Research Article

Efficient Use of Variation in Evolutionary Optimization

John W. Pepper

Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501, USA

Correspondence should be addressed to John W. Pepper, jpepper@santafe.edu

Evolutionary algorithms face a fundamental trade-off between exploration and exploitation. Rapid performance improvement tends to be accompanied by a rapid loss of diversity from the population of potential solutions, causing premature convergence on local rather than global optima. However, the rate at which diversity is lost from a population is not simply a function of the strength of selection but also its efficiency, or rate of performance improvement relative to loss of variation. Selection efficiency can be quantified as the linear correlation between objective performance and reproduction. Commonly used selection algorithms contain several sources of inefficiency, some of which are easily avoided and others of which are not. Selection algorithms based on continuously varying generation time instead of discretely varying number of offspring can approach the theoretical limit on the efficient use of population diversity.

1. Introduction

"Premature convergence", or the loss of diversity before a satisfactory solution is found, is a persistent problem in evolutionary optimization [1]. This reflects the fundamental trade-off between exploration and exploitation, or between thoroughness and speed in evolutionary search [2]. If selection is too weak, progress is slow and many generations are required to find a solution. On the other hand, if selection is too strong, the population rapidly loses diversity and may become stranded on a local fitness peak. A wide variety of techniques have been proposed to address this problem, but it has generally been approached on an *ad hoc* empirical basis, and little theory has been available to guide the design of selection algorithms.

While the trade-off between improving performance and preserving diversity cannot be avoided, it can be ameliorated through the efficient use of variation. Diversity within a population acts as the fuel of the selection process: it is required for selection to act, but is itself consumed in the process. However, selection algorithms differ not only in speed, but also in "fuel efficiency", or rate of improvement relative to loss of variation. In the following sections, I develop a method for quantifying the efficiency of fitness functions, defined here as mappings from objective performance to reproduction. (Such mappings are sometimes referred to as "selection methods".) The approach is based on the powerful formalism from evolutionary biology known as the "Price equation", which is increasingly used in evolutionary genetics [3]. I next compare several widely-used selection methods to characterize their sources of inefficiency, and to illustrate the advantages of more efficient selection. I also consider whether less efficient algorithms have any offsetting advantages that justify their use. Finally, I discuss the design of fast and efficient fitness functions, and propose a new kind of algorithm, based on varying generation time instead of number of offspring, which can approach perfect efficiency in the use of genetic variation.

2. Quantifying Selection Efficiency

The ultimate goal of evolutionary optimization is to maximize some objective measure of performance on a given task. Here I measure progress toward optimization in terms of the mean performance level of the population (In evolutionary computation applications, the ultimate interest may be in the highest performance level in a population of candidate solutions, rather than the mean. However, mathematical theory is only available to quantify change in population mean through selection rather than change in population maximum. As a practical matter, maximizing mean

performance will also maximize best performance, all else being equal). The goal of improving performance conflicts partially with a subsidiary goal: maintaining the diverse population of candidate solutions or "individuals" needed to thoroughly explore search spaces and find the best possible solutions. The conflict arises because the unequal reproduction that drives improvement in average performance also reduce population diversity. Unequal contributions to the next generation's gene pool by different individuals always reduces diversity except in the special case of negative frequency-dependent selection (which increases diversity). If selection is frequency-independent, unequal reproduction reduces diversity, in direct proportion to the reproductive variance among individuals (see the appendix).

Although selection cannot improve a population's average performance in the next generation without unequal reproduction, the converse is not true. Unequal reproduction and resulting loss of diversity need not improve average performance. Variance in reproduction that is uncorrelated with performance can reduce genetic diversity (though genetic drift) just as quickly as can effective selection, but without increasing mean performance. Because correlation between performance and reproduction is what makes selection effective at optimization, I focus on the strength of this correlation to quantify the efficiency of fitness functions.

In addition to selection, genetic operators such as mutation and recombination can also change a population's mean performance (although in an unpredictable direction). Here I focus exclusively on the effects of selection, or differential reproduction, because this is the source of premature convergence in evolutionary optimization. Let each individual in the population (indexed by i) have a measured performance level p_i. The average population performance before selection is $\overline{p} = \sum p_i/N$, where $N = $ population size. After one generation of selection, average population performance will be the average of the parent performances weighted by the contribution of each parent to the next generation: $\overline{p}' = \sum p_i w_i / \sum w_i$, where $w_i = $ the number of offspring produced by the i'th individual. (Note that this assumes perfect heritability of performance from parent to offspring.) To simplify the notation, it is convenient to replace absolute reproduction w_i with relative reproduction, $\widetilde{w}_i = w_i/\overline{w}$, so that mean performance in the offspring generation is $\overline{p}' = \text{ave}(p_i\widetilde{w}_i)$. The change in average performance caused by one round of selection is then $\Delta\overline{p} = \overline{p}' - \overline{p}$, or

$$\Delta\overline{p} = \text{ave}(p\widetilde{w}) - \text{ave}(p). \tag{1}$$

As a result of selection, performance improvement across one generation is exactly $\Delta\overline{p}$ above. We can rewrite (1) in a useful form by using two identities: firstly, $\text{ave}(p\widetilde{w}) = \text{ave}(p) \cdot \text{ave}(\widetilde{w}) + \text{cov}(p\widetilde{w})$, where "cov" represents covariance. Secondly, $\text{ave}(\widetilde{w}) = 1$ by definition. With these substitutions, the improvement in performance from parent to offspring generation is

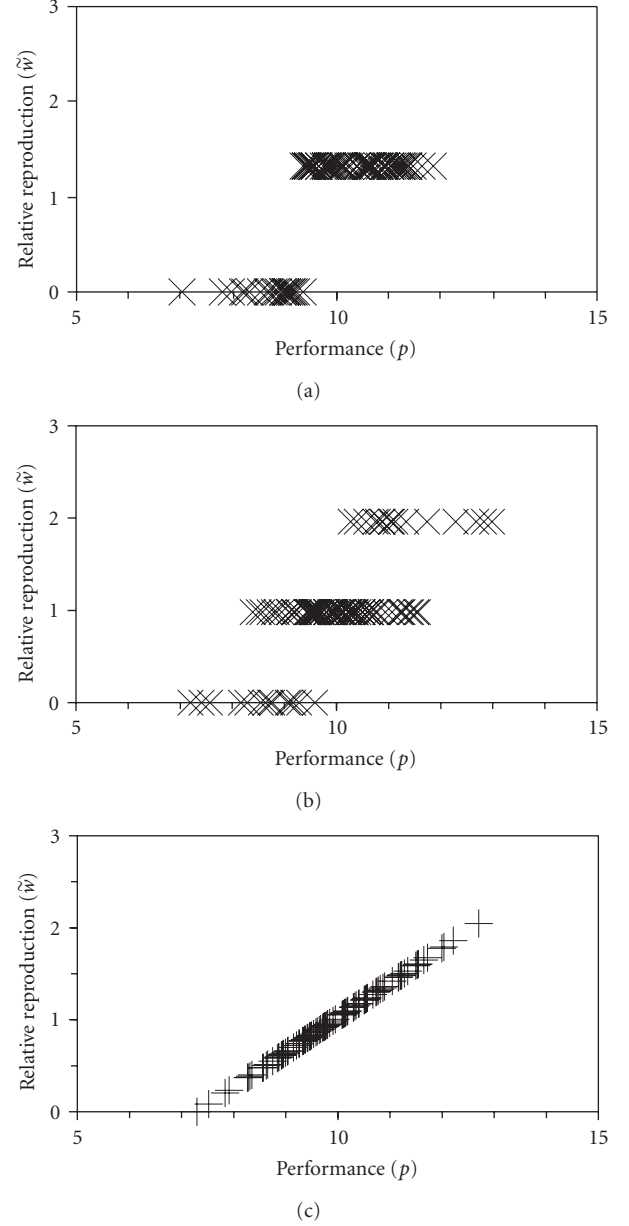$$\Delta\overline{p} = \text{cov}(p, \widetilde{w}) \tag{2}$$

(a)

(b)

(c)

FIGURE 1: Three fitness functions illustrated using the same set of 100 simulated individuals with performance values drawn from a normal distribution with mean = 10 and standard deviation = 1. (a) threshold selection (b) stochastic proportionate selection (SPS), (c) deterministic proportionate selection (using (8)). Each mark represents one individual.

(see [4]). To highlight the factors affecting optimization rate, it is useful to use another identity to rewrite this covariance as a product of its three factors:

$$\Delta\overline{p} = \sigma_p \cdot \sigma_{\widetilde{w}} \cdot \rho_{p\widetilde{w}}, \tag{3}$$

where σ is a standard deviation among individuals in performance (p) or relative reproduction ($\widetilde{w}$), and $\rho_{p\widetilde{w}}$ is the linear correlation coefficient between the two [4].

Equation (3) provides insight into how to maximize selection efficiency, or the ratio of performance improvement
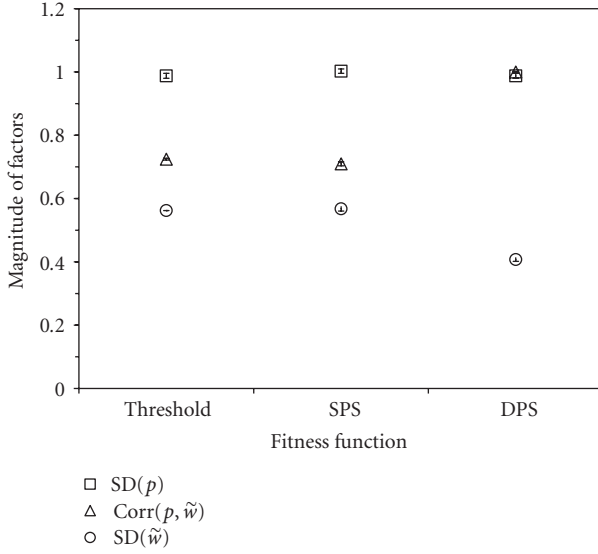
FIGURE 2: The three factors contributing to performance improvement compared over 1 round of selection across three fitness functions using numerical simulations: threshold selection, stochastic proportionate selection (SPS), and deterministic proportionate selection (DPS). Each sample consisted of 100 simulated individuals with performance values drawn from a normal distribution with mean = 100, SD = 1. Markers show means, and bars show ± standard error over 100 samples. (Note that error bars are too small to extend beyond marker symbols.)
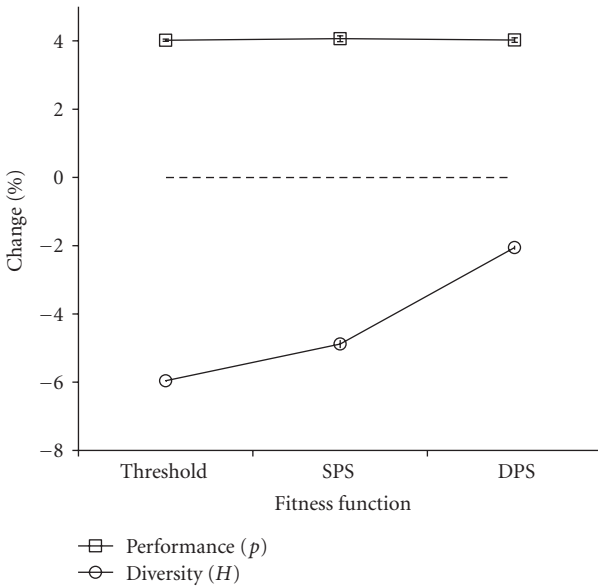

FIGURE 3: The same three fitness functions shown in Figure 2 compared for the one-generation change produced in mean performance, and in population diversity. Bars show standard errors. (Note that error bars are too small to extend beyond marker symbols).

to loss of diversity. Deviation in individual performance (σ_p) is fixed for a given population, but $\sigma_{\widetilde{w}}$ and $\rho_{p\widetilde{w}}$ depend on the selection method. Deviation in reproduction ($\sigma_{\widetilde{w}}$) varies with the strength of selection. Increasing $\sigma_{\widetilde{w}}$ can increase performance improvement, but at the cost of faster loss of diversity. The linear correlation between performance and reproduction ($\rho_{p\widetilde{w}}$) corresponds to the efficiency of selection, in the sense that increasing this term increases performance improvement *without* increasing loss of diversity and performance variation. When $\rho_{p\widetilde{w}} = 0$, selection is completely inefficient: it consumes variation without improving average performance. In the language of evolutionary theory, this is termed "drift" instead of "selection". At the other extreme of $\rho_{p\widetilde{w}} = 1$, the ratio of performance increase to variance reduction is maximized. Thus the rate at which variation is lost from a population is not simply a function of selection strength ($\sigma_{\widetilde{w}}$), as is sometimes assumed, but also of selection efficiency ($\rho_{p\widetilde{w}}$).

3. Sources of Inefficiency in Fitness Functions

The perfectly linear fitness function ($\rho_{p\widetilde{w}} = 1$) is an ideal of efficiency that is not realized by any algorithm in general use. All standard fitness functions depart from linear correlation either through deterministic nonlinearities, fluctuating stochastic nonlinearities, or both. An example of a deterministically nonlinear fitness function is threshold selection, in which reproduction is an all-or-nothing step function of performance (Figure 1(a)). Any such highly nonlinear fitness function will necessarily have a linear correlation well below 1. Fitness functions without any deterministic nonlinearity are termed "fitness-proportionate selection" because expected reproduction is directly proportional to performance [1]. However, these functions introduce fluctuating stochastic nonlinearity in converting expected to actual reproduction, so that expected reproduction has perfect linear correlation with performance, but actual reproduction does not. This is hard to avoid because unlike the expected number of offspring, the actual number of offspring is constrained to whole numbers and so must vary stochastically around the expected number. For example, the commonly used "stochastic universal sampling" algorithm [5] works as follows: an expected reproduction of ω is partitioned into a fractional portion ($\omega\%1$) and a whole-number portion [$\omega-(\omega\%1)$], where % is the modulo operator. The algorithm produces the whole number of offspring, plus one additional offspring with a fractional probability of ($\omega\%1$). Despite its lack of deterministic nonlinearity, the correlation between performance and actual number of offspring is less than 1 because of stochastic fluctuations (e.g., Figure 1(b), where $\omega = 1$ for each individual, but w varies stochastically). I will refer to this algorithm as "stochastic proportionate selection" (SPS).

Such stochastic fluctuations in actual reproduction are larger in other implementations of fitness-proportionate selection, such as "roulette wheel" sampling [2]. Still other algorithms, such as tournament selection [1], include both deterministic and stochastic sources of nonlinearity. Here the

selection of a pair of individuals to compare is stochastic, while the choice of which of the two reproduces depends on their relative performance rank, which is a deterministic nonlinear function of performance. Both deterministic and stochastic nonlinearities in fitness functions reduce the correlation between performance and actual reproduction, and thereby reduce selection efficiency.

To examine the effect of selection efficiency on diversity, I used a numerical simulation consisting of a population of 100 individuals (candidate solutions) with performance values drawn from a normal distribution with mean = 10 and standard deviation = 1. I compared the effects of a single round of selection using threshold selection (Figure 1(a)), stochastic proportionate selection (SPS) (Figure 1(b)), and deterministic proportionate selection (DPS) ((8), Figure 1(c)). The numerical simulation allowed fractional offspring, but the problem of how deterministic proportionate selection can be implemented with whole numbers of individuals is deferred to Section 6 below. To tune the threshold fitness function to give the same performance improvement as the other two functions, I allowed reproduction only by the best-performing 76% of the population. Deterministic proportionate selection generated less variance in reproduction than the other two, but reproduction was more highly correlated with performance (Figure 2). These two differences resulted in an equal performance increase in the offspring generation for all three fitness functions (Figure 3). Thus the deterministic proportionate selection function consumed less performance variation while producing the same performance improvement. I next investigated whether DPS also preserved more genotype diversity while producing the same performance improvement.

To quantify diversity, I used the Shannon-Weiner diversity index from evolutionary biology, which is equivalent to the entropy of the genotypes in the population:

$$H = -\sum_g f_g \log_2 f_g, \qquad (4)$$

where g indexes the genotypes in the population, and f_g is the population frequency of genotype g. Entropy is maximized when each individual is unique, and minimized when all individuals share the same genotype. To simplify calculations I assumed that each individual in the population was unique prior to selection, but violating this assumption would not change the outcome qualitatively. Selection reduced diversity several-fold less under the deterministic proportionate function than under either the stochastic proportionate or threshold functions, while improving performance at the same rate (Figure 3).

4. Is Inefficient Selection Ever Useful?

I have focused here on the advantages of linear fitness functions for conserving genetic diversity. However, both deterministic nonlinearities and stochastic effects have some potential advantages. Might these justify the use of nonlinear fitness functions despite their lower efficiency?

Deterministically nonlinear fitness functions permit stronger selection (higher $\sigma_{\tilde{w}}$) than linear functions. At the extreme, reproduction by only the individual(s) with the highest performance increases average performance by $\Delta \bar{p} = p_{\max} - \bar{p}$. More generally, larger one-generation improvements are possible with nonlinear than with linear fitness functions. However, this rapid short-term improvement comes at the cost of the variation required for longer-term improvement. Genetic variation could be created anew in each generation, but this is computationally expensive and reduces evolutionary search algorithms to inefficient hill-climbers. For this reason, deterministic nonlinearity in fitness functions is unlikely to be helpful in most applications.

Stochastic fitness functions offer a different potential advantage by helping populations escape from local performance peaks. Slightly deleterious mutations can persist or spread under stochastic selection, making it possible for populations to cross low-performance fitness valleys requiring multiple mutations. Stochastic effects also allow the population to drift among different genotypes with equal performance. This may facilitate the exploration of "neutral networks" in genotype space, leading to the discovery of higher performance peaks [6]. However, stochastic effects on reproduction also have drawbacks. They can push populations away from global as well as local peaks. In some algorithms, they may also slow the discovery of higher-performance peaks by allowing beneficial new mutations to be lost. It remains an open question how often stochastic fitness functions improve evolutionary optimization, and how much stochasticity is desirable. To investigate these questions, it will help to have algorithms in which stochastic effects can be directly controlled by the experimenter rather than being a by-product of the particular algorithm used. This is easily achieved by adding a stochastic term to a deterministic linear fitness function. This approach has the additional advantage that stochastic effects can be reduced to any desired magnitude without incurring a computational cost. In contrast, intrinsically stochastic algorithms require very large population sizes to drive stochastic effects to low levels.

5. Fast and Efficient Fitness Functions

How can a fitness function be designed to maximize the rate of performance increase while also optimizing efficiency? Efficiency defined as the linear correlation $\rho_{p\tilde{w}}$ is maximized when reproduction is a linear function of performance. It is convenient to represent such fitness functions in the standard linear form:

$$w_i = a(p_i + b), \qquad (5)$$

where p_i and w_i are individual performance and reproduction, respectively, and a and b are system parameters. With discrete generations, it is usually desirable to maintain a stable population size across generations, which constrains

the average number of offspring per individual ($\overline{w}$) to 1. This constrains the value of a to

$$a = \frac{1}{\text{ave}(p_i + b)} = \frac{1}{\overline{p} + b}. \tag{6}$$

Substituting (6) into (5) gives us a linear fitness function yielding a stable population size:

$$w_i = \frac{(p_i + b)}{(\overline{p} + b)}. \tag{7}$$

What value of b will maximize the rate of performance improvement? Recall from (3) that the one-generation improvement in average performance due to selection is a product of three quantities: σ_p, $\rho_{p\widetilde{w}}$, and $\sigma_{\widetilde{w}}$. The first of these is a fixed property of the population. The second is already maximized at 1 under linear fitness functions. This leaves only variance in individual reproduction $\sigma_{\widetilde{w}}$ to be maximized in order to maximize the performance improvement Δp. When w_i is a linear function of p_i, its variance σ_w is maximized by maximizing the slope of the fitness function, which is defined in (5) as a. Equation (6) shows that a increases as b approaches $-\overline{p}$, so that b should be as close as possible to $-\overline{p}$ to maximize improvement. However, there is a constraint that individual reproduction (w_i) cannot be negative, which means that $b \geq -p_i$ for all i (5). If the worst performance in the population is denoted as $p_{\min}$, then the lowest possible value for b is $-p_{\min}$, which results in the individual(s) with the lowest performance having exactly zero offspring. Substituting this value for b into (7) yields the stable linear fitness function with the maximum rate of performance increase:

$$w_i = \frac{(p_i - p_{\min})}{(\overline{p} - p_{\min})}. \tag{8}$$

6. A Variable-Generation Algorithm for Efficient Selection

If a deterministic linear fitness function is the theoretical ideal, how can it be implemented in practice? As discussed above, inefficiency in commonly used fitness functions arises in part from easily avoidable sources of nonlinearity. However, all standard algorithms also contain nonlinearities arising from the fact that performance is a continuous variable, while the number of offspring is discrete. Stochastically converting real numbers of expected offspring to whole numbers of actual offspring reduces the linear correlation between performance and actual reproduction.

We can overcome this problem by recognizing that selection on genotypes acts through their rate of reproduction per unit time. Instead of varying the number of offspring, one can independently vary the generation time for each individual [7]. This requires an algorithm incorporating overlapping generations and a continuous representation of time. Individual reproductive rates can then vary continuously rather than discretely, and can correlate perfectly with individual performance.

To implement this idea, individual reproduction is treated as a growth rate, by analogy with population growth rates. A population growth rate tells us how large a population will be after a given time:

$$s_t = s_0 w^t, \tag{9}$$

where s_0 is initial population size, s_t is population size after t time units, and w is growth rate. Rearranging (9) tells us how long it will take the population size to change by a given factor s_t/s_0 under a given growth rate w:

$$t = \frac{\ln(s_t/s_0)}{\ln(w)}. \tag{10}$$

Our current problem concerns individuals rather than populations, but we can use the same reasoning to ask how long it will take an individual to die (equivalent to shrinking to size zero) or reproduce (equivalent to doubling in size) as a function of its individual growth rate w_i. Because individuals are discrete, we round off individual "size" to the nearest whole number. Thus for $w_i < 1$, we can ask how long it will take for the individual to fall below half its initial size, given its negative growth rate. At this point, the individual's size is closer to zero than one, and we recognize this by removing it from the population. Similarly, if an individual's growth rate is greater than one, we ask how long it will take for its size to rise above 1.5. At this point it is closer to being two individuals than one, and we recognize this by doubling it via reproduction. (Note that unlike rounding the number of offspring under stochastic fitness-proportionate algorithms, rounding individual size to whole numbers is not stochastic and does not introduce stochastic nonlinearity into the fitness function. Because waiting times vary continuously, genotype growth rates also vary continuously as a deterministic linear function of performance.)

For $w < 1$, waiting time to death is found by substituting 0.5 for s_t/s_0 in (10), giving

$$t_d = \frac{-0.693}{\ln(w)}. \tag{11}$$

For $w > 1$, waiting time to reproduction is found by substituting 1.5 for s_t/s_0, giving:

$$t_r = \frac{0.405}{\ln(w)}. \tag{12}$$

When an individual's reproductive rate is evaluated, its future death or reproduction is scheduled for a time point in the future designated as a real number on a time line. These events will be scheduled in the distant future when the reproductive rate is close to 1, and in the near future when it is far from 1 (Figure 4).

At the beginning of a run, each individual's performance is evaluated and its reproduction or death is scheduled. After this, the algorithm simply consists of repeatedly cycling through the following steps: (1) carry out the first event on the schedule. (2) If the event was a birth, evaluate the new individual's performance. (3) recalculate all waiting times
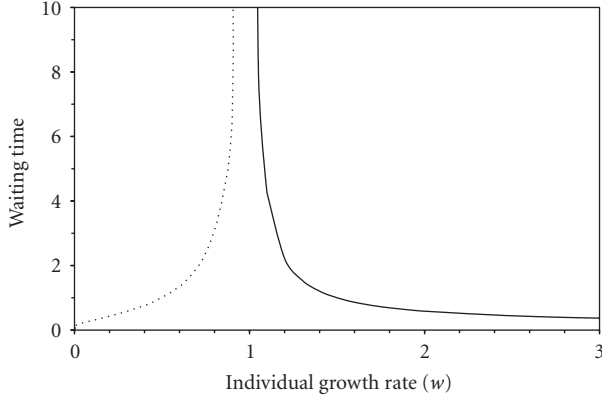
FIGURE 4: Waiting time to death (dotted line) or reproduction (solid line) as a function of individual growth rate. (From (11) and (12).)

to reflect the new average performance, and update the schedule. In practice, it might be useful to recalculate waiting times less often in order to reduce the computational load. For example, each individual's waiting time could be calculated at birth and then not recalculated until its scheduled event was within some specified time horizon.

7. Conclusions

In these results, truly linear fitness functions, in the form of deterministic proportionate selection, reduced population diversity and performance variation less than other fitness functions that improve performance the same amount in one round of selection. This strongly suggests that over multiple generations, the same rate of performance improvement would be sustained with less loss of diversity. Consequently, DPS should yield better solutions, particularly for tasks where premature convergence is otherwise a problem. The variable-generation algorithm outlined above allows actual reproductive rates to be exactly proportional to performance, providing one way to implement DPS. Although stochastic fitness functions may eventually prove useful on some fitness landscapes, intrinsically linear fitness functions provide the best foundation for designing them because they allow stochastic terms to be added in a controlled fashion.

One important caveat is that these conclusions are based on consideration of a single round of selection in isolation. Longer-term selection is also affected by the genetic operators that create variation, such as mutation and recombination, and by their interactions with selection. In particular, this paper does not address the issue of how selection interacts with recombination among epistatic loci (e.g., [8]). While I am not aware of any reason the conclusions reached here would not hold in the broader context of long-term evolution with recombination; this remains to be investigated.

Appendix

The purpose of this appendix is to quantify the extent to which unequal reproduction reduces diversity in a population. It will show that when selection is frequency-independent, unequal reproduction reduces diversity in direct proportion to the reproductive variance among individuals. I follow Section 3 above in quantifying diversity with the Shannon-Wiener diversity index, which is equivalent to the entropy of genotypes.

Before selection, variance in the frequencies of alternative genotypes is

$$\text{var}(f) = \text{E}(f^2) - \text{E}(f)^2, \tag{A.1}$$

and after one round of selection it is

$$\text{var}(\widetilde{w}f) = \text{E}(\widetilde{w}^2 f^2) - \text{E}(\widetilde{w}f)^2, \tag{A.2}$$

where variance (var) and expectation (E) operate across genotypes, f is the frequency of each genotype, and $\widetilde{w}$ is the reproduction of each genotype relative to the population mean. If selection is frequency-independent, then w and f are independent, so that (A.2) can be rewritten as

$$\text{var}(\widetilde{w}f) = [\text{E}(\widetilde{w}^2) \cdot \text{E}(f^2)] - \left(\text{E}(\widetilde{w})^2 \cdot \text{E}(f)^2\right). \tag{A.3}$$

Because $\text{E}(\widetilde{w}) = 1$ by definition, (A.3) simplified to

$$\text{var}(\widetilde{w}f) = \text{E}(\widetilde{w}^2) \cdot \text{E}(f^2) - \text{E}(f)^2. \tag{A.4}$$

Let $\Delta\text{var}(f)$ represent the change in var(f) caused by one round of selection. Subtracting (A.1) from (A.4) gives

$$\Delta\text{var}(f) = \text{E}(f^2) \cdot [\text{E}(\widetilde{w}^2) - 1]. \tag{A.5}$$

Because $\text{E}(\widetilde{w}) = 1$, the second term on the right, $\text{E}(\widetilde{w}^2) - 1 = \text{E}(\widetilde{w}^2) - [\text{E}(\widetilde{w})]^2 = \text{var}(\widetilde{w})$. Substituting var$(\widetilde{w})$ for $\text{E}(\widetilde{w}^2) - 1$ gives

$$\Delta\text{var}(f) = \text{E}(f^2) \cdot \text{var}(\widetilde{w}). \tag{A.6}$$

Thus the decrease in the variance of genotype frequencies is proportional to the variance in reproduction. Thus minimizing variance in reproduction also minimizes loss of diversity (H).

Acknowledgments

References

[1] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, UK, 1996.

[2] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Mich, USA, 1975.

[3] S. A. Frank, "George Price's contributions to evolutionary genetics," *Journal of Theoretical Biology*, vol. 175, no. 3, pp. 373–388, 1995.

[4] G. R. Price, "Selection and covariance," *Nature*, vol. 227, pp. 520–521, 1970.

[5] J. E. R. Baker, "Reducing bias and inefficiency in the selection algorithm," in *Proceedings of the 2nd International Conference on Genetic Algorithms and Their Applications*, J. J. Grefenstette, et al., Ed., pp. 14–21, Erlbaum Associates, Hillsdale, NJ, USA, 1987.

[6] W. Fontana and P. Schuster, "Continuity in evolution: on the nature of transitions," *Science*, vol. 280, no. 5368, pp. 1451–1455, 1998.

[7] J. H. Holland, "Building blocks, cohort genetic algorithms, and hyperplane-defined functions," *Evolutionary Computation*, vol. 8, no. 4, pp. 373–391, 2000.

[8] J. W. Pepper, "The evolution of evolvability in genetic linkage patterns," *BioSystems*, vol. 69, pp. 115–126, 2003.