The Schema Theorem and Price's Theorem

Lee Altenberg
Institute of Statistics and Decision Sciences
Duke University, Durham, NC, USA 27708-0251
Internet: `altenber@acpub.duke.edu`

Abstract

Holland's Schema Theorem is widely taken to be the foundation for explanations of the power of genetic algorithms (GAs). Yet some dissent has been expressed as to its implications. Here, dissenting arguments are reviewed and elaborated upon, explaining why the Schema Theorem has no implications for how well a GA is performing. Interpretations of the Schema Theorem have implicitly assumed that a correlation exists between parent and offspring fitnesses, and this assumption is made explicit in results based on Price's Covariance and Selection Theorem. Schemata do not play a part in the performance theorems derived for representations and operators in general. However, schemata re-emerge when recombination operators are used. Using Geiringer's recombination distribution representation of recombination operators, a "missing" schema theorem is derived which makes explicit the intuition for when a GA should perform well. Finally, the method of "adaptive landscape" analysis is examined and counterexamples offered to the commonly used correlation statistic. Instead, an alternative statistic — the transmission function in the fitness domain — is proposed as the optimal statistic for estimating GA performance from limited samples.

1 INTRODUCTION

Although it is generally stated that the Schema Theorem (Holland, 1975) explains the power of genetic algorithms (GAs), dissent to this view has been expressed a number of times (Grefenstette and Baker 1989, Mühlenbein 1991, Radcliffe 1992). Mühlenbein points out that "the Schema Theorem is almost a tautology, only describing proportional selection," and that "the question of why the genetic algorithm builds better and better substrings by crossing-over is ignored." Radcliffe points out that

1. The Schema Theorem holds even with random representations, which cannot be expected to perform better than random search, whereas it has been used to claim that GAs perform better than random search;
2. The Schema Theorem holds even when the schemata defined by a representation may not capture the properties that determine fitness; and
3. The Schema Theorem extends to arbitrary subsets of the search space regardless of the kind of genetic operators, not merely the subsets defined by Holland schemata (Grefenstette 1989, Radcliffe 1991, Vose 1991).

The Schema Theorem, in short, does not address the search component of genetic algorithms on which performance depends, and cannot distinguish genetic algorithms that are performing well from those that are not. How, then, has the Schema Theorem been interpreted as providing a foundation for understanding GA performance?

What the Schema Theorem says is that schemata with above-average fitness (especially short, low order schemata), increase their frequency in the population each generation at an exponential rate when rare. The mistake is to conclude that this growth of schemata has any implications for the quality of the search carried out by the GA. The Schema Theorem's implication, as many have put it, is that the genetic algorithm is focusing its search on promising regions of the search space, and thus increasing the likelihood that new samples of the search space will have higher fitness. But the phrase "promising regions of the search space" is a construct through which hidden assumptions are introduced which are not implied by the Schema Theorem. What is a "region", and what makes it "promising"?

The regions are schemata, and "promising regions" are schemata with above-average fitness. Offspring produced by recombination will tend to be drawn from the same "regions" as their parents, depending on the disruption rate from recombination. The common interpretation of the Schema Theorem implicitly assumes that any member of an above-average schema is likely to produce offspring of above-average fitness, i.e. that there is a correlation between membership in an above-average schema and production of fitter offspring. But the existence of such correlations is logically independent of the validity of the Schema Theorem.

For example, consider a population with a needle-in-a-haystack fitness function, where exactly one genotype (the "needle") has a high fitness, and all the other genotypes in the search space (the "hay") have the same low fitness. Consider a population in which the "needle" has already been found. The needle will tend to increase in frequency by selection, while recombination will most likely generate more "hay". The Schema Theorem will still be seen to operate, in that short schemata with above-average fitness (those schemata containing the needle) will increase in frequency, even though the fitness of new instances of the schemata (more hay) will not be any more likely to have the high fitness of the needle.

It is the quality of the search that must be used to characterize the performance of a genetic algorithm. One basis for evaluation is to compare the ability of a GA to generate new, highly fit individuals with the rate at which they are generated by random search. A direct approach to measuring GA performance is to analyze the change in the fitness distribution as the population evolves. For a GA to perform better than random search, the upper tail of the fitness distribution has to grow in time to be larger than the tail produced by random search. Some initial efforts at characterizing the growth of the upper tail of the fitness distribution were provided in Altenberg (1994), where a notion of "evolvability" — the ability to produce individuals fitter than any existing — was introduced as a measure of GA performance. A basic result is that for a GA to perform better than random search, there has to be a correlation between the fitness of parents and the upper tail of the fitness distribution of their offspring. This was obtained by using Price's Covariance and Selection Theorem (Price 1970, 1972) with a particular measurement function that extracts the fitness distribution from the population.

In this paper, I first review the application of Price's Theorem to GA performance analysis. Then I show how Price's Theorem can be used to obtain the Schema Theorem by employing a measurement function that extracts the frequency of a schema from the population. The difference between the theorem that measures GA performance, and the Schema Theorem, which does not, is shown to be simply a choice of measurement functions.

In the process of deriving results that relate the parent-offspring correlations to the performance of the GA under a generalized transmission function, schemata disappear as pertinent entities. Therefore, "schema processing" is not a requirement for performance in evolutionary algorithms in general. However, under recombination operators, schemata reappear in the formula for the change in the fitness distribution. This "missing" schema theorem shows explicitly that there must be correlations between schema fitnesses and offspring fitness distributions for good GA performance. It gives a quantitative expression to the Building Blocks Hypothesis (Goldberg 1989) and suggests ways to modify recombination operators to improve genetic algorithm performance.

2 GENETIC ALGORITHM ANALYSIS USING PRICE'S THEOREM

The strategy I take here (see Altenberg (1994) for details) is to start with a general formulation of the "canonical" genetic algorithm dynamics, for arbitrary representations, operators, and fitness functions. *Measurement functions* are then introduced to extract macroscopic features of the population. The evolution of these features can be shown, using Price's Covariance and Selection Theorem, to depend on the covariance between the measurement function and fitness. The choice of one measurement function gives us the Schema Theorem, while the choice of another measurement function gives us the evolution of the fitness distribution in the population, which I refer to as the *Local Performance Theorem*. Thus, the inability of the Schema Theorem to distinguish GA performance can be seen simply as the consequence of the measurement function that was chosen.

2.1 A GENERAL MODEL OF THE CANONICAL GENETIC ALGORITHM

A "canonical" model of genetic algorithms has been generally used since its formulation by Holland (1975), which incorporates assumptions common to many evolutionary models in population genetics: discrete, non-overlapping generations, frequency-independent selection, and infinite population size. The algorithm iterates three steps: selection, random mating, and production of offspring to constitute the population in the next generation.

Definition: Canonical Genetic Algorithm
The dynamical system representing the "canonical" genetic algorithm is:

$$p(x)' = \sum_{y,z \in \mathcal{S}} T(x \leftarrow y, z) \frac{w(y)w(z)}{\overline{w}^2} p(y)p(z), \tag{1}$$

where

- $p(x)$ *is the frequency of chromosome* x *in the population, and* $p(x)'$ *is the frequency in the next generation;*
- $\mathcal{S}$ *is the search space of* n *chromosomal types;*
- $T(x \leftarrow y, z)$, *the* transmission function, *is the probability that offspring genotype* x *is produced by parental genotypes* y *and* z *as a result of the action of genetic operators on the representation, with* $T(x \leftarrow y, z) = T(x \leftarrow z, y)$, *and* $\sum_x T(x \leftarrow y, z) = 1$ *for all* $y, z \in \mathcal{S}$;
- $w(x)$ *is the fitness of chromosome* x; *and*
- $\overline{w} = \sum_x w(x)p(x)$ *is the mean fitness of the population;*

This general form of the transmission-selection recursion was used by Slatkin (1970), and has been used subsequently for a variety of quantitative genetic and complex transmission systems (Cavalli-Sforza and Feldman 1976, Karlin 1979, Altenberg and Feldman 1987), and has been derived independently in genetic algorithm analysis (Vose 1990, Vose and Liepins 1991).

No assumptions are made about the structure of the chromosomes — e.g. the number of loci, the number of alleles at each locus, or even the linearity of the chromosome. The specific structure of the transmission function $T(x \leftarrow y, z)$ will carry the information about the chromosomal structure and genetic operators that is relevant to the dynamics of the GA.

As a cross-reference, the "mixing matrix" defined by Vose (1990) is the n by n matrix

$$\left\| T(\mathbf{0} \leftarrow y_i, z_j) \right\|_{i,j=1}^{n},$$

where $\mathbf{0}$ is the chromosome with all 0 alleles in the case of binary chromosomes, and chromosomes y and z are indexed from 1 to n. This is sufficient to characterize the transmission function in the case where mutation and recombination are symmetric with respect to either allele at each locus, by using n permutations of the arguments.

2.1.1 A Note on Fitness

The term "fitness" has undergone a semantic shift in its migration from population biology to evolutionary computation. In population biology, fitness generally refers to the actual rate that an individual type ends up being sampled in contributing to the next generation. So the fitness coefficient $w(x)$ lumps together all the disparate influences from different traits, intraspecific competition, and environmental interaction that produce it. In the evolutionary computation literature, fitness has come to be used synonymously with one or more objective functions (e.g. Koza (1992)). Under this usage is there is no longer a word that refers specifically to the reproductive contribution of a genotype. Here I will keep the distinction between objective function and fitness, and use "fitness" in its sense in population biology.

The term "fitness proportionate selection" refers to fitnesses that are independent of chromosome frequencies. Many selection schemes, such as tournament and rank-based selection, truncation selection, fitness sharing, and other population-based rescaling, are examples of *frequency-dependent* selection (Altenberg (1991) contains further references). In frequency-dependent selection, the fitness $w(x)$ is a function not only of x but of the composition of the population as well. All the theorems and corollaries in this paper apply to frequency-dependent selection. This is because they are all local, i.e. they apply to changes over a single generation based on the current composition of the population, so that any frequency-dependence in the fitness function $w(x)$ does not enter into the result.

The results on GA performance in this paper are defined directly in terms of the fitness distribution of the population. However, fitness functions are often defined in terms of an underlying objective function for the elements in the search space. This is the case with tournament selection, in which an individual's fitness equals the rank of their objective function in the population ($w = 1/N$ for the worst and $w = 1$ for the best individual in a population of size N). In these cases, GA performance ultimately is concerned with the distributions of objective function values in the population. The map from objective function to fitness would add an additional layer to the analysis of GA performance, and is not investigated here. However, numerous empirical studies have been undertaken to ascertain the effects of different selection schemes, with GA performance defined on underlying objective functions. So in the future such an analysis would be worthwhile.

2.1.2 Toward a Macroscopic Analysis

In the evolution of a population, individual chromosomes come and go, and their frequencies follow complex trajectories. These microscopic details are not the usual subject of interest when considering the performance of the GA (the one exception being the frequency of the fittest member of the search space). Rather, it is macroscopic properties, such as the population's mean fitness or fitness distribution, whose evolutionary trajectory is of interest. This is similar to the case of statistical mechanics, where one is interested not in the trajectories of individual molecules, but in the distribution of energies in the material.

It would be very useful if the evolutionary dynamics of the population could be defined solely at the macroscopic level — i.e. if the macroscopic description were *dynamically sufficient*. In GAs this will generally not be the case. However, let us consider one special condition when it is possible to describe the evolution of the fitness distribution solely in terms of the fitness distribution: when the fitness function $w(x)$ is invertible, i.e. no two genotypes have the same fitness. Then (1) can be transformed into a recursion in fitness domain:

$$f(w)' = \int_0^\infty T(w \leftarrow u, v)\frac{uv}{\overline{w}^2} f(u)\, f(v)\, du\, dv, \tag{2}$$

where $f(w)$ is the probability density of fitness w in the population (integration may be over discrete measure), and $T(w \leftarrow u, v) = T(x \leftarrow y, z)$ when $w = w(x)$, $u = w(y)$, and $v = w(z)$.

For the purposes of statistical estimation of the performance of a GA, which will be an imprecise task to begin with, it may be sufficient to proceed as though the GA dynamics

Table 1: Measurement functions, $F(x)$ (some taking arguments), and the population properties measured by their mean in the population, $\overline{F}$.

Population Property Measured by $\overline{F}$:	Measurement Function:
(1) Fitness distribution upper tail:	$F(x,w) = \begin{cases} 1 & w(x) > w \\ 0 & w(x) \leq w \end{cases}$
(2) Frequency of schema $\mathcal{H}$:	$F(x,\mathcal{H}) = \begin{cases} 1 & x \in \mathcal{H} \\ 0 & x \notin \mathcal{H} \end{cases}$
(3) Mean fitness:	$F(x) = w(x)$
(4) Fitness distribution's n-th non-central moment:	$F(x) = w(x)^n$
(5) Mean phenotype (vector valued):	$F(x) \in \mathbb{R}^n$
(6) Mean objective function:	$F(x) \in \mathbb{R}$

could be represented as in (2). That will be the strategy I suggest for statistically predicting the performance of a GA based on a limited sample from a GA run: an empirically derived estimate of $T(w \leftarrow u, v)$ may be used in (2) to approximate the dynamics of (1), in order to make predictions about GA performance. This is taken up in Section 4 on "adaptive landscape" analysis.

2.2 MEASUREMENT FUNCTIONS

A means of extracting macroscopic dynamics of a population from its microscopic dynamics (1) is the use of the appropriate *measurement functions*.

The fitness $w(x)$ is an example of a measurement function. Measurement functions need not be restricted to fitnesses, nor even scalar values. In general, let the measurement function $F(x)$ represent some property of genotype x, with $F : \mathcal{S} \mapsto \mathcal{V}$, where $\mathcal{V}$ is a vector space over the real numbers (e.g. $\mathbb{R}^k$ or $[0,1]^k$ for some positive integer k). The change in the population average of a measurement function is a measure of how the population is evolving:

$$\overline{F} = \sum_{x} F(x)\, p(x), \quad \overline{F}' = \sum_{x} F(x)\, p(x)' \tag{3}$$

A measurement function can be defined to indicate when a genotype instantiates a particular schema $\mathcal{H}$, by adding $\mathcal{H}$ as a parameter: $F(x,\mathcal{H}) = 1$ if $x \in \mathcal{H}$ and 0 otherwise. In general we can let $F : \mathcal{S} \times \mathcal{P} \mapsto \mathcal{V}$ be a parameterized family of measurement functions, for some parameter space $\mathcal{P}$.

Examples of different measurement functions and the population properties measured by $\overline{F}$ are shown in Table 1. Measurement functions (1) and (2) are the focus her: (1) extracts the fitness distribution of the population, and (2) extracts the frequency of a schema in the population.

2.3 PRICE'S THEOREM

Price (1970) introduced a theorem that partitions the effect of selection on a population in terms of covariances between fitness and the property of interest (allele frequencies were the property considered by Price) and effects due to transmission. Price's theorem has been applied in a number of different contexts in evolutionary genetics, including kin selection (Grafen 1985, Taylor 1988), group selection (Wade 1985), the evolution of mating systems (Uyenoyama 1988), and quantitative genetics (Frank and Slatkin 1990). Price's theorem gives the one-generation change in the population mean value of F:

Theorem 1 (Covariance and Selection, Price, 1970)
For any parental pair $\{y, z\}$*, let* $\phi(y, z)$ *represent the expected value of* F *among their offspring. Thus:*

$$\phi(y,z) = \sum_{x} F(x)\, T(x \leftarrow y, z). \tag{4}$$

Then the population average of the measurement function in the next generation is

$$\overline{F}' = \overline{\phi} + \mathrm{Cov}[\phi(y,z),\ w(y)w(z)/\overline{w}^2] \tag{5}$$

$$\overline{\phi} = \sum_{y,z} \phi(y,z)\, p(y)p(z)$$

is the average offspring value in a population reproducing without selection, and

$$\mathrm{Cov}[\phi(y,z),\ w(y)w(z)/\overline{w}^2] = \sum_{y,z} \phi(y,z)\, \frac{w(y)w(z)}{\overline{w}^2}\, p(y)p(z) - \overline{\phi} \tag{6}$$

is the population covariance (i.e. the covariance over the distribution of genotypes in the population) between the parental fitness values and the measured values of their offspring.

Proof. One must assume that for each y and z, the expectation $\phi(y,z)$ exists (for measurement functions (1) and (2), the expectation always exists). Substitution of (1), (4), and (6) into (3) directly produces (5). ■

Price's theorem shows that the covariance between parental fitness and offspring traits is the means by which selection directs the evolution of the population. Several corollaries follow:

Corollary 1 *Let* $C(y,z) = \phi(y,z) - [F(y) + F(z)]/2$ *represent the difference between the mean of* F *among parents* y *and* z*, and the mean of* F *in their offspring. Then*

$$\overline{F}' - \overline{F} = \mathrm{Cov}[F(x), w(x)/\overline{w}] + \overline{C} + \mathrm{Cov}[C(x,y), w(x)\, w(y)/\overline{w}^2],$$

where $\overline{C} = \sum_{y,z} C(y,z)\, p(y)\, p(z)$.

Proof.

$\overline{w}$

Corollary 2 (Fisher's Fundamental Theorem, 1930)
Consider a population evolving in the absence of a genetic operator, so

$$T(x \leftarrow y, z) = [\delta(x,y) + \delta(x,z)]/2,$$

where

$$\delta(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y. \end{cases}$$

Then $C(y,z) = 0$. *For* $F(x) = w(x)$, *Corollary 1 gives:*

$$\overline{w}' - \overline{w} = \overline{w}\ \mathrm{Var}[w(x)/\overline{w}].$$

2.4 A LOCAL PERFORMANCE MEASURE FOR GENETIC ALGORITHMS

Price's theorem can be used to extract the change in the *distribution* of fitness values in the population by using the measurement function (1) from Table 1. Then

$$\overline{F}(w) = \sum_x F(x,w)\,p(x) = \sum_{x:\,w(x)>w} p(x)$$

is the proportion of the population that has fitness greater than w. Price's Theorem gives:

Corollary 3 (Evolution of the fitness distribution)
The fitness distribution in the next generation is:

$$\overline{F}(w)' = \overline{\phi}(w) + \mathrm{Cov}[\phi(y,z,w),\ w(y)w(z)/\overline{w}^2], \tag{7}$$

where $\phi(y,z,w)$ *is the proportion of offspring from parents* y *and* z *that with fitness greater than* w.

Note that $\phi(y,z,w)$ always exists, even when the distribution of fitnesses among the offspring of y and z has no expectation, i.e. when $\sum_x w(x)\,T(x \leftarrow y,z)$ is infinite.

The expression (7) can be made more informative by rewriting $\phi(y,z,w)$ as the sum of a random search term plus a *search bias* term that gives how parents y and z compare with random search in their offspring fitnesses. Let $\mathcal{R}(w)$ be the probability that random search produces an individual fitter than than w, and let the *search bias*, $\beta(y,z,w)$, be:

$$\beta(y,z,w) = \phi(y,z,w) - \mathcal{R}(w) = \sum_x F(x,w)\,T(x \leftarrow y,z) - \mathcal{R}(w).$$

The average search bias for a population before selection is $\overline{\beta}(w) = \sum_{y,z} \beta(y,z,w)\ p(y)\,p(z)$. The coefficient of regression of $\beta(y,z,w)$ on $w(y)\,w(z)/\overline{w}^2$ is

$$\mathrm{Reg}[\beta(y,z,w) \rightarrow w(y)\,w(z)/\overline{w}^2] = \mathrm{Cov}[\beta(y,z,w),\ w(y)\,w(z)/\overline{w}^2]\ /\ \mathrm{Var}[w(y)\,w(z)/\overline{w}^2].$$

It measures the magnitude of how $\beta(y,z,w)$ varies with $w(y)w(z)/\overline{w}^2$ in the population.

Theorem 2 (Local Performance Measure)
The probability distribution of fitnesses in the next generation is

$$\overline{F}(w)' = \mathcal{R}(w)\ +\ \overline{\beta}(w)\ +\ \mathrm{Reg}[\beta(y,z,w) \rightarrow w(y)\,w(z)/\overline{w}^2]\ \mathrm{Var}[w(y)\,w(z)/\overline{w}^2].$$

Theorem 2 shows that in order for the GA to perform better than random search in producing individuals fitter than than w, the average search bias, plus the parent-offspring regression scaled by the fitness variance,

$$\overline{\beta}(w) + \mathrm{Reg}[\beta(y,z,w) \rightarrow w(y)\,w(z)/\overline{w}^2]\ \mathrm{Var}[w(y)\,w(z)/\overline{w}^2], \tag{9}$$

must be positive. As in the Schema Theorem, this is a local result because the terms in (8) other than $\mathcal{R}(w)$ depend on the composition of the population and thus change as it evolves.

Both the regression and the search bias terms require the transmission function to have "knowledge" about the fitness function. Under random search, the expected value of both these terms would be zero. Some knowledge of the fitness function must be incorporated in the transmission function for the expected value of these terms to be positive. It is this knowledge — whether incorporated explicitly or implicitly — that is the source of power in genetic algorithms.

2.5 THE SCHEMA THEOREM

Holland's Schema Theorem (Holland 1975) is classically given as follows. Let

$\mathcal{H}$ represent a particular schema as defined by Holland (1975),
L be the length of the chromosome, and $L(\mathcal{H}) \leq L-1$ be the defining length of the schema;
$p(\mathcal{H}) = \sum_{x \in \mathcal{H}} p(x)$ be the frequency of schema $\mathcal{H}$ in the population, and
$\overline{w}(\mathcal{H}) = \sum_{x \in \mathcal{H}} w(x)p(x)/p(\mathcal{H})$ be the marginal fitness of schema $\mathcal{H}$.

Theorem 3 (The Schema Theorem, Holland 1975)
In a genetic algorithm using a proportional selection algorithm and single point crossover occurring with probability r, the following holds for each schema $\mathcal{H}$:

$$p(\mathcal{H})' \geq p(\mathcal{H}) \frac{\overline{w}(\mathcal{H})}{\overline{w}} \left(1 - r\frac{L(\mathcal{H})}{L-1}\right) \tag{10}$$

Now, Price's Theorem can be used to obtain the Schema Theorem by using:

$$F(x, \mathcal{H}) = \begin{cases} 1 & \text{if } x \in \mathcal{H} \\ 0 & \text{if } i \notin \mathcal{H} \end{cases}$$

and $\phi(y, z, \mathcal{H}) = \sum_x F(x, \mathcal{H}) T(x \leftarrow y, z)$, which represents the fraction of offspring of parents y and z that are in schema $\mathcal{H}$. Then $p(\mathcal{H}) = \overline{F}(\mathcal{H})$, and

Corollary 4 (Schema Frequency Change)

$$p(\mathcal{H})' = \overline{\phi}(\mathcal{H}) + \mathrm{Cov}[\phi(y, z, \mathcal{H}),\ w(y)w(z)/\overline{w}^2],$$

Two sources can be seen to contribute to a change in schema frequency:

1. linkage disequilibrium, i.e. the schema frequency minus the product of the frequencies of the alleles comprising the schema. Negative linkage disequilibrium would produce $\overline{\phi}(\mathcal{H}) > p(\mathcal{H})$; and
2. covariance between parental fitnesses and the proportion of their offspring in the schema.

Equation (11) can be made more informative by rewriting $\phi(y, z, \mathcal{H})$ in terms of a "disruption" coefficient. A value $\alpha_{\mathcal{H}} \in [0, 1]$ can be defined that places a lower bound on the faithfulness of transmission of any schema $\mathcal{H}$:

$$\phi(y, z, \mathcal{H}) \geq \frac{1}{2}(1 - \alpha_{\mathcal{H}})[F(y, \mathcal{H}) + F(z, \mathcal{H})] \tag{12}$$

and

$$\alpha_{\mathcal{H}} = 1 - \min_{y \in \mathcal{H} \text{ or } z \in \mathcal{H}} \left[\phi(y,z,\mathcal{H}) \frac{2}{F(y,\mathcal{H}) + F(z,\mathcal{H})} \right]$$

Actually, $\alpha_{\mathcal{H}}$ can be defined for any subset of the search space ("predicate" in Vose (1991) or "forma" in Radcliffe (1991)). For Holland schemata under single-point crossover, $\alpha_{\mathcal{H}} = r\,L(\mathcal{H})/(L-1)$ (the rate that crossover disrupts schema $\mathcal{H}$). Using (12) we obtain:

Theorem 4 (Schema, Covariance Form)
The change in the frequency of any subset $\mathcal{H}$ of the search space (i.e. a schema) over one generation is bounded below by:

$$p(\mathcal{H})' \geq \{p(\mathcal{H}) + \mathrm{Cov}[F(y,\mathcal{H}), w(x)/\overline{w}]\}\,(1-\alpha_{\mathcal{H}}). \tag{13}$$

Therefore, if

$$\mathrm{Cov}\left[F(y,\mathcal{H}), \frac{w(x)}{\overline{w}}\right] > \frac{\alpha_{\mathcal{H}}}{1-\alpha_{\mathcal{H}}}$$

then schema $\mathcal{H}$ will increase in frequency.

Proof.

$$\begin{aligned}
\overline{F}(\mathcal{H})' &= \sum_{x,y,z} F(x,\mathcal{H}) T(x \leftarrow y,z) \frac{w(y)w(z)}{\overline{w}^2} p(y)p(z) \\
& \sum_{y,z} \phi(y,z,\mathcal{H}) \frac{w(y)w(z)}{\overline{w}^2} p(y)p(z) \\
&\geq \frac{1}{2}(1-\alpha_{\mathcal{H}}) \sum_{y,z} [F(y,\mathcal{H}) + F(z,\mathcal{H})] \frac{w(y)w(z)}{\overline{w}^2} p(y)p(z) \\
& (1-\alpha_{\mathcal{H}}) \sum_{y} F(y,\mathcal{H}) w(y) p(y)/\overline{w} = (1-\alpha_{\mathcal{H}}) \left[\overline{F}(\mathcal{H}) + \mathrm{Cov}[F(y,\mathcal{H}), w(x)/\overline{w}]\right]
\end{aligned}$$

■

Thus, if there is a great enough covariance between fitness and being a member of a schema, the schema will increase in frequency.

Although both applications of Price's Theorem — to schema frequency change and change in the fitness distribution — involve covariances with parental fitness values, the crucial point is that the covariance term (from (13)), $\mathrm{Cov}[F(y,\mathcal{H}), w(x)/\overline{w}]$, and the covariance term (from (7)), $\mathrm{Cov}[\phi(y,z,w),\ w(y)w(z)/\overline{w}^2]$, are independently defined. So conditions that produce growth in the frequencies of different schemata are independent of conditions that produce growth in the upper tails of the fitness distribution.

For example, consider a fitness function with a random distribution being the one-sided stable distribution of index 1/2 (Feller, 1971): $R(w) = 2\mathcal{N}(a/\sqrt{w}) - 1$, where $\mathcal{N}(y)$ is the Normal distribution and a is a scale parameter. This distribution is a way of generating "needles in the haystack" on all length scales. A GA with this fitness function will generically have schemata that obey (10), even though it is still random search.

3 RECOMBINATION AND THE RE-EMERGENCE OF SCHEMATA

In the local performance measure for the genetic algorithm, schemata disappear as relevant entities. No summations over hyperplanes or other subsets of the search space appear in Theorem 2. Schemata are therefore not informative structures for operators and representations in general. However, it is the recombination operator for which schemata have been hypothesized to play a special role. What I show in this section is that when one examines (7) using recombination operators specifically, schemata re-emerge in the local performance theorem, and they appear in a way that offers possible new insight into how schemata enter into GA performance. This "missing" schema theorem makes explicit the intuition, missing from the Schema Theorem, about what makes a good building block.

Recombination operators in a multiple-locus genetic algorithm can be generally characterized using the recombination distribution analysis introduced by Geiringer (1944), and developed independently by Syswerda (1989) (see also Karlin and Liberman 1978, Booker 1993, and Vose and Wright 1994). Consider a system of L loci. Any particular recombination event can be described by indicating which parent the allele for each locus came from. This can be done with a mask, a vector $\boldsymbol{r} \in \{0,1\}^L$, of binary variables $r_i \in \{0,1\}$, which indicate the loci that are transmitted together from either parent. So all loci with $r_i = 0$ are transmitted from one parent, while the remainder of the loci, with $r_i = 1$, are transmitted from the other parent. The vectors $\boldsymbol{r} = \boldsymbol{0} = (0 \ldots 0)$ and $\boldsymbol{r} = \boldsymbol{1} = (1 \ldots 1)$ correspond to an absence of recombination in transmission. With $\boldsymbol{r}$ representing the recombination event that occurred in transmission, the offspring $\boldsymbol{x}$ of parental chromosomes $\boldsymbol{y}$ and $\boldsymbol{z}$ can be expressed as:

$$\boldsymbol{x} = \boldsymbol{r} \circ \boldsymbol{y} + (\boldsymbol{1} - \boldsymbol{r}) \circ \boldsymbol{z},$$

where $\circ$ is the Schur product: $\boldsymbol{u} \circ \boldsymbol{v} = (u_1 v_1 \ldots u_L v_L)$ (allele multiplication and addition is just for the convenience of notation; it is defined only with 0 as the other operand).

The action of any particular recombination operator can be represented as a probability distribution, $R(\boldsymbol{r})$, over the set $\boldsymbol{r} \in \{0,1\}^L$. Thus $\sum_{\boldsymbol{r} \in \{0,1\}^L} R(\boldsymbol{r}) = 1$. Using $R(\boldsymbol{r})$ the transmission probabilities can be written:

$$T(\boldsymbol{x} \leftarrow \boldsymbol{y}, \boldsymbol{z}) = \sum_{\boldsymbol{r} \in \{0,1\}^L} R(\boldsymbol{r})\, \delta(\boldsymbol{x},\ \boldsymbol{r} \circ \boldsymbol{y} + (\boldsymbol{1} - \boldsymbol{r}) \circ \boldsymbol{z}).$$

Because the order of the parents is taken to be irrelevant, $\boldsymbol{r}$ and $\boldsymbol{1} - \boldsymbol{r}$ represent the same recombination event, hence $R(\boldsymbol{r}) = R(\boldsymbol{1} - \boldsymbol{r})$, which gives $T(\boldsymbol{x} \leftarrow \boldsymbol{y}, \boldsymbol{z}) = T(\boldsymbol{x} \leftarrow \boldsymbol{z}, \boldsymbol{y})$.

Often with genetic algorithms, the genetic operator is applied to only a proportion, α, of the population. In this case one would have:

$$T(\boldsymbol{x} \leftarrow \boldsymbol{y}, \boldsymbol{z}) = (1 - \alpha)[\delta(\boldsymbol{x}, \boldsymbol{y}) + \delta(\boldsymbol{x}, \boldsymbol{z})]/2 + \alpha \sum_{\boldsymbol{r} \in \{0,1\}^L} R(\boldsymbol{r})\, \delta(\boldsymbol{x},\ \boldsymbol{r} \circ \boldsymbol{y} + (\boldsymbol{1} - \boldsymbol{r}) \circ \boldsymbol{z}).$$

Examples. Uniform crossover (Ackley 1987, Syswerda 1989), i.e. free recombination (Charlesworth *et al.* 1992, Goodnight 1988), is described by $R(\boldsymbol{r}) = 2^{-L}$. Single-point crossover is described (Karlin and Liberman 1978) by:

$$R(\boldsymbol{r}) = \begin{cases} 1/(L-1) & \text{if } \sum_{i=1}^{L-1} |r_{i+1} - r_i| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Single-point shuffle crossover (Eshelman *et al.* 1989) is described by:

$$R(\boldsymbol{r}) = \begin{cases} 1/(L-1)\binom{L}{n(\boldsymbol{r})} & \text{if } n(\boldsymbol{r}) = 1, \ldots, L-1, \\ 0 & \text{if } n(\boldsymbol{r}) = 0 \text{ or } L, \end{cases}$$

where $n(\boldsymbol{r}) = \sum_{i=1}^{L} r_i$ is the number of 1s in $\boldsymbol{r}$.

Note that each $\boldsymbol{r}$ partitions the loci into two sets. Let us collect from $\boldsymbol{x}$ the loci with $r_i = 0$ to make a vector $\boldsymbol{x}_0(\boldsymbol{r})$, and similarly collect the loci with $r_i = 1$ to make a vector $\boldsymbol{x}_1(\boldsymbol{r})$. Let $\mathcal{H}(\boldsymbol{r})$ denote the set of schemata with defining positions $\{i : r_i = 1\}$. Thus the vectors $\boldsymbol{x}_0(\boldsymbol{r}) \in \mathcal{H}(\boldsymbol{1}-\boldsymbol{r})$ and $\boldsymbol{x}_1(\boldsymbol{r}) \in \mathcal{H}(\boldsymbol{r})$ represent Holland schemata. For notational brevity I henceforth write simply $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$, with the dependence on $\boldsymbol{r}$ being understood.

The marginal fitnesses of the schemata are:

$$\overline{w}_0(\boldsymbol{x}_0) = \sum_{\boldsymbol{x}_1 \in \mathcal{H}(\boldsymbol{r})} w(\boldsymbol{x}_0, \boldsymbol{x}_1)\, p(\boldsymbol{x}_0, \boldsymbol{x}_1)/p_0(\boldsymbol{x}_0),$$

$$\overline{w}_1(\boldsymbol{x}_1) = \sum_{\boldsymbol{x}_0 \in \mathcal{H}(\boldsymbol{1}-\boldsymbol{r})} w(\boldsymbol{x}_0, \boldsymbol{x}_1)\, p(\boldsymbol{x}_0, \boldsymbol{x}_1)/p_1(\boldsymbol{x}_1),$$

where

$$p_0(\boldsymbol{x}_0) = \sum_{\boldsymbol{x}_1 \in \mathcal{H}(\boldsymbol{r})} p(\boldsymbol{x}_0, \boldsymbol{x}_1) \quad \text{and} \quad p_1(\boldsymbol{x}_1) = \sum_{\boldsymbol{x}_0 \in \mathcal{H}(\boldsymbol{1}-\boldsymbol{r})} p(\boldsymbol{x}_0, \boldsymbol{x}_1).$$

At this point we can express (7) in Corollary 3 using the marginal fitnesses of schemata defined by each $\boldsymbol{r}$.

Theorem 5 (Evolution of the fitness distribution under recombination)
The change in the fitness distribution over one generation under the action of selection and recombination is:

$$\begin{aligned} \overline{F}(w) = &\sum_{\boldsymbol{r} \in \{0,1\}^L} R(\boldsymbol{r}) \operatorname{Cov}[F(\boldsymbol{x}, w), \frac{\overline{w}_0(\boldsymbol{x}_0)\overline{w}_1(\boldsymbol{x}_1)}{\overline{w}^2}] \qquad (14) \\ &- \sum_{\boldsymbol{r} \in \{0,1\}^L} R(\boldsymbol{r}) \sum_{\substack{\boldsymbol{x}_0 \in \mathcal{H}(\boldsymbol{1}-\boldsymbol{r}) \\ \boldsymbol{x}_1 \in \mathcal{H}(\boldsymbol{r})}} [p(\boldsymbol{x}) - p_0(\boldsymbol{x}_0)\, p_1(\boldsymbol{x}_1)]\, [F(\boldsymbol{x}, w) - \overline{F}(w)]\, \frac{\overline{w}_0(\boldsymbol{x}_0)\overline{w}_1(\boldsymbol{x}_1)}{\overline{w}^2} \end{aligned}$$

where the partition of $\boldsymbol{x}$ into vectors $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ is understood to be determined by each transmission vector $\boldsymbol{r}$ in the sum.

The proof is given in the Appendix.

Theorem 5 is what I have referred to as the "missing" schema theorem. Equation (14) shows a number of features:

The covariance term. The change in the fitness distribution $\overline{F}(w)$ depends on the covariance between the schema fitnesses $\overline{w}_0(\boldsymbol{x}_0)\overline{w}_1(\boldsymbol{x}_1)$ and $F(\boldsymbol{x}, w)$. Thus a positive covariance between the fittest schemata and the fittest offspring will contribute toward an increase in the upper tail of the fitness distribution.

Not all schemata are "processed". Not all possible Holland schemata appear in (5), but only the ones for which the recombination event r occurs with some probability (i.e. $R(r) > 0$). In the case of classical single-point crossover, only $L-1$ recombination events may occur out of the $2^{L-1}-1$ possible recombination events (subtracting transmission of intact chromosomes and symmetry in the parents). Thus, the schemata from only $L-1$ different configurations of defining positions contribute to (14). So, with two alleles at each locus, only $2(2^1+2^2+\ldots+2^{L-1}) = 2^{L+1}-4$ schemata are involved in (14) under single-point crossover. This is compared to a possible 3^L-2^L schemata (subtracting the highest order schemata, i.e. chromosomes) that could result from a recombination event in the case of uniform crossover.

Schemata enter as complementary pairs. Schema fitnesses always occur in complementary pairs whose defining positions encompass all the loci.

Disruption is quantified by the linkage disequilibrium. The *linkage disequilibrium* between schemata x_0 and x_1 is the term $p(x) - p_0(x_0)p_1(x_1)$. It is a measure of the co-occurrence of schemata x_0 and x_1 in the population. If $p(x) > p_0(x_0)\,p_1(x_1)$, then recombination event r disrupts more instances of genotype x than it creates. If in addition, $F(x,w) > \overline{F}(w)$, then this term contributes negatively toward the change in $\overline{F}(w)$. Conversely, if a combination of schemata has a deficit in the population (i.e. $p(x) < p_0(x_0)\,p_1(x_1)$), and the measurement function for this combination is greater than the population average (i.e. $F(x,w)-\overline{F}(w)$), then the recombination event r will contribute toward in increase in $\overline{F}(w)$.

If all loci were in linkage equilibrium, exhibiting *Robbins proportions* $p(x) = \prod_{i=1\ldots L} p_i(x_i)$ (Robbins 1918, Christiansen 1987, Booker 1993), then (14) reduces to:

$$\overline{F}(w)' - \overline{F}(w) = \sum_{r\in\{0,1\}^L} R(r)\,\mathrm{Cov}[F(x,w),\ \frac{\overline{w}_0(x_0)\,\overline{w}_1(x_1)}{\overline{w}^2}]. \tag{15}$$

Robbins proportions are assumed in much of quantitative genetic analysis, both classically (Cockerham 1954), and more recently (Bürger 1993), because linkage disequilibrium presents analytical difficulties. Asoh and Muhlenbein (1994) and Mühlenbein and Schlierkamp-Vosen (1993) assume Robbins proportions in their quantitative-genetic approach to GA analysis. Using $F(x) = w(x)$ as the measurement function, they show that under free recombination, a term similar to (15) evaluates to a sum of variances of epistatic fitness components derived from a linear regression.

Except under special assumptions, however, selection will generate linkage disequilibrium that produces departures from the results that assume Robbins proportions (Turelli and Barton 1990). The only recombination operator that will enforce Robbins proportions in the face of selection is Syswerda's "simulated crossover" (Syswerda 1993). Simulated crossover produces offspring by independently drawing the allele for each locus from the entire population after selection. One may even speculate that the performance advantage seen in simulated crossover in some way relates to it producing a population that exhibits "balanced design" from the point of view of analysis of variance, allowing estimation of the epistasis components (Reeves and Wright, this volume).

The epistasis variance components from Asoh and Muhlenbein (1994) figure into the parent-offspring covariance in fitness. In their covariance sum, higher order schemata appear with exponentially decreasing weights. Thus, the lowest order components are most important in determining the parent-offspring correlation. These epistasis variance components, it should be noted, appear implicitly in the paper by Radcliffe and Surry (this volume). They constitute the increments between successive forma variances shown in their Figure 2. Radcliffe and Surry find that the rate of decline in the forma variances as forma order increases is a good predictor of the GA performance of different representations. This is equivalent to there being large epistasis components for low order schemata, which produces the highest parent-offspring correlation in fitness in the result of Asoh and Muhlenbein (1994).

Guidance for improving the genetic operator. The terms

$$\begin{aligned} &\mathrm{Cov}[F(x,w), \frac{\overline{w}_0(x_0)\overline{w}_1(x_1)}{\overline{w}^2}] \\ &\quad - \sum_{x} [p(x) - p_0(x_0)p_1(x_1)]\,[F(x,w) - \overline{F}(w)]\,\frac{\overline{w}_0(x_0)\overline{w}_1(x_1)}{\overline{w}^2}, \end{aligned} \tag{16}$$

for each recombination event, r, provide a rationale for modifying the recombination distribution to increase the performance of the GA. Probabilities $R(r)$ for which terms (16) are negative should be set to zero, and the distribution $R(r)$ allocated among the most positive terms (16). The best strategy of modifying $R(r)$ presents an interesting problem: I propose that a good strategy would be to start with uniform recombination and progressively concentrate it on the highest terms in (16).

4 ADAPTIVE LANDSCAPE ANALYSIS

The "adaptive landscape" concept was introduced by Wright (1932) to help describe evolution when the actions of selection, recombination, mutation, and drift produce are multiple attractors in the space of genotypes or genotype frequencies. Under the rubric of "landscape" analysis, a number of studies have employed covariance statistics as predictors of the performance of evolutionary algorithms (Weinberger 1990, Manderick *et al.* 1991, Weinberger 1991a,b, Mathias and Whitley 1992, Stadler and Schnabl 1992, Stadler and Happel 1992, Stadler 1992, Menczer and Parisi 1992, Fontana *et al.* 1993, Weinberger and Stadler 1993, Kinnear 1994, Stadler 1994, Grefenstette, this volume). I consider first some general aspects of the landscape concept, and then examine the use of covariance statistics to predict the performance of the GA.

4.1 THE LANDSCAPE CONCEPT

The "adaptive landscape" is a visually intuitive way of describing how evolution moves through the search space. A search space is made into a landscape by defining closeness relations between its points, so that for each point in the search space, neighborhoods of "nearby" points are defined. The purpose of doing this is to represent the attractors of the evolutionary process as "fitness peaks", with the premise that selection concentrates a population within a domain of attraction around the fittest genotype in the domain. The concepts of local search, multimodal fitness functions, and hill climbing are all landscape concepts.

Definitions of closeness relations are often derived from metrics that are seemingly natural for the search space, for example, Hamming distances for binary chromosomes, and Euclidean distance in the case of search spaces in $\mathbb{R}^n$. However, in order for closeness relations to be relevant to the evolutionary dynamics, they must be based on the transmission function, since it is the transmission function that connects one point in the search space to another by defining the transition probabilities between parents and offspring. In the adaptive landscape literature, this distinction between extrinsically defined landscapes and landscapes defined by the transmission function is frequently omitted.

Application of the landscape metaphor is difficult, if not infeasible, for sexual transmission functions. For this reason, some authors have implicitly used mutation to define their adaptive landscape even when recombination is the genetic operator acting. The definition of closeness becomes problematic because the distribution of offspring of a given parent depends on the frequency of other parents in the population. For example, consider a mating between two complementary binary chromosomes when uniform recombination is used. The neighborhood of the chromosomes will be the entire search space, because recombinant offspring include every possible chromosome. Since the neighborhood of a chromosome depends on chromosomes that it is mated with, the adaptive landscape depends on the composition of the population, and could thus be described as frequency-dependent. The sexual adaptive landscape will change as the population evolves on it.

The concept of multimodality illustrates the problem of using metrics extrinsic to the transmission function to define the adaptive landscape. Consider a search space in $\mathbb{R}^n$ with a multimodal fitness function. The function is multimodal in terms of the Euclidean metric on $\mathbb{R}^n$. But the Euclidean neighborhoods may be obliterated when the real-valued phenotype is encoded into a binary chromosome and neighborhoods are defined by the action of mutation or recombination. For example, let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ be encoded into binary chromosomes $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^L$. The Hamming neighborhoods $H(\boldsymbol{x}, \boldsymbol{y}) \leq k$ may have no correspondence to Euclidean neighborhoods $|\boldsymbol{a} - \boldsymbol{b}| \leq c$. Thus multimodality under the Euclidean metric is irrelevant to the GA unless the transmission function preserves the Euclidean metric. Multimodality should not be considered a property of the fitness function alone, but only of the relationship between the fitness function and the transmission function.

4.1.1 An Illustration of Multimodality's Relation to Transmission

Consider the fitness function from p. 34 in Michalewicz (1994):

$$w(x_1, x_2) = 21.5 + x_1 \sin(4\pi x_1) + x_2 \sin(20\pi x_2),$$

defined on the variables x_1, x_2. In terms of the normal Euclidean neighborhoods about (x_1, x_2), $w(x_1, x_2)$ is highly multimodal, as can be seen in Figure 1. There are over 500 modes on the area defined by the constraints

$$-3 \leq x_1 \leq 12.1 \text{ and } 4.1 \leq x_2 \leq 5.8.$$

A transmission function that could be said to produce the Euclidean neighborhoods is a Gaussian mutation operator that perturbs (x_1, x_2) to $(x_1 + \epsilon_1,\ x_2 + \epsilon_2)$ with probability density

$$C \exp[-(\epsilon_1^2 + \epsilon_2^2)/2\sigma^2], \tag{17}$$

with σ small and C the normalizing constant. The adaptive landscape could be said to be multimodal with respect to this genetic operator.

Suppose we change the representation into four new variables, integers n_1, n_2 and fractions $\phi_1, \phi_2 \in [0, 1)$:

$$n_1 \quad \mathrm{Int}(2x_1), \text{ and } \phi_1 = 2x_1 \quad n_1,$$

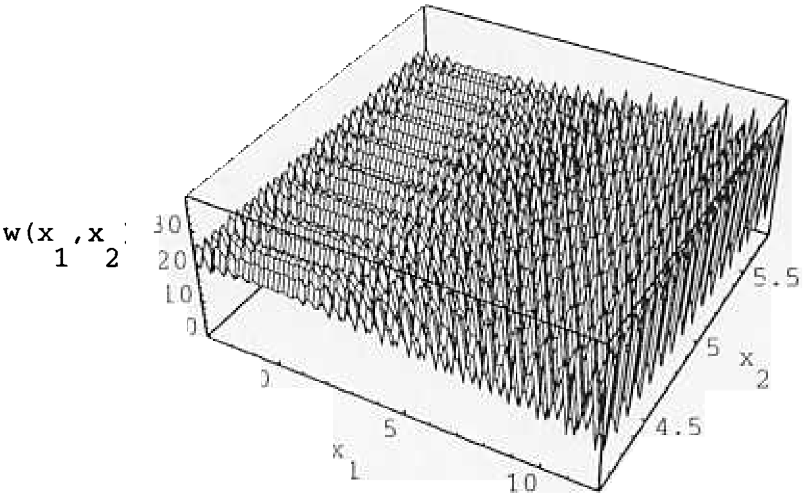

Figure 1: The fitness function $w(x_1, x_2) = 21.5 + x_1 \sin(4\pi x_1) + x_2 \sin(20\pi x_2)$ is highly multimodal in terms of the Euclidean neighborhoods on (x_1, x_2).

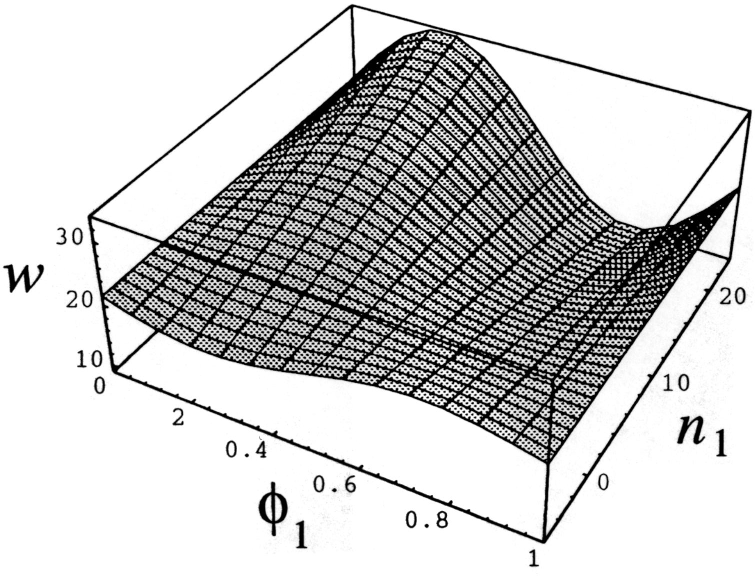

Figure 2: The "adaptive landscape" produced from mutation operators acting on the transformed representation $(n_1, \phi_1, n_2, \phi_2)$, where $x_1 = (n_1 + \phi_1)/2$, and $x_2 = (n_2 + \phi_2)/10$. Over the same region as in Figure 1, w has few modes, as seen in this slice through the 4 dimensional space, setting $n_2 = 50$, $\phi_2 = 0$.

$$n_2 = \mathrm{Int}(10x_2), \text{ and } \phi_2 = 10x_2 - n_2,$$

where $\mathrm{Int}(v)$ is the largest integer not greater than v. Thus $x_1 = (n_1 + \phi_1)/2$, and $x_2 = (n_2 + \phi_2)/10$.

This transformation of variables uses our *a priori* knowledge about the fitness function to produce a smoother adaptive landscape. Neighborhoods for the new representation are produced by using a mutation operator that increases or decreases n_1 or n_2 by 1, or perturbs ϕ_1 or ϕ_2 in a Gaussian manner. In this new topography, the fitness function has very few modes, as shown in Figure 2.

Instead of changing the representation to produce a smooth landscape, one can keep the native variables x_1 and x_2, but change the mutation operator. The new mutation operator perturbs (x_1, x_2) to $(x_1 + \epsilon_1 + \nu_1,\ x_2 + \epsilon_2 + \nu_2)$ with probability densities $f(\nu_1) = 1/2$ for $\nu_1 = 1/2$ or $-1/2$, and $f(\nu_2) = 1/2$ for $\nu_2 = 1/10$ or $-1/10$, and

$$f(\epsilon_1, \epsilon_2) = C \exp[-(\epsilon_1^2 + \epsilon_2^2)/2\sigma^2],$$

with σ small and C the normalizing constant. This change in the genetic operator produces evolutionary dynamics identical to that produced by the change in the representation. This exemplifies the duality between representations and operators (Altenberg 1994).

Rather than trying to push the landscape metaphor further, it may be more fruitful to return to the roots of the concept, which is the existence of multiple attractors in evolutionary dynamics (or metastable states, in the case of stochastic evolutionary systems). The task of producing a smooth adaptive landscape is, in effect, to design operators and representations that yield a single domain of attraction, where all populations converge to the fittest member of the search space. In order to evaluate an adaptive landscape that contains multiple attractors, one needs a way of characterizing the attractors. This is the goal of adaptive landscapes statistics that have been developed.

4.2 LANDSCAPE STATISTICS

It would be useful to be able to predict the performance of a GA, or of particular representations or operators used by a GA, based on a limited number of sample points. I will review previous work toward this goal, pose some counterexamples to the statistics that have been developed, and offer a new statistic that solves some of the difficulties.

A number of studies have employed the statistical technique introduced by Weinberger (1990) toward predicting the performance of a GA. They rely on the autocorrelation statistic:

$$\rho_{op}(\tau) = \frac{\mathrm{Cov}[w(\boldsymbol{x}_\tau), w(\boldsymbol{x}_0)]}{(\mathrm{Var}[w(\boldsymbol{x}_\tau)]\ \mathrm{Var}[w(\boldsymbol{x}_0)])^{1/2}}, \tag{18}$$

where $\boldsymbol{x}_\tau$ is derived from $\boldsymbol{x}_0$ by τ iterations of the genetic operator, Cov and Var are taken over some measure, $m(\boldsymbol{x}_\tau, \boldsymbol{x}_0)$, on the search space $\mathcal{S}$:

$$\mathrm{Cov}[w(\boldsymbol{x}_\tau), w(\boldsymbol{x}_0)] = \int_{\mathcal{S}} w(\boldsymbol{x}_\tau)\, w(\boldsymbol{x}_0)\, dm(\boldsymbol{x}_\tau, \boldsymbol{x}_0) - \int_{\mathcal{S}} w(\boldsymbol{x}_\tau)\, dm(\boldsymbol{x}_\tau, \boldsymbol{x}_0) \int_{\mathcal{S}} w(\boldsymbol{x}_0)\, dm(\boldsymbol{x}_\tau, \boldsymbol{x}_0).$$

The measure $m(\boldsymbol{x}_\tau, \boldsymbol{x}_0)$ derives from the way samples of the search space are taken. Weinberger uses random walks over the search space generated by iteration of asexual genetic operators. Manderick *et al.* (1991) point out that only asexual genetic operators allow one to generate the random walks that produce the sequence x_τ, precluding this technique from sexual genetic operators.

In order to use $\rho_{op}(\tau)$ as a predictor of evolution on the landscape, Weinberger notes that one must assume that the landscape is statistically isotropic, and that the fitness distribution of the sequence of points in the search space is stationary. He points out, however, that stationarity is violated in the presence of selection, and that landscapes will depart to varying degrees from isotropy.

Others have avoided this problem by using a single-generation correlation statistic, where $\tau = 1$. One can then incorporate sexual genetic operators by defining a function $g(u, v)$ that combines the fitnesses (u, v) of the two parents. Typically, $g(u, v) = (u + v)/2$, giving:

$$\rho_{op} = \frac{\mathrm{Cov}[w, (u+v)/2]}{(\mathrm{Var}[w]\ \mathrm{Var}[u]/2)^{1/2}},$$

where w is the fitness of offspring from parents with fitness u and v, and Cov and Var are with respect to some measure over the search space.

No one has claimed that the autocorrelation statistic would be an exact estimator of the performance of complex GAs, which is why empirical studies of its applicability have been undertaken. However, as shown in Theorem 2, the properly defined covariance statistics are exact estimators of evolutionary change in the population. The autocorrelation function uses the fitness, $w(x)$, as a measurement function. When $w(x)$ is used as the measurement function in Price's Theorem, one obtains the change in the population mean fitness (see Table 1). However, what is more important to the performance of the GA is the change in the upper tail of the fitness distribution, which is obtained by using $F(x, w)$ as in Theorem 2. This suggests circumstances in which ρ_{op} may be fooled.

4.2.1 Counterexamples for ρ_{op}

I give two constructed examples of landscapes in which ρ_{op} fails to predict GA performance. In them I will define the transmission functions $T(w \leftarrow u, v)$ directly in the fitness domain, as in (2). Thus $T(w \leftarrow u, v)$ contains all the information about the fitness landscape.

It should be noted that the value of ρ_{op} is derived not from the total transmission function $T(w \leftarrow u, v)$, but from just the portion of $T(w \leftarrow u, v)$ that results from application of the genetic operator. In typical GAs, the genetic operator acts with probability $\alpha < 1$. The canonical recursion in the fitness domain (2) is then written:

$$p(w)' = (1 - \alpha)\, p(w)\, w/\overline{w} + \alpha \int_0^\infty P(w \leftarrow u, v)\, \frac{uv}{\overline{w}^2}\, p(u)\, p(v)\, du\, dv, \tag{19}$$

where the transmission function $P(w \leftarrow u, v)$ represents the action of the genetic operator and is referred to as the *search kernel* (Altenberg 1994). Therefore, the values of α and $P(w \leftarrow u, v)$ contain all the information about the landscape that affects performance. The statistic ρ_{op} is always taken with respect to the search kernel $P(w \leftarrow u, v)$.

The following are two cases in which ρ_{op} errs in describing the evolutionary performance of the GA:

1. One case with a maximal parent-offspring correlation, but which gives poor GA performance, because parents never produce offspring fitter than what already exists, and

2. A second case with no correlation between the fitnesses of parents and their offspring, but which nevertheless gives excellent GA performance, because the proportion of offspring that are fitter than their parents is a constant, even as the parental fitnesses increase.

High parent-offspring correlation but no evolvability. Suppose the fitness of an offspring produced by the genetic operator is always the average of the parental fitnesses. If one uses $q(u,v) = (u+v)/2$, so that ρ_{op} is the correlation between mean parental fitness and offspring fitness, then $\rho_{op} = 1$.

However, this is classical blending inheritance, in which the fitness variance of the population rapidly decays. Furthermore, the fittest member of the population will never be greater than the fittest in the first generation. Thus, although the mean fitness of the population will increase initially, it has no evolvability.

Zero parent-offspring correlation but high evolvability. In this example there is no correlation between parent and offspring fitness, yet there is high evolvability because each pair of parents has the same chance of producing still fitter offspring, no matter how fit the parents are within a certain limit. This is achieved with a lognormal distribution:

$$P(w \leftarrow u,v) = \frac{1}{w\,\sigma(u,v)\sqrt{2\pi}} \exp\left\{ \frac{[\ln[w] - \mu(u,v)]^2}{2\sigma(u,v)^2} \right\}$$

where $\mu(u,v)$ and $\sigma(u,v)$ are scalar functions of u and v, derived as follows.

To obtain the desired value $\rho_{op} = 0$, the mean offspring fitness $\phi(u,v)$ is set to a constant for all parents:

$$\phi(u,v) = \int_0^\infty w\,P(w \leftarrow u,v)\,dw = \exp\left[\mu(u,v) + \sigma(u,v)^2\right] \quad \hat{w} \quad \text{for all } u,v.$$

This requires:

$$\mu(u,v) = \ln[\hat{w}] - \sigma(u,v)^2/2.$$

The desired high evolvability is obtained by ensuring that a constant proportion of offspring are fitter than some function $g(u,v)$ of the parents' fitnesses. This requires:

$$\mathcal{N}\left[\frac{\ln[g(u,v)] - \mu(u,v)}{\sigma(u,v)}\right] = C, \tag{22}$$

where $\mathcal{N}[\,]$ is the normal distribution. The choice of $g(u,v)$ is arbitrary, but reasonable examples for defining evolvability would include $g(u,v) = (u+v)/2$, or $g(u,v) = \sqrt{uv}$, or $g(u,v) = \max[u,v]$. Equation (22) gives the condition:

$$\mu(u,v) = \ln[g(u,v)] - s\,\sigma(u,v) \tag{23}$$

where s is the value giving $\mathcal{N}[s] = C$. Together conditions (21) and (23) are solved by:

$$\sigma(u,v) = s - \sqrt{s^2 - 2\ln[g(u,v)/\hat{w}]}. \tag{24}$$

This requires $g(u,v)$ be constrained to $g(u,v) \leq \hat{w}\,e^{s^2/2}$.

Let ρ_{op} be computed as $\mathrm{Cov}[w, q(u,v)]/(\mathrm{Var}[w]\ \mathrm{Var}[q(u,v)])^{1/2}$ for some arbitrary function $q(u,v)$ of the parents' fitnesses u, v, where w is the offspring fitness. Condition (20) gives

$\rho_{op} = 0$, since

$$\begin{aligned}
\mathrm{Cov}[w, q(u,v)] &= \int_0^\infty w\, q(u,v)\, P(w \leftarrow u,v)\, p(u)\, p(v)\, du\, dv\, dw \\
&\quad - \int_0^\infty w\, P(w \leftarrow u,v)\, p(u)\, p(v)\, du\, dv\, dw \int_0^\infty q(u,v)\, p(u)\, p(v)\, du\, dv \\
&\quad \int_0^\infty q(u,v)\, p(u)\, p(v) \left[\int_0^\infty w\, P(w \leftarrow u,v)\, dw\right] du\, dv \\
&\quad - \int_0^\infty p(u)\, p(v) \left[\int_0^\infty w\, P(w \leftarrow u,v)\, dw\right] du\, dv \int_0^\infty q(u,v)\, p(u)\, p(v)\, du\, dv \\
&\quad \hat{w} \int_0^\infty q(u,v)\, p(u)\, p(v)\, du\, dv - 1 * \hat{w} * \int_0^\infty q(u,v)\, p(u)\, p(v)\, du\, dv \\
&= 0.
\end{aligned}$$

Yet with a reasonably small value of α in (19), the fitness distribution will keep increasing in the upper tail, up to fitnesses of at least $\hat{w}\, e^{d^2/2}$, because of the constant rate of producing offspring fitter than $g(u,v)$ even as u, v grow. Even though $\rho_{op} = 0$, this landscape could be described as very smooth, because below the limit $g(u,v) \leq \hat{w}\, e^{s^2/2}$, the neighborhood of any genotype (i.e. its offspring) includes a portion that are fitter than it. Therefore, none of the lower fitnesses can be "local" optima, and the population evolves toward the global limit. So ρ_{op} in this example is not providing an estimate of landscape smoothness either.

4.2.2 A New Statistic

In order to predict the performance of a GA based solely on sampled fitness values, one must assume that the fitnesses are dynamically sufficient, as described in Section 2.1.2. However, in general this assumption will be rendered only an approximation by the occurrence of noninvertibility in the fitness function. In many actual cases, though, it may be a good enough approximation to yield good predictions.

The most complete description of the transmission function in the fitness domain — one which loses no information (assuming the invertibility of $w(x)$) — is simply $T(w \leftarrow u, v)$. Other statistics such as ρ_{op} involve averages that already lose information. Therefore I propose the following:

Conjecture: *When attempting to predict the performance of a genetic algorithm using the fitnesses of a limited sample of points, the best statistic to use should be an estimate of the search kernel*

$$\hat{P}(w \leftarrow u,v) \approx P(w \leftarrow u,v),$$

produced using the values of $w(y)$, $w(z)$ of parents and $w(x)$ of offspring sampled during the GA.

A simple way to proceed in predicting the future course of a GA is to take the estimate of the search kernel $\hat{P}(w \leftarrow u,v)$ and insert it in recursion (19) to simulate the progress of the GA, and predict the evolution of the fitness distribution (an approach also taken by Grefenstette, this volume). It may be possible to analyze the search kernel more directly to predict the performance of the GA. I would propose in addition:

Conjecture: *The critical determinant of GA performance is how rapidly the evolvability — i.e., the likelihood of parents being able to produce offspring fitter than themselves — decays with increasing parental fitness.*

One could classify different representations and operators by the decay rates of the search kernels they produce (e.g. exponential, hyperbolic, etc.). Search kernels with the least decay ought to exhibit the best GA performance.

If there are multiple domains of attraction in the dynamics of the GA, different initial populations may yield divergent estimates of $P(w \leftarrow u, v)$, even when $P(w \leftarrow u, v)$ is dynamically sufficient. A general caveat, therefore, is that the feasibility of predicting the performance of the GA depends on some level of regularity in the adaptive landscape.

The estimation of $P(w \leftarrow u, v)$ based on the fitnesses of a limited sample of points in a run of a genetic algorithm is a problem of generalization. Inference must be made on the sample search kernel. A Bayesian approach toward producing the search kernel estimator would be to begin with a prior distribution over a family of functions $\hat{P}(w \leftarrow u, v)$, and use the sampled data to form a posterior distribution. The function with the maximum posterior likelihood could be taken as the best estimator of $P(w \leftarrow u, v)$. The ability to generalize from the sampled data depends on ones prior distribution (Solla 1990). Generalization requires some knowledge that allows one to narrow the prior distribution to a smaller universe of distribution functions $\hat{P}(w \leftarrow u, v)$, within which one believes the actual function $P(w \leftarrow u, w)$ is likely to be found.

5 CONCLUSIONS

This paper begins with a critique of the Schema Theorem describing why it does not come to bear on the performance of genetic algorithms. The Schema Theorem does not capture the intuitive idea about what makes a GA work — that offspring with above-average fitness can be produced by recombining schemata with above-average fitness. There is a "missing" theorem needed to capture this intuition, and this is Price's Covariance and Selection Theorem. Price's theorem is used to show how changes in different macroscopic properties of populations in a genetic algorithm can be derived by using the microscopic dynamics of the GA combined with the appropriate measurement function. When the measurement function is a fitness indicator function, one obtains the evolution of the fitness distribution over one generation. When the measurement function is a schema indicator function, one obtains the evolution of the schema frequency. Thus, the Schema Theorem can be expressed using Price's theorem. However, the fact that schemata with above-average fitness increase in frequency says nothing about the performance of the GA. The ability for a GA to increase the upper tail of the fitness distribution is necessary for good performance.

This is expressed in a local performance theorem for genetic algorithms. Schemata do not appear in this performance theorem for general representations and operators. When one examines recombination operators specifically, however, schemata reappear in the performance theorem in a way that shows some qualitatively novel aspects of schema processing.

This "missing" schema theorem is obtained by using the recombination distribution representation of transmission introduced by Geiringer (1944). It makes explicit the intuition about how schema processing can provide a GA with good performance, namely: (1) that the recombination operator determines which schemata are being recombined; and (2) that there needs to be a correlation between complementary schemata of high fitness and the fitness distributions of their recombinant offspring in order for the GA to increase the chance of sampling fitter individuals. It also shows the influence of linkage disequilibrium on the performance of the GA.

Finally, the "adaptive landscape" approach to understanding GA performance is discussed. I examine some of the problems that ensue when one defines the landscape using metrics extrinsic to the transmission function. While the properly defined covariance statistics give quantitative estimates for the change in the fitness distribution, as shown in the local performance theorem, the autocorrelation statistics commonly used in landscape analysis do not, and this is illustrated with two examples of landscapes that give GA performance exactly contrary to that predicted by these statistics. I propose that the best estimator for predicting the behavior of a GA is simply the approximation of the transmission function in the fitness domain, and that it is the rate of decay of evolvability as parents increase in fitness that is the critical feature of the transmission function for GA performance. With these statistics calculated for the transmission functions produced by different operators and representations, one may be able to better design genetic algorithms.

APPENDIX

PROOF OF THEOREM 5

The general recursion (1) with the recombination operator becomes:

$$p(\boldsymbol{x})' = \sum_{\boldsymbol{r} \in \{0,1\}^L} R(\boldsymbol{r})\, q(\boldsymbol{x}|\boldsymbol{r}), \tag{25}$$

where

$$q(\boldsymbol{x}|\boldsymbol{r}) = \sum_{\boldsymbol{y},\boldsymbol{z}} \delta(\boldsymbol{x},\, \boldsymbol{r} \circ \boldsymbol{y} + (\boldsymbol{1}-\boldsymbol{r}) \circ \boldsymbol{z}) \frac{w(\boldsymbol{y})\, w(\boldsymbol{z})}{\overline{w}^2}\, p(\boldsymbol{y})\, p(\boldsymbol{z}). \tag{26}$$

The representation of the recursion for the general multi-locus, selection-recombination system was first obtained by Karlin and Liberman (1979) in a form similar to (25). Vose (1990) and Vose and Liepins (1991) independently developed a different representation for (25) assuming two alleles at each locus, which has been called the "exact schema theorem" (Juliany and Vose 1994).

The term $q(\boldsymbol{x}|\boldsymbol{r})$ evaluates to:

$$\begin{aligned}
q(\boldsymbol{x}|\boldsymbol{r}) \quad & \sum_{\boldsymbol{y},\boldsymbol{z}} \delta(\boldsymbol{x},\, \boldsymbol{r} \circ \boldsymbol{y} + (\boldsymbol{1}-\boldsymbol{r}) \circ \boldsymbol{z}) \frac{w(\boldsymbol{y})\, w(\boldsymbol{z})}{\overline{w}^2}\, p(\boldsymbol{y})\, p(\boldsymbol{z}) \\
& \sum_{\substack{\boldsymbol{y}_0 \in \mathcal{H}(\boldsymbol{1}-\boldsymbol{r}) \\ \boldsymbol{y}_1 \in \mathcal{H}(\boldsymbol{r})}} \frac{w(\boldsymbol{x}_0, \boldsymbol{y}_1)\, w(\boldsymbol{y}_0, \boldsymbol{x}_1)}{\overline{w}^2}\, p(\boldsymbol{x}_0, \boldsymbol{y}_1)\, p(\boldsymbol{y}_0, \boldsymbol{x}_1) \\
& \sum_{\boldsymbol{y}_1 \in \mathcal{H}(\boldsymbol{r})} \frac{w(\boldsymbol{x}_0, \boldsymbol{y}_1)}{\overline{w}} p(\boldsymbol{x}_0, \boldsymbol{y}_1) \sum_{\boldsymbol{y}_0 \in \mathcal{H}(\boldsymbol{1}-\boldsymbol{r})} \frac{w(\boldsymbol{y}_0, \boldsymbol{x}_1)}{\overline{w}}\, p(\boldsymbol{y}_0, \boldsymbol{x}_1) \\
& \overline{w}_0(\boldsymbol{x}_0)\, p_0(\boldsymbol{x}_0)\, \overline{w}_1(\boldsymbol{x}_1)\, p_1(\boldsymbol{x}_1) / \overline{w}^2.
\end{aligned}$$

Equation (26) gives:

$$\overline{F}(w)' = \sum_{\boldsymbol{x}} F(\boldsymbol{x}, w)\, p(\boldsymbol{x})' = \sum_{\boldsymbol{r} \in \{0,1\}^L} R(\boldsymbol{r}) \sum_{\boldsymbol{x}} F(\boldsymbol{x}, w)\; q(\boldsymbol{x}|\boldsymbol{r}). \tag{27}$$

The sum on the right evaluates to:

$$\begin{aligned}
\sum_{\boldsymbol{x}} F(\boldsymbol{x}, w)\; q(\boldsymbol{x}|\boldsymbol{r}) &= \sum_{\substack{\boldsymbol{x}_0 \in \mathcal{H}(\mathbf{1}-\boldsymbol{r}) \\ \boldsymbol{x}_1 \in \mathcal{H}(\boldsymbol{r})}} F(\boldsymbol{x}, w) \frac{\overline{w}_0(\boldsymbol{x}_0)\, \overline{w}_1(\boldsymbol{x}_1)}{\overline{w}^2} p_0(\boldsymbol{x}_0)\, p_1(\boldsymbol{x}_1) \\
&= \overline{F}(w) + \mathrm{Cov}[F(\boldsymbol{x}, w), \frac{\overline{w}_0(\boldsymbol{x}_0)\, \overline{w}_1(\boldsymbol{x}_1)}{\overline{w}^2}] \\
&\quad - \sum_{\substack{\boldsymbol{x}_0 \in \mathcal{H}(\mathbf{1}-\boldsymbol{r}) \\ \boldsymbol{x}_1 \in \mathcal{H}(\boldsymbol{r})}} [F(\boldsymbol{x}, w) - \overline{F}(w)][p(\boldsymbol{x}) - p_0(\boldsymbol{x}_0) p_1(\boldsymbol{x}_1)] \frac{\overline{w}_0(\boldsymbol{x}_0)\, \overline{w}_1(\boldsymbol{x}_1)}{\overline{w}^2}
\end{aligned}$$

Substitution into (27) gives the result.

Acknowledgements

I thank Roger Altenberg, Wayne Altenberg, and Darrell Whitley for their consideration during the completion of this paper. Thanks to Heinz Mühlenbein and Michael Vose for incisive comments, Marcy Uyenoyama, Eric Jakobsson, the Hawaii Institute of Geophysics, and the Maui High Performance Computing Center for infrastructural support, and the American Dance Festival for inspiration.

References

Ackley, D. H. 1987. *A Connectionist Machine for Genetic Hillclimbing.* Kluwer Academic Publishers, Boston, MA.

Altenberg, L. 1991. Chaos from linear frequency-dependent selection. *American Naturalist* 138: 51–68.

Altenberg, L. 1994. The evolution of evolvability in genetic programming. In K. E. Kinnear, editor, *Advances in Genetic Programming*, pages 47–74. MIT Press, Cambridge, MA.

Altenberg, L. and M. W. Feldman. 1987. Selection, generalized transmission, and the evolution of modifier genes. I. The reduction principle. *Genetics* 117: 559–572.

Asoh, H. and H. Mühlenbein. 1994. Estimating the heritability by decomposing the genetic variance. Technical Report 94-02-12, GMD, Sakt Augustin, Available by ftp from anonymous@omega.gmd.de under /gmd/as/ga/paper.

Booker, L. B. 1993. Recombination distributions for genetic algorithms. In L. D. Whitley, editor, *Foundations of Genetic Algorithms 2*, pages 29–44. Morgan Kaufmann, San Mateo, CA.

Bürger, R. 1993. Predictions of the dynamics of a polygenic character under directional selection. *Journal of Theoretical Biology* 162: 487–513.

Cavalli-Sforza, L. L. and M. W. Feldman. 1976. Evolution of continuous variation: direct approach through joint distribution of genotypes and phenotypes. *Proceedings of the National Academy of Science U.S.A.* 73: 1689–1692.

Charlesworth, D., M. T. Morgan, and B. Charlesworth. 1992. The effect of linkage and population size on inbreeding depression due to mutational load. *Genetical Research* 59(1): 49–61.

Christiansen, F. B. 1987. The deviation from linkage equilibrium with multiple loci varying in a stepping-stone cline. *Journal of Genetics* 66: 45–67.

Cockerham, C. C. 1954. An extension of the concept of partitioning hereditary variance for analysis ofcovariances among relatives when epistasis is present. *Genetics* 39: 859–882.

Eshelman, L. J., R. A. Caruana, and J. D. Schaffer. 1989. Biases in crossover landscape. In J. D. Schaffer, editor, *Proceedings of the Third International Conference on Genetic Algorithms*, pages 10–19, San Mateo, CA. Morgan Kaufmann.

Feller, W. 1971. *An Introduction to Probability Theory and Its Applications.* John Wiley and Sons, New York, page 27.

Fisher, R. A. 1930. *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford, pages 30–37.

Fontana, W., P. F. Stadler, E. G. Bornberg-Bauer, T. Griesmacher, I. L. Hofacker, M. Tacker, P. Tarazona, E. D. Weinberger, and P. Schuster. 1993. RNA folding and combinatory landscapes. *Physical Review E* 47(3): 2083–2099.

Frank, S. A. and M. Slatkin. 1990. The distribution of allelic effects under mutation and selection. *Genetical Research, Cambridge* 55: 111–117.

Geiringer, H. 1944. On the probability theory of linkage in mendelian heredity. *Annals of Mathematical Statistics* 15: 25–57.

Goldberg, D. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison Wesley.

Goodnight, C. J. 1988. Epistasis and the effect of founder events on the additive genetic variance. *Evolution* 42(3): 441–454.

Grafen, A. 1985. A geometric view of relatedness. *Oxford Surveys in Evolutionary Biology* 2: 28–89.

Grefenstette, J. 1989. Conditions for implicit parallelism. In G. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 252–261. Morgan Kaufmann, San Mateo, CA.

Grefenstette, J. J. 1995. Using fitness distributions of genetic operators to predict performance. In D. Whitley and M. D. Vose, editors, *Foundations of Genetic Algorithms 3.* Morgan Kaufmann, San Mateo, CA.

Grefenstette, J. J. and J. E. Baker. 1989. How genetic algorithms work: a critical look at implicit parallelism. In J. D. Schaffer, editor, *Proceedings of the Third International Conference on Genetic Algorithms*, pages 20–27, San Mateo, CA. Morgan Kaufmann.

Holland, J. H. 1975. *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor.

Juliany, J. and M. D. Vose. 1994. The genetic algorithm fractal. *Evolutionary Computation* 2(2): 165–180.

Karlin, S. 1979. Models of multifactorial inheritance: I, Multivariate formulations and basic convergence results. *Theoretical Population Biology* 15: 308–355.

Karlin, S. and U. Liberman. 1978. Classifications and comparisons of multilocus recombination distributions. *Proceedings of the National Academy of Sciences of the U.S.A.* 75(12): 6332–6336.

Karlin, S. and U. Liberman. 1979. Central equilibria in multilocus systems. I. Generalized nonepistatic selection regimes. *Genetics* 91: 777–798.

Kinnear, K. E. 1994. Fitness landscapes and difficulty in genetic programming. In J. D. Schaffer, H. P. Schwefel, and H. Kitano, editors, *Proceedings of the IEEE World Congress on Computational Intelligence*, pages 142–147, Piscataway N.J.

Koza, J. R. 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* MIT Press, Cambridge, MA.

Manderick, B., M. de Weger, and P. Spiessens. 1991. The genetic algorithm and the structure of the fitness landscape. In R. K. Belew and L. B. Booker, editors, *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 143–150, San Mateo, CA. Morgan Kaufmann Publishers.

Mathias, K. and D. Whitley. 1992. Genetic operators, the fitness landscape and the traveling salesman problem. In R. Männer and B. Manderick, editors, *Parallel Problem Solving from Nature, 2*, pages 219–228, Amsterdam. North-Holland.

Menczer, F. and D. Parisi. 1992. Evidence of hyperplanes in the genetic learning of neural networks. *Biological Cybernetics* 66(3): 283–289.

Michalewicz, Z. 1994. *Genetic Algorithms + Data Structures = Evolution Programs.* Springer-Verlag, Berlin.

Mühlenbein, H. 1991. Evolution in time and space — the parallel genetic algorithm. In G. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 316–338. Morgan Kaufmann, San Mateo, CA.

Mühlenbein, H. and D. Schlierkamp-Vosen. 1993. The science of breeding and its application to the breeder genetic algorithm (BGA). *Evolutionary Computation* 1(4): 335–360.

Price, G. R. 1970. Selection and covariance. *Nature* 227: 520–521.

Price, G. R. 1972. Extension of covariance selection mathematics. *Annals of Human Genetics* 35: 485–489.

Radcliffe, N. 1991. Equivalence class analysis of genetic algorithms. *Complex Systems* 5(2): 183–205.

Radcliffe, N. J. 1992. Non-linear genetic representations. In R. Männer and B. Manderick, editors, *Parallel Problem Solving from Nature, 2*, pages 259–268, Amsterdam. North-Holland.

Radcliffe, N. J. and P. D. Surry. 1995. Fitness variance of formae and performance prediction. In D. Whitley and M. D. Vose, editors, *Foundations of Genetic Algorithms 3*. Morgan Kaufmann, San Mateo, CA.

Reeves, C. and C. Wright. 1995. An experimental design perspective on genetic algorithms. In D. Whitley and M. D. Vose, editors, *Foundations of Genetic Algorithms 3*. Morgan Kaufmann, San Mateo, CA.

Robbins, R. B. 1918. Some applications of mathematics to breeding problems III. *Genetics* 3: 375–389.

Slatkin, M. 1970. Selection and polygenic characters. *Proceedings of the National Academy of Sciences U.S.A.* 66: 87–93.

Solla, S. A. 1990. Supervised learning and generalization. In *Neural Networks: Biological Computers or Electronic Brains*, pages 21–28. Springer-Verlag, Paris, France.

Stadler, P. F. 1992. Correlation in landscapes of combinatorial optimization problems. *Europhysics Letters* 20(6): 479–482.

Stadler, P. F. 1994. Linear operators on correlated landscapes. *Journal de Physique I* 4(5): 681–696.

Stadler, P. F. and R. Happel. 1992. Correlation structure of the landscape of the graph-bipartitioning problem. *Journal of Physics A: Math. Gen.* 25: 3103–3110.

Stadler, P. F. and W. Schnabl. 1992. The landscape of the traveling salesman problem. *Physics Letters A* 161: 337–344.

Syswerda, G. 1989. Uniform crossover in genetic algorithms. In J. D. Schaffer, editor, *Proceedings of the Third International Conference on Genetic Algorithms*, pages 2–9, San Mateo, CA. Morgan Kaufmann.

Syswerda, G. 1993. Simulated crossover in genetic algorithms. In L. D. Whitley, editor, *Foundations of Genetic Algorithms 2*, pages 239–255. Morgan Kaufmann, San Mateo, CA.

Taylor, P. D. 1988. Inclusive fitness models with two sexes. *Theoretical Population Biology* 34: 145–168.

Turelli, M. and N. H. Barton. 1990. Dynamics of polygenic characters under selection. *Theoretical Population Biology* 38: 1–57.

Uyenoyama, M. K. 1988. On the evolution of genetic incompatibility systems: incompatibility as a mechanism for the regulation of outcrossing distance. In R. E. Michod and B. R. Levin, editors, *The Evolution of Sex*, pages 212–232. Sinauer Associates, Sunderland, MA.

Vose, M. D. 1990. Formalizing genetic algorithms. In *Proceedings of the IEEE workshop on Genetic Algorithms, Neural Networks, and Simulated Annealing Applied to Problems in Signal and Image Processing*, Glasgow, UK.

Vose, M. D. 1991. Generalizing the notion of schema in genetic algorithms. *Artificial Intelligence* 50(3): 385–396.

Vose, M. D. and G. E. Liepins. 1991. Punctuated equilibria in genetic search. *Complex Systems* 5(1): 31–44.

Vose, M. D. and A. Wright. 1994. The walsh transform and the theory of the simple genetic algorithm. *Pattern Recognition* In press.

Wade, M. J. 1985. Soft selection, hard selection, kin selection, and group selection. *American Naturalist* 125: 61–73.

Weinberger, E. D. 1990. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological Cybernetics* 63: 325–336.

Weinberger, E. D. 1991a. Local properties of Kauffman's N-k model, a tuneably rugged energy landscape. *Physical Review A* 44(10): 6399–6413.

Weinberger, E. D. 1991b. Fourier and Taylor series on fitness landscapes. *Biological Cybernetics* 65: 321–330.

Weinberger, E. D. and P. F. Stadler. 1993. Why *some* fitness landscapes are fractal. *Journal of Theoretical Biology* 163: 255–275.

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the Sixth International Congress on Genetics* 1: 356–366.