EVOLVING SENSORS IN ENVIRONMENTS OF CONTROLLED COMPLEXITY

F. Menczer and R. K. Belew

Cognitive Computer Science Research Group
Computer Science & Engineering Department
University of California, San Diego
La Jolla, CA 92093-0114 USA
Email:{fil, rik}@cs.ucsd.edu
Phone: (619)534-8187, Fax: (619)534-7029

Abstract[1]**.** Sensors represent a crucial link between the *evolutionary* forces shaping a species' relationship with its environment, and the individual's *cognitive* abilities to behave and learn. We report on experiments using a new class of "latent energy environments" (LEE) models to define environments of carefully controlled complexity which allow us to state bounds for random and optimal behaviors that are independent of strategies for achieving the behaviors. Using LEE's analytic basis for defining environments, we then use neural networks (NNets) to model individuals and a steady-state genetic algorithm to model an evolutionary process shaping the NNets, in particular their sensors. Our experiments consider two types of "contact" and "ambient" sensors, and variants where the NNets are not allowed to learn, learn via error correction from internal prediction, and via reinforcement learning. We find that predictive learning, even when using a larger repertoire of the more sophisticated ambient sensors, provides no advantage over NNets unable to learn. However, reinforcement learning using a small number of crude contact sensors does provide a significant advantage. Our analysis of these results points to a trade-off between the genetic "robustness" of sensors and their informativeness to a learning system.

1. Introduction

A defining characteristic of the artificial life (ALife) perspective, in common with earlier ethological research, is a holistic view of organisms and their environment, as part of a single system to be modeled. A critical aspect of all ALife models, then, is how the organism/environment "interface" is defined. This specification effectively cleaves the modeler's problem into two, a model of the internal "cognitive" system the organism uses to control its behavior, and a model of environmental change, due in part to these behaviors. In many ALife models, the complexity of the behaviors displayed by the organism is evaluated without taking into account the role played by the sensory system and motor system engineered by the modeler. Rather than assuming any such *a priori* division, we propose to investigate the evolution of the organism/environment interface itself.

The sensory system in particular provides a strong coupling between environmental complexity and difficulty of the survival task. Intelligent behavior may result from complex sensors and trivial processing of this information, or simple sensors and clever processing (Menczer and Parisi 1992; Cliff and others 1993). In fact, experiments by Miglino et al. in which the arbitrary specification of the sensory interface appeared to have important consequences for learning helped to motivate our work (Miglino and Parisi 1991).

Sensors — transducers from external environmental signals to the internal cognitive system — therefore represent a crucial link between two distinct adaptive forces. They are simultaneously phenotypic features shaped by evolutionary forces to define a species' relationship with its environment, and the "input" channels on which a cognitive system's abilities to act and learn are wholly dependent. Features of the sensory system such as determinism, reliability, information content, computational complexity,

[1]To appear in R. Brooks and P. Maes (eds.) *Artificial Life IV*. Cambridge, MA: MIT Press (1994)

signal/noise level, locality, etc. determine the difficulty of the task to be performed by the organism. The same holds for the motor system, but in this paper we will limit our attention to the sensory system.

In Section 2 we begin by sketching a new class of "latent energy environments," and an ALife simulator called LEE underlying our experiments. LEE allows environments of carefully controlled complexity to be defined in terms of the temporal and spatial distribution of a small number of "atoms" and simple "chemical reactions" between atoms that together define the "latent energy" available to support one or more species. This construction lets us make useful analytic statements (e.g., bounding the range of potential solutions), while still permitting the sort of "open-ended" evolutionary experimentation missing from genetic algorithm (GA) simulations using a pre-specified, exogenous fitness function. Section 3 uses LEE as the basis of an analysis of minimal and optimal behaviors possible in these environments.

In Section 4, we use LEE as the basis of experiments using neural networks (NNets) to model individuals and the GA to model an evolutionary process shaping the NNets, in particular their sensors. We report on experiments using two classes of sensors, "contact" (tactile) sensors which allow an organism to sense only directly adjacent space, and "ambient" (olfactory) sensors which provide information from a broader spatial range. We consider variants where the NNets are not allowed to learn, learn via error correction from internal prediction, and via reinforcement learning.

2. Latent Energy Environments

The same ALife coupling of organisms with environments mentioned above brings with it a major methodological problem within ALife research: results reporting behaviors of different organisms in different environments are incommensurate. It is therefore difficult to assess whether an apparently superior behavior is the consequence of more sophisticated adaptive techniques or due to relative *complexity* of the environments. For example, should we be more impressed by the abilities of "amoebas" that successfully evolve to forage for foods (as in (Nolfi, Elman et al. 1990; Parisi, Cecconi et al. 1990)), by "ants" capable of following "trails" (as in (Collins and Jefferson 1992)), or by "agents" that avoid predators (as in (Ackley and Littman 1992))? It would be very desirable to be able to define artificial environments of controlled complexity, within which a wide range of ALife techniques might be directly compared.

On the other hand, traditional genetic algorithms can be viewed as having gone too far towards analytically tractable models of the environment. The typical GA models every environment as a "fitness" function returning a value f(x) for each phenotype x, and the process of evolution as finding the value x* optimizing f(). GA theory has progressed considerably within this frame, for example stating conditions on f() which make a search for the optimum "GA-hard" or "GA-easy," but casting evolution as a search for optimality with respect to a fixed, exogenously defined function leaves much to be desired in terms of biological veridicality. An ALife environment must also support the specification of strategies of arbitrary and unanticipated complexity, allowing creative, "open-ended" evolution.

Our work builds on several other attempts to define generic conditions on environmental complexity that are analytically tractable without constraining evolution's creative potential. The physicist Otto Rössler observed that varying food density could be used as a simple metric on the complexity of an environment facing an organism (Rössler 1974). Even a random walk is an adequate foraging technique if food is abundant, but more coherent movement is required if the foodstuffs are more widely spaced; in even more scarce environments, the foraging organism may need to depend on landmarks, cognitive maps, etc. Wilson has pursued a similar analysis of environmental complexity as a function of spatial food density (Wilson 1991).

Food density is clearly a useful dimension for ALife simulation, but this single dimension of environmental variability must be extended to include other factors if we are to be able to test the full repertoire of ALife models. Holland's recent ECHO model describes an extraordinarily rich domain in which many intra- and inter-specic evolutionary questions might be explored (Holland 1992; Jones and Mitchell 1993). A small set of basic elements and rules by which these are combined, metabolized and replicated are defined, and conventional constructs such as "organism," "species," "predator," etc. emerge indirectly. In fact the range of possible experiments, or outcomes to a single experiment in ECHO is so broad that, like real biological systems, its behavior is often extremely difficult to analyze.

"Latent energy environments" (LEE) are a modest step from the single dimension of food density towards the richness of the ECHO model. The first step is to break up the conventional notion of "food." Food is typically modeled as a spatially localized element of the environment, consumption of which is directly correlated with an organism's fitness. Instead of food, LEE environments are filled with "atoms" (e.g., A, B and C) that must be combined by organisms in order to realize the "energy" given off by various "chemical reactions".[2] These reactions may be exothermic or endothermic (i.e., give off positive or negative energy) and may give off other atoms: a reaction can thus be represented as $A + B \rightarrow Energy + Byproducts$. A LEE environment is specified by defining the set of atoms, their spatial and temporal distribution, and the chemical reactions over these atoms. The set of possible reactions can be represented by the entries of a symmetric reaction table representing the "physics" of the environment, which typically remains constant throughout a simulation.

We can now define two critical features of a LEE environment. The *latent energy* of an environment is the total potential energy available from all reactions over the atoms in the environment.[3] By controlling the rate at which atoms of each type are introduced, we can regulate the amount of potential energy available in its latent form, while actual energy can ultimately be realized only upon dynamically choosing a reaction sequence. Second, the amount of *work* required to release the latent energy is defined in terms of the distance one of the atoms in a reaction must be moved in order to be brought into contact with the other. This is just the correct metric for evaluating organisms who must harvest energy to survive, and whose fundamental behavior, as will be shown in the next section, is movement. By controlling the spatial distributions of atom types in the world, we can dynamically regulate the amount of work required to combine them.

In short, we have quantitative knowledge of and control over the latent energy available in a world, how much work must be done in order to realize this energy, and how difficult it therefore is for an organism (or ecology of organisms) to survive, without specifying just how the latent energy is to be realized by one or more species. The monitoring of population dynamics makes it possible to compare ecologies in different environments, in terms of how efficiently each exploits the energy latent in its own environment.

2.1. Modeling individuals' behavior

The defining feature of the LEE model is that energy in an environment can only be released through the *behaviors* of organisms. Behaviors induce reaction sequences, thus catalyzing the transformation of latent energy into actual, life-giving energy. The notion of behavior -- actions taken by the organism that change the world and/or the organism's relationship to the world -- immediately binds an organism to its environment in an intrinsic way. Thus we take *movement* to be our canonical behavior, in part because this overt action is most easily observed.

In the experiments reported here we model the mapping from *sensory* states to *motor* actions using well-studied feed-forward *neural networks.* An aspect of real organisms that is often forgotten by cognitive scientists is that these three components of an organism's cognitive apparatus are but a subset of a more inclusive *soma*, with adaptive consequences in its own right. We briefly

[2] It is important to caution that our use of physical and chemical terminology ("atoms," "reactions," "energy," "work," etc.) is metaphoric. In particular, in the context of our investigation of sensors, we are *not* trying to model the low-level physiological physics of receptors.

[3] In fact, latent energy is not well-defined for arbitrarily specified reactions. If the reactions are carefully defined, however, the amount of energy latent in an environment can be derived directly from the atomic distributions.

describe each of these four components of LEE organisms below.

Soma: There are two features of an organism's body that are relevant to the experiments reported here. First, each organism is assumed to have an internal cavity we call its *gut*. We will assume that every gut is capable of holding a single atom. An organism automatically ingests any atom over which it travels, causing this atom to be placed in its gut. If the gut already contains an atom and the new atom causes them to be involved in a reaction, the reaction is performed and the organism either accrues (exothermic reaction) or loses (endothermic reaction) energy. The second feature of a body is that it determines the placement and orientation of an organism's sensors and motors. While none of these somatic characteristics are varied or allowed to evolve in the experiments reported here, the gut will be shown to function as an import form of "memory" in Section 3.2, and two different sensor configurations are considered below. More generally, LEE's somatic model has been designed to explore what often seems ill-founded discrimination into "cognitive" and "non-cognitive" features.

Sensors: The sensory system receives information from both the world "outside" the organismic membrane and from "inside" its internal soma. Sensors define the mapping from these sources of information and the brain's input. In general, LEE sensors may differ in range, directionality, sensitivity, resolution, specificity, accuracy, etc. Here we will consider only two types of sensor, *contact* and *ambient*. Contact sensors provide a binary presence/absence indication of a single atom type in the space directly adjacent to the sensor. Contact sensors are sufficient to support avoidance behaviors, but not approach. Ambient sensors provide an indication of a particular atom's presence, summed over cells in a neighborhood of the sensor and weighted inversely according to their distance[4]. Ambient sensors can underlie approaching behavior only if the brain possesses either a memory to compare *temporal* differences in a single sensor, or multiple ambient sensors placed and oriented differently to compare *spatial* differences.

Motors: Motors function as the output effectors of the organism, changing its location or orientation with respect to a fixed environment referential frame. In general, they may differ in range, energetic efficiency, accuracy, etc. In these experiments we use exclusively a very simple motor system used by others (Nolfi, Elman et al. 1990; Parisi, Cecconi et al. 1990). Two binary output channels (left and right) control a single motor allowing one of four moves: stand still (00), turn 90 degrees right (10) or left (01), or go one cell forward (11).

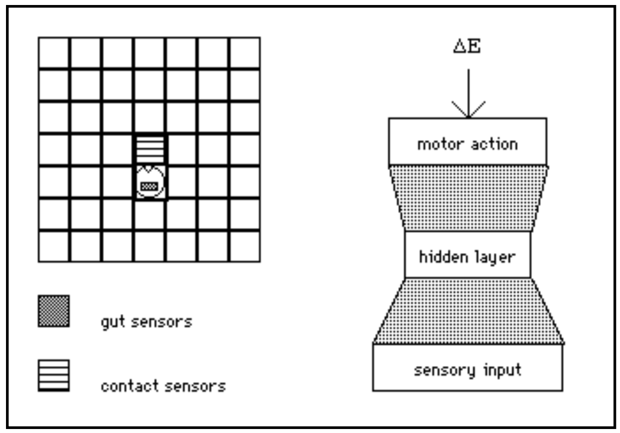

Figure 1: sensory system of LEE organisms for the avoidance task (left); architecture of the corresponding NNets (right). All connections are modifiable by reinforcement learning (the arrow indicates the source of the reward-penalty signal).

Neural network: The signals from each sensor are used as the inputs to a feed-forward neural network with a single hidden layer of units, and the NNet's output units are used to control the motors. We will consider two variants of this basic NNet, each using a different source of feedback for learning. In the experiments of Section 4.3, we follow the work of Nolfi et al. (Nolfi, Elman et al. 1990). The NNet's output is extended to produce not only a motor action, but also a *prediction* of the sensory input it will receive after that action. This prediction is compared with the actual sensory information following the action and prediction error is minimized by a standard back-propagation technique. In the experiments of Sections 4.4, prediction-based error-correction learning is replaced by

[4] In these experiments the metric is defined behaviorally, in terms of the number of moves required of the organism to reach this location. Given the simple motor system, this distance is well-specified.

reinforcement learning. NNet weights are modified only immediately after actions catalyzing a non-zero energy reaction using a simplified version of associative reward-penalty.

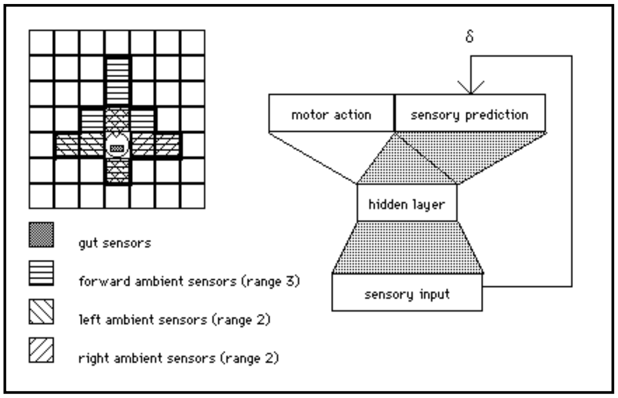

Figure 2: sensory system of LEE organisms for the approaching task (left); architecture of the corresponding NNets (right). The connections shown in dark are modifiable by prediction learning (the arrow indicates the source of the teaching signal for back-propagation).

Putting two organisms together: Figures 1 and 2 show two examples of fully specified organisms. Both organisms use a pair of sensors for the gut and the same motor system. Figure 1 shows an organism with a pair of contact sensors, using a hidden layer of two units and reinforcement learning. Figure 2 shows an organism with three pairs of ambient sensors, using prediction learning. This second organism may seem unnecessarily complicated, but recall that predictive learning demands a difference, in either space or time. To allow a spatial comparison of signals, we use two sensors directed towards opposite sides of the organism, each with a range of 2 moves; to brake the lateral symmetry of these two sensors, we add a third sensor with range 3, oriented frontally. Each sensor is replicated twice to allow full information, for a total of six ambient sensors and two gut sensors. This network has eight hidden units. Note also that the weights from the hidden layer to the motor units cannot be modified by learning. If these NNet characteristics are to change, it must be due to evolutionary forces.

2.2. Modeling populations' evolution

The same model of an organism can and must be viewed from two distinctly different levels of analysis. The description above was in terms of the *individual's* behaviors, the soma, sensors, etc. that generated them, and the relevant time scale was that of the behaviors. However, this same organism is simultaneously a member of an evolving

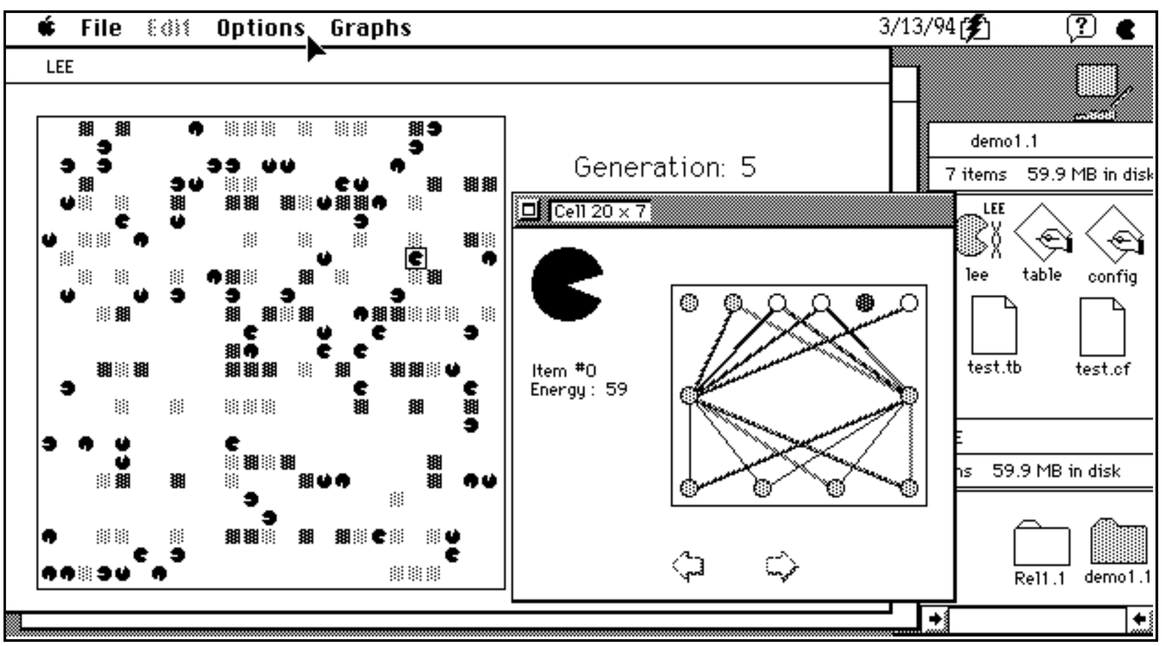

Figure 3: interactive screen during a LEE simulation on the Macintosh.

population , with the relevant time scale being that of generations. These two perspectives on LEE are suggested by the two windows from the Macintosh interface to the LEE simulator shown in Figure 3.[5] One shows a population of organisms scattered over an environment containing a uniform distribution of two atom types, and the other shows an expanded view of an individual's NNet.

In LEE the construct tying an individual's behaviors to its evolutionary fitness is the energy it accrues in its gut by "metabolizing" reactions. If ever the organism runs out of energy, it dies. If it collects more than a fixed threshold of energy in this way, it asexually reproduces. A critical observation is that this form of fitness is intrinsic to LEE environments, resulting directly from competition for resources within the population rather than being extrinsically applied as an evaluation on individuals (Mitchell and Forrest 1993).

After a LEE environment is specified, a simulation is begun by randomly placing an initial population of organisms within it. For each time step, every organism executes a basic life cycle:

```
for each cycle {
  for each alive organism {
    sense world;
    feed forward activations;
    move;
    digest;         /* catalyze
                       reactions */
    learn;          /* change NNet
                       weights */
    if (energy > α) {
      reproduce;    /* copy genotype */
      mutate;       /* new genotype */
    }
    else if (energy < ω) die;
  }
  replenish world;
}
```

[5]C source code and documentation for release 1.* of LEE (Menczer and Belew 1993) is available by anonymous ftp from cs.ucsd.edu (132.239.51.3) in the pub/LEE directory. From the World Wide Web, use URL http://www-cse.ucsd.edu/users/fil.

This cycle is repeated indefinitely, with organisms foraging independently in a shared world until some die, other reproduce, and their offspring repeat the process. An important consequences of this "steady-state" version of a genetic algorithm (DeJong and Sarma 1993) is that the population size does *not* remain constant throughout an experiment. Analysis of this dynamic variable, and its connection to individual behaviors, is considered in the next section.

In addition to specification of selective pressures deriving from relative fitness, a Darwinian process must also have a mechanism for generating variation. In these experiments we will consider only random mutations (i.e., there is no sexual recombination/cross-over) of two classes of genotypic characters. The first are the types of atoms the sensors sense (A,B,C, etc.). With small probability (0.05) an offspring's sensor may signal a different atom type than the corresponding sensor in its parent. The second class of genotypic features are the NNet's weights. In all of the experiments reported here, weights are represented as floating-point numbers, and mutation performed by adding random, uniformly distributed noise.

3. Analysis

3.1. Fitness

How can we measure how well behaviors are adapted in the environment? The characteristics of an environment allow us to establish a connection between the size of an evolving population and the average adaptation of the observed behaviors in the population. The finiteness of environmental resources imposes a limit on how many individuals can be sustained: we identify this maximum population size with the "carrying capacity" of the environment. For example, extinction indicates that the carrying capacity of the environment is insufficient to support a population large enough to withstand stochastic fluctuations.

To see how to quantitatively estimate optimality of behaviors by monitoring the population size throughout an experiment, consider the relationship between latent energy and population size in a particular, very simple and regular reaction table for three atomic atoms, A, B and C:

$$\begin{array}{c|ccc} & A & B & C \\ \hline A & -\beta E & E & - \\ B & E & -\beta E & - \\ C & - & - & - \end{array} \tag{1}$$

where $E>0$, $|\beta|<1$ and a dash indicates no reaction. These conditions insure the non-zero-sum energetic content on the environment necessary to avoid extinction at the onset of the evolutionary process.

Notice that the matrix (1) is symmetric, so order of ingested atoms does not matter. We can characterize two different strategies, one for combining A and B atoms catalyzing exothermic reactions, and one combining homogeneous atoms catalyzing endothermic reactions. In fact, a single parameter can be used to measure the relative frequency of these two strategy types (see below). Since we know C atoms to be causally irrelevant, we leave them out of the environment entirely. The rates of replenishments for the other two atoms is $r(A)=r(B)=1$.

Table (1) needs to be augmented with knowledge about the spatiotemporal distributions of the atoms in order to determine the environmental complexity of the survival task. Let us assume the environmental resources are introduced with uniform distribution in space and time, at rate r (measured in atoms per unit time and area). It is important to note that these resources are shared by the population: these conditions lead to density-dependence of fitness (Stearns 1992). Finally let us define a small constant cost, ε, incurred every time that an organism makes a move. It is then easy to show (Menczer and Belew 1994) that at equilibrium, the expected population size p is given by:

$$p = \frac{rE}{\varepsilon}(\eta + \eta\beta - \beta) \tag{2}$$

where η is the time-average of the fraction of catalyzed reactions yielding energy E. Equation (2) provides an important link between optimality of behavioral strategies, expressed through η, and population size. In particular, the case of optimal behavior ($\eta=1$) corresponds to the maximum sustainable population, i.e., the carrying capacity:

$$p_{\max} = \frac{rE}{\varepsilon}. \tag{3}$$

Conversely, to estimate the optimality of a population's behavior at equilibrium by measuring the population size, we can solve (2) for η:

$$\eta = \frac{\frac{p\varepsilon}{rE} + \beta}{1+\beta}. \tag{4}$$

Another useful bound results from behaviors in which individuals are as likely to catalyze exothermic as endothermic reactions. This results in a "random choice" population level:

$$p_{\eta=1/2} = \frac{rE}{2\varepsilon}(1-\beta). \tag{5}$$

Equations (3) and (5) give us an upper bound and a baseline, respectively, to measure the average fitness of the population. We have shown elsewhere (Menczer and Belew 1994) how to use these equations to make accurate predictions about the outcomes of simulations with simple environments, and to compare behaviors in different environments.

When evolved behaviors allow organisms to make a more efficient use of the latent energy, an increase in η follows and there is a transition to a new equilibrium. How can we characterize at the individual level the increase in fitness observed at population level? During these phases density dependence is no longer the sole fitness factor, and thus carrying capacity is neither a sufficient predictor of population size nor a satisfactory fitness measure. Population size and average age, on the other hand, remain very well correlated at all times (the correlation coefficient being μ~0.8). This suggests a connection between age and individual fitness. To clarify the connection, let us introduce a fitness measure that can be used without bias. *Expected reproductive success* is typically defined as:

$$R_0 \equiv E\left[\frac{\#\,offsprings}{lifetime}\right]. \qquad (6)$$

To find the relationship between R_0 and population size, we can use two sensible assumptions. First, the number of offspring generated up to age a grows linearly with a:

$$n(a) \approx \rho a \qquad (7)$$

where ρ is a constant called *reproductive rate*. Second, in agreement with measured age distributions in LEE (Menczer and Belew 1994), the number of organisms having age a at any given instant follows the Euler-Lotka equation (in the continuous limit):

$$N(a) \approx \begin{cases} pe^{-a} & a < a_{\max} \\ 0 & a \geq a_{\max} \end{cases}. \qquad (8)$$

Using (7) we find

$$R_o = \int_0^{a_{\max}} n(a)da = \frac{\rho}{2} a_{\max}^2. \qquad (9)$$

To find a_{max} notice that, from (8),

$$\begin{aligned} N(a_{\max}) &\sim 1 \\ pe^{-a_{\max}} &\sim 1 \\ a_{\max} &\sim \ln p \end{aligned} \qquad (10)$$

and thus we conclude:

$$R_0 \propto (\ln p)^2. \qquad (11)$$

Equation (11) establishes a link between expected individual fitness and population size which remains valid even when the behavior of the organisms is improving and thus equation (4) is of no assistance. Furthermore, it provides a way to measure selective pressures between different behaviors from differences in population size.

3.2. A priori strategies

Using only contact sensors, it is not possible to approach atoms. Atoms must be encountered by some type of "blind" foraging pattern. Given the toroidal structure of the LEE world, always moving straight ahead is typical. With only blind foraging allowed, and for the environment described by table (1), the optimal behavior is:

```
IF the gut is empty, eat.
ELSEIF the adjacent cell's content
     is the same as the content of
     the gut, turn
ELSE, go forward.
```

To implement this *optimal* strategy, the sensory system should ignore C's and discriminate A from B with both gut and external sensors. Assuming the availability of gut and contact sensors giving accurate discriminating information, an (example of) optimal NNet function can be represented by the following table:

Gut	Envt	Left	Right
A	A	1	0
A	B	1	1
B	A	1	1
B	B	0	1

Note that this table simplifies the space of potential behaviors from the entire repertoire of available sensors in order to focus on the four consequential decisions. To make the discriminations necessary for this optimal strategy, one of the gut's two sensors must sense A and the other B, and the same must be true of the two environmental sensors.

If some sensors give information that is only partially discriminating, then only suboptimal strategies are possible. For example assume that the gut's sensors are as above but the environmental sensors both sense A. With this limited information, the best possible strategy is:

```
IF the gut is empty, eat.
ELSEIF both the gut and the
     adjacent cell's contents
     are A, turn.
ELSE, go forward.
```

Critically, the NNet implementation of this sub-optimal strategy:

Gut	Envt	Left	Right
A	A	1	0
A	?	1	1
B	A	1	1
B	?	1	1

depends on learning a function very much like "inclusive OR," which is known to be much simpler than the "exclusive OR" function underlying the optimal strategy above.

4. Experiments

4.1. Controls

The first experiment is made of two control simulations aimed at measuring the baseline and best performance obtainable by evolving the network connections alone. There is no learning during life, and no evolution of sensors. Later experiments can be compared with these controls to see how much evolved sensors are better (worse) than the worst (best) possible sensors we can design.

The worst possible sensors are those that leave the organism blind, i.e., without any information whatsoever about the gut or the external environment. This is done by setting all sensors to C, so that the inputs of the NNet are always zero: the adapted weights must implement constant actions, independent of the environment, i.e., a "random choice" strategy. The best possible sensors are those allowing organisms to discriminate between A and B atoms. This is done by setting both sensor pairs to A,B: the adapted weights may use the "full information" signals to implement the optimal strategy.

The results are shown in Figure 4. Population averages are over different simulation runs. Random choice, as expected, cannot show any improvement past the very initial phase in which the strategy to always move sets in. Full information sensors, on the other hand, allow near-optimal convergence. To see this, note that the population size at the end of the simulations corresponds to an average fraction η~0.87 of positive energy reactions (from equation (4)). Another way to assess the optimality of the evolved population is by the selective pressure in terms of reproductive success ratio: $R_0/R_0(p_{max})$~0.93 (from equation (11)). These measures match the observation that the majority of organisms at the end of the simulations exhibit behaviors corresponding to the optimal strategy (cf. previous section).

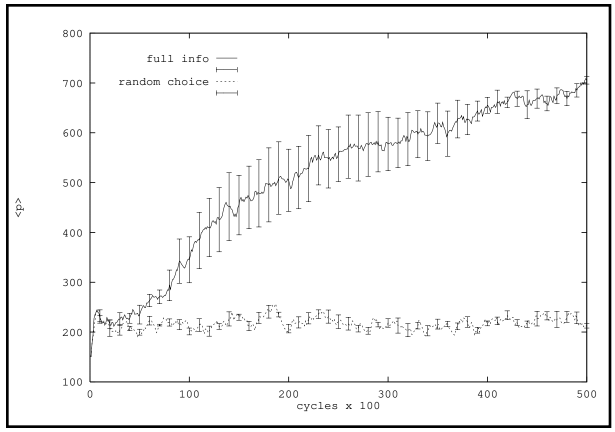

Figure 4: population size in control simulations.

4.2. Evolution alone

We now want to test whether evolution alone can adapt sensory systems while coadapting connection weights with the information provided by the sensors in order to produce good behaviors. Thus no learning occurs during life in the experiments reported in this subsection, and the sensory system evolves as illustrated in the previous section. The only gut and contact sensor pairs which afford the information necessary for the optimal strategy are A,B and B,A. In either case, we expect the population to converge to optimal percentages of 50% for A (and B) sensors, and 0% for C sensors.

Figure 5 reports the resulting evolution of sensors in this experiment by plotting percentages of A and C summed over all sensors. The B percentage is not shown for clarity (it can be obtained by subtracting those of A and C from 100%, and its expected optimum is identical to that of A). Although the tendency is as predicted, the large error bars indicate the importance of random genetic drift in these experiments. Upon repeating the simulation for different runs, sensor percentages converge to highly dispersed values.

Our interpretation of the large drift is that only about 5% of the possible sensor configurations (4 out of the total 3^4=81) allow the optimal strategy to evolve. Furthermore, these configurations do not all share the same coadapted connection weights, so the resulting behavior is very fragile in the face of mutations on the sensory system. In contrast, a suboptimal strategy like the one shown in the previous section can be implemented with 28/81~35% of the sensor configurations. While this strategy results in somewhat inferior performance, it is more robust because not only more configurations can achieve it, but more importantly, many of them function with the same network weights. This follows naturally from some of the sensors being redundant with respect to the suboptimal strategy: these sensors can be mutated without affecting the network behavior. Such redundance is also the source of the observed genetic drift: *robustness* and *drift* are two faces of the same selective pressure against optimality. To be sure, optimal sensors are generated from mutations at all times, but they are destroyed before the connections evolve to the coadapted weights. The fact that the suboptimal strategy is the winning one is also confirmed by looking at how organisms at the end of a run behave in the environment.

One way to increase the selective pressure toward the optimal strategy is to increase the environmental complexity. This is easily done with LEE by dynamic changes in the parameters describing the environment, such as the reaction table entries or the spatiotemporal distributions of resources. We have compared the experiment above with one in which the environment becomes increasingly hostile by simply increasing the β parameter linearly with time, so that $d\beta/dt=1.1\times10^{-4}$ cycle^{-1}. To see that this corresponds to increasing complexity, notice that the random choice behavior ceases to be a viable strategy, resulting in inevitable extinction as soon as $\beta=1$ (zero-sum environment).

The population levels for constant ($d\beta/dt=0$) and variable environments are plotted in figure 6. Notice that the minimum for the changing environments is reached when $\beta=1$ (3000 cycles). After that, the population average performance is consistently better in the more complex environments. It is important that, although the population levels are about the same at the and of the simulation, they do not correspond to equal performances: for example, recall from (4) that p~700 corresponds to η~0.87 in a fixed environment, but to η~0.97 in a changing environment. In fact, almost all the organisms evolved in complex environments exhibit the optimal behavior at the end of the simulation. However, the population reaches the same size in the two experiments: evolution has reached its reproductive success goal, which, in agreement with (11), does not depend on environmental complexity.

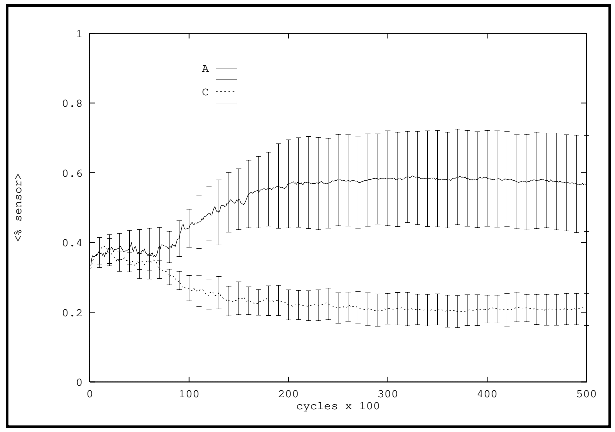

Figure 5: sensor percentages in simulations with evolution alone (no learning).

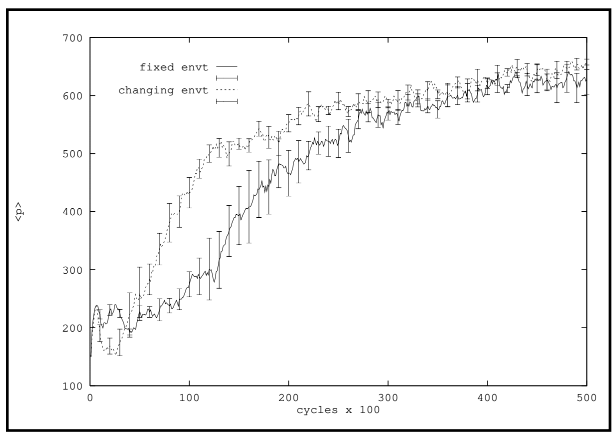

Figure 6: population size in simulations with evolution alone (no learning) and different environmental complexity.

4.3. Prediction learning

Following the hypothesis that learning may facilitate the evolution of informative sensors, we have explored the Baldwin effect's potential to impose an extra selective pressure on the sensory system (without making the environment more complex) and consequently reduce the drift. A requirement for any learning mechanism in our model is the assumption that there be no additional "teaching" input available to the organism, beyond the relatively neutral stimulus of the organism's environment. This would seem to preclude many successful "supervised" learning techniques, e.g., back-propagation (Rumelhart, Hinton et al. 1986). However, if the organism is simply forced to *predict* the expected outcome of its actions, differences between its expectations and the actual outcome can generate the same sort of error information, without any additional teacher (Nolfi, Elman et al. 1990; Parisi, Cecconi et al. 1990). This motivates our first choice, prediction learning using a standard back-propagation algorithm[6]. If an organism's learning system can converge on a consistent explanation of the world that can be of aid in the survival task, we may expect that the Baldwin effect will facilitate the evolution of sensors.

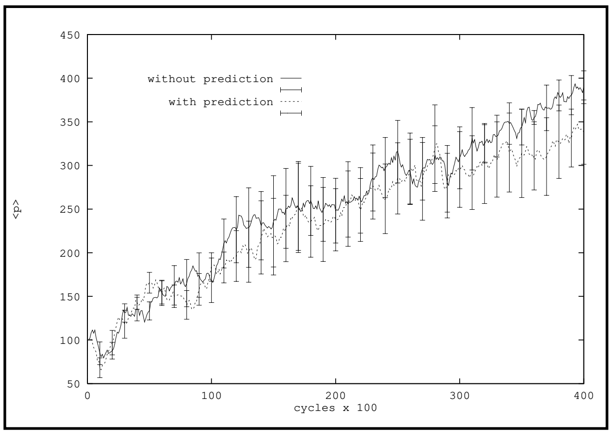

Figure 7: population size in simulations with ambient sensors, with and without prediction learning.

To increase the selective pressure against the random choice strategy without changing the corresponding carrying capacity, we have modified the environments in this experiment by clustering the A and B sources with gaussian distributions. To avoid initial extinction, we have compensated this added complexity using β=1/3. Results are plotted in figures 7 and 8 for population size and sensor percentages, respectively. The observed performance is consistently generated by evolved behaviors which perform some approaching, but as far as combining atoms correspond to the suboptimal strategy discussed above. Good approaching strategies can improve fitness only to a limited extent in our environments, therefore these results basically show that prediction learning provides no significant advantage over evolution alone, neither for fitness improvement nor for drift reduction.

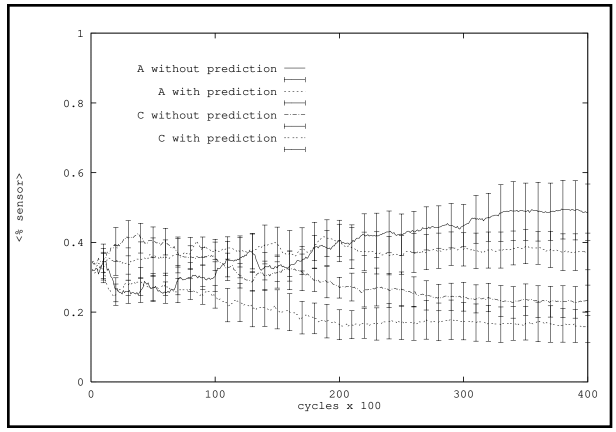

Figure 8: evolved sensor percentages in simulations with ambient sensors, with and without prediction learning.

This apparently disappointing result is actually of assistance in understanding the nature of the information afforded by ambient sensors and their interaction with the prediction learning algorithm. One difficulty related to these sensors is the *ambiguity* of their signals, because the function of the environment state that they compute integrates signals over the neighborhood and thus is non-invertible (many-to-one). In this sense, the information they provide is harder to use than for contact sensors, even though theoretically it allows prediction because less local. Another difficulty related to prediction learning is that the prediction task is *not well correlated* with the survival task. One way to predict well is not to move, but this is clearly a suicidal behavior. Another reliable prediction in these environments is the absence of C atoms, i.e., a zero signal from all inputs corresponding to C sensors. Therefore the learning process drives phenotypes toward behaviors that are not useful because based on blind (C) sensors. Figure 9 shows a scatter plot of R_0 fitness and sum-squared-error during a typical run with fixed, full

[6]Learning rate=0.2, no momentum.

information sensors. The prediction error is positively correlated with reproductive success (μ~0.86), confirming that evolution selects against good predictors and thus this type of learning is unlikely to facilitate evolution in this type of experiment.

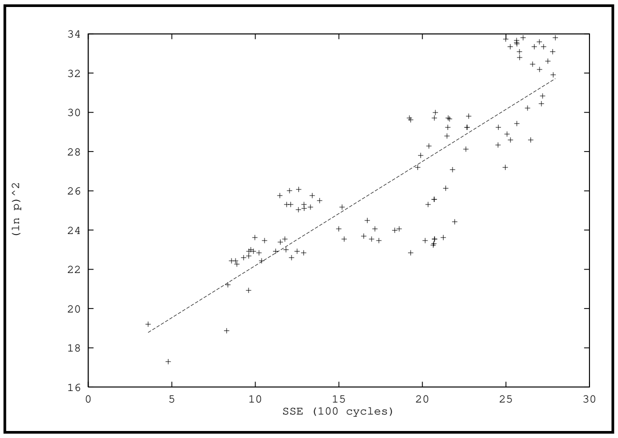

Figure 9: scatter plot of prediction error and expected reproductive success (from (11)). Linear regression fit is also shown.

4.4. Reinforcement learning

From the result of the previous experiment we certainly cannot draw the conclusion that learning in general cannot interact constructively with evolution, but rather the observation that prediction learning is not appropriate for our simulations. Therefore we need another unsupervised learning algorithm, with the same motivation as for prediction learning: to determine whether the Baldwin effect can facilitate the evolution of informative sensors. However, we seek a learning algorithm whose error criterion is better correlated with the fitness of the survival task, and for this we turn to reinforcement learning. A number of authors have explored the use of *associative reinforcement learning* in conjunction with NNets, GAs, and ALife (Ackley and Littman 1990, 1992; Barto and Anandan 1985; Whitley et al. 1991). Evolution itself can be viewed as a form of reinforcement search algorithm, but the reinforcement signal (death or reproduction) is heavily delayed and of relatively little use during life (Ackley and Littman 1992). If however a useful reinforcement signal is made available by an organism's interaction with its environment, then it can be used to determine an error information directly from the association between inputs and actions. Such a signal is indeed present during the life of organisms in LEEs: the changes in internal energy following their actions. An increase in energy corresponds to a reward, a decrease to a penalty, and we expect the corresponding error criterion to be more closely related to the survival task than prediction error.

To use reinforcement learning in experiments on the evolution of sensors, we go back to the simpler situation in which only contact sensors are used. So the only actions that determine the fitness are those choosing whether or not to eat a food when it is sensed by the contact sensor, based on the gut content. Notice that this setting avoids the problem of delayed reinforcement altogether. Therefore we use reinforcement to modify weights only immediately after actions catalyzing reactions. The implemented algorithm is a simplified version of associative reward-penalty (Barto and Anandan 1985). When a change in energy $|\Delta E|>\varepsilon$ occurs, a two-valued reinforcement signal

$$r = sgn(\Delta E) \tag{12}$$

is generated (this does not violate our assumption that there be no external teacher, since the reinforcement can be generated internally, without intervention of cognitive skills). This is used to produce a teaching action pattern from the last output:

$$o_i^t = \begin{cases} H\left(o_i - \frac{1}{2}\right) & r = +1 \\ 1 - H\left(o_i - \frac{1}{2}\right) & r = -1 \end{cases} \tag{13}$$

where H is Heaviside step function and o_i is the activation of output i (before being clipped to 0/1). The classic delta rule is then used for weight correction:

$$\Delta w_{ij} = \gamma o_j \delta_i, \tag{14}$$

where γ=0.1 is the learning rate. The delta function for output units is $\delta_i = o_i^t - o_i$ and for hidden units it is obtained by standard error back-propagation.

Figures 10 and 11 show the results of adapting sensors with evolution alone and with reinforcement learning, respectively. A significant reduction in drift and increased selective pressure toward informative sensors (e.g., against C sensors) is observed in the simulations with learning. This reflects the fact that in a greater fraction of the runs the population converges to full information sensors (approximately 50% A, 50% B and no C) and optimal behaviors as defined in the previous section.

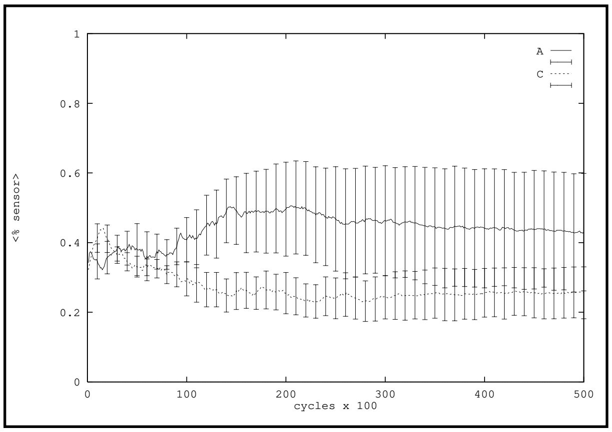

Figure 10: evolved sensor percentages in simulations without reinforcement learning.

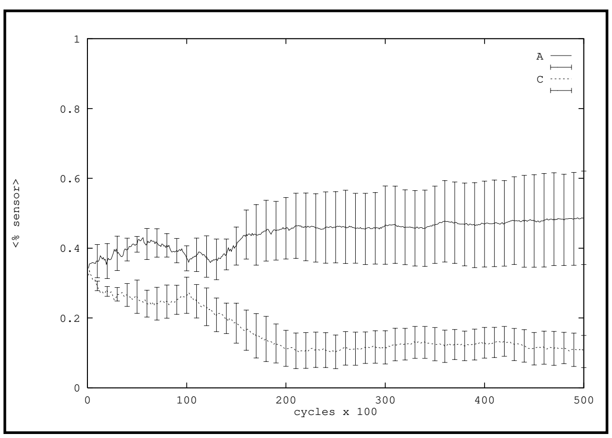

Figure 11: evolved sensor percentages in simulations with reinforcement learning.

5. Conclusion

We believe that LEE represents both a rich theoretical framework and a useful simulation tool for the ALife community. In this paper we have demonstrated how it can be used to model biologically realistic selection pressures and thereby measure the complexity of a survival task. An important emerging feature of the model is its emphasis toward behaviors which are *adapted* as opposed to *optimal* in their environments. This adresses a common criticism of biologists toward adaptationist models of evolution as an optimization procedure (Gould and Lewontin 1979).

We have reported on the use of LEE to investigate the evolution of sensors, something that has turned out to be an unexpectedly difficult problem due to the interdependence between the NNet implementing a behavior and the sensory interface producing the information upon which the behavior must be based. We have shown that, if the cognitive capabilities provided by learning are well correlated with those required for survival in a particular environment, then phenotypic leaning can assist adaptation in that environment by facilitating the evolution of appropriate genotypes.

The model has been successfully applied to other issues, such as assessing the adaptive advantages of random individual plasticity in non-stationary environments (Menczer 1994) and studying the evolution of age at maturity in the presence of cultural transmission by imitation (Cecconi et al. in preparation). In the near future we intend to use LEE for research on several other issues, including: development and morphology of both somatic and cognitive characteristics; sexual reproduction; emergence of speciation; multiple species ecologies; and predator-prey population dynamics. Further work is necessary to better understand the relationships between different types of learning and evolution in environments of controlled complexity.

Acknowledgments

This project originated from conversations with Stefano Nolfi and Domenico Parisi at the Institute of Psychology of the National Research Center in Rome, Italy. Stefano Nolfi, Jeff Elman, Greg Linden, and Federico Cecconi have contributed to the code. We also thank Federico Cecconi and everyone in the Cognitive Computer Science Research Group, in the CSE Department at UCSD, for helpful discussions on many modeling and experimental issues.

References

Ackley, D. and M. Littman (1990). Generalization and Scaling in Reinforcement Learning. Advances in Neural Information Processing Systes (NIPS2), San Mateo, CA: Morgan Kaufmann.

Ackley, D. and M. Littman (1992). Interactions between learning and evolution. Artificial Life II Eds. C. Langton, C. Taylor, J. Farmer and S. Rasmussen. Reading, MA, Addison Wesley. 487-507.

Barto, A. and P. Anandan (1985). Pattern Recognizing Stochastic Learning Automata. IEEE Transactions on Systems, Man, and Cybernetics **15**: 360-375.

Cecconi, F., F. Menczer, and R. Belew (in preparation). Learning and the evolution of age at maturity. To be submitted to Adaptive Behavior.

Cliff, D., P. Husband, and I. Harvey (1993). Analysis of evolved sensory-motor controller. In Second European Conference on Artificial Life (ECAL93), Brussels.

Collins, R. and D. Jefferson (1992). AntFarm: Toward Simulated Evolution. Artificial Life II Eds. C. Langton, C. Taylor, J. Farmer and S. Rasmussen. Reading, MA, Addison Wesley. 579-601.

DeJong, K. and J. Sarma (1993). Generation Gaps Revisited. Foundations of Genetic Algorithms 2, San Mateo, CA: Morgan Kaufmann.

Gould, S. and R. Lewontin (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptionist programme. Proceedings of the Royal Society of London **B205**: 581-598.

Holland, J. (1992). Adaptation in natural and artificial systems, Second Edition. Cambridge, MA, Bradford Books (MIT Press).

Jones, T. and M. Mitchell (1993). Introduction to the ECHO model. Santa Fe Institute. Working paper 93-12-074.

Menczer, F. (1994). Changing Latent Energy Environments: A Case for the Evolution of Plasticity. University of California, San Diego. Technical report CS94-336.

Menczer, F. and R. Belew (1993). LEE: A Tool for Artificial Life Simulations. University of California, San Diego. Technical report CS93-301.

Menczer, F. and R. Belew (1994). Latent Energy Environments. Plastic Individuals in Evolving Populations Eds. R. Belew and M. Mitchell. Reading, MA, Addison Wesley.

Menczer, F. and D. Parisi (1992). Recombination and unsupervised learning: effects of crossover in the genetic optimization of neural networks. Network **3**: 423-442.

Miglino, O. and D. Parisi (1991). Evolutionary stable and unstable strategies in neural networks. IJCNN, Piscataway, NJ: IEEE.

Mitchell, M. and S. Forrest (1993). Genetic Algorithms and Artificial Life. Santa Fe Institute. Working paper 93-11-072.

Nolfi, S., J. E. Elman, and D. Parisi (1990). Learning and Evolution in Neural Networks. University of California, San Diego. Technical report CRL-TR-9019.

Parisi, D., F. Cecconi, and S. Nolfi (1990). Econets: Neural networks that learn in an environment. Network **1**: 149-168.

Rössler, O. (1974). Adequate Locomotion Strategies for an Abstract Environment -- A Relational Approach. Physics and Mathematics of the Nervous System Eds. M. C. M, W. G. W and M. Cin. New York, NY, Sringer-Verlag. 399-418.

Rumelhart, D., G. H. Hinton, and R. Williams (1986). Learning internal representations by error propagation. Parallel Distributed Processing: Explorations in the Microstructure of Cognition Eds. D. R. DE and J. McClelland. Cambridge, MA, Bradford Books (MIT Press).

Stearns, S. (1992). The Evolution of Life Histories. New York, NY, Oxford University Press.

Whitley, L., S. Dominic, and R. Das (1991). Genetic reinforcement learning with multilayer neural networks. Fourth International Conference on Genetic Algorithms, San Mateo, CA: Morgan Kaufmann.

Wilson, S. (1991). The Animat Path to AI. From Animals to Animats: First Intl. Conf. on Simulation of Adaptive Behavior, Cambridge, MA: Bradford Books (MIT Press).