	20	3	10	11378	8.6	4.5	69	15

[a] $\mathcal{N}$ is the number of neurons of each individual, $\mathcal{P}$ the number of individuals, $\mathcal{V}$ the number of vital informations, $\mathcal{I}$ the number of individual informations, t the number of time steps needed to convergence, $\mathcal{I}(\infty)$ the mean number of individual informations after the convergence, $\mathcal{G}$ the mean number of generations, **C** the correlation, and ΔC_{ij} the dispersion of **C** in a given population, as described in the text. The averages are taken from five different set of populations.

have either exactly the same informations or a large set of them. We can say that the synaptic matrices of individuals in the same class have a macroscopic overlap with each other. These classes explain why the larger the number of generations, the smaller the dispersion, a tendency shown in Table I. Deaths allow mating of parents in the same class in order to generate young which are identical to the parents (we did not include social stratification; hence individuals from the same or different classes can fit together and generate young). Note that if we apply the recipe in which an individual with a large number of informations has reproductive advantages, the growth of the class with the largest number of informations is also enhanced, because the probability of mating in this same class is increased. Similar distributions of overlaps were also detected in ref. 15.

The presence of classes also has other implications. As cited before, each class will suffer the influence of a foreign information in a proper manner and in addition whole classes can disappear or it is even possible that some of them are not altered by the presence of a new individual with completely different informations. As changes occur in different scales because there are classes of different sizes, we suggest that the equilibrium state may be a critical one, and this model of a population may present "self-organizing criticality."(16) As this point is not the main subject of this work, we will not explore it, although we believe that it is worth noting. If it did really occur, it would be a nice realization of self-organizing criticality in systems with a high connectivity.

5. CONCLUSIONS

In this paper, we presented a self-interacting population of neural networks, introducing culture. We have shown that learning plays a role in evolution, even when no mutations are included. The phase space of the system presents multivalleys, a characteristic feature of complex systems. Other interesting phenomena—formation of classes, possibility of self-organizing criticality, etc.—also appear in our model.

ACKNOWLEDGMENTS

We are grateful to L. Peliti and J. J. Arenzon for suggestions and discussions, J. F. Fontanari for providing some fundamental references, and D. Stauffer for reading the manuscript in its most primitive form and for valuable suggestions to improve the final version. This work is partially supported by Brazilian agencies FINEP, FAPERJ, CNPq, and CAPES.

REFERENCES

1. C. B. Harley, *J. Theor. Biol.* **89**:611 (1981).
2. T. A. Crowl and A. P. Covich, *Science* **247**:949 (1990).
3. J. M. Baldwin, *Am. Nat.* **30**:441 (1896).
4. C. Darwin, *On the origin of species* (1859).
5. G. E. Hinton and S. J. Nowlan, *Complex Systems* **1**:495 (1987).
6. J. F. Fontanari and R. Meir, *Complex Systems* **4**:401 (1990).
7. D. O. Hebb, *The Organization of Behavior* (Wiley, 1949).
8. W. S. McCulloch and W. A. Pitts, *Bull. Math. Biophys.* **5**:115 (1943).
9. J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**:2554 (1982).
10. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Ann. Phys.* (N.Y.) **173**:30 (1987); G. A. Kohring, *J. Stat. Phys.* **59**:1077 (1990).
11. T. J. P. Penna and P. M. C. de Oliveira, *J. Phys. A* **22**:L719 (1989); P. M. C. de Oliveira, *Computing Boolean Statistical Models* (World Scientific, 1991).
12. C. Boesch and H. Boesch-Achermann, *La Recherche* **233**:724 (1991), and references therein.
13. W. Brian Arthur and S. A. Kauffman, in *The Economy as an Evolving Complex System*, P. W. Anderson, K. J. Arrow, and D. Pines, eds. (Addison-Wesley, 1988); W. Brian Arthur, *Sci. Am.* **262**:80 (Feb. 1990); P. Bak and K. Chen, *Sci. Am.* **264**:46 (Jan. 1991); M. Mézard and G. Toulouse, *La Recherche* **232**:616 (1991); S. A. Kauffman, *Sci. Am.* **265**:78 (Aug. 1991).
14. G. Kohring, *J. Phys. A* **23**:2237 (1990); T. J. P. Penna, P. M. C. de Oliveira, J. J. Arenzon, R. M. C. de Almeida, and J. R. Iglesias, *Int. J. Mod. Phys. C* **2**:711 (1991).
15. B. Derrida and L. Peliti, *Bull. Math. Biol.* **53**:355 (1991); M. Serva and L. Peliti, *J. Phys. A* **24**:L705 (1991); P. Higgs and B. Derrida, *J. Phys. A* **24**:L785 (1991).
16. P. Bak, C. Tang, and K. Wiesenfeld, *Phys. Rev. Lett.* **59**:381 (1987); *Phys. Rev. A* **38**:364 (1988).

Communicated by D. Stauffer