Selective Pressure in Evolutionary Algorithms: A Characterization of Selection Mechanisms

Thomas Bäck
University of Dortmund, Department of Computer Science, LS XI
D–44221 Dortmund, Germany
baeck@ls11.informatik.uni-dortmund.de

Abstract— **Due to its independence of the actual search space and its impact on the exploration-exploitation tradeoff, selection is an important operator in any kind of Evolutionary Algorithm. In this paper, all important selection operators are discussed and quantitatively compared with respect to their selective pressure. The comparison clarifies that only a few really different and useful selection operators exist: Proportional selection (in combination with a scaling method), linear ranking, tournament selection, and (μ,λ)-selection (respectively $(\mu+\lambda)$-selection). Their selective pressure increases in the order as they are listed here. The theoretical results are confirmed by an experimental investigation using a Genetic Algorithm with different selection methods on a simple unimodal objective function.**

I. Introduction

Evolutionary Algorithms (EAs) are a class of direct probabilistic search algorithms based on the model of organic evolution. Currently, *Genetic Algorithms* (GAs) [17; 12], *Evolution Strategies* (ESs) [21; 23], and *Evolutionary Programming* (EP) [10; 9] are the most prominent representatives of these algorithms.

In general, all Evolutionary Algorithms are characterized by the fact that they work on a *population* $P = \{a_1, \ldots, a_\lambda\} \in I^\lambda$ of *individuals* $a_i \in I$, each of which represents a search point in the space of potential solutions to a given optimization problem. The population undergoes subsequent modification steps by means of randomized genetic operators that are intended to model *recombination*, *mutation*, and *selection* $s : I^\lambda \rightarrow I^\lambda$. After initialization, the algorithms execute a loop through these operators until a termination criterion is satisfied, and each of these cycles is called a *generation* ($P(t)$ is used here to denote a population at generation t). The optimization problem to which the EA is applied gives a quality information (*fitness*) Φ for the individuals by evaluating each individual as a candidate solution of the *objective function* $f : I \rightarrow \mathbb{R}$. Normally, Φ and f are identical, but some EAs make use of a *scaling* function $\delta : \mathbb{R} \rightarrow \mathbb{R}^+$ that maps objective function values to strictly positive fitness values, i.e. $\Phi = \delta \circ f$. The selection operator uses the fitness information to favor individuals of higher quality to transfer their information to the next generation of the evolution process.

The allocation of responsibilities to the genetic operators is relatively well understood: Mutation introduces innovations into the population, recombination changes the context of already available, useful information, and selection directs the search towards promosing regions of the search space. Acting together, mutation and recombination *explore* the search space while selection *exploits* the information represented within the population. The balance between exploration and exploitation or, in other words, between the creation of diversity and its reduction by focusing on the individuals of higher fitness, is critical in order to achieve a reasonable behavior of EAs in case of complicated optimization problems. The selection operator provides a mechanism to affect this balance towards exploitation — by increasing emphasis on the better individuals — or towards exploration — by providing similar chances to survive even for worse individuals. Informally, the term *selective pressure* is widely used to characterize the strong (high selective pressure) respectively weaker (smaller selective pressure) emphasis of selection on the best individuals.

The selection operator becomes even more important due to the fact that it is completely independent of the remainder of the EA. In particular, any selection operator from an EA A can be transferred to an arbitrary different EA B, regardless of the structure of the search space I. This general feature of selection distinguishes it from the much more specific recombination and mutation operators and characterizes selection as a universal, problem- and algorithm-independent operator. An overview of the variety of different operators developed for search spaces such as binary vectors (GAs), real-valued vectors (ESs and EP), permutations (GAs), and even more complex ones was presented by Michalewicz [20].

Due to the strong impact of selection on the evolutionary search process, it is very desirable to quantify the selective pressure of particular selection mechanisms. Goldberg and Deb started work into this direction by presenting the concept of *takeover time* as a measure of selective pressure [14]. In section II, we will summarize and extend their results in order to analyze and compare the most important selection methods which are nowadays used in EAs. Furthermore, probabilistic selection mechanisms are characterized in section II by the *selection probabilities* $p_i = \mathcal{P}(a_i \in s(P(t)) \mid a_i \in P(t))$ of individuals $a_i \in P(t)$. For the selection probabilities, $\sum_{i=1}^{\lambda} p_i = 1$ is required to hold. Some properties of selection methods seem to be particularly desirable in order to control the characteristics of the search process easily by means of varying the control parameter(s) of selection mechanisms:

(1) The impact of the control parameter(s) on selective pressure should be simple and predictable.
(2) One single control parameter for selective pressure is

This research is supported by grant Schw 361/5-2 from Deutsche Forschungsgemeinschaft (DFG).

0-7803-1899-4/94 $4.00 ©1994 IEEE

preferable.

(3) The range of selective pressure that can be realized by varying the control parameter should be as large as possible.

Unfortunately, some selection methods fail to obey even to the first property, which clarifies the still existing lack of a clear understanding of the effect of parameter variations. Section III presents an experimental investigation of the impact of selective pressure in case of a simple but illustrative search problem to be optimized by a GA. The results, interpreted under consideration of the theoretically predicted selective pressure, can be used to assess the usefulness of selection mechanisms with respect to the properties (1)–(3).

II. Selection Mechanisms

The concept of takeover time quantifies selective pressure by the number of generations required by repeated application of selection — and no other genetic operators — to fill the complete population with copies of the single initially best individual [14]. At the beginning, the population contains exactly one best individual $\hat{a}$ and $\lambda - 1$ worse ones, and selection is iterated until the population consists of λ copies of $\hat{a}$:

Definition 1
Let $P(0)$ denote the initial population of size λ and $\hat{a} \in P(0)$ its best individual. Let $\iota(P(t))$ denote the predicate $\iota(P(t)) = (a_1(t) = a_2(t) = \ldots = a_\lambda(t) = \hat{a})$. Then

$$\tau^* = \min\{t \in I\!N \mid \iota(s^t(P(0))) = \text{true}\}$$

is called *takeover time* of the selection operator $s : I^\lambda \to I^\lambda$.

In order to assure that $\hat{a}$ does not go extinct by chance, which would cause τ^* to become infinite, a guaranteed survival of one copy of $\hat{a}$ is assumed (an 1-elitist selection method according to the terminology used in [1]).

The basic idea of the takeover time concept consists in the assumption that smaller (larger) takeover times correspond with stronger (weaker) selective pressure. The minimal value of τ^* is one and indicates a complete dominance of the best individual, i.e. after a single application of selection the population consists of λ copies of $\hat{a}$. On the other extreme, τ^* becomes infinite as soon as all individuals are assigned identical selection probabilities $p_i = 1/\lambda$, a situation which corresponds to random walk. In the following, the most prominent selection methods are classified according to their absolute and relative positions within this range of selective pressures.

A. Proportional Selection

The *proportional selection* method as introduced to GAs by Holland assigns selection probabilities according to the relative fitness of individuals [17]:

$$p_i = \frac{\Phi(a_i)}{\sum_{j=1}^{\lambda} \Phi(a_j)} \quad . \tag{1}$$

Since the selection probabilities are directly determined by the objective function f, takeover time results are obtainable only with respect to a particular f. Goldberg and Deb succeeded for the assumption $\Phi = f$ and the two one-dimensional functions $f_1(x) = x^c$, $f_2(x) = \exp(c \cdot x)$, and obtained the results $\tau_1^* = c^{-1} \cdot (\lambda \ln \lambda - 1)$ respectively $\tau_2^* = c^{-1} \cdot \lambda \ln \lambda$ [14].

Equation (1), however, fails to work in case of negative objective function values or minimization tasks (if $\Phi = f$ is assumed). For these reasons, a variety of scaling functions δ are nowadays used in combination with proportional selection, including techniques such as *linear static scaling* [16], *linear dynamic scaling* [16], *exponential scaling* [16], *logarithmic scaling* [16], and *sigma truncation* [11; 4]. Here, we focus on linear dynamic scaling, the method which seems to be most widely used. For a minimization task, its functional form $\delta(y_i) = y_i - \min\{f(a_j) \mid a_j \in P(t-\omega)\}$ (maximization: $\delta(y_i) = -y_i + \max\{f(a_j) \mid a_j \in P(t-\omega)\}$) scales objective function values $y_i = f(a_i)$ according to the worst individual that occurred in the population ω generations before; ω is called the *scaling window* (e.g. the setting $\omega = 5$ is often recommended). If we use the abbreviation $\Phi_i = f_i - c$ $(c > 0)$ to denote the scaled fitness of individual a_i for a maximization task, the selection probabilities are given by

$$p_i' = \frac{f_i - c}{S_f - \lambda \cdot c} \quad ,$$

where $S_f = \sum_{j=1}^{\lambda} f_j$. Comparing p_i' to the unscaled selection probabilities $p_i = f_i / S_f$, we obtain the result $p_i' > p_i \Leftrightarrow f_i > \bar{f}$ respectively $p_i' < p_i \Leftrightarrow f_i < \bar{f}$. In other words, linear scaling increases the selective pressure towards better than average individuals at the cost of worse than average individuals. A quantification of this increase, however, is impossible due to the impact of the fitness distribution of the whole population on the selection probabilities.

B. Tournament Selection

The *tournament selection* method as used e.g. in messy GAs [15] selects a single individual by choosing some number $q \geq 1$ of individuals at random from the population and copying the best individual from this q-group to the next generation. The process is repeated λ times to fill the population. A common *tournament size* is $q = 2$ [14]. Assuming (without loss of generality) the individuals to be ordered according to their objective function values ($\Phi = f$) such that $f(a_1) \leq f(a_2) \leq \ldots \leq f(a_\lambda)$ (for a minimization task), the selection probabilities can be derived.

Theorem 1
For q-tournament selection, the selection probabilities are given by

$$p_i = \lambda^{-q} \cdot ((\lambda - i + 1)^q - (\lambda - i)^q) \quad . \tag{2}$$

Proof: The individual a_i survives the tournament, if it is sampled at least once and *only* index values from the set $\{i, i+1, \ldots, \lambda\}$ are sampled at all within the q trials. The number of combinatorial possibilities to realize this is given by the number of q-permutations with repetitions

of indices from $\{i, i+1, \ldots, \lambda\}$ such that index i occurs with a multiplicity from $\{1, 2, \ldots, q\}$ and any index different from i occurs with a multiplicity from $\{0, 1, \ldots, q-1\}$. Following the theory of exponential generating functions, this number is given by the coefficient of the term $x^q/q!$ in the corresponding exponential generating function [6]:

$$(\exp(x)-1)\cdot\exp(x)^{\lambda-i} = 1+\sum_{q=1}^{\infty}((\lambda-i+1)^q-(\lambda-i)^q)\cdot\frac{x^q}{q!}\,,$$

and the coefficient turns into the desired probability p_i by observing that each of the q tournament members has a uniform sampling probability $1/\lambda$. □

For this method, Goldberg and Deb derived the takeover time $\tau_q^* = (\ln\lambda + \ln(\ln\lambda))/\ln q$, an expression which clarifies that selective pressure is much stronger (about a factor of λ) than for proportional selection (except the random walk setting $q = 1$, which corresponds to $p_i = 1/\lambda$).

C. Linear Ranking

The *linear ranking* selection method by Baker uses a linear function to map indices i to selection probabilities p_i [2; 3; 16]. Again, individuals are assumed to be sorted according to fitness. Then, the selection probabilities are given by

$$p_i \;=\; \left(\eta^+ - (\eta^+ - \eta^-)\cdot\frac{i-1}{\lambda-1}\right)/\lambda\,. \qquad (3)$$

The constraint $\sum_{i=1}^{\lambda} p_i = 1$ requires that $1 \le \eta^+ \le 2$ and $\eta^- = 2 - \eta^+$ are fulfilled. The constants η^+ and η^- are called maximum respectively minimum *expected value* and determine the slope of the linear function. Normally, a value of $\eta^+ = 1.1$ is recommended. According to Goldberg and Deb, the takeover time is well approximated by $\tau^*_{\eta^+=2} \approx (\ln\lambda + \ln(\ln\lambda))/\ln 2$ (for $\eta^+ = 2$) respectively $\tau^*_{\eta^+} \approx 2\cdot\ln(\lambda-1)/(\eta^+-1)$ (for $1 < \eta^+ < 2$) [14].

Whitley presented a linear ranking method equivalent to Baker's, which allows for a direct computation of the index $j = i - 1 \in \{0, \ldots, \lambda-1\}$ that designates the selected individual [26]: $j = \lfloor\frac{\lambda}{2(c-1)}\cdot(c-\sqrt{c^2-4(c-1)\chi})\rfloor$, where $\chi \in [0,1]$ denotes a uniform random variable. For $1 < c \le 2$, the method is practically identical to linear ranking with $\eta^+ = c$. This can be derived by considering $\chi = \mathcal{P}(j(\chi) \le x) =: F(x)$ as continuous distribution function (this requires to omit the floor operator), solving for $F(x) = \frac{x}{\lambda}(c - \frac{x}{\lambda}(c-1))$ and taking the derivative with respect to x. This results in the continuous analogue $dF(x)/dx = (c - 2(c-1)\frac{x}{\lambda})/\lambda$ of equation (3).

D. (μ,λ)- and (μ+λ)-Selection

Both (μ,λ)- and $(\mu+\lambda)$-selection [23] are used in ESs, but the (μ,λ)-method is preferred because it facilitates the self-adaptation of strategy parameters (see e.g. [24]). Originally, these methods are designed to reduce an offspring population (created by recombination and mutation) of size $\lambda \ge \mu$ to a new parent population of size μ. While (μ,λ)-selection achieves this by choosing the μ best offspring individuals as parents of the next generation, $(\mu+\lambda)$-selection chooses the μ best out of offspring and old parents, thus guaranteeing a monotonous course of evolution (a μ-elitist selection method according to the terminology used in [1]). Though both methods are completely deterministic and intended to work by changing population sizes, (μ,λ)-selection can be transferred to a probabilistic method with constant population size by assigning equal selection probabilities to the μ best individuals in a population of size λ:

$$p_i \;=\; \begin{cases} 1/\mu & ,\; 1 \le i \le \mu \\ 0 & ,\; \mu < i \le \lambda \end{cases}\,. \qquad (4)$$

Concerning takeover times, the number N_t of copies of the best individual at generation t follows the simple recurrence $N_{t+1} = N_t\cdot\lambda/\mu$ where $N_0 = 1$; i.e. $N_t = (\lambda/\mu)^t$. For $N_{\tau^*} = \lambda$ a takeover time $\tau^*_{(\mu,\lambda)} = \ln\lambda/\ln(\lambda/\mu)$ is obtained.

In case of $(\mu+\lambda)$-selection the recurrence becomes more complicated because the previos parents have to be considered in addition to generation t. This results in a recurrence of the form $N_{t+1} = \frac{\lambda}{\mu}\cdot(N_t + N_{t-1})$, where $N_0 = 1$ and $N_1 = \lambda/\mu$. The recurrence (a generalized Fibonacci sequence) can be solved by standard methods (e.g. [18]) and yields

$$N_t \;=\; (\alpha_1^{t+1} - \alpha_2^{t+1})/\sqrt{\tfrac{\lambda}{\mu}\cdot\left(\tfrac{\lambda}{\mu}+4\right)}\,, \qquad (5)$$

where $\alpha_{1,2} = \frac{\lambda}{2\mu} \pm \frac{1}{2}\cdot(\frac{\lambda}{\mu}\cdot(\frac{\lambda}{\mu}+4))^{1/2}$ are the roots of the characteristic equation $\alpha^2 - \alpha\lambda/\mu - \lambda/\mu = 0$. The closed form (5) does not admit an analytical solution for $\tau^*_{(\mu+\lambda)}$, but numerical evaluations confirm the expectation that τ^* is smaller than for (μ,λ)-selection and approaches the latter for $\lambda/\mu \gg 4$.

E. Other Methods

In the following, we briefly consider four further methods that have gained some interest in EAs, namely EP-selection, nonlinear ranking, Boltzmann tournament selection, and Boltzmann selection.

Evolutionary Programming

The selection operator used in modern variants of EP [8; 9] can be regarded as a hybrid of tournament and $(\lambda+\lambda)$-selection. Working on the union $U(t) = P(t) \cup P'(t)$ of parents $P(t) \in I^\lambda$ and offspring $P'(t) \in I^\lambda$, the method proceeds as follows: For each individual $a_k \in U(t)$ ($k \in \{1, \ldots, 2\lambda\}$), q (tournament size; a common value for EP is $q = 10$) individuals are chosen at random from $U(t)$ and compared to a_k with respect to fitness values. This process results in a score $w_k \in \{0, \ldots, q\}$ that indicates how many of the q sampled individuals are worse than a_k: $w_k = \sum_{j=1}^{q} \mathbf{1}_{\mathbb{R}^+}(\Phi(a_{\chi_j}) - \Phi(a_k))$, where $\chi_j \in \{1, \ldots, 2\lambda\}$ denotes a uniform random variable sampled anew for each value of j and $\mathbf{1}_A$ is the indicator function of a set A ($\mathbf{1}_A(x) = 1$ if $x \in A$ and zero otherwise).

This selection method can be analyzed with respect to its limit behavior for $q \to \infty$. In the following, we assume the 2λ individuals in $U(t)$ to be ordered according to increasing fitness values. Considering a particular fixed individual a_k,

we use $\mathcal{B}_k = (\Phi(a_k) \leq \Phi(a_\chi))$ to denote the event that a randomly sampled individual a_χ performs at least as poor as a_k. For one trial, the probability $p := \mathcal{P}(\mathcal{B}_k)$ is just given by $p = (2\lambda - k + 1)/(2\lambda)$. As q independent trials are performed, the resulting fitness score w_k is binomially distributed, i.e. $\forall k \in \{1, \ldots, 2\lambda\}$:

$$\mathcal{P}(w_k = v) = \binom{q}{v} p^v (1-p)^{q-v} \quad .$$

In addition, the absolute frequencies $H_k(\mathcal{B}_k)$ of $\mathcal{B}_k$ are also random variables, following a binomial distribution with parameters q and p [7], and the same is true for the relative frequencies $h_k = H_k/q$ with expectation $\mathbf{E}(h_k) = p$ and variance $\mathbf{Var}(h_k) = p \cdot (1-p)/q$. As $\lim_{q\to\infty} \mathbf{Var}(h_k) = 0$, the expectations are experimentally realized with increasing precision as q grows. Since $\mathbf{E}(h_1) = 1 > \ldots > \mathbf{E}(h_\lambda) = \frac{\lambda+1}{2\lambda}$ are the λ largest expectations, we have given a proof of the following theorem 2:

THEOREM 2
Let $s_{\{q\}} : I^{2\lambda} \to I^\lambda$ and $s_{(\mu+\lambda)} : I^{\mu+\lambda} \to I^\mu$ denote selection in Evolutionary Programming respectively $(\mu+\lambda)$-selection. Then, for any $P \in I^{2\lambda}$:

$$\mathcal{P}(\lim_{q\to\infty} s_{\{q\}}(P) = s_{(\lambda+\lambda)}(P)) = 1 \; .$$

As justified by this theorem, it makes sense to call this selection method a probabilistic $(\lambda+\lambda)$-selection.

Nonlinear Ranking

Of course, a rank-based selection method not nessessarily has to use a linear function to map ordered indices to selection probabilities. An example is presented by Michalewicz who proposes the expression

$$p_i \quad = \quad c \cdot (1-c)^{i-1} \qquad (6)$$

to calculate selection probabilities (see [20], p. 57). The constant c denotes the selection probability of the best individual. These probabilities, however, fail to sum to unity ($\sum_{i=1}^{\lambda} p_i = 1 - (1-c)^\lambda$). Furthermore, the method is practically identical to tournament selection if $c = 1 - (1 - 1/\lambda)^q$ is chosen, i.e. it does not introduce a qualitatively new selection mechanism. The equivalence with tournament selection can be derived as follows: Using the indicated value of c in equation (6), we obtain $p_i = ((1 - 1/\lambda)^{i-1})^q - ((1-1/\lambda)^i)^q$, which can be approximated by $p_i = (1-(i-1)/\lambda)^q - (1-i/\lambda)^q$ (since $(1-x)^i \approx 1 - ix$ for small x). The latter expression is equivalent to (2). For example, assuming $\lambda = 100$ the value $c = 0.04$ as proposed by Michalewicz corresponds directly with 4-tournament selection. From an algorithmic point of view, the latter method is surely preferable because it does not require to sort the population according to fitness values.

Boltzmann Tournament Selection

This selection method was developed by Goldberg in order to carry over the asymptotic global convergence theory of *Simulated Annealing* (SA) [25] to GAs and to create a new form of niching mechanism that maintains stable subpopulations of similar objective function values [13]. The method works by performing pairwise competitions between individuals that have to differ from each other in objective function values by a threshold amount Θ. In order to select one individual, three such candidates a_1, a_2, a_3 undergo two pairwise competitions, an *anti-acceptance* competition between a_2 and a_3 and an *acceptance* competition between the winner $a_w \in \{a_2, a_3\}$ and a_1. The resulting logistic probabilities $p' = \exp(-f(a_3)/T)/[\exp(-f(a_2)/T) + \exp(-f(a_3)/T)]$ for a_2 to win against a_3 respectively $p = \exp(-f(a_1)/T)/[\exp(-f(a_1)/T) + \exp(-f(a_w)/T)]$ for a_1 to win against a_w achieve the desired Boltzmann distribution according to the model of simulated annealing. The parameter T is analogous to temperature in the physical systems modeled by SA and controls the probability of the worse individual to survive the competition. The necessary gradual reduction of T during the search — the *cooling schedule* — is a critical issue in SA and Boltzmann tournament selection.

Though Goldberg presented the source code for an implementation of this selection method, neither the cooling schedule nor the adjustment of Θ are indicated or obvious [13]. Recently, Mahfoud derived the rather complicated expressions for selection probabilities p_i [19], but asking for takeover times does not make sense due to the intended niching properties of the method. Currently, the impact of T and Θ on selective pressure seems to be at best qualitatively understood, and the method is practically not used except for some experimental purpose.

Boltzmann Selection

Not to be confused with the method discussed in the previous section, Boltzmann selection is a misleading name for yet another scaling method for proportional selection, using a scaling function $\delta(y_i) = \exp(y_i/T)$ [5]. The authors indicate selective pressure to be low (high) when the control parameter T is high (low). The method, however, suffers from the drawback common to all combinations of proportional selection with a scaling function: It incorporates both an uncontrollable component and a controllable one, such that the overall impact on selective pressure is only qualitatively predictable.

III. COMPARISON AND EXPERIMENTAL RESULTS

In this section we focus on the most important of the selection mechanisms discussed so far, namely on (μ^+,λ)-, tournament, linear ranking, and proportional selection. For a population size $\lambda = 100$ and standard control parameter settings ($\mu = 15$, $q = 2$, $\eta^+ = 1.1$) of the rank-based selection methods, figure 1 shows the dependence of selection probabilities on the rank. 10-tournament selection is also added to clarify the nonlinearity of this method.

The corresponding takeover times according to the expressions presented in section II amount to $\tau^*_{(15+100)} = 2.33$, $\tau^*_{(15,100)} = 2.43$, $\tau^*_{q=2} = 8.85$, $\tau^*_{\eta^+=1.1} = 91.9$, and $\tau^* = 461$ for proportional selection. In view of the fact that

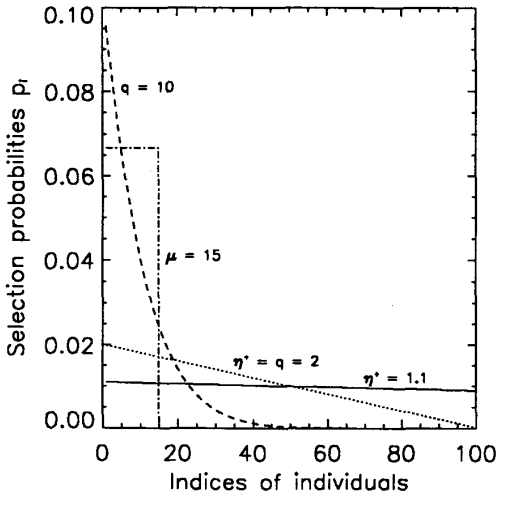

Fig. 1: Graphical visualization of selection probabilities for linear ranking ($\eta^+ = 1.1$ and $\eta^+ = 2$), tournament selection ($q = 2$ and $q = 10$), and (μ,λ)-selection ($\mu = 15$). The population size is $\lambda = 100$.

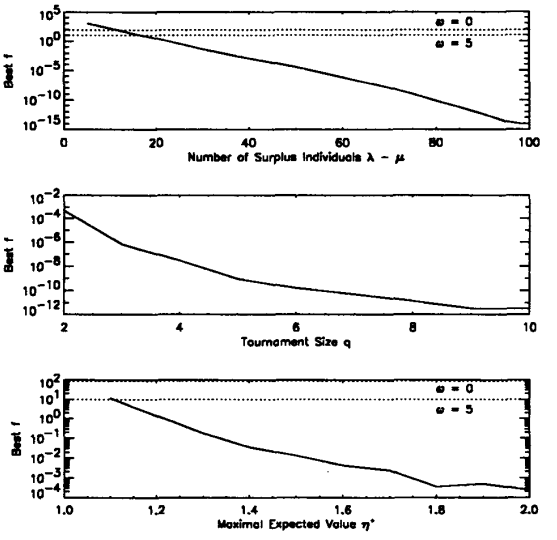

Fig. 2: The effect of varying selective pressure on the sphere model. Upper graphic: (μ,100)-selection with varying values of μ (expressed in terms of $\lambda-\mu$). Middle graphic: Tournament selection with varying values of q. Lower graphic: Linear ranking with varying values of η^+.

linear ranking with $\eta^+ = 1.1$ distributes selection probabilities almost equally among the 100 individuals (see figure 1), it is remarkable how much weaker selective pressure of proportional selection is. Indeed, there is much evidence to argue that scaling functions have been developed mainly in order to make proportional selection practically useful.

In order to check the selective pressure of these methods also experimentally, we performed test runs on the sphere model objective function $f(\vec{x}) = \sum_{i=1}^{n} x_i^2$, $\vec{x} \in I\!R^n$, with a dimension $n = 30$. The experiment uses a GA with 30 bits encoding each object variable in the range $[-40, 60]$ (using a Gray code), a mutation rate $p_m = 0.001$ per bit, a two-point crossover operator with application probability $p_c = 0.6$, and a population size $\lambda = 100$. These settings represent a kind of standard GA if proportional selection with linear dynamic scaling ($\omega = 5$) is added.

For the experimental runs we used (μ,100)-selection with $\mu \in \{1, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95\}$, linear ranking with $\eta^+ \in \{1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}$, tournament selection with $q \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$, and proportional selection with linear dynamic scaling and $\omega \in \{0, 5\}$. For each of these settings, 20 independent runs (but with identical initial population) were performed for 10^5 function evaluations (1000 generations) per run. The average final best objective function values obtained from these experiments are shown in figure 2 as a function of the parameter that determines selective pressure. For proportional selection, the results for both settings of ω are indicated by the dotted lines.

The experimental results provide a good confirmation of the theoretically predicted order of selective pressure. First, one observes that in general the quality of the results improves as selective pressure increases. Proportional selection has weakest selective pressure and yields worst results which improve only slightly when the scaling window grows. A further improvement by up to five orders of magnitude is achieved by linear ranking as η^+ approaches a value of 2.0. The corresponding function value in the vicinity of 10^{-4} is also reached by tournament selection (middle graphic) with $q = 2$ as expected from the theory. By increasing tournament size to a value of $q = 10$, the results improve by eight additional orders of magnitude. Finally, the upper graphic reveals a (1,100)-selection to work optimal on this simple problem, reaching a result of about 10^{-14}. Furthermore, (μ,λ)-selection provides the widest range of selective pressure as indicated by 17 orders of magnitude difference in function values between a (95,100)- and a (1,100)-selection. Comparing the takeover times of (μ,λ)- and tournament selection, we can predict their identity for $\mu = \lambda^{1-\ln q/(\ln\lambda+\ln(\ln\lambda))}$, which results in $\mu \approx 18$ for $q = 10$ respectively $\mu \approx 59$ for $q = 2$ ($\lambda = 100$). By comparing the objective function values in figure 2, these results are well confirmed, and linear ranking can be related to (μ,λ)-selection in a similar way.

IV. Conclusions

Based on the concept of takeover time as a quantitative measure of selective pressure, all important selection mechanisms of EAs are analyzed and compared in this paper. The results clearly demonstrate that selective pressure increases in the order proportional selection, linear ranking, tournament selection (which is essentially just an exponential ranking method), (μ^+,λ)-selection. Except proportional selection and Boltzmann tournament selection, all methods provide a single control parameter (η^+, q, μ) that

has a clearly predictable effect on selective pressure. Various scaling methods developed to be used in combination with proportional selection seem to serve the purpose of artificially creating a minimal amount of selective pressure in order to have proportional selection working. Concerning the range of selective pressure that can be realized by a variation of the control parameter, (μ,λ)-selection turns out to be most flexible.

We must remember, however, that selective pressure is only half of the story. Genotypic diversity that becomes extinct by high selective pressure has to be constantly generated anew by sufficiently powerful genetic operators. It is no accident that in GAs, algorithms that rely mostly on recombination and use mutation only alongside, selection methods with small selective pressure are used preferably, while strong selective pressure was invented in the field of ESs and EP, where mutation dominates the search. A formal quantification of the balance between creation and extinction of information will surely provide an important step towards a deeper understanding of the working principles of EAs.

References

[1] Th. Bäck and F. Hoffmeister. Extended selection mechanisms in genetic algorithms. In R. K. Belew and L. B. Booker, editors, *Proceedings of the 4th International Conference on Genetic Algorithms*, pages 92–99. Morgan Kaufmann Publishers, San Mateo, CA, 1991.

[2] J. E. Baker. Adaptive selection methods for genetic algorithms. In J. J. Grefenstette, editor, *Proceedings of the 1st International Conference on Genetic Algorithms and Their Applications*, pages 101–111. Lawrence Erlbaum Associates, Hillsdale, NJ, 1985.

[3] J. E. Baker. Reducing bias and inefficiency in the selection algorithm. In J. J. Grefenstette, editor, *Proceedings of the 2nd International Conference on Genetic Algorithms and Their Applications*, pages 14–21. Lawrence Erlbaum Associates, Hillsdale, NJ, 1987.

[4] F. Z. Brill, D. E. Brown, and W. N. Martin. Fast genetic selection of features for neural network classifiers. *IEEE Transactions on Neural Networks*, 3(2):324–328, March 1992.

[5] M. de la Maza and B. Tidor. An analysis of selection procedures with particular attention paid to proportional and boltzmann selection. In S. Forrest, editor, *Proceedings of the 5th International Conference on Genetic Algorithms*, pages 124–131. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[6] W. Dörfler. *Mathematik für Informatiker*, volume 1. Carl Hanser, München, 1 edition, 1977.

[7] W. Dörfler. *Mathematik für Informatiker*, volume 2. Carl Hanser, München, 1 edition, 1978.

[8] D. B. Fogel. An analysis of evolutionary programming. In D. B. Fogel and W. Atmar, editors, *Proceedings of the 1st Annual Conference on Evolutionary Programming*, pages 43–51. Evolutionary Programming Society, San Diego, CA, 1992.

[9] D. B. Fogel. *Evolving Artificial Intelligence.* PhD thesis, University of California, San Diego, CA, 1992.

[10] L. J. Fogel, A. J. Owens, and M. J. Walsh. *Artificial Intelligence through Simulated Evolution.* Wiley, New York, 1966.

[11] S. Forrest. *Documentation for Prisoner's Dilemma and Norms Programs that use the Genetic Algorithm.* The University of New Mexico, Albuquerque, NM, 1985.

[12] D. E. Goldberg. *Genetic algorithms in search, optimization and machine learning.* Addison Wesley, Reading, MA, 1989.

[13] D. E. Goldberg. A note on Boltzmann tournament selection for genetic algorithms and population–oriented simulated annealing. *Complex Systems*, 4:445–460, 1990.

[14] D. E. Goldberg and K. Deb. A comparative analysis of selection schemes used in genetic algorithms. In G. J. E. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 69–93. Morgan Kaufmann Publishers, San Mateo, CA, 1991.

[15] D. E. Goldberg, B. Korb, and K. Deb. Messy genetic algorithms: Motivation, analysis, and first results. *Complex Systems*, 3(5):493–530, October 1989.

[16] J. J. Grefenstette and J. E. Baker. How genetic algorithms work: A critical look at implicit parallelism. In Schaffer [22], pages 20–27.

[17] J. H. Holland. *Adaptation in natural and artificial systems.* The University of Michigan Press, Ann Arbor, MI, 1975.

[18] C. L. Liu. *Introduction to Combinatorial Mathematics.* Computer Science Series. McGraw-Hill, New York, 1968.

[19] S. Mahfoud. An analysis of Boltzmann tournament selection. IlliGAL Report 91007, University of Illinois at Urbana-Champaign, IL, 1991.

[20] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs.* Artificial Intelligence. Springer, Berlin, 1992.

[21] I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Frommann–Holzboog, Stuttgart, 1973.

[22] J. D. Schaffer, editor. *Proceedings of the 3rd International Conference on Genetic Algorithms.* Morgan Kaufmann Publishers, San Mateo, CA, 1989.

[23] H.-P. Schwefel. *Numerical Optimization of Computer Models.* Wiley, Chichester, 1981.

[24] H.-P. Schwefel. Natural evolution and collective optimum–seeking. In A. Sydow, editor, *Computational Systems Analysis: Topics and Trends*, pages 5–14. Elsevier, Amsterdam, 1992.

[25] P. J. M. van Laarhoven and E. H. L. Aarts. *Simulated Annealing: Theory and Applications.* Mathematics and Its Applications. Kluwer, Dordrecht, 1987.

[26] D. Whitley. The GENITOR algorithm and selection pressure: Why rank–based allocation of reproductive trials is best. In Schaffer [22], pages 116–121.