Evolutionary Computation — A Study on Collective Learning

Hans–Paul Schwefel
Department of Computer Science, University of Dortmund
D - 44221 Dortmund, Germany

ABSTRACT

On the one hand, people admire the often strikingly efficient results of organic evolution. On the other hand, however, they decry mutation and selection to be a rather prodigal, inefficient trial-and-error strategy. Taking into account the parallel information processing in a heterogeneous population and sexual propagation with recombination as well as the genetic control of the reproduction accuracy, computer simulated evolution reveals a couple of interesting, sometimes surprising, properties of nature's learning-by-doing algorithm. *Survival of the fittest*, Spencer's résumé of Darwin's view, turns out to be a bad advice if taken literally. Individual death, forgetting, and even regression show up to be necessary ingredients of the life game. Whether the process should be named gradualistic or punctualistic, is a matter of the observer's point of view.

Keywords: Evolutionary algorithms, hillclimbing, survival of the fittest, self-adaptation, internal model of the *environment*, requisite variety, groping in the dark.

1. INTRODUCTION

Evolution can be looked at from a large variety of positions. Beginning with the closest physico-analytic viewpoint, one might focus attention to the molecular and cellular processes. A more distant point of view centers on the behavior of populations and species. Another difference emerges from whether one emphasizes the homeostatic aspect of some kind of adaptation to a given environment, which is more relevant in the short term, or the macroscopic view of a development to the more complex, sometimes called *higher*, in the long term. Not many dare to use the euphemistic word *better* here, and some even do not like at all the term *progress* in connection with organic evolution. But how does such an attitude match with the silent presupposition that homo sapiens sapiens is the crown of creation and imitating a single brain of that species could help in solving most difficult real-world problems?

The instruments used here, will be a macroscope and a time accelerator. Moreover, for methodological reasons, an optimistic point of view will be shared by comparing macro-evolution with iterative optimization, or, more cautiously, with stepwise meliorization techniques, i.e., hill-climbing or ridge-following procedures. By means of a simple algorithmic formulation of the main evolutionary principles, it is possible to reveal some properties of the process that in some cases are striking at the first glance. These findings may not only be helpful for better understanding *natural intelligence* but also be beneficial for global long-term planning and other groping-in-the-dark situations.

2. ORGANIC EVOLUTION

Eigen [1] once modeled the replication of RNA matrices in the primordial soup by means of the following set of nonlinear ordinary differential equations. Let x_i be the concentration of species $i \in [1, n = 4^N]$ with N denoting the number of nucleotide bases (there are 4 different types of them) that form the backbone of an RNA macromolecule. The course over time of the concentrations is most simply described by

$$\dot{x}_i = x_i(A_i Q - E) + \sum_{\substack{j=1 \\ j \neq i}}^{n} w_{ij} x_j, \qquad (1)$$

where the coefficients A_i may be termed fertilities, E the mortality, $Q = (1-p)^N$ the probability that string i is correctly replicated despite mutation rate p per base, and w_{ij} the transition probabilities from sort j to sort i. The sum over all n concentrations being $\sum_{i=1}^{n} x_i = 1$, leads to nonlinear dynamics with all its consequences.

Maynard Smith [2] reduced this system of n differential equations down to just two by considering a best fitted master species (index m) and the averaged mutants (index j) only. Neglecting reverse mutations (thus all improvements!) from type j to type m, he got two simple relations:

$$\dot{x}_m = x_m(A_m Q - E);$$
$$\dot{x}_j = x_j(A_j - E) + x_m A_m(1 - Q). \qquad (2)$$

The system as a whole, called *quasi-species* if $0 \le x_j < 1$, is in equilibrium if $E = x_m A_m + x_j A_j$, because then $\dot{x}_m + \dot{x}_j = 0$. Asking for conditions that yield a stable quasi-species with $\dot{x}_m = 0$ and $\dot{x}_j = 0$, and using the approximation $\ln(1-z) \approx -z$ for $z \ll 1$, one finds the inequality

$$N < \frac{\ln \frac{A_m}{A_j}}{p} = \frac{\ln s}{p}. \qquad (3)$$

Assuming a *selection advantage* (as Maynard Smith calls the ratio of the two reproduction rates) of $s = 3$ and a mutation rate of $p = 0.01$ as observed (at least) for interactions between paired nucleotides, the model above constrains stable RNA lengths to $N < 110$, not enough for encoding enzymes additionally to some proteins. But enzymes are necessary to control the reproduction process, especially to repair defective codons (base triplets encoding an amino acid each according to the redundant genetic code) and thus reduce the environmentally caused pre-mutations to suitable values for larger genomes. The Qβ phage, a very simple virus with $N \approx 4500$, is able to reduce the actual mutation rate to $p \approx 0.0005$, thus operating close to its error bound for $s = 10$. The obvious gap in explaining its emergence has been termed Eigen's paradox [2].

A simple question may be allowed here. Why do people tend to translate the term fitness into relative reproduction rates A_i instead of relative selection rates E_i? If the environment plays the rôle of the selector, and that is how I interpret Darwin's *natural selection* principle, then it tests the newborn individuals and gives them or gives them not a chance to become parents of the next generation. Equal death rates E, but different birth rates A_i, may model what is called mating selection in case of sexual reproduction, but are not adequate for modeling environmental selection. And before introducing sexual propagation one should not omit the earlier achievement of multicellularity. The latter may be looked at as a natural method of avoiding Hamming cliffs in the genotype/phenotype mapping, thus smoothing the fitness landscape. Think of a tissue originating from a mother cell by several consecutive cell divisions. Even if the first cell were genetically either black or white, the tissue could give rise to a grey phenotype by means of somatic mutations occurring during the copy processes. Genetically controlled somatic mutation rates would be able to smooth a transition from genetically black to genetically white, or vice versa, by first increasing the error rate and later, after the corresponding gene has switched, decreasing it again.

Just slightly changing Eq. (1) into

$$\dot{x}_i = x_i(AQ - E_i) + \sum_{\substack{j=1 \\ j \ne i}}^{n} w_{ij} x_j, \qquad (4)$$

i.e., using identical fertilities and different mortalities instead of the opposite, asking the same questions, and imposing the same assumptions as above, yields an at first glance similar result:

$$N < \frac{\ln \frac{E_j}{E_m}}{p} = \frac{\ln \tilde{s}}{p}. \qquad (5)$$

But now it has a different meaning. The fraction $\tilde{s}$, now really resembling a *selection pressure*, may become much larger than the reproduction ratio s. Just imagine two spawners carrying many thousands of eggs. None of the one, but nearly all of the other female fish may become adult and have progeny themselves. The stronger selection now supports longer stable genomes, encoding of enzymes, and thus endogenously adapting the error rate. And, by the way, the model according to Eq. (4) leads away from the often pinpointed tautology within the maxim *survival of the fittest*. The fact of a larger (surviving) progeny of fitter ancestors has become an effect of (environmental) selection instead of a cause.

3. ARTIFICIAL EVOLUTION

At the beginning of the computer era, people expected miracles from the sheer number of iterations that should be possible in the near future to solve even hard problems. Ashby's homeostat [3] was an automaton finding back to a feasible state by a sequence of random trials, uniformly distributed over a given parameter space. Many people have made the mistake of thinking of mutations as such *pure* random trials. A couple of them was malignant. They wanted to show that evolution theory never will be able to explain how *nature* found a way to complex living beings (with $N \approx 10^9$) within only 10^{17} seconds, the age of our globe. Montroll's random walk [4] paradigm, on the other hand, neglects the selection principle of evolution. Both mutation and selection (chance and necessity, or exploration and exploitation) are the first principles, which, of course, have to be programmed properly.

Broadly accepted hereditary evidence has led to saying: *The apple does not fall far off of the tree.* A better model of variations from one generation to the next, at least for multi-cellular individuals – an early evolutionary achievement with no simple genotype/phenotype mapping – may be a normal or Gaussian distribution for phenotypic changements between generations, its maximum being centered at the respective ancestor's position. The rôle of chance in such a model is not explicative, however, but only descriptive. An important question now is the suitable size of the standard deviation σ of the changes, which may be addressed as *mean mutation step size* from one generation to the next. This kind of question arises with all optimization or meliorization procedures.

Modeling the selection principle is far more easy, as it seems first. Some evolution programmers, taking *survival of the fittest* literally, forgot to limit their artificial individuals' life span: According to a given selection criterion, a descendant is rejected if its vitality is less than that of its ancestor, the ancestor otherwise. This scheme may be called a (1+1) or two membered Evolution Strategy (ES in the following). Technically, the evolving entity is a feasible solution to an optimization problem, which corresponds to a vital individual in biology. Rechenberg [5] has derived theoretical results for the convergence velocity of that process in an n-dimensional continuous parameter space, though the first applications used discrete parameters. Most important was his finding that for an endless ridge following situation as well as for a minimum (or maximum) approaching situation the convergence rate is inversely proportional to the number n of decision variables. Distances growing with the square root of n, the number of iterations or generations needed to proceed along the shortest path between two arbitrary points in the search space, increases with $O(n^{3/2})$ only, and not exponentially as in the case of simple Monte-Carlo strategies.

This type of creeping random search strategy (see, e.g., Brooks [6], Schumer and Steiglitz [7], or Rastrigin [8]) was first devised for experimental optimization, where measurement inaccuracies drop out one-variable-at-a-time and gradient-following procedures due to their inability of non-local operation. Some kind of pole-vaulting strategy would be more helpful in case of multimodal, noisy, and fractal landscapes.

Bremermann's *simulated evolution* [9] does not differ so much from Rechenberg's as, e.g., G.E.P. Box's *Evolutionary Operation* EVOP [10] does, an experimental design technique, and the so-called Simplex and Complex strategies of Nelder and Mead [11] and M.J. Box [12] for numerical optimization. Whereas random trials are vividly rejected by G.E.P. Box, he centers several experiments (principally at the same time) in a deterministic way around the position of the current best point in a low-dimensional parameter sub-space. The best of all then is taken as the center of the next trial series. Nelder and Mead, and M.J. Box, using a polyhedron for placing the trials, reject the worst position and find a new one by reflecting the worst with respect to the center of the remaining points of the simplex or complex, thus making use of the knowledge gathered by several (parallel) trial solutions.

The first concept above may be called a $(1+\lambda)$, the latter a $(\mu+1)$ evolutionary scheme (also called *steady-state* or *extinction of the worst* strategy), μ denoting the number of parents, λ the number of children within one generation. More general, therefore, is a $(\mu+\lambda)$ ES with μ ancestors having λ descendants altogether, the μ best of all $\mu+\lambda$ artificial beings becoming parents of the next generation. The fact that individual lives are limited in time, is reflected by the (μ, λ) version, introduced in the early Seventies [22]. Now the μ parents are no longer considered in forming the next generation, thus $\lambda > \mu$, a surplus of births, is a must. It was Malthus' pessimistic consequence of that fact, which Darwin contrasted with his view of *natural selection*, by the way. Of course, the extinctive or non-elitist comma-version ES may loose the so far best position and thus diverge, at least temporarily. Mean convergence in the long run becomes more difficult to prove, but theoretical results about ESs have been steadily improved during the last decade (see, e.g., Bäck [14], Beyer [15], and Rudolph [16]). We now know that the selection pressure, i.e. λ/μ, supports the progress rate only logarithmically. Linear speedup with the population size can be achieved only if the problem to be solved is complex enough, recombination by means of mimicking sexual propagation is used, and the mutation variance σ^2 is optimally adapted. There are missing links, however, especially with respect to the self-adaptation of the mutation strength(s). All observations in the following, therefore, were found by computer simulation only.

It must be mentioned here of course that there exist other Evolutionary Algorithms (EA), i.e., Genetic Algorithms (GA) and Evolutionary Programming (EP), to name at least the most prominent and often applied instantiations. Two observations are interesting, if not striking: Firstly, the basic concepts of ES, GA, and EP have been created independently, more than thirty years ago, and at different places. Secondly, though the founders meanwhile have found together during an ever increasing number of international conferences – Evolutionary Computation now forms one of the pillars of the discipline *Computational Intelligence* [13], the other pillars being artificial Neural Networks and Fuzzy Logic – the underlying models of evolution still differ from each other. EP [17] [18] intends to model the birth and death of species and thus cannot include recombination. Exploration is modeled by Gaussian mutations, and selection is performed by tournaments among random subsets of μ old and μ new species, resembling the competition for a common pool of limited resources, i.e., the *struggle for life*. The evolving entity within GAs [19] [20], on the other hand, is the genome, typically represented by a binary string. The main source of variation is crossover of two parental strings (bit flip mutations are rare additional events to recover single zeros or ones that may get lost in small populations). Exploitation is performed similar to Eigen's RNA model above, i.e., by means of fitness proportional mating selection alone. No additional environmental selection can take place, since λ equals μ in the canonical GA. The current state of affairs is best summarized in the Handbook of Evolutionary Computation [21].

4. SELF-ADAPTATION OF STRATEGY PARAMETERS

As for all optimization techniques, the appropriate step size adjustment is of crucial importance. Rechenberg found that there is a *window* of one decade only for σ, within which the (1+1) ES shows a reasonable convergence velocity. He devised a simple rule for exogenously adjusting a near optimum performance of the process, i.e., to control the success frequency p_s such that it lies in the vicinity of one success among five trials – a clear sign of pole-vaulting. If the observed rate shows up with $p_s > 1/5$, then σ (the same for all dimensions) should be increased, otherwise decreased if $p_s < 1/5$. This advice is good for many but not all situations. Moreover, it does not give any hints to adapt the standard deviations of the phenotypic changements individually. Some may be too large, others too small, at the same time. Only within the multimembered strategy, one can include the step size σ or even different standard deviations σ_i (mutation rates in case of discrete variables) into the set of the individuals' genes and adapt them on-line by mutating and recombining the corresponding σ_i as well as the object variables x_i (for more details see[22]). There is some evidence that besides repair enzymes also mutator genes exist that enhance mutations at some gene loci.

Let us look first at the case of just one common step size σ for changing all x_i. Within a $(1,\lambda)$ ES the correct step size turns out to be even more important than within a $(1+\lambda)$ version. In the first case regression takes place instead of progress when the step size is too large, whereas stagnation is the worst case in the latter. At a first glance, therefore, *survival* of an ancestor (elitist selection) might be a good advice. Simulation results, however, teach the contrary to be true. This is the first surprise.

Fig. 1 demonstrates the difference between a (1+10) and a (1,10) ES when minimizing the function $F_1 = \sum_{i=1}^{n} x_i^2$. The number of variables was taken as large as $n = 30$ in order to avoid improper conclusions. In lower dimensional cases nearly every strategy may achieve good results. One common σ for varying all x_i is changed itself by mutation, i.e., by multiplying the ancestor's value with a random number drawn from a logarithmic normal distribution in order to avoid exogenous drift. The $(1,10)$ strategy with extinctive (some call it truncation) selection turns out to be superior. An explanation for this surprising fact is the following: An individual may, by chance, inherit a good set of x_i together with a σ that is not suitable for the next generation transition.

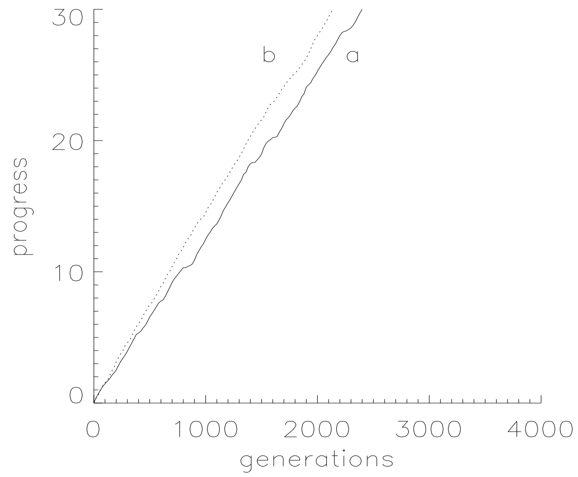

Figure 1: Endogenous learning of the mutation strength (one common standard deviation σ) for function F_1 with $n = 30$ variables.
a) (1+10) Evolution Strategy (ES); b) (1,10) ES.
The *progress* is measured in terms of $\lg(F^{(0)}/F^{(g)})$, where $F^{(0)}$ denotes the start value, $F^{(g)}$ the current value of the objective function at generation g. Displayed is just one typical run for each variant, since averaging over many runs smoothes away the stagnation and recession wiggles.

The $(1+\lambda)$ scheme tends to preserve such genome and thus leads to periods of stagnation. Within a $(1,\lambda)$ ES the good position, occasionally won with an unsuitable step size, is lost, together with the latter, during the next generation. This short term regression, however, enhances the long term velocity of the whole process by a stronger selection with respect to the suitable step size (strategy parameter). Simply speaking: Forgetting is as important as learning, or the first must be seen as a necessary integral part of the latter. One might interpret the fact of an inherent finite life span (preprogrammed maximum number of cell divisions) of living beings as an appropriate measure of nature to overcome the difficulties of undeserved success. Forgetting obsolete *knowledge* is even more crucial in a changing environment, the normal case in nature.

5. COLLECTIVE LEARNING OF PROPER SCALINGS

In many cases it is not sufficient to adapt one common step size for all object variables. For an objective function like $F_2 = \sum_{i=1}^{30} i x_i^2$, for example, individual standard deviations σ_i, appropriately scaled, could speed up the progress rate considerably. To achieve this kind of flexibility within the multimembered evolution strategy, each individual is characterized by a set of n step sizes in addition to the n object variables. They are mutated by multiplication with two random factors, one being common for all step sizes as before, the other acting individually. Thus, a general and specific scalings can be learned at the same time. Operating with a $(1, \lambda)$ strategy – however large λ may be – leads to a second surprise: This kind of process does not work at all, it gets stuck prematurely by approaching a relative optimum in a lower dimensional space. The reason is rather simple: As said above, the convergence rate is inversely proportional to n, the dimension of the search space. Descendants operating in a subspace by sharply reducing some of the step sizes have a short term advantage. Selecting the fittest descendant to become the one and only parent of the next generation, is counterproductive in the long term, as was the survival of the ancestor.

Figures 2 and 3 demonstrate how to overcome the difficulty. If more than one – not only the best, but also inferior – descendants get a chance to have progeny and recombination by sexual propagation takes place, i.e., mixing of the information gathered by different individuals during the course of evolution, then stagnation can be overcome. Now the convergence rate steeply goes up with the population size. On a conventional one-processor computer the parallelism of that scheme cannot be realized, but multi-processor machines and networks of computers operating in parallel are more and more widespread. That is why all figures show the progress over the number of generations, not over sequential computing time.

The overwhelming success of recombination (plus requisite variety) demonstrated here, may explain the early appearance of sexual propagation on earth – in the realm of plants already. But it is unlikely that only the additional variability provokes the success.

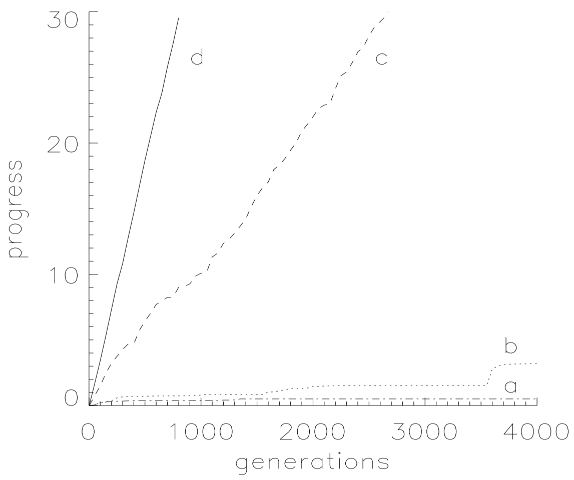

Figure 2: Learning of the scaling I.
a) (1, 10) ES, no recombination,
b) (3, 10) ES with recombination,
c) (6, 30) ES with recombination,
d) (15,100) ES with recombination.

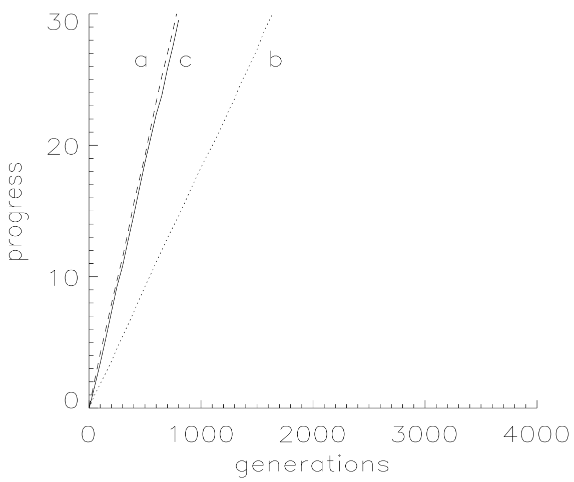

Figure 3: Learning of the scaling II.
a) (1,100) ES, no recombination,
b) (15,100) ES without recombination,
c) (15,100) ES with recombination,
a) and b) with prefixed optimum scalings,
c) with autoadaptive scaling (3c equals 2d).

A better explanation might be the following: The typical situation during the meliorization process is ridge following. Within a population some individuals have a position on one side, others on the other side of the ridge. Mixing genetic information is a means of riding the ridge more efficiently. A similar argument holds for mixing the step sizes: Individuals on one side of the ridge have *internal models* (made up of the set of step size relations) of the response surface that are different from those on the other side. Even if both models are wrong for the long term, since both may be locally adapted only (if the *model* learned is not a law of nature), some *mean* model (or better: hypothesis) may turn out to be more useful for the future. Simply speaking, one may say that *natural intelligence* is distributed.

Now the question of the appropriate selection pressure prevails: How many of the descendants should be selected as new parents. The answer – at least for the objective function chosen – is given by Fig. 4. All other conditions being held constant, including the number of descendants $\lambda = 100$ within one generation, only μ, the number of parents and with them the selection pressure, was changed. Four cases were investigated for function F_2:

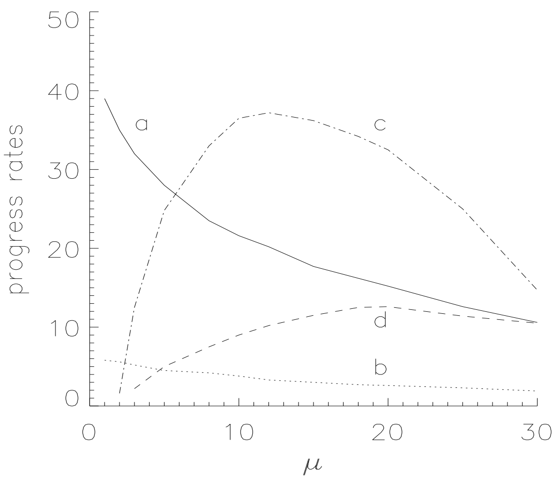

Figure 4: Comparing progress rates per 1000 generations of different $(\mu, 100)$ ESs over μ, the number of parents.
a) Prefixed optimum scaling ($\sigma_i^{(0)} = c/\sqrt{i}$),
b) Prefixed arbitrary scaling ($\sigma_i^{(0)} = \tilde{c}$),
c) Adaptive scaling with recombination,
d) Same as c), but without recombination.

Whereas in both cases a) and b) $\mu = 1$ is the best choice, it is better, even necessary, to increase μ far beyond one in situations where the proper scaling of the variables has to be learned by adapting different mutation strengths (case c). The diagram moreover demonstrates the effectivity of the collective learning process. Under proper conditions nearly the same convergence rate as with total knowledge of the optimum scaling can be achieved – counterintuitively by relying on individuals and their internal models that are by far not the *best* ones. This is the third surprise. Too strong selection pressure not only slows down the meliorization process, it may even lead to divergence (not shown in Fig. 4) under non-elitist selection conditions.

6. LEARNING OF A METRIC AND THE EPIGENETIC APPARATUS

Topologies of vitality response surfaces normally are not as simple as assumed above. The next possible complication is to incline the main axes of the hyper-ellipses which form the subspaces $F = const.$ Now scaling alone does not help in achieving optimum performance. It becomes necessary to change two or more, sometimes even all, variables at the same time (GA researchers have called this awkward situation *epistasis*). What can nature do, what has it done, to overcome the difficulty? In many cases one has found that a single gene influences several phenotypic characteristics (pleiotropy) and vice versa (polygeny). These are the two sides of the same coin, which is correlation, the perhaps best known example of it being allometric growth. The transmission mechanism between genotype and phenotype, called epigenetic apparatus, in a first order may be approximated by linear correlation. In addition to individual step sizes, now correlation coefficients or angles of inclination of an n-dimensional ellipse, forming the surface of constant probability density of a mutation, have to be learned.

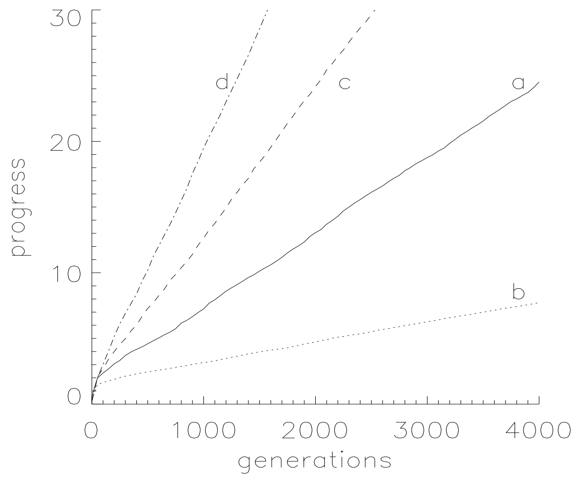

Figure 5: Learning of a metric.
a) (1,100) ES with constant non-optimal scaling,
b) (15,100) ES under same conditions, i.e., without recombination,
c) (15,100) ES with recombination and adaptive individual step sizes,
d) (15,100) ES as before, with additional learning of linear correlations between the phenotypic mutations.

Fig. 5 shows first results for the objective function $F_3 = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j \right)^2$. Four cases were simulated, three of them corresponding to those of Fig. 3. In both cases c) and d), recombination, i.e., averaging both object and strategy parameters of two parents, was used. It is obvious that these sampling conditions bear a variety of possibilities with respect to the mutabilities of step sizes and correlation angles so that simulations c) and d) might not yet represent the best choices. Nevertheless, the results show how much may be gained in terms of progress velocity by on-line learning of a simple *internal model* of the topology of the environment, the *real world* for the simulated evolutionary process. In most cases such a model will not be a correct *theory*, but simply a useful local or temporal hypothesis. A closer look reveals that even none of the better individuals ever learns the correct model of its environment. But if the population size is not too small, they behave as if – collectively: The fourth surprise.

7. GROPING IN THE DARK

Up until now all figures showing the evolutionary progress over time or generations were drawn for objective function values only every 50th generation and for the mean of the population, moreover. If we take a closer look by picking out one of the decision variables, e.g. x_{15} in function F_2, and look at it at every generation (Fig. 6), then the picture reveals more details.

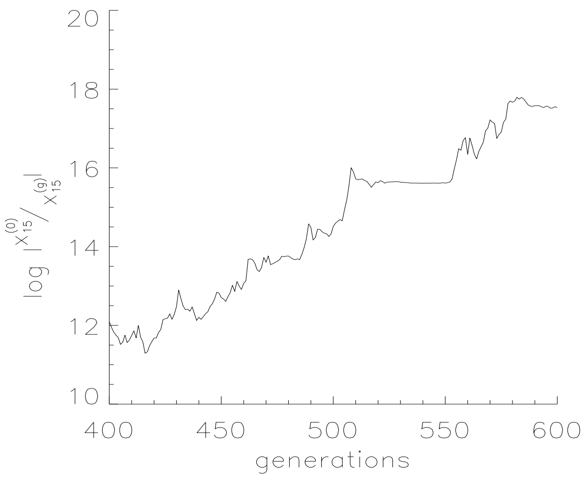

Figure 6: Time cut of one of the variables.

It depends on the density of the historical record whether we may speak of a gradual or a punctualistic process (see, e.g., Stanley [23]). Due to the fact that the objective function depends on many variable attributes, a single one of them must not resemble the progress as a whole. Great success in one direction forgives regression in others – at least in a nonlinear world. And, for sure, looking at some arbitrary cut of the time record, one might get the impression of stochastically disturbed *long waves* with a more or less fixed period. Even more aggregated subobjectives, like the GNP of a nation, could exhibit such behavior, if the underlying process operates left of the maximum of curve c) in Fig. 4. A similar argument holds in case of several competing criteria between which a compromise has to be found. EAs have been applied to such multicriteria optimization problems, too. In nature the situation arises as soon as there are different predators of one prey species.

8. CONCLUSIONS

Many people today, when speaking about long term planning, environmental forecasting, technology assessment etc., are embarrassed by the degree of our ignorance. Very often then they speak of the uncertainties involved. But looking more precisely, isn't it a fact that there are, at the same time and with access to the same data, different certainties, i.e., different interpretations of the past and different aspirations for the future, or, in other words, different *internal models* of the world? In the light of the simulation results above, one should appreciate, not regret, that. Due to the findings of a rather new field of science, i.e. nonlinear dynamics, we must admit that knowledge about the long-term future is principally unavailable for an open dissipative system operating far off of equilibria. We really are groping in the dark [24]. Therefore we should not try to establish one best model of the world, but make the best of the different individual ones of the ridge we are trying to follow without clairvoyance. Even if all of them were wrong they might as well be worthwhile to be recombined with each other. Instead of relying upon too strong competition, which leads to stagnation as we have seen, we better should agree upon cooperative exploration and exploitation.

References

[1] M. Eigen, Self-organization of matter and the evolution of biological macromolecules, *Naturwissenschaften* **58**, 1971, pp. 465-523.

[2] J. Maynard Smith, Models of evolution, *Proc. Royal Soc. London B* **219**, 1983, pp. 315-325.

[3] W.R. Ashby, *Design for a Brain*, New York: Wiley, 2nd edition, 1960.

[4] E.W. Montroll and K.E. Shuler, Dynamics of technological evolution: Random walk model for the research enterprise, *Proc. Natl. Acad. Sci. USA* **76**, 1979, pp. 6030-6034.

[5] I. Rechenberg, *Evolutionsstrategie '94*, Stuttgart: Verlag Frommann-Holzboog, 1994 (enlarged edition of his PhD thesis of 1971).

[6] S.H. Brooks, A discussion of random methods for seeking maxima, *Oper.Res.* **6**, 1958, pp. 244-251.

[7] M.A. Schumer and K. Steiglitz, Adaptive step size random search, *IEEE Trans. AC* **13**, 1968, pp. 270-276.

[8] L.A. Rastrigin, *Sluchainyi poisk v zadachakh optimisatsii mnogoparametricheskikh sistem*, Riga: Zinatne, 1965.

[9] H.J. Bremermann, Numerical optimization procedures derived from biological evolution processes, in: H.L. Oestreicher and D.R. Moore (eds.), *Cybernetic Problems in Bionics*, New York: Gordon and Breach, 1968.

[10] G.E.P. Box and N.R. Draper, *Evolutionary Operation: A Statistical Method for Process Improvement*, New York: Wiley, 1969.

[11] N.A. Nelder and R. Mead, A simplex method for function minimization, *The Computer Journal* **7**, 1965, pp. 308-313.

[12] M.J. Box, A new method of constrained optimization and a comparison with other methods, *The Computer Journal* **8**, 1965, pp. 42-52.

[13] J.M. Zurada et al. (eds.), *Computational Intelligence—Imitating Life*, Piscataway, NJ: IEEE Press, 1994.

[14] T. Bäck, *Evolutionary Algorithms in Theory and Practice*, New York: Oxford University Press, 1996.

[15] H.-G. Beyer, Toward a theory of evolution strategies: On the benefit of sex – the $(\mu/\mu, \lambda)$ theory, *Evolutionary Computation* **3**, 1996, pp. 81-111.

[16] G. Rudolph, *Convergence Properties of Evolutionary Algorithms*, Hamburg: Verlag Dr. Kovač, 1997.

[17] L.J. Fogel et al., *Artificial Intelligence through Simulated Evolution*, New York: Wiley, 1966.

[18] D.B. Fogel, *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, New York: IEEE Press, 1995.

[19] J.H. Holland, *Adaptation in Natural and Artificial Systems*, Ann Arbor, MI: The University of Michigan Press, 1975.

[20] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading, MA: Addison-Wesley, 1989.

[21] T. Bäck et al. (eds.), *Handbook of Evolutionary Computation*, New York: Oxford University Press, 1997.

[22] H.-P. Schwefel, *Evolution and Optimum Seeking*, New York: Wiley, 1995 (enlarged translation of the PhD thesis of 1974/75).

[23] S.M. Stanley, *The New Evolutionary Timetable*, New York: Basic Books, 1981.

[24] D. Meadows, J. Richardson, and G. Bruckmann, *Groping in the Dark—The First Decade of Global Modeling*, Chichester: Wiley, 1982.