Understanding the Biases of Generalised Recombination: Part II

Riccardo Poli rpoli@essex.ac.uk
Department Computer Science, University of Essex, Colchester, CO4 3SQ, UK

Christopher R. Stephens csteph@essex.ac.uk
Department of Computer Science, University of Essex, UK, and
Instituto de Ciencias Nucleares, UNAM, A. Postal 70-543, México, D.F. 04510

Abstract

This is the second part of a two-part paper where we propose, model theoretically and study a general notion of recombination for fixed-length strings where homologous recombination, inversion, gene duplication, gene deletion, diploidy and more are just special cases. In Part I, we derived both microscopic and coarse-grained evolution equations for strings and schemata for a selecto-recombinative GA using generalised recombination, and we explained the hierarchical nature of the schema evolution equations. In this part, we provide a variety of fixed points for evolution in the case where recombination is used alone, thereby generalising Geiringer's theorem. In addition, we numerically integrate the infinite-population schema equations for some interesting problems, where selection and recombination are used together to illustrate how these operators interact. Finally, to assess by how much genetic drift can make a system deviate from the infinite-population-model predictions we discuss the results of real GA runs for the same model problems with generalised recombination, selection and finite populations of different sizes.

1 Introduction

All search algorithms and search operators, except totally trivial ones, such as random search with resampling, explore the space of tentative solutions with some form of bias. That is, at any particular stage in the search the probability of sampling certain solutions is higher than for others. Understanding these biases is vital if we want to understand how search algorithms behave and how to adjust their behaviour to better fit the problem at hand.

In Evolutionary Algorithm (EA) theory, one important way of analysing the search biases of operators is to determine where, in the space of all possible populations, they would lead a population if they were the only operations used in an EA. To do this, we need to analyse the fixed points of the EA evolution equations.

One area that has been particularly studied, in this respect, is the fixed points of the dynamics of recombination, leading to results related to the classical Geiringer's theorem (Geiringer, 1944) familiar from population genetics. For instance, in (Poli et al., 2002b; Poli et al., 2003) Geiringer's theorem was generalised to the case of homologous crossover on variable-length linear structures, while in (Poli et al., 2002a) this was generalised to the case of subtree crossover. A different approach, based on an analysis of the underlying Markov chain, was taken in (Mitavskiy and Rowe, 2006), where a rather general version of Geiringer's theorem for a *finite* population model was derived. This

result is applicable to non-linear structures such as program trees as exemplified in (Mitavskiy and Rowe, 2005).

The work described in this two-part paper was motivated by the objective of understanding the search biases induced by a large and powerful set of genetic operators. These operators are all special cases of a general operator, called *generalised recombination*, which allows for *any* redistribution of the parental genes to the offspring.

In Part I, we began our analysis by providing some basic definitions relating to the notion of schema and characterising the notion of generalised recombination. We then derived exact microscopic and macroscopic probabilistic models for the evolution of a population of fixed length, ℓ, strings and schemata undergoing selection and generalised recombination, illustrating the features of these models with simple examples. We then explained the hierarchical nature of the schema evolution equations and showed how the theory generalises past work in EAs. In this part, we will use these theoretical results as a starting point for our analysis of the biases of generalised recombination.

Due to the richness of this more general form of recombination, the intrinsic biases of the corresponding operator are much richer and interesting too. To understand these biases, independent of those due to selection or genetic drift, we consider an infinite population version of the model without selection. We show that the bias of generalised recombination, as with homologous recombination, is to destroy correlations, thereby implying that the asymptotic dynamics is governed by that of the order-1 schemata. However, in distinction to the homologous case, here, the dynamics of the ℓ one-schemata themselves is highly non-trivial, there being a mixing between them that can lead to interesting phenomena, such as periodic oscillations, and lateral diffusion of single alleles from one locus to another. To combat the common criticisms that infinite population models and flat fitness landscapes have nothing in common with real GAs, we show how such predicted qualitative behaviour may be observed in real GA runs with selection, investigating in some particular models how the biases of generalised recombination appear and change as a function of population size and fitness landscape, i.e., how the biases of generalised recombination compete with those of selection and genetic drift.

By considering the asymptotic large t behaviour of the model we show also how concrete quantitative predictions can be derived within the model. In particular, we analytically derive the functional form of the fixed point distribution for the population, showing that it takes a product form. Once again, to show the relevance of this result to real GAs we show how, for the latter, the population asymptotes to a product form, showing that generalised recombination has destroyed correlations. Further, in the case where an allele in one locus can eventually migrate to any other, we derive an explicit analytic expression for the fixed point. The validity of this fixed point is verified by explicitly integrating the infinite population equations using an integrator for them-the "Schemulator". Once again, these results are verified for a real GA in that the average behaviour of the runs approaches that of the infinite population for the early part of a run before genetic drift begins to dominate.

This part II is organised as follows. In Section 2 we report a small subset of key results from Part I. We study the biases of generalised recombination by finding the fixed points of the evolution equations for order-1 schemata in Section 3 for different classes of recombination distributions and under the standard assumption of infinite populations. As mentioned in Part I, understanding these biases is important, as the order-1 schemata determine the existence and stability of the fixed-points for higher-

order schemata and strings as well. These are dealt with in Section 4. The infinite-population results given in these two sections allow us to construct a clear picture of the biases of generalised recombination for a general class of recombination distributions and to generalise Geiringer's theorem in Section 5. We discuss the expected behaviour of an evolutionary system under generalised recombination and selection in Section 6. There we also consider how finite-population effects, such as drift, can alter the dynamics of the system. In Section 7 we test our quantitative and qualitative predictions by directly integrating the evolution equations for an infinite population with generalised recombination, with and without selection. We corroborate the technical analysis of generalised recombination by performing real runs of a GA with finite populations of different sizes, again, with and without selection (Section 8). Although modelling natural evolution is beyond the scope of this paper, to illustrate the representational power of generalised recombination and the theory presented in the paper, in Section 9 we show how one can generate models for non-binary strings, for systems with diploid representations, and even systems where multiple chromosomes coexist. We discuss our findings and provide some conclusions and indications of future work in Section 10. Finally, to improve the exposition of this material, most proofs are relegated to Appendix A, for the benefit of theoreticians and more mathematically oriented readers.

2 Some Key Results from Part I

We first need to re-introduce again some notation that was defined in Part I:

- $r = (m, v)$ is a Generalised Crossover Mask (GCM) where $m = m_1 \cdots m_\ell$ is an ℓ-component bit vector and $v = (v_1, \cdots, v_\ell)$ is a vector of integers whose components are in $\mathcal{N}_\ell = \{1, \cdots, \ell\}$.
- $\bar{r} = (\overline{m}, v)$, where $\overline{m}$ is the bitwise negation of the bit-string m, is the negation of GCM r.
- $\mathcal{R}_\ell^\ell$ is the set of all possible GCMs.
- $p_c(r)$ is the probability of choosing any particular GCM r for crossover (Generalised Recombination Distribution, GRD).
- $I_r = \{i : m_i = 1\}$ represents the genes picked out from a parent by r that go to form part of the offspring.
- H_s^a denotes a schema of order 1 with its single defining symbol a at position s.
- $\Gamma(h, I_r) = \bigcap_{i \in I_r} H_{v_i}^{h_i}$ and $\Gamma(h, I_{\bar{r}}) = \bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}$,
- $E[\cdot]$ is the expectation operator.
- $\Phi(h, t)$ is the proportion of strings or schemata of type h in the population at generation t.
- $p(h, t)$ is the probability of selecting a string or a schema of type h as a parent from the population at generation t.
- $D(h) = \{i | h_i \neq *\}$.

- $\mathcal{R}_1(h) = \{r|D(h) \subseteq I_r\},\ \mathcal{R}_2(h) = \{r|D(h) \subseteq I_{\bar{r}}\},\ \mathcal{R}_3(h) = \mathcal{R}_\ell^\ell \setminus \mathcal{R}_1(h) \setminus \mathcal{R}_2(h)$.
- $\mathcal{R}_1'(h) = \{r \in \mathcal{R}_1(h) : \forall i,j \in D(h), i \neq j \implies v_i \neq v_j\}$ and $\mathcal{R}_1''(h) = \mathcal{R}_1(h) \setminus \mathcal{R}_1'(h)$.
- x^y indicates pattern x repeated y times.
- $\bigotimes_i$ is a concatenation operator which concatenates its arguments.
- The order-1 mixing graph associated to a GRD is the graph represented by the connection matrix $C = (c_{ij})$ with $c_{ij} = \delta\big(p_c(*^\ell, *^{i-1}j*^{\ell-i}) > 0\big)$ where $p_c(*^\ell, *^{i-1}j*^{\ell-i})$ is a schema-based coarse graining of the GRD (see Part I) and $\delta(x) = 1$ if x is true, while $\delta(x) = 0$ otherwise.
- A recombination is order-1 mixing if its order-1 mixing graph is strongly connected.
- A recombination component is a strongly connected component of the order-1 recombination graph.
- Similarly, it is possible to define higher-order mixing graphs, etc. (see Part I).

Theorem 1 (Schema evolution equation)**.** *The expected frequency of a schema h at the next generation in a generational GA with any type of selection with replacement and generalised recombination is given by*

$$E[\Phi(h,t+1)] \quad = \quad \sum_{r \in \mathcal{R}_\ell^\ell} p_c(r) p(\Gamma(h,I_r),t) p(\Gamma(h,I_{\bar{r}}),t). \tag{1}$$

Theorem 2 (Schema hierarchy)**.** *The evolution of a string or schema h under selection and generalised recombination is governed by the following equation*

$$E[\Phi(h,t+1)] \quad = \quad \sum_{r \in \mathcal{R}_1'(h)} (p_c(r) + p_c(\bar{r})) p\left(\bigotimes_{k=1}^{|I_r|} \left(*^{v_{i_k} - v_{i_{k-1}} - 1} h_{i_k} \right) *^{\ell - v_{i_{|I_r|}}}, t \right) + \mathfrak{b}_h(t) \tag{2}$$

the first component of which is linear in the selection probabilities of schemata obtained by assigning the positions of the defining characters in h in all possible ways. The forcing term

$$\begin{aligned} \mathfrak{b}_h(t) \quad = \quad & \sum_{r \in \mathcal{R}_1''(h)} (p_c(r) + p_c(\bar{r})) p\Big(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t \Big) \\ + \quad & \sum_{r \in \mathcal{R}_3(h)} p_c(r) p\Big(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t \Big) p\Big(\bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}, t \Big) \end{aligned} \tag{3}$$

is linear-quadratic in the selection probabilities of schemata of lower order than h.

3 Dynamics of Order-1 Schemata

As is clear from Theorem 2 and as discussed in Part I, order-1 schemata play a privileged role in recombination, let us focus on schemata H_s^a where only one allele is specified. By coarse-graining on the recombination distribution (see Part I), the schema

evolution equation (Equation (1) for these schemata transforms into:

$$\begin{aligned} E[\Phi(H_s^a, t+1)] &= \sum_{(m_s, v_s) \in \mathcal{R}_\ell} p_c(*^{s-1} m_s *^{\ell-s}, *^{s-1} v_s *^{\ell-s}) p\left(H_{v_s}^a, t\right) \\ &= \sum_{k=1}^{\ell} p_c(*^{\ell}, *^{s-1} k *^{\ell-s}) p\left(H_k^a, t\right). \end{aligned}$$

That is, the evolution of order-1 schemata is governed by systems of ℓ linear equations. There are as many such systems as the arity of the alphabet adopted for strings. In the binary case $a \in \{0, 1\}$ and so there are two such systems.

So, in general, unlike the case for homologous crossovers, with generalised recombination, order-1 schemata may evolve even on a flat landscape (where $p(h, t) = \Phi(h, t)$ for any schema h). The flat landscape case is interesting as its analysis unveils the biases of genetic operators (McPhee et al., 2001; Poli et al., 2002b; Poli et al., 2003; Langdon and Poli, 2002). These biases become very important whenever selection is not dominating, as, for example, when the algorithm is exploring an area rich in neutral networks.

Let us consider the case of an infinitely large population and a flat landscape. In this case, in vector notation, the system of equations becomes

$$\vec{\Phi}^a(t+1) = A \vec{\Phi}^a(t) \tag{4}$$

where $\vec{\Phi}^a(t) = [\Phi(H_1^a, t), \cdots, \Phi(H_\ell^a, t)]^T$ and $A = (a_{sk})$ is a matrix with elements $a_{sk} = p_c(*^{\ell}, *^{s-1} k *^{\ell-s})$. Since $\sum_{k=1}^{\ell} p_c(*^{\ell}, *^{s-1} k *^{\ell-s}) = p_c(*^{\ell}, *^{\ell}) = 1$ the matrix A is row stochastic, but it is not necessarily column stochastic.

Naturally, the explicit solution for the dynamics of order-1 schemata in the infinite-population flat-landscape case is given by

$$\vec{\Phi}^a(t) = A^t \vec{\Phi}^a(0). \tag{5}$$

However, whether $\lim_{t \to \infty} \vec{\Phi}^a(t)$ exists or not depends on the properties of A.

For the case $\ell = 2$ in (Stephens and Poli, 2005b) we found that, except in special conditions, a fixed point for the proportions of order-1 schemata $\Phi(H_s^a, t)$ exists. This is generally the case for any ℓ. Let us denote such a fixed point by $\Phi^*(H_s^a)$.

3.1 Fixed Points

Let us look for fixed points for the dynamical system defined by Equation 4. They will have to be eigenvectors of the matrix A with an associated eigenvalue $\lambda = 1$.

Because of the row stochasticity of A, it is easy to see that $\mathbf{1} = [1, \cdots, 1]^T$ is an eigenvector of the matrix. That is, for order-1 schemata, a fixed point always exists of the form

$$\Phi^*(H_s^a) = c(a)$$

for $s = 1, \cdots, \ell$, where $c(a)$ is a constant (possibly a different one for each a). Naturally the constants $c(a)$ must obey the conservation of probability for the ℓ sets of order-1 schemata partitioning the search space. That is, we require that, for all s and t, $\sum_a \Phi(H_s^a, t) = 1$. When evaluated at the fixed point, this leads to the following constraint on the values of the $c(a)$'s:

$$\sum_a c(a) = 1.$$

Generally, analytically finding other fixed points may not be simple. Also, determining whether a fixed point is a global attractor for the system is non-trivial.[1] There are, however, some fairly general classes of generalised recombinations where we can probe deeper. We will consider some of these in the following sections.

3.1.1 Detailed Balance

One important general property of fixed points for order-1 schemata, is that if a fixed point exists, then a form of *allele detailed balance* must hold at the fixed point. Detailed balance requires that the sum of the probability of moving the alleles in one locus (schema) to all others be the same as the sum of the probabilities of moving that allele from all other loci to the locus in question. Formally, this requires:

$$\sum_{k\neq s} p_c(*^\ell, *^{s-1}k*^{\ell-s})\Phi^*(H_k^a) = \Phi^*(H_s^a)\sum_{k\neq s} p_c(*^\ell, *^{k-1}s*^{\ell-k}).$$

Naturally, we can add $p_c(*^\ell, *^{s-1}s*^{\ell-s})\Phi^*(H_s^a)$ to both sides of the equation obtaining

$$\sum_{k} p_c(*^\ell, *^{s-1}k*^{\ell-s})\Phi^*(H_k^a) = \Phi^*(H_s^a)\sum_{k} p_c(*^\ell, *^{k-1}s*^{\ell-k}).$$

If we define $\alpha_s = \sum_k p_c(*^\ell, *^{k-1}s*^{\ell-k})$, then we can rewrite the equations for all s in vector notation as follows:

$$A\vec{\Phi}^{a*} = \mathrm{diag}(\alpha_1, \cdots, \alpha_\ell)\vec{\Phi}^{a*}.$$

That is, the fixed point must be a right eigenvector of the matrix $\mathrm{diag}(\alpha_1^{-1}, \cdots, \alpha_\ell^{-1})A$ for detailed balance to hold. (Naturally the fixed point must also be a right eigenvector of A.)

3.1.2 Order-1 Mixing Recombination

Theorem 3. *In a GA with an infinite population, with a flat fitness landscape and using an order-1 mixing GRD $p_c(r)$ (see Part I), for any s the frequencies of order-1 schemata converge asymptotically to*

$$\Phi^*(H_s^a) = \frac{1}{\sqrt{\ell}} u_m \cdot \vec{\Phi}^a(0),$$

where u_m is the normalised left eigenvector corresponding to the largest eigenvalue of the matrix A with elements $a_{sk} = p_c(*^\ell, *^{s-1}k*^{\ell-s})$, and where $\vec{\Phi}^a(0)$ is a vector containing the order-1 schema frequencies in the initial generation. (See Appendix A.1 for a proof.)

This result is compatible with the family of fixed points $\Phi^*(H_s^a) = c(a)$ mentioned earlier, and indicates on which fixed point the system will settle as a function of the initial conditions $\vec{\Phi}^a(0)$. Namely,

$$c(a) = \frac{1}{\sqrt{\ell}} u_m \cdot \vec{\Phi}^a(0).$$

Corollary 1. *Under the conditions of the previous theorem, if the GRD is duplication-free, order-1 schema frequencies converge to the following fixed point*

$$\Phi^*(H_s^a) = \frac{1}{\ell}\sum_i \Phi(H_i^a, 0).$$

(See Appendix A.2 for a proof.)

[1] Naturally, if the GRD is known, one can easily find *numerical* answers to these questions simply by using standard linear algebra techniques.

Naturally, if A is symmetric, i.e. if $a_{sk} = p_c(*^\ell, *^{s-1}k*^{\ell-s}) = p_c(*^\ell, *^{k-1}s*^{\ell-k}) = a_{ks}$, *all* the normalised left eigenvectors coincide with the normalised right eigenvectors, that is $u_i = v_i$ for all i. So, a symmetric recombination distribution leads to the fixed point provided in this corollary (whether or not the GRD is duplication-free).

3.1.3 Homologous Crossover

One other class of recombination distributions where we can say something more explicit about fixed points is the class of homologous crossovers. These are characterised by the fact that only GCMs of the form $r = (m, (1, 2, \cdots, \ell))$ have non-zero probability. So, $a_{sk} = p_c(*^\ell, *^{s-1}k*^{\ell-s}) = \delta(s = k)$ and, so, A is the identity matrix. In this case, as expected, any initial condition is a fixed point for order-1 schemata. That is

$$\Phi^*(H^a_s) = \Phi(H^a_s, 0).$$

3.1.4 Fully Disconnected Recombination Components

Let $Q(p_c)$ be the set of recombination components induced by the GRD $p_c(r)$. The elements of $Q(p_c)$ are (disjoint) sets of integers. Their union is $\{1, \cdots, \ell\}$.

The homologous crossover case is a special one in which the recombination component graph includes ℓ disconnected nodes (i.e., $|Q(p_c)| = \ell$). The order-1 mixing case is one where all nodes belong to a single component (i.e., $|Q(p_c)| = 1$). Let us consider what happens in other cases similar to these, where the loci can be grouped into a number of components, but where the components themselves are completely disconnected. In other words, we consider the case where the recombination component DAG (see Part I) includes $q = |Q(p_c)|$ nodes with $1 < q < \ell$ and *no arcs*.

In this case, with an appropriate renaming of the nodes in the graph, the matrix A is block diagonal, with q blocks. So, effectively we can decompose the vector $\vec{\Phi}^a$ into q sub-vectors $\vec{\Phi}^a_n$ and the matrix A into q squared sub-matrices A_n (the blocks along the diagonal of A) and rewrite the evolution equations for order-1 schemata as:

$$\vec{\Phi}^a_n(t+1) = A_n \vec{\Phi}^a_n(t)$$

for $n \in Q(p_c)$. It is then easy to see that each of these smaller dynamical systems has an eigenvalue $\lambda_n = 1$ with an associated eigenvector $\mathbf{1}$ (where the vector $\mathbf{1}$ has $|n|$ components). So, a fixed point exists of the form

$$\vec{\Phi}^{a^*}_n = c(n, a)\mathbf{1}$$

for $n \in Q(p_c)$, where $c(n, a)$ are constants which depend only on the component n and the allele a. These, again, must respect the conservation of probability and so

$$\sum_a c(n, a) = 1.$$

If each component is order-1 mixing then the results presented in Section 3.1.2 can trivially be generalised to fully disconnected components. In particular, if there is no form of duplication, then we have the following attractors for the components:

$$\Phi^{a*}_s = c(n, a) = \frac{1}{|n|} \sum_{i \in n} \Phi(H^a_i, 0)$$

for all $n \in Q(p_c)$ and for all $s \in n$.

4 Dynamics of Higher-Order Schemata and Strings

In this section we want to analyse the evolution equations for schemata of order higher than 1. Theorem 2 gives us a relatively detailed decomposition of the equations for strings and schemata of higher order. In order to understand the biases of generalised recombination, we will specialise Equation 2 to the case of an infinite population and a flat landscape. In addition, as we have done for order-1 schemata, we will need to consider the evolution equations of entire groups of schemata. For this reason we will need again to rewrite these in vector notation.

To start this process, let us first rewrite the equation using a notation for schemata that makes explicit what are the defining symbols and where they are. As in Part I, we use the notation $H^{a_1,a_2,\cdots,a_n}_{l_1,l_2,\cdots,l_n}$ to represent a generic schema of order $n \leq \ell$ and with defining symbols $a_1, a_2, \cdots, a_n$ at positions $l_1, l_2, \cdots, l_n$. Then, for an infinite population and the schema $h = H^{a_1,a_2,\cdots,a_n}_{l_1,l_2,\cdots,l_n}$, Equation 2 can be rewritten as

$$\Phi(h, t+1) \quad = \quad \sum_{r \in \mathcal{R}'_1(h)} (p_c(r) + p_c(\bar{r}))p\left(H^{a_1,a_2,\cdots,a_n}_{v_1,v_2,\cdots,v_n}, t\right) + \mathfrak{b}_h(t), \tag{6}$$

where, by definition,

$$R'_1(h) = \{r = (m, v) | D(h) \subseteq I_r \text{ and } \forall i, j \in D(h), i \neq j \implies v_i \neq v_j\}$$

and $D(h) = \{l_1, l_2, \cdots, l_n\}$ for the schema in question. Collecting terms in this equation leads to the following rewrite

$$\begin{aligned} &\Phi(h, t+1) \qquad\qquad (7)\\ &= \sum_{v_1 \neq v_2 \neq \cdots \neq v_n = 1}^{\ell} \left[p_c(H^{0,0,\cdots 0}_{l_1,l_2,\cdots,l_n}, H^{v_1,v_2,\cdots,v_n}_{l_1,l_2,\cdots,l_n}) + p_c(H^{1,1,\cdots 1}_{l_1,l_2,\cdots,l_n}, H^{v_1,v_2,\cdots,v_n}_{l_1,l_2,\cdots,l_n})\right] \\ &\times \quad p\left(H^{a_1,a_2,\cdots,a_n}_{v_1,v_2,\cdots,v_n}, t\right) + \mathfrak{b}_h(t) \end{aligned}$$

where appropriate schemata have been used to coarse grain the m and v parts of the GRD $p_c(r)$.

The right-hand side of Equation 7 contains schemata of the same order as h and having the same defining characters $a_1, a_2, \cdots, a_n$ as h but not necessarily at the same loci as h. Indeed, there are $\binom{\ell}{n}$ different ways of placing the alleles $a_1, a_2, \cdots, a_n$ in the ℓ available loci, and, so, for a generic recombination distribution, this is the number of schemata of order n in the right-hand side of the equation. Therefore, in order to determine the dynamics of the system, for each choice of alleles $a_1, a_2, \cdots, a_n$, we need to track $\binom{\ell}{n}$ equations of the same form as Equation 7. In vector notation the system of equations turns into the following dynamical system

$$\vec{\Phi}^{a_1,a_2,\cdots,a_n}(t+1) = A^{a_1,a_2,\cdots,a_n}\vec{\Phi}^{a_1,a_2,\cdots,a_n}(t) + \vec{\mathfrak{b}}^{a_1,a_2,\cdots,a_n}(t), \tag{8}$$

where the element in row $l_1, l_2, \cdots, l_n$ and column $v_1, v_2, \cdots, v_n$ of the matrix $A^{a_1,a_2,\cdots,a_n}$ is given by $p_c(H^{0,0,\cdots 0}_{l_1,l_2,\cdots,l_n}, H^{v_1,v_2,\cdots,v_n}_{l_1,l_2,\cdots,l_n}) + p_c(H^{1,1,\cdots 1}_{l_1,l_2,\cdots,l_n}, H^{v_1,v_2,\cdots,v_n}_{l_1,l_2,\cdots,l_n})$. Naturally, for each n and set of alleles $a_1, a_2, \cdots, a_n$, we have a different dynamical system with a different A matrix and a different $\vec{\mathfrak{b}}$ vector. Before we look at the general fixed points for these systems, let us consider an example.

4.1 Example: Order-2 Schemata

Let us focus on order-2 schemata of the form $H^{a,b}_{s,u}$ where exactly two alleles are specified and $s \neq u$. By applying Theorem 2 and using the more explicit notation introduced in Part I, we can write the schema evolution equations for these schemata as:

$$\begin{aligned} E[\Phi(H^{a,b}_{s,u}, t+1)] &= \sum_{v_s \neq v_u = 1}^{\ell} \left(p_c(H^{0,0}_{s,u}, H^{v_s,v_u}_{s,u}) + p_c(H^{1,1}_{s,u}, H^{v_s,v_u}_{s,u})\right) p\left(H^{a,b}_{v_s,v_u}, t\right) \\ &+ \delta(a=b) \sum_{v_s=1}^{\ell} \left(p_c(H^{0,0}_{s,u}, H^{v_s,v_s}_{s,u}) + p_c(H^{1,1}_{s,u}, H^{v_s,v_s}_{s,u})\right) p\left(H^{a}_{v_s}, t\right) \quad (9) \\ &+ \sum_{v_s,v_u=1}^{\ell} \left(\sum_{m_s \neq m_u \in \{0,1\}} p_c(H^{m_s,m_u}_{s,u}, H^{v_s,v_u}_{s,u}) \right) p\left(H^{a}_{v_s}, t\right) p\left(H^{b}_{v_u}, t\right). \end{aligned}$$

So, the evolution of order-2 schemata is governed by systems of equations with a linear part which depends on the selection probability of schemata of order 2, and a non-linear forcing term which depends on order-1 schemata. For an infinitely large population and a flat landscape, the order-2 schema equations become

$$\vec{\Phi}^{a,b}(t+1) = A^{a,b}\vec{\Phi}^{a,b}(t) + \vec{\mathfrak{b}}^{a,b}(t)$$

where $\vec{\Phi}^{a,b}(t) = [\Phi(H^{a,b}_{1,2}, t), \Phi(H^{a,b}_{1,3}, t), \cdots, \Phi(H^{a,b}_{\ell-1,\ell}, t)]^T$. The number of such systems depends on the arity of the alphabet adopted for strings. Because of the symmetry $H^{a,b}_{i,j} = H^{b,a}_{j,i}$, in the binary case ($a, b \in \{0,1\}$) there are three such systems: one for $a = b = 0$, one for $a = b = 1$ and a joint one for $a = 0, b = 1$ and $a = 1, b = 0$.

4.2 Fixed Points

In order to say something more about the properties of the dynamical systems in Equation 8 we need to introduce the following general result (see Appendix A.3 for a proof):

Lemma 1. *Consider a system of linear difference equations of the form*

$$x(t+1) = Ax(t) + b(t) \quad (10)$$

where $A = (a_{ij})$ is a square matrix such that $a_{ij} \geq 0$ and $\sum_j a_{ij} < 1$, while $b(t)$ is a non-negative vector with $|b(t)| \leq B < \infty$ such that $\lim_{t\to\infty} b(t) = b^*$ exists. If there exist an n such that all the elements of A^n are positive, then the system has the unique, global attractor $x^* = (I - A)^{-1}b^*$ and such an attractor is non-negative.

As we have already seen, for a schema of order n, the right-hand side of Equation 3 contains schemata of order at most $n-1$. Also, the source term $\mathfrak{b}_h(t)$ is a probability and so $\mathfrak{b}_h(t) \in [0,1]$. If the recombination distribution is one where order-1 through to order $n-1$ schemata converge, then the vector $\vec{\mathfrak{b}}^{a_1,a_2,\cdots,a_n}(t)$ converges to a non-negative vector $\vec{\mathfrak{b}}^{a_1,a_2,\cdots,a_n *}$. If, in addition,

$$\sum_{v_1 \neq v_2 \neq \cdots \neq v_n = 1}^{\ell} \left[p_c(H^{0,0,\cdots 0}_{l_1,l_2,\cdots,l_n}, H^{v_1,v_2,\cdots,v_n}_{l_1,l_2,\cdots,l_n}) + p_c(H^{1,1,\cdots 1}_{l_1,l_2,\cdots,l_n}, H^{v_1,v_2,\cdots,v_n}_{l_1,l_2,\cdots,l_n}) \right] < 1$$

and the recombination distribution is order-n mixing (see Part I), then, by Lemma 1, there is a unique, non-negative fixed point

$$\vec{\Phi}^{a_1,a_2,\cdots,a_n *} = (I - A^{a_1,a_2,\cdots,a_n})^{-1}\ \vec{\mathfrak{b}}^{a_1,a_2,\cdots,a_n *}. \quad (11)$$

The elements of the vector $\vec{\mathfrak{b}}^{a_1,a_2,\cdots,a_n*}$ are polynomials in the fixed-point proportions of schemata whose defining symbols form proper subsets of $\{a_1, a_2, \cdots, a_n\}$. So, the fixed points for one order are hierarchically constructed from the $A^{a_1,a_2,\cdots,a_n}$ matrices and $\vec{\mathfrak{b}}^{a_1,a_2,\cdots,a_n*}$ vectors for lower orders.

Note, we called the result in Equation 11 a "unique fixed-point" or a "unique global attractor". By this we mean that $\vec{\Phi}^{a_1,a_2,\cdots,a_n*}$ is uniquely determined by the limit $\vec{\mathfrak{b}}^{a_1,a_2,\cdots,a_n*}$. Under the appropriate conditions on the GRD this limit exists and, therefore, is unique *for any given set of initial conditions*. So, Equation 11 is really a *family of fixed-points*, different fixed-points in the family being reached depending on the initial conditions.

5 Generalised Geiringer Manifold and Geiringer Theorem

We now want to restrict our attention to the case where $p_c(m, v) = 0$ for all v such that $\exists i \neq j, v_i = v_j$, that is let us assume no allele duplication can take place. We still assume a flat landscape and an infinite population.

Let us analyse this case with a simple example first. For this purpose we consider order-2 schemata again. Under the assumptions mentioned above we can simplify Equation 9 obtaining:

$$\begin{aligned}\Phi(H^{a,b}_{s,u}, t+1) &= \sum_{v_s \neq v_u = 1}^{\ell} \left(p_c(H^{0,0}_{s,u}, H^{v_s,v_u}_{s,u}) + p_c(H^{1,1}_{s,u}, H^{v_s,v_u}_{s,u})\right) \Phi\left(H^{a,b}_{v_s,v_u}, t\right) \\ &+ \sum_{v_s \neq v_u = 1}^{\ell} \left(\sum_{m_s \neq m_u \in \{0,1\}} p_c(H^{m_s,m_u}_{s,u}, H^{v_s,v_u}_{s,u}) \right) \Phi\left(H^{a}_{v_s}, t\right) \Phi\left(H^{b}_{v_u}, t\right)\end{aligned} \tag{12}$$

where we can limit the second summation in v_s and v_u to $v_s \neq v_u$ because the GRD is duplication free. What would a fixed point for this equation look like? Do we really need to invert a (potentially big) matrix to find it (c.f. Equation 11)?

As we have seen in the previous section the fixed point for a schema of order n depends on the fixed points for schemata of order $n-1$ or lower. In this case $n = 2$ and, so, the fixed-points for Equation 12 will depend on the fixed points for order-1 schemata only. We know that fixed points of the form $\Phi^*(H^x_i) = c(x)$ exist for such schemata. Could there be fixed points for order-2 schemata that are of a simple form like this? The answer is "yes". Indeed, it is trivial to verify by direct substitution in Equation 12 that, if $\Phi^*(H^x_i) = c(x)$ is a fixed-point for order-1 schemata, then

$$\Phi^*(H^{x,y}_{i,j}) = c(x)c(y)$$

is a fixed point for the order-2 schemata (irrespective of what function $c(x)$ is).

This result generalises to any n as shown by the following result (see Appendix A.4 for a proof):

Theorem 4 (Generalised Geiringer manifold)**.** *A fixed point distribution for the proportion of a string or a schema $h_1h_2\cdots h_\ell$ under generalised crossover with a duplication-free recombination distribution for an infinite population operating on a flat fitness landscape is given by*

$$\Phi^*(h_1 \cdots h_\ell) = \prod_{q \in Q(p_c)} \prod_{i \in q} c(q, h_i) \tag{13}$$

where $c(q, *) = 1$.

This result is important because *it provides a generalisation of the manifold described, for homologous crossover, by Geiringer* (Geiringer, 1944). All points on our generalised Geiringer manifold are fixed points for a genetic system under generalised recombination. Naturally, the result also covers all the fixed points for order-1 schemata described in Section 3.

It is interesting to rewrite Equation 13 in a slightly different form. If $\nu(h,n,a)$ represents the number of times symbol a appears in one of the loci in component n of the string or schema h, and Ω represents our alphabet, then

$$\Phi^*(h) = \prod_{n \in Q(p_c)} \prod_{a \in \Omega} (c(n,a))^{\nu(h,n,a)}. \tag{14}$$

So, for example, if our alphabet is $\Omega = \{0,1,2,3\}$, $|Q(p_c)| = 1$ and we set $c(n,0) = c(n,1) = 1/3$ and $c(n,2) = c(n,3) = 1/6$, then $\Phi^*(0102) = (1/2)^2 \times (1/2) \times (1/3) \times (1/3)^0 = 1/24$. Interestingly, in the case of a binary alphabet, for a fixed $c(n,0)$, the probability of sampling a given string is only a function of the unitation value (the number of ones) of the string.[2]

Naturally, although any choice of $c(n,a)$ will provide a formal fixed point for the evolution equations, we are only interested in choices which respect the conservation of probability constraint $\sum_a c(n,a) = 1$. Despite this constraint, we still have a huge family of potential fixed points. An important question is whether any of these fixed points would be a global attractor for the system and whether this would depend on initial conditions and, if so, how.

In (Stephens and Poli, 2005b) we presented an exact and general solution for the dynamics for the case $\ell = 2$ and a complete analysis of the corresponding fixed points. The techniques used there can provide exact answers also for $\ell > 2$. However, the complexity of the solutions grows very quickly with ℓ. Fortunately, for order-n mixing recombination distributions we can say something more without the need for a complete eigenvalue/eigenvector analysis.

Theorem 5 (Generalised Geiringer Theorem)**.** *Under the conditions of the previous theorem, if, additionally, the GRD is order-1 through to order-ℓ mixing and*

$$\sum_{v_1 \neq v_2 \neq \cdots \neq v_n = 1}^{\ell} \left[p_c(M^{0,0,\cdots 0}_{l_1,l_2,\cdots,l_n}, V^{v_1,v_2,\cdots,v_n}_{l_1,l_2,\cdots,l_n}) + p_c(M^{1,1,\cdots 1}_{l_1,l_2,\cdots,l_n}, V^{v_1,v_2,\cdots,v_n}_{l_1,l_2,\cdots,l_n}) \right] < 1$$

for all n, then

$$\Phi^*(h_1 \cdots h_\ell) = \prod_{i=1}^{\ell} c(h_i) \tag{15}$$

with

$$c(a) = \frac{1}{\ell} \sum_i \Phi(H_i^a, 0) \tag{16}$$

is a global attractor for the proportions of schema (or string) $h_1 \cdots h_\ell$.

Proof. We know that, for an order-1 to -ℓ mixing crossover with the property stated above, there is a unique fixed point (for each set of initial conditions): $(I - A)^{-1} b^*$. The previous theorem has provided us with one (in this case there can only be one

[2] For binary strings $c(n,1) = 1 - c(n,0)$.

component in the order-1 recombination graph), so the unique global attractor must be of the form $\Phi^*(h_1 \cdots h_\ell) = \prod_{i=1}^{\ell} c(h_i)$. The remaining question is to identify the function $c(x)$. Corollary 1 gives us the answer. □

6 Non-flat Landscapes and Finite Populations

In the previous sections we have analysed the behaviour of generalised crossover in isolation, that is in the absence of selection or other operators. This is clearly an important first step in order to understand how crossover samples the search space. However, these results, although of theoretical interest, would be of little practical use if we could not relate them to the behaviour exhibited by crossover *in the presence* of selection and finite populations.

6.1 Recombination and Selection

Let us try to understand what we should expect to see when generalised recombination and selection are used in conjunction.

Over the years a large body of theoretical evidence has been gathered for GAs operating on fixed length strings (often borrowing from population genetics). For example, we know that homologous crossovers (such as uniform crossover or one-point crossover) try to push the GA towards populations where the alleles in the initial population are perfectly mixed — the classical Geiringer manifold (Geiringer, 1944). This is illustrated in Figure 1. Also, the case where selection is the only operator acting on a population has been studied extensively. We know, for example, that in general, selection tries to push the GA towards populations containing only copies of the best string in the population (assuming that only one type of individual has the highest fitness in the initial generation). Importantly, all homogeneous populations belong also to the Geiringer manifold for homologous crossovers (with these operators, when one crosses an individual with itself, one gets that individual back again). This is illustrated in Figure 2. Finally, we know that mutation pushes the population towards a particular point in population space where all possible strings are equally represented in the population (Vose, 1999) and that this point also belongs to the Geiringer manifold for homologous crossovers. In addition, we have quite clear characterisations of the behaviour of each combination of two operators.

Clearly in a system operating under generalised recombination, the selection and mutation biases will remain the same.[3] The crossover bias, however, should generally be expected to be different. In Section 5 we have clarified what the Geiringer manifold for a duplication-free generalised recombination is, or, in other words, what it means for a population to be perfectly mixed. One important point to notice here is that, unlike for homologous crossover, not all possible homogeneous populations are necessarily on the Geiringer manifold for a generalised recombination. For example, with a fully-mixing (see Part I), duplication-free generalised recombination, and for a binary-string representation, only two homogeneous populations are on the manifold, namely the population where all strings are $11\cdots1$ and the population where all strings are $00\cdots0$. It is easy to understand why other homogeneous populations are not. Consider, for example, a population made up of copies of the string 001. If we use a GCM of the form $r = (m, v)$ where v is a permutation with $v_3 \neq 3$, then crossover could (and would) produce something different from 001. For example, $r = (m, (3, 2, 1))$ would produce

[3] We have not included mutation in our infinite population model in this paper. However, as shown in (Stephens and Poli, 2005a), extending it to include mutation is not hard.

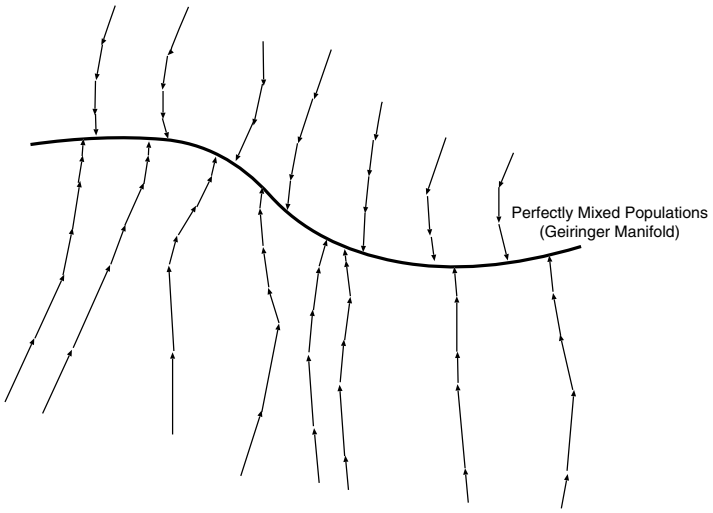

Figure 1: Crossover bias. When crossover alone acts on the population, the system moves towards the Geiringer manifold. The trajectory of the system and the fixed-point reached on the manifold depend on the initial conditions.

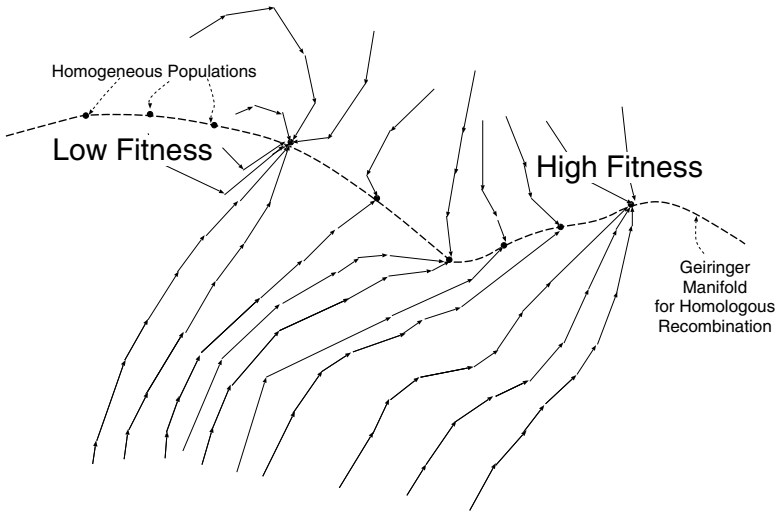

Figure 2: Selection bias. The plane represents the space of all possible populations. Circles represent homogeneous populations. Arrows represent the possible trajectories of a GA initialised in different areas of the population space. The system tends to converge towards high fitness populations, but this may or may not happen (depending on its initial state).

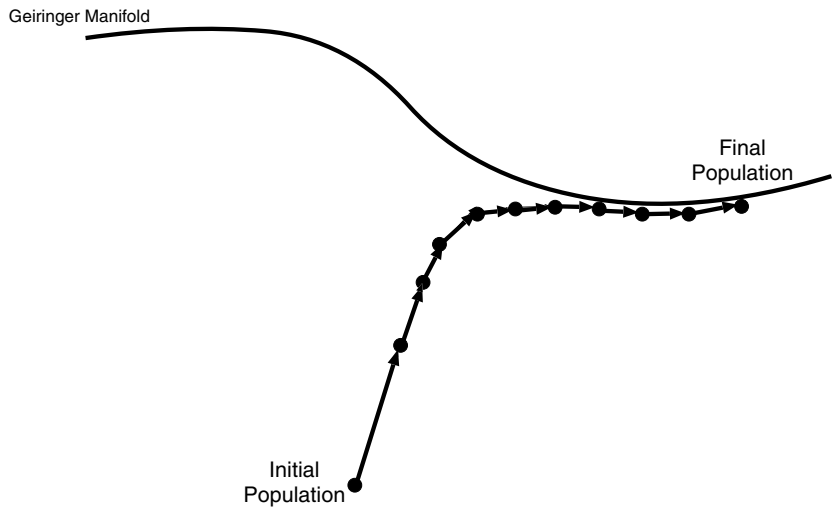

Figure 3: Typical interaction between the selection bias and the bias of generalised recombination. Crossover tries to push the population towards the Geiringer manifold (the sub-space of all perfectly mixed populations). Selection tries to drive the population towards highly-fit homogeneous populations. Since most (if not all) of these populations are not on the generalised Geiringer manifold, we should expect to see diphasic trajectories for the system similar to the one shown.

instances of 100. So, the system could move away from the original homogeneous population.

In previous work (Stephens and Poli, 2005c), we showed (integrating the schema evolution equations for a 3-bit system solving a one-max problem) that, in the presence of weak selection, the behaviour shown by crossover in the absence of selection is still present and effectively dominates the behaviour in the early phases of a run. Here we want to argue that this effect is general and we should expect to see crossover bring the system nearer to the Geiringer manifold in virtually all cases unless the system is near it from the start.

In fact, unless the GA is initialised near the manifold we should generally expect to see trajectories like the ones illustrated in Figure 3. That is, we expect that there would be a strong initial crossover bias and that that could be even stronger than the pressure exerted by selection. So, initially the system should move quickly towards the Geiringer manifold. As it gets near it, the pressure exerted by crossover should decrease and so the selection pressure would start really driving the system towards high fitness areas. In this second phase, we would expect the trajectory to remain close to the Geiringer manifold since any significant deviation from it would produce an increased reaction from crossover. As selection succeeds in driving the system towards high-fitness areas and diversity decreases, the selection pressure is expected to drop, thereby allowing the system to get nearer the Geiringer manifold. There is a limit, however, to how close it can get: if the highest-fitness homogenous population towards which selection is driving the system is not on the manifold, then the system will settle for areas where the crossover and selection biases are minimised and in "equilibrium".[4]

[4]There are cases where selection does not oppose at all the crossover bias. We know, for example, that

In the absence of selection, the manifold is described by the factorisation in Equation 15. However, in the presence of selection the average allele frequencies $c(a)$ are expected to vary over time according to the equation

$$c(a,t) = \frac{1}{\ell} \sum_i \Phi(H_i^a, t)$$

and, so, after a rapid transient phase we should expect the string distribution of the population to become approximately factorised in the form

$$\Phi(h_1 \cdots h_\ell, t) = \prod_{i=1}^{\ell} c(h_i, t), \tag{17}$$

where, unlike Equation 16, the factors are time varying. This applies to both strings and schemata of any order (for schemata for any $h_i = *$ one can simply set $c(*, t) = 1$).[5]

As we will show later, there is empirical support for this conjectured behaviour. Nonetheless, all of this is expected to happen in the case of infinitely large populations. So, an important question is: what should we expect in the case of finite populations?

6.2 Genetic Drift and Crossover Drift

In finite populations there are two sources of randomness that are not present in infinite ones. Firstly, there can be sampling errors in the selection process. That is, because of the finite number of selection steps performed to create each new generation, it is almost never the case that parents are selected with frequencies that match exactly their selection probabilities. Secondly, because we only have a finite number of possible crossover events, in each particular run GCMs are chosen with frequencies that do not match those in the chosen GRD.

6.2.1 Genetic Drift

The effects of the randomness due to selection can be evaluated by studying the changes in string/gene frequencies in a system where reproduction is performed randomly without selection (or with selection but assuming all individuals have equal fitness). In these conditions such variations are termed *genetic drift* (Ridley, 1993; Bäck et al., 2000). We know that, in the absence of selection, genetic drift would normally lead to a loss of diversity and eventually to the convergence to a homogeneous population containing only copies of the same string (Bürger, 2000; Rogers and Prügel-Bennett, 1999). In this sense drift behaves similarly to selection. What is different is that, in the absence of fitness, under drift, any element of the search space is equally likely to eventually take over the population (at least as long as the population is initialised uniformly at random).

Naturally, this effect is present irrespective of whether or not other genetic operators are used and, so, we should expect to see genetic drift in finite populations undergoing generalised crossover. However, how quickly genetic drift leads a population to

in the case of a multiplicative landscape (where the fitness function can be factorised as $f(h_1 \cdots h_\ell) = \prod_i f_i(h_i)$) and with homologous crossover, selection does not induce correlations between loci. So, if the population is on the Geiringer manifold for homologous crossover, it will stay there, while if it isn't it will quickly get on it and stay there for the rest of the run (typically "sliding" towards fitter areas of the search space). For the case of generalised recombination, there is evidence suggesting that functions of unitation (such as one-max) and, more generally, functions of multitation play the same role with respect to the generalised Geiringer manifold. We will explore this topic in a future paper.

[5] Equation 17 is a generalisation of what was found for homologous crossover and selection in (Stephens, 2001).

fixation and whether or not all homogeneous populations are equally likely as fixation points depends on the search biases of the operators chosen.

In fitness proportionate selection, the selection drift can be reduced, by using Stochastic Universal Sampling (SUS) instead of the standard roulette-wheel selection (Baker, 1987). Conceptually, SUS can be seen as a form of roulette-wheel selection where the roulette has M equally spaced pointers instead of one, M being the population size. Then, we can imagine that this new wheel is spun only once, rather than M times, to select the individuals to place in the mating pool. This clearly reduces the noise in the selection process and, so, we will adopt this form of fitness-proportionate selection in the experiments reported in Section 8.

6.2.2 Crossover Drift

We are not aware of any studies where the effects of the randomness due to recombination not picking GCMs with frequencies that match those in the GRD have been evaluated. Certainly we expect, that like with genetic drift, this would cause random variations in the gene/string frequencies with respect to those expected in an infinite population. We will term such variations *crossover drift*.

In order to assess the effects of crossover drift one needs to know the probability distribution for finite-population GCM frequencies. A quantitative analysis of crossover drift is beyond the scope of this paper. Suffice it to say here that this is clearly doable. For example, if the crossover operator is such that it can be modelled *exactly* by drawing GCMs from an urn, then, clearly, the masks would follow a multinomial distribution with success probabilities given by the GRD. So, by using our exact schema equations we could certainly predict the variance in string and schema frequencies caused by crossover drift.

Interestingly, standard implementations of one-point crossover, two-point crossover, uniform crossover, etc. produce multinomial mask distributions. So, in the experimental results in Section 8, we have implemented our crossover so that GCMs were multinomially distributed. Achieving this is straightforward — it is sufficient to apply the roulette-wheel algorithm on the GRD (rather than the population fitness vector).[6]

6.2.3 Combined Effects

We expect selection and crossover drifts to have an effect on the applicability to finite populations of some of our infinite population results. In particular, we note that perfect detailed balance (see Section 3.1.1) can only be achieved in infinitely large populations since it requires that all possible recombination and selection events happen with frequencies that match exactly their corresponding expected probabilities. So, any property of a population that depends crucially on detailed balance may not hold in the presence of drift (e.g., in finite populations). In particular, some fixed points available for infinite populations may disappear altogether in finite populations.

In the case of generalised crossover, as we have noted before, if generic duplication is allowed, then crossover will push the population towards a population containing only one of the alleles in Ω (in the binary case, only 0's or only 1's). However, with finite populations and in the absence of selection, even a fully mixing *duplication-free* generalised crossover produces the same effect irrespective of the initial conditions, contrary

[6] Obviously, one could pick GCMs by applying some other algorithm. A form of SUS could, for example, significantly reduce crossover drift, particularly in cases where the number of GCMs having non-zero probability is small compared with the number of individuals created by crossover. We intend to explore this idea in future research.

to what is stated in Theorem 5 which holds for infinite populations. This is because the fixed-points for order-1 schemata provided in Section 3 depend on precise detailed balance. However, due to drift this is not guaranteed. As a result, in finite populations and for a flat landscape, we should observe variations in the average frequency of alleles of any given type (the quantities $c(a)$ in Equation 16) even for a fully-mixing duplication-free GRD. Crossover is still expected to drive and keep the population near the generalised Geiringer manifold, but the drift on the $c(a)$'s is then expected to cause the population to randomly slide along the manifold until it reaches a population where only alleles of one type are present.

In the presence of selection, we should see drift, selection and crossover constantly trying to find a mutually agreeable equilibrium point. As a result, with finite populations we cannot expect to see the predictions of our schema-theoretic infinite population model to hold exactly. We should, however, see behaviours that are consistent with the drift-free, infinite-population model. We will see examples of these effects happening in Section 8.

7 "Schemulator" Runs

In order to better understand the infinite-population dynamics of a genetic system under selection and generalised recombination we have implemented an integrator written in Java (we call it the "*Schemulator*"–a contraction of "schema simulator") which expands and then numerically integrates the string (and schema) evolution equations for any choice of recombination distribution, of fitness function and of initial conditions.

To corroborate our theoretical results we want to first verify our predictions as to the existence and location of fixed points for the flat fitness landscape case. Figure 4 shows the dynamics of some schemata and strings in a population with $\ell = 3$ and a duplication-free recombination distribution where $p_c(m, v) \neq 0$ for all the 48 GCMs where v is a permutation vector, and $p_c(m, v) = 0$ for the remaining 168 GCMs. The non-zero entries of the GRD were randomly generated and then normalised so that $\sum p_c(r) = 1$. The resulting recombination distribution had only one component, $\mathcal{N}_\ell = \{1, \cdots, \ell\}$, which includes all ℓ loci. In order to be able to distinguish between the dynamics of different schemata, we used unequal initial proportions for strings, namely: $\Phi(000, 0) = 0.3$, $\Phi(001, 0) = 0.25$, $\Phi(010, 0) = \Phi(011, 0) = \Phi(100, 0) = 0.1$, $\Phi(101, 0) = 0.05$, $\Phi(110, 0) = 0.02$ and $\Phi(111, 0) = 0.08$.

As shown in Figure 4, the order-1 schemata $1**$, $*1*$ and $**1$ rapidly converge to a fixed point where $\Phi^*(1**) = \Phi^*(*1*) = \Phi^*(**1)$. This is exactly what is predicted by Equation 13. The order-one-schema fixed point proportion, 0.343, suggests that $c(\mathcal{N}_\ell, 1) = \frac{103}{300} \approx 0.343$ and $c(\mathcal{N}_\ell, 0) = 1 - c(\mathcal{N}_\ell, 1) = \frac{197}{300} \approx 0.657$.

Order-2 schemata also converge to identical values, i.e. $\Phi^*(11*) = \Phi^*(*11) = \Phi^*(1*1)$. The fixed-point proportion is (within numerical errors) exactly $c(\mathcal{N}_\ell, 1)^2 = \left(\frac{103}{300}\right)^2 \approx 0.119$, which is what Equation 13 predicts.

The predictions of our generalised Geiringer theorem also hold for strings. For example, the strings 110 and 011 converge to their predicted fixed point $\Phi^*(110) = \Phi^*(011) = c(\mathcal{N}_\ell, 1)^2 c(\mathcal{N}_\ell, 0) = \frac{103 \times 197^2}{300^3} \approx 0.077$ and 111 converges towards the predicted $\Phi^*(111) = c(\mathcal{N}_\ell, 1)^3 = \left(\frac{103}{300}\right)^3 \approx 0.040$ within numerical errors.

Where did the magic value $\frac{103}{300}$ come from? The GRD used in this example is one covered by Corollary 1, i.e. it is one for which $c(n, a) = \frac{1}{|n|} \sum_{i \in n} \Phi(H_i^a, 0)$. Since in this

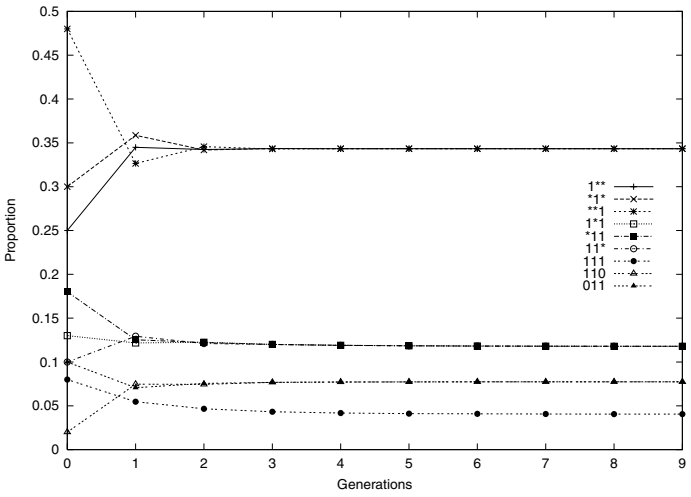

Figure 4: Infinite-population dynamics of strings and schemata for $\ell = 3$ and a duplication-free, order-1 mixing, random recombination distribution.

particular example we only have one component,

$$\begin{aligned} c(\mathcal{N}_\ell, 1) &= \frac{1}{3}\left(\Phi(1{*}{*}, 0) + \Phi(*1*, 0) + \Phi(**1, 0)\right) \\ &= \frac{1}{3}\left(\frac{1}{4} + \frac{3}{10} + \frac{12}{25}\right) = \frac{103}{300} \approx 0.343. \end{aligned}$$

To understand the behaviour of systems where crossover is used in conjunction with selection we used Schemulator runs using both a linear problem (the one-max problem for $\ell = 3$) and a problem with the maximum degree of epistasis (the needle-in-a-haystack problem for $\ell = 4$).

Let us first consider one-max. Also in this case we used all GCMs where v is a permutation list, that is the GRD was duplication-free. Each of the 48 allowed masks had a probability of being chosen of $1/48$. In order to distinguish between the dynamics of different schemata, once again we started with an asymmetric initial population: $\Phi(000, 0) = 0.2$, $\Phi(001, 0) = \Phi(010, 0) = \Phi(011, 0) = \Phi(101, 0) = \Phi(110, 0) = 0.1$, $\Phi(100, 0) = 0.3$, and $\Phi(111, 0) = 0.0$. Figure 5 shows plots of the proportions of four representative strings (000, which has fitness 0, 001, which has fitness 1, 110, with fitness 2, and the optimal string 111, which has fitness 3). For comparison, the figure also reports the value of $\prod_{i=1}^{\ell} c(h_i, t)$ for the four strings. It is apparent how, within very few generations, the string proportions approach the factorised form $\prod_{i=1}^{\ell} c(h_i, t)$ predicted in the previous section (Equation 15). That is, after a transient, the population moves along the Geiringer manifold for generalised crossover.

Let us now turn to the needle-in-a-haystack problem. In this case we considered a 4 bit problem and again we used a duplication-free GRD, thereby using 384 GCMs out of the possible 4096. Each mask had a probability of being chosen of $1/384$. Also in this case we initialised the population asymmetrically giving an initial proportion of 0.1 to strings 0000, 0010, 0111, 1000, 1001 and 1011, a proportion of 0.2 to strings 0001 and 1100, and zero proportions to all other strings. Note that the optimum string 1111 was

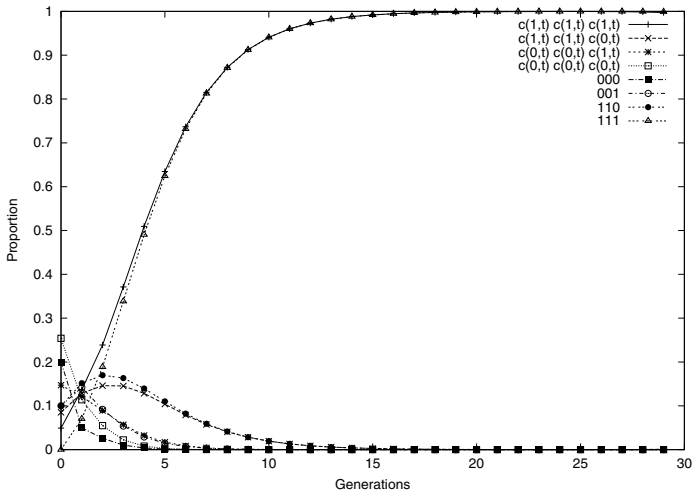

Figure 5: Infinite-population dynamics of a representative sample of strings in a 3-bit GA solving the one-max problem in the presence of generalised recombination and selection. The predicted trajectory along the Geiringer manifold for the strings is also shown (Equation 15).

not present in the initial population. This string had fitness 2, while all other strings had fitness 1.

Figure 6 shows five representative strings (0000 which has Hamming distance 4 from the needle, 0010 which has Hamming distance 3 from the needle, 0110 with distance 2, 0111 with distance 1 from the needle and the needle 1111 itself). For comparison, the figure also reports the value of $\prod_{i=1}^{\ell} c(h_i, t)$ for the five strings. Even if in this case string proportions approach the factorised form predicted in the previous section (Equation 15) more slowly than in the one-max problem, the similarity between the dynamics of the two is apparent.[7] This suggests that even in the most extreme epistatic conditions, a GA quickly moves near the Geiringer manifold and then "slides" along it under the effect of selection.

8 Real GA Runs

Let us now verify experimentally how good the predictions of our infinite population model are. To do this we considered the same three problems we used in the previous section: a 3-bit flat-landscape problem with and without uniform initialisation, a 3-bit one-max problem, and a 4-bit needle-in-a-haystack problem. In order to assess the effects of population size, M, we studied the cases $M = 4$, $M = 10$ and $M = 100$. For all problems and settings we did 1,000 independent runs. In the runs we used generalised recombination and fitness proportionate selection. Recombination returned one offspring per parent pair and it was applied with 100% probability. In these conditions we need to perform M crossovers to create a new generation, and so we need a total of $2M$ selection steps per generation. Selection used SUS.

[7] Of course, had we initialised the population in the standard way, i.e. uniformly at random, rather than with asymmetric proportions, the population would have been on the Geiringer manifold from the beginning, so, both for one-max and for the needle-in-a-haystack, there would have been virtually no difference between the actual dynamics and the factorised form $\prod_{i=1}^{\ell} c(h_i, t)$.

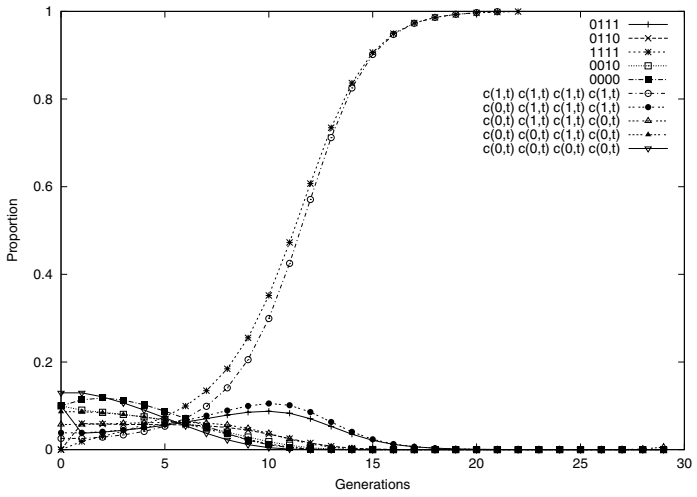

Figure 6: Infinite-population dynamics of a representative sample of strings in a 4-bit GA solving the needle-in-a-haystack problem in the presence of generalised of recombination and selection. The predicted trajectory along the Geiringer manifold for the strings is also shown (Equation 15).

Figure 7 shows the results of running our GA with a population of $M = 4$ strings using all possible GCMs which do not involve duplication with equal probability on a 3-bit flat-landscape. The population was initialised picking random strings from the search space with uniform probability. With infinite populations this would correspond to having equal proportions of all strings. The figure shows the proportions of a subset of interesting strings averaged across all runs. In these runs the proportions of all strings except the strings 000 and 111 tend asymptotically to 0. The proportions of 000 and 111 eventually would reach 50%. This does not mean that the two strings end up coexisting. It simply means that in 50% of the cases the population converges to 111 and in 50% to 000.[8] This happens because, the drift and the crossover biases are both present and the system is trying to settle on a point that is both a homogeneous population (which is a fixed point for drift) and is on the Geiringer manifold (which is a fixed point for crossover). So, runs ended up with either a population formed by only copies of 000, or copies of 111, because they are the only two homogeneous populations on the Geiringer manifold. Under homologous crossover one would see a very different picture where all homogeneous populations would be reached with equal probability. As a result we would observe no dynamics in average-proportion plots such as Figure 7.

The situation is qualitatively similar in the case of populations of size $M = 10$. In this case, however, as shown in Figure 8 the movement towards homogeneous populations is slower since larger populations present less drift. This is confirmed in Figure 9 which shows the dynamics for populations of size 100. In this case the behaviour predicted by our infinite population model (which would predict no dynamics) and that shown by averaging actual runs are very similar.

To illustrate what happens when the population is not initialised uniformly (i.e.,

[8] This is clarified by looking at the standard deviations for string frequencies. These are very large, asymptotically reaching 0.5. We have measured them in all experiments, but we don't report them to avoid cluttering the figures.

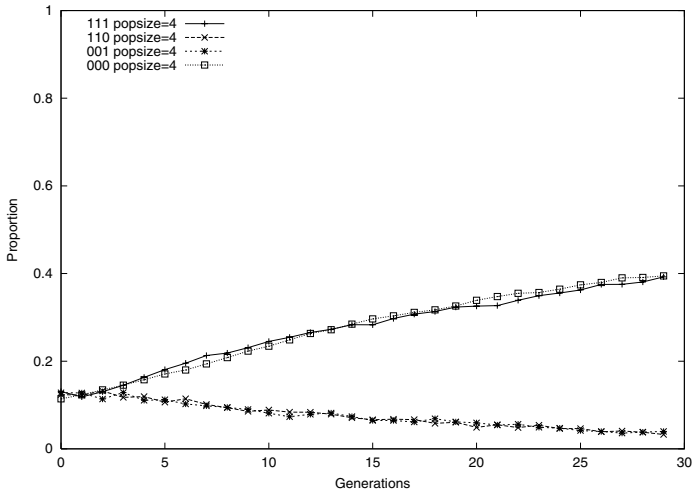

Figure 7: Finite-population dynamics of a population of $M = 4$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen uniformly at random.

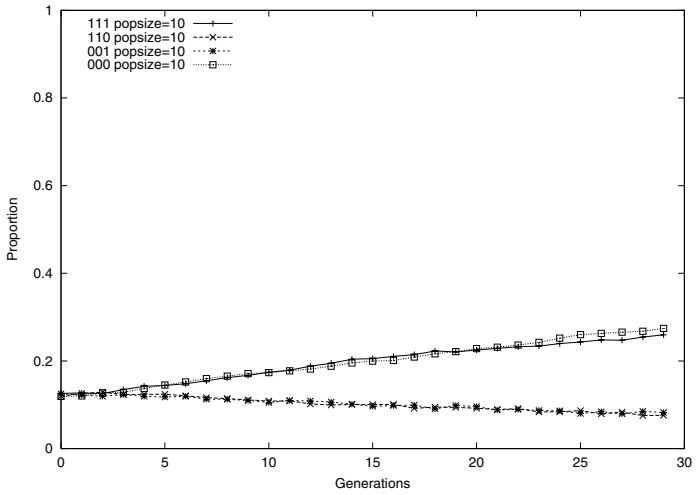

Figure 8: Finite-population dynamics of a population of $M = 10$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen uniformly at random.

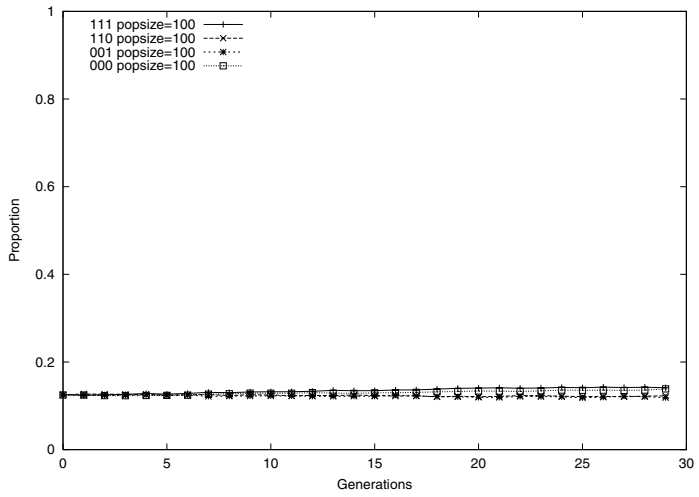

Figure 9: Finite-population dynamics of a population of $M = 100$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen uniformly at random.

with all strings having the same probability of being in the population), we repeated the 3-bit flat-landscape runs but this time starting from the same initial proportions used in the Schemulator run in Figure 4. Figures 10, 11 and 12 show the results obtained with $M = 4$, $M = 10$ and $M = 100$, respectively. These figures illustrate how the dynamics predicted by the infinite population model is effectively superimposed on (and in the case of small populations, hijacked by) the bias of drift, with the GA with larger populations effectively behaving as predicted by the Schemulator within the time span shown (compare the plots for 111 and 110 in Figures 12 and 4). In longer runs we would expect drift eventually to lead the population to one of the two possible homogeneous populations even in the case of large populations. In all cases the proportions for 111 and 000 reach a limit, $\Phi^*(111)$ and $\Phi^*(000)$, respectively, where $\Phi^*(111) + \Phi^*(000) = 1$ but $\Phi^*(111) \neq \Phi^*(000)$ when these stings are not initialised with equal proportions.

The results for the 3-bit one-max problem are shown in Figures 13, 14 and 15. In this case the similarity between the predictions of the infinite-population model (Figure 5) and actual runs is striking, the behaviour of actual runs having been captured by the infinite population model. Indeed, in the case of $M = 100$ the plots in Figure 15 are almost indistinguishable from the corresponding plots in Figure 5.

The results for the 4-bit needle-in-a-haystack problem are shown in Figures 16, 17 and 18. In this case the similarity between the predictions of the theory (Figure 6) and actual runs is good only in the case of $M = 100$. Because the needle-in-a-haystack landscape is effectively flat everywhere except at the needle, for smaller population sizes we see a behaviour that is effectively a mixture of the behaviour shown on a flat-landscape (with drift) and the infinite population Schemulator prediction. Drift effects are particularly strong also because we gave the needle a fitness, 2, which is only twice the fitness of the "hay" (the rest of the search space).[9]

In all cases, for the first few generations, where the crossover bias is strongest, the

[9]Drift has much less marked effects with longer needles (if the fitness of the needle is 100, for example, even runs with 10 individuals start being rather similar to the infinite population model).

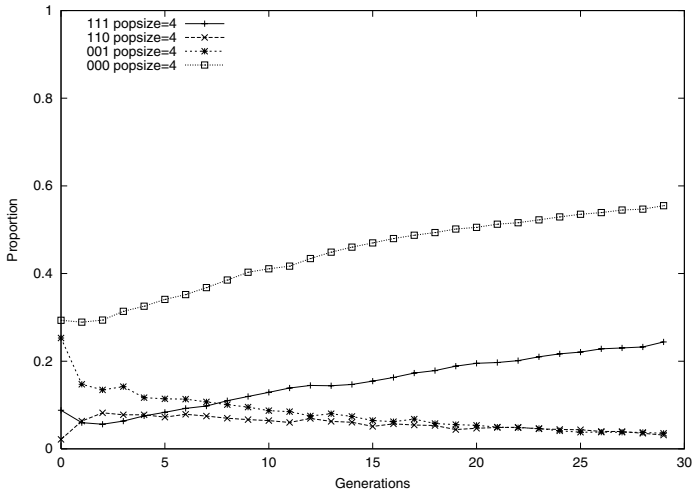

Figure 10: Finite-population dynamics of a population of $M = 4$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

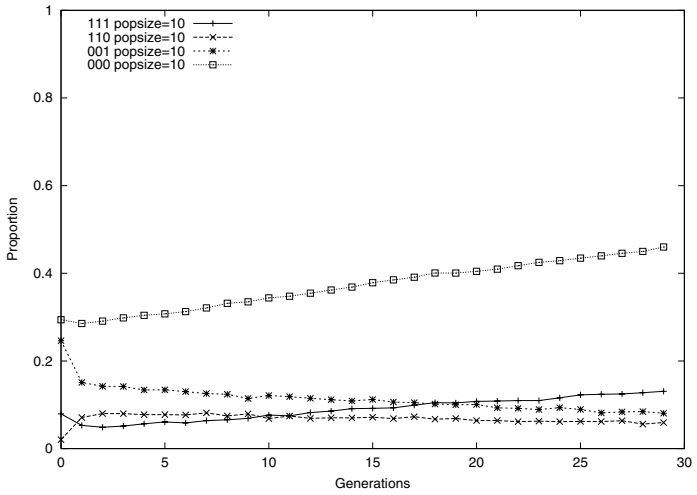

Figure 11: Finite-population dynamics of a population of $M = 10$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

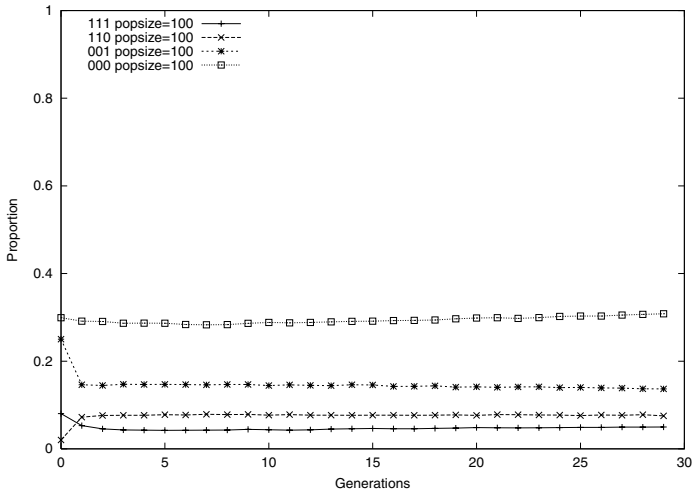

Figure 12: Finite-population dynamics of a population of $M = 100$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

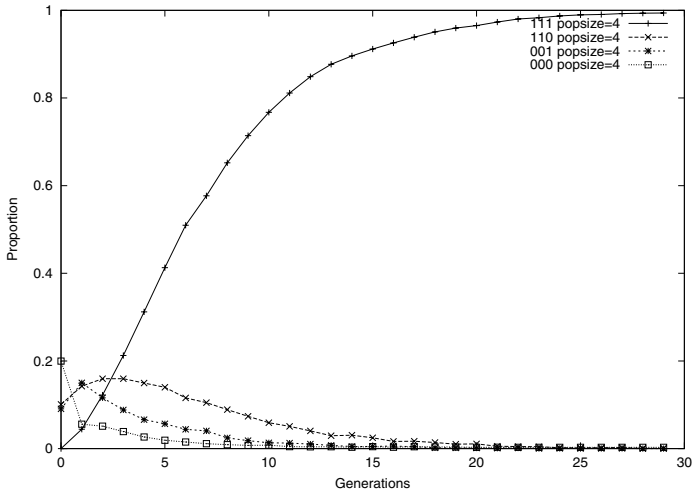

Figure 13: Finite-population dynamics of a population of $M = 4$ strings of length $\ell = 3$ evolving on the one-max landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

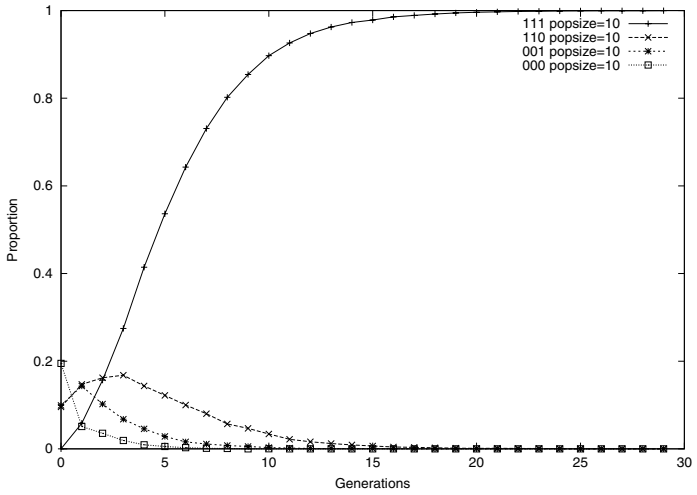

Figure 14: Finite-population dynamics of a population of $M = 10$ strings of length $\ell = 3$ evolving on the one-max landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

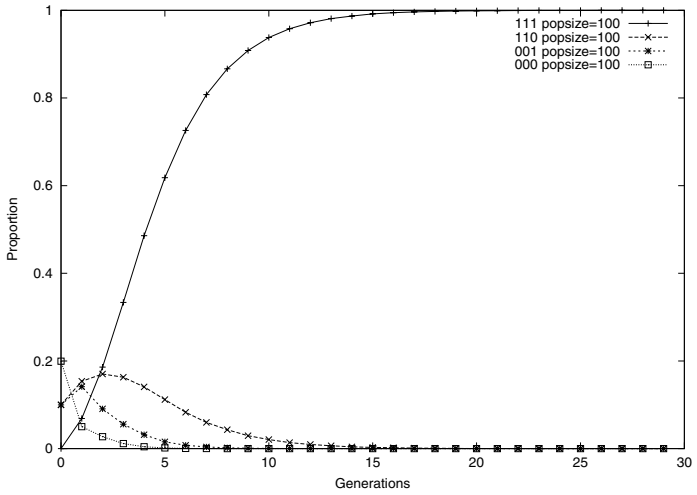

Figure 15: Finite-population dynamics of a population of $M = 100$ strings of length $\ell = 3$ evolving on the one-max landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

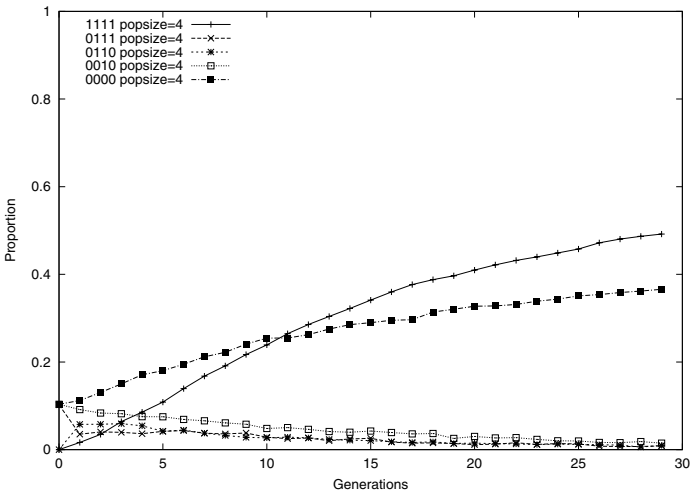

Figure 16: Finite-population dynamics of a population of $M = 4$ strings of length $\ell = 4$ evolving on a needle-in-a-haystack landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

fit between infinite population model and actual runs is very good.

9 Closing the Gap with Nature: Non-binary Alphabets, Diploidy and Multiple Chromosomes

The examples and simulations presented so far in this paper have been for binary alphabets. However, the theory is more general than that and is, indeed, *applicable to alphabets of any cardinality*, including the quaternary alphabet $\{A, C, G, T\}$ of DNA. In this section we want to see to what extent the theory can help to model and understand natural evolution and EAs which borrow additional ideas, such as diploidy, from it.

Let us start by modelling a diploid system under selection and recombination, where, for simplicity, we assume that only one pair of homologous chromosomes is present in each individual. Diploid systems of this kind have been studied in population genetics for decades (see for example (Bürger, 2000) and references therein). Also, GAs with diploid representations and corresponding dominance mechanisms have been used for many years in EC. However, diploid GAs have been modelled theoretically only recently in (Liekens et al., 2003), where a microscopic dynamical-system approach was used. Here we will show how easily this can be achieved using generalised recombination and a schema-theoretic approach.

We will concentrate on gene frequencies in fertilised eggs, i.e., in adult individuals. So, we will use strings and schemata of the form $h = h'_1 \cdots h'_\ell h''_1 \cdots h''_\ell$ with twice as many loci as those in a (haploid) chromosome, with the first ℓ loci representing the maternal and the second ℓ loci representing the paternal chromosome of a homologous pair. The frequency of a diploid string $h = h'_1 \cdots h'_\ell h''_1 \cdots h''_\ell$ is then given by Equation 1. That is

$$E[\Phi(h, t+1)] \quad = \quad \sum_{r \in \mathcal{R}^\ell_\ell} p_c(r) p(\Gamma(h'_1 \cdots h'_\ell h''_1 \cdots h''_\ell, I_r), t) p(\Gamma(h'_1 \cdots h'_\ell h''_1 \cdots h''_\ell, I_{\bar{r}}), t),$$

where $p_c(r)$ is a particular type of GRD. Let us see what form $p_c(r)$ takes.

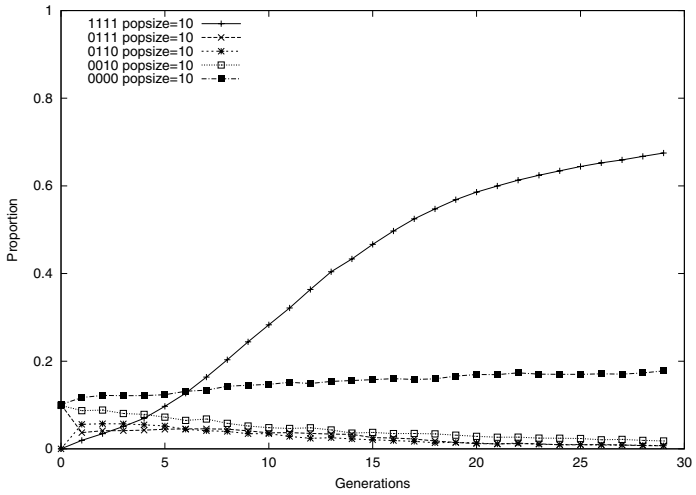

Figure 17: Finite-population dynamics of a population of $M = 10$ strings of length $\ell = 4$ evolving on a needle-in-a-haystack landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

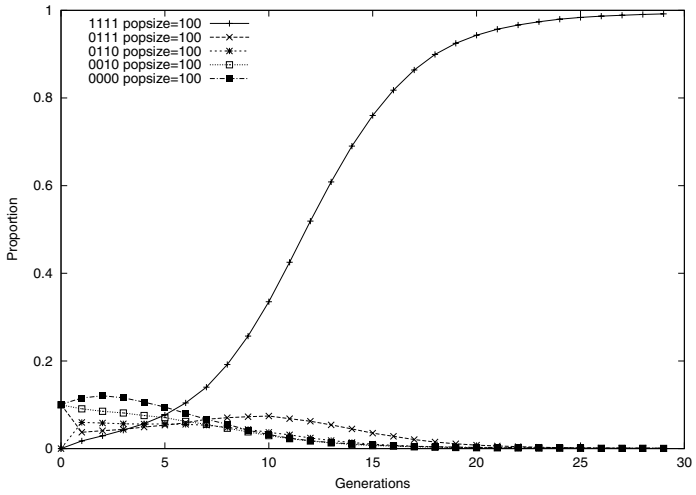

Figure 18: Finite-population dynamics of a population of $M = 100$ strings of length $\ell = 4$ evolving on a needle-in-a-haystack landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

We start by imagining that each cell has only one pair of chromosomes and we ignore gender. Given two parents $a'_1 \cdots a'_\ell a''_1 \cdots a''_\ell$ and $b'_1 \cdots b'_\ell b''_1 \cdots b''_\ell$, in nature we have the following steps:

- When an egg or a sperm cell are created, a form of homologous crossover may happen between the two homologous chromosomes, $a'_1 \cdots a'_\ell$ and $a''_1 \cdots a''_\ell$, in a parent producing a mixed chromosome $a'''_1 \cdots a'''_\ell$ which ends up in a gamete. If crossover does not happen then either $a'''_1 \cdots a'''_\ell = a'_1 \cdots a'_\ell$ or $a'''_1 \cdots a'''_\ell = a''_1 \cdots a''_\ell$ with equal probability.
- Similarly homologous crossover may happen between the two homologous chromosomes, $b'_1 \cdots b'_\ell$ and $b''_1 \cdots b''_\ell$, in a second parent producing a gamete with a mixed chromosome $b'''_1 \cdots b'''_\ell$. Again, if crossover does not happen then either $b'''_1 \cdots b'''_\ell = b'_1 \cdots b'_\ell$ or $b'''_1 \cdots b'''_\ell = b''_1 \cdots b''_\ell$ with equal probability.
- The egg is fertilised by a sperm cell producing a fertilised egg (and, eventually, an adult individual) whose genetic makeup is either $a'''_1 \cdots a'''_\ell b'''_1 \cdots b'''_\ell$ or $b'''_1 \cdots b'''_\ell a'''_1 \cdots a'''_\ell$.

Note that, in nature, there is always a fair chance that the chromosomes transmitted to the offspring by one parent are an identical copy of one of the chromosomes of that parent. In fact, typically, the probability of that happening is at least $\frac{1}{2}$ (homologous chromosomes are duplicated before crossing over can happen, and crossover, if it happens, typically involves only one copy of each chromosome).

For the sake of simplicity, let us imagine that in the first parent a recombination event between $a'_1 \cdots a'_\ell$ and $a''_1 \cdots a''_\ell$ happens in 50% of the cases and when it happens it takes the form of one-point crossover where the l.h.s. of the offspring comes from $a'_1 \cdots a'_\ell$ and the r.h.s. comes from $a''_1 \cdots a''_\ell$ in half of the cases, while the l.h.s. of the offspring comes from $a''_1 \cdots a''_\ell$ and the r.h.s. comes from $a'_1 \cdots a'_\ell$ in the other half of the cases.

In order to model the three operations mentioned above, we can use GCMs. As an example let us consider the case $\ell = 2$. In order to model the diploid recombination process we use the generalised crossover masks in Table 1. The GCMs in the table should all be invoked with probability $p_c(r) = \frac{1}{32}$.

The recombination distribution in this example has two recombination components: one formed by loci 1 and 3, and the second formed by loci 2 and 4. This is easily verified by computing the connection matrix for the order-1 mixing graph

$$C = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

which can be transformed into the block diagonal form

$$C' = \sigma C \sigma^T = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

where σ is a suitable permutation matrix which corresponds to a $2 \leftrightarrow 3$ locus renaming. This is a general property: the GRD for a diploid genetic system with ℓ loci and crossing

Table 1: A set of GCMs that implement diploidy in the case of chromosomes of length $\ell = 2$.

r	Offspring	r	Offspring
$(1100,(1,2,1,2))$	$a'_1a'_2b'_1b'_2$	$(0011,(1,2,1,2))$	$b'_1b'_2a'_1a'_2$
$(1100,(1,2,1,4))$	$a'_1a'_2b'_1b''_2$	$(0011,(1,2,1,4))$	$b'_1b'_2a'_1a''_2$
$(1100,(1,2,3,2))$	$a'_1a'_2b''_1b'_2$	$(0011,(1,2,3,2))$	$b'_1b'_2a''_1a'_2$
$(1100,(1,2,3,4))$	$a'_1a'_2b''_1b''_2$	$(0011,(1,2,3,4))$	$b'_1b'_2a''_1a''_2$
$(1100,(1,4,1,2))$	$a'_1a''_2b'_1b'_2$	$(0011,(1,4,1,2))$	$b'_1b''_2a'_1a'_2$
$(1100,(1,4,1,4))$	$a'_1a''_2b'_1b''_2$	$(0011,(1,4,1,4))$	$b'_1b''_2a'_1a''_2$
$(1100,(1,4,3,2))$	$a'_1a''_2b''_1b'_2$	$(0011,(1,4,3,2))$	$b'_1b''_2a''_1a'_2$
$(1100,(1,4,3,4))$	$a'_1a''_2b''_1b''_2$	$(0011,(1,4,3,4))$	$b'_1b''_2a''_1a''_2$
$(1100,(3,2,1,2))$	$a''_1a'_2b'_1b'_2$	$(0011,(3,2,1,2))$	$b''_1b'_2a'_1a'_2$
$(1100,(3,2,1,4))$	$a''_1a'_2b'_1b''_2$	$(0011,(3,2,1,4))$	$b''_1b'_2a'_1a''_2$
$(1100,(3,2,3,2))$	$a''_1a'_2b''_1b'_2$	$(0011,(3,2,3,2))$	$b''_1b'_2a''_1a'_2$
$(1100,(3,2,3,4))$	$a''_1a'_2b''_1b''_2$	$(0011,(3,2,3,4))$	$b''_1b'_2a''_1a''_2$
$(1100,(3,4,1,2))$	$a''_1a''_2b'_1b'_2$	$(0011,(3,4,1,2))$	$b''_1b''_2a'_1a'_2$
$(1100,(3,4,1,4))$	$a''_1a''_2b'_1b''_2$	$(0011,(3,4,1,4))$	$b''_1b''_2a'_1a''_2$
$(1100,(3,4,3,2))$	$a''_1a''_2b''_1b'_2$	$(0011,(3,4,3,2))$	$b''_1b''_2a''_1a'_2$
$(1100,(3,4,3,4))$	$a''_1a''_2b''_1b''_2$	$(0011,(3,4,3,4))$	$b''_1b''_2a''_1a''_2$

over, but no gene duplication, inversion, etc., has ℓ components. So, clearly the GRD for a diploid system is not order-1 mixing.

This way of proceeding may look artificial: typically genetics focuses is on gametic frequencies. In our model we use a finer level of detail. However, we can calculate gametic frequencies, $\Phi(h_1 \cdots h_\ell, t)$, by using marginals of $\Phi(h'_1 \cdots h'_\ell h''_1 \cdots h''_\ell, t)$. That is:

$$\Phi(h_1 \cdots h_\ell, t) = \frac{1}{2}\left(\Phi(h_1 \cdots h_\ell *^\ell, t) + \Phi(*^\ell h_1 \cdots h_\ell, t)\right).$$

Generalised recombination is powerful enough to model a diploid system *with* crossover and additional operations such as duplication, etc.. For example, one could simulate unequal crossing over by giving a non-zero probability of occurrence to GCMs which perform a shift left or a shift right operation on a fragment of genetic material. For instance, for $\ell = 4$, given parents $a = a'_1a'_2a'_3a'_4a''_1a''_2a''_3a''_4$ and $b = b'_1b'_2b'_3b'_4b''_1b''_2b''_3b''_4$ the GCM $r = (11110000,(1,2,2,3,1,2,3,4))$ would produce the offspring $a'_1a'_2a'_2a'_3b'_1b'_2b'_3b'_4$ which would correspond to a gene duplication event for gene 2. This, of course, can only be considered a first-order approximation of what happens in nature, since with a fixed-length representation any gene duplication must be accompanied by a corresponding gene deletion and vice versa (indeed gene 4 of parent a got deleted in the example above). We have started exploring a variable length extension of this work which avoids this problem altogether (Stephens and Poli, 2005a) but this is beyond the scope of this article.

Similarly, it is possible to model the case where more than one pair of diploid chromosomes are present in the representation. If we have n pairs of chromosomes a, b, c, $\cdots$ of lengths ℓ_a, ℓ_b, ℓ_c, $\cdots$ this can be obtained by using a representation of the form $a'_1 \cdots a'_{\ell_a} a''_1 \cdots a''_{\ell_a} b'_1 \cdots b'_{\ell_b} b''_1 \cdots b''_{\ell_b} c'_1 \cdots c'_{\ell_c} c''_1 \cdots c''_{\ell_c} \ldots$. Because different chromosome pairs duplicate and recombine independently, in the multi-chromosome case the GRD is effectively a product of the GRDs associated with each chromosome pair.

10 Discussion and Conclusions

In this two-part paper we have provided, within the context of a fixed length representation, a single, unified theoretical framework that is powerful enough to exactly model genetic systems that exhibit a rich array of genetic operators, far beyond those of the canonical GA. These include, for instance: gene duplication, gene deletion, inversion, homologous recombination, permutations, diploidy, multiple chromosomes etc. that are not only known to happen in nature but that have also been fruitfully used in EAs.

A good theory should provide qualitative insights and understanding, as well as the possibility of a quantitative comparison. We have endeavoured in this paper to show that our theory does both. To summarise some of the highlights:

- We showed that the dynamics of genetic systems with generalised recombination are more naturally written in terms of building block schemata rather than strings, leading to an exponential reduction in the complexity of the equations.
- We showed that there is a hierarchy of building blocks wherein those of higher order are constructed from others of lower order, the hierarchy terminating at order-1 schemata
- We showed that, in the absence of selection, and in the infinite-population limit, one can solve for the asymptotic $t \to \infty$ behaviour of the hierarchy, where the fixed-point proportions for any string or schema can be written purely as polynomials of the order-1 schema frequencies at $t = \infty$.
- We further showed that the order-1 schemata form closed ℓ-dimensional sets of coupled linear equations, the fixed points of which can be found using standard analysis of the corresponding eigensystems.
- In the case of a duplication free GRD, we found the fixed point for the order-1 schemata.
- In the case of a duplication free GRD, we found that the functional form of the fixed point for an arbitrary string or schema is factorised, thereby generalising the notion of the Geiringer manifold.
- In the case of a duplication free GRD that is fully mixing, we showed that the Geiringer manifold is a global attractor for the system.
- We verified our qualitative and quantitative analytical predictions in the case of an infinite population in the absence of selection by comparing with an explicit integration of the dynamical equations.
- We conjectured that, also in the presence of selection, the bias of generalised recombination is to push the population distribution towards a factorised form, where each factor is potentially time dependent. We subsequently verified this for some model fitness landscapes using the Schemulator.
- We conjectured how the biases of recombination, selection and drift would interact and verified this in real GA runs.

Our overall analysis was motivated by the objective of understanding the search biases induced by such a large and powerful set of genetic operators. This allowed us to

formulate a generalisation of Geiringer's theorem. As usual, analysis of the equations in the presence of selection is much harder to do mathematically for any non-trivial landscape. However, the availability of an exact probabilistic model has allowed for the implementation of an evolution equation integrator (the Schemulator) with which we can numerically explore the interaction between the recombination and selection biases for arbitrary fitness functions and small string lengths under the assumption of an infinite population. Although a standard theoretical assumption, this has often been criticised, hence, its validity was tested by comparing with finite population simulations. As might be expected, these simulations were in good agreement with the theoretical predictions whenever genetic drift due to sampling errors is a marginal phenomenon. This, typically, is the case if one has a sufficiently large population, sampling errors typically scaling as $1/\sqrt{M}$. However, the results can give useful insight into the dynamics for smaller populations as long as one does not integrate the equations for too many time steps (e.g., in the case of short runs), as we have confirmed experimentally.

In the future we intend to study the evolution equations for diploid recombination distributions and to extend the results presented in this paper to the case of variable length strings, thereby, hopefully, contributing further results to theoretical population genetics as well as EC.

From a practitioner's point of view, where could one expect to find that generalised recombination operators perform better than standard recombination operators? We already have some answers. Let us consider, for example, the effects of the lateral diffusion process typically present in generalised recombination. With this process, every time the population reaches an area of flat fitness, lateral diffusion in combination with homologous mixing will start destroying the correlations induced by selection and will effectively re-randomise the population (using unequal allele frequencies) in the neighbourhood of the best solutions found so far and in proximity of the Geiringer manifold.[10] This can have a very beneficial impact, both in realising open ended evolutionary systems, and in exploring, in an unbiased way, neutral networks. As another example, let us consider the effects of duplication. In many systems the function of an allele is not fully (in some, not even partly) determined by its locus. This is the case, for example, in nature, but also in practical EAs such as the messy-GA (Goldberg et al., 1989) and certain types of linear genetic programming systems (which evolve programs for register based CPUs in fixed length chromosomes). In these systems gene duplication may be an excellent mechanism to promote reuse of useful instructions. Naturally, generalised recombination is expected to be beneficial also in problems where solutions present a high degree of genotypic self-similarity (a trivial example is the one-max problem, which, as we have empirically verified (data not reported), is solved more quickly when using generalised recombination than with homologous crossover). Finally, we should note that the availability of exact schema equations (such as those for generalised recombination provided in this paper) allows one to study the interactions of multiple operators and to determine their optimal parameter settings (McPhee and Poli, 2002).

Acknowledgments

We thank Bill Langdon, Marc Schoenauer, the anonymous referees and the associate editor for their help and comments. EPSRC is thanked for financial support (grant

[10] In this sense generalised recombination has some features of mutation.

number GR/T24616/01). Leverhulme Trust is thanked for a visiting professorship (F/OO213/J) for CRS. CRS thanks DGAPA of the UNAM for a Sabbatical Fellowship and Conacyt project 30422-E.

A Proofs

A.1 Theorem 3

Proof. Let us denote with a^n_{sk} the elements of the matrix A^n. In the case of an order-1 mixing recombination distribution, there must exist an n such that $a^n_{sk} > 0$ for all s and k. Therefore, in this case we can apply the Perron-Frobenius theorem (see, for example, (Davis and Principe, 1993)) to understand the dynamics of order-1 schemata. In particular, this theorem guarantees that there is a real positive eigenvalue λ_m of A which dominates all other eigenvalues (i.e., $|\lambda| < \lambda_m$) and that it has multiplicity one. In addition the eigenvector v_m corresponding to λ_m is positive. The theorem does not tell us the value of λ_m, but we can infer that using the following simple probabilistic argument: Because A is row-stochastic, we know that $\lambda = 1$ is an eigenvalue. Let us assume that this is distinct from λ_m and so $\lambda_m > 1$. In this case, assuming v_m is normalised (so all components are ≤ 1), if we chose $\vec{\Phi}^a(0) = v_m$ we would get $\vec{\Phi}^a(t) = A^t v_m = \lambda^t_m v_m$, which, for sufficiently large t, would eventually lead some component of $\vec{\Phi}^a(t)$ to become bigger than 1. However, this cannot happen because these represent probabilities. So, we must have $\lambda_m = 1$ and $v_m = \mathbf{1}/\sqrt{\ell}$.

Let us λ_i be the eigenvalues of A and v_i the corresponding right eigenvectors (normalised to be unit vectors). Also, let u_i be the normalised left eigenvectors of A. Then we can express the initial conditions $\vec{\Phi}^a(0)$ as a linear combination of the v_i's, i.e. $\vec{\Phi}^a(0) = \sum_i w_i v_i$, where $w_i = \vec{\Phi}^a(0) \cdot u_i$ is the projection of $\vec{\Phi}^a(0)$ along each left eigenvector u_i. Then

$$\vec{\Phi}^a(t) = A^t \vec{\Phi}^a(0) = A^t \sum_i w_i v_i = \sum_i w_i A^t v_i = \sum_i w_i A^{t-1} \lambda_i v_i = ... = \sum_i w_i \lambda^t_i v_i$$

So, because $|\lambda_i| < 1$ except for $i = m$,

$$\lim_{t\to\infty} \vec{\Phi}^a(t) = \sum_i w_i \left(\lim_{t\to\infty} \lambda^t_i \right) v_i = w_m v_m = \left(\vec{\Phi}^a(0) \cdot u_m \right) v_m.$$

Therefore, an attractor for $\vec{\Phi}^a(t)$ exists and this depends only on the initial conditions. Since, $v_m = \mathbf{1}/\sqrt{\ell}$, the fixed point is one where all components are identical and equal to a weighted average of the initial schema frequencies, namely

$$\Phi^*(H^a_s) = \frac{1}{\sqrt{\ell}} \sum_i u_{m,i} \Phi(H^a_i, 0) = \frac{1}{\sqrt{\ell}} u_m \cdot \vec{\Phi}^a(0),$$

where $u_{m,i}$ is the i-th component of the left eigenvector u_m. □

A.2 Corollary 1

Proof. If no form of duplication is allowed in the recombination distribution of a particular form of generalised crossover, then alleles can only be moved and shuffled but can be neither created nor destroyed. Since we know that in the case of order-1 mixing GRDs a fixed point exists, then detailed balance must hold for this case. We want to investigate what values of $\alpha_i = \sum_k p_c(*^\ell, *^{k-1} i *^{\ell-k})$ (Section 3.1.1) are compatible with

this type of GRD. Let us assume that $\alpha_i < 1$ for some i. Because the A matrix is row stochastic, the sum of all the elements in the matrix must be ℓ. So, matrix elements not in column i must add up to $\ell - \alpha_i > \ell - 1$. Thus, there must exist (at least) one column j for which $\alpha_j > 1$. This does not violate the detailed balance in general. However, if there is no duplication of alleles, it is not possible for α_j to be greater than 1, since it would mean that locus j is donating to all loci (including itself) more that 100% of its alleles! So, if there is no duplication, that is if the recombination distribution allows only permutations v in $p_c(m, v)$, then the matrix A must also be column stochastic.

Since A is column stochastic (as well as row stochastic), i.e., $\sum_s a_{sk} = \sum_s p_c(*^\ell, *^{s-1}k*^{\ell-s}) = 1$, the normalised left eigenvector associated to the eigenvalue $\lambda_m = 1$ is $u_m = \mathbf{1}/\sqrt{\ell}$ (that is $u_m = v_m$). Therefore, the global attractor for the system becomes

$$\Phi^*(H_s^a) = c(a) = \frac{1}{\ell}\sum_i \Phi(H_i^a, 0).$$

□

A.3 Lemma 1

Proof. The explicit solution of Equation 10 is

$$x(t) = A^t x(0) + \sum_{\tau=0}^{t-1} A^{(t-1-\tau)} b(\tau). \tag{18}$$

The matrix A satisfies the conditions in Perron-Frobenius theorem, so its largest eigenvalue is simple, real and dominates all other eigenvalues. In addition, because $\sum_j a_{ij} < 1$ for all j (i.e., $||A|| < 1$), the largest eigenvalue of A must be smaller than 1. This guarantees that $(I - A)^{-1}$ exists.

Let us take the limit for $t \to \infty$ of the right hand side of Equation 18. We have that $\lim_{t\to\infty} A^t x(0) = [0 \cdots 0]^T$. Computing $\lim_{t\to\infty} \sum_{\tau=0}^{t-1} A^{(t-1-\tau)} b(\tau)$ requires making use of the definition of limit (that is, $\lim_{t\to\infty} f(t) = L$ if and only if, given $\epsilon > 0$, there exists a T such that $t > T$ implies $|f(t) - L| < \epsilon$).

We want to prove that

$$\lim_{t\to\infty} \sum_{\tau=0}^{t-1} A^{(t-1-\tau)} b(\tau) = \left(\lim_{t\to\infty} \sum_{\tau=0}^{t-1} A^\tau\right) b^* = (I - A)^{-1} b^*.$$

We prove this by showing that $\lim_{t\to\infty} \left(\sum_{\tau=0}^{t-1} A^{(t-1-\tau)} b(\tau) - \sum_{\tau=0}^{t-1} A^\tau b^*\right) = \mathbf{0}$. So, let us fix an $\epsilon_a > 0$ and see if we can find a T_a such that $\left|\sum_{\tau=0}^{t-1} A^{(t-1-\tau)} b(\tau) - \sum_{\tau=0}^{\infty} A^\tau b^*\right| < \epsilon_a$ for any $t > T_a$.

Because $\lim_{t\to\infty} b(t) = b^*$, then given an $\epsilon_b > 0$, there exists a T_b such that $t > T_b$ implies $|b(t) - b^*| < \epsilon_b$.

For $t > T_b$ we have

$$\begin{aligned}
&\left|\sum_{\tau=0}^{t-1} A^{(t-1-\tau)}b(\tau) - \sum_{\tau=0}^{t-1} A^{\tau}b^*\right| \\
&= \left|\sum_{\tau=0}^{t-1} A^{(t-1-\tau)}b(\tau) - \sum_{\tau=0}^{t-1} A^{(t-1-\tau)}b^*\right| \\
&= \left|\sum_{\tau=0}^{T_b-1} A^{(t-1-\tau)}(b(\tau)-b^*) + \sum_{\tau=T_b}^{t-1} A^{(t-1-\tau)}(b(\tau)-b^*)\right| \\
&\leq \left|\sum_{\tau=0}^{T_b-1} A^{(t-1-\tau)}(b(\tau)-b^*)\right| + \left|\sum_{\tau=T_b}^{t-1} A^{(t-1-\tau)}(b(\tau)-b^*)\right| \\
&\leq \sum_{\tau=0}^{T_b-1} \left|A^{(t-1-\tau)}(b(\tau)-b^*)\right| + \sum_{\tau=T_b}^{t-1} \left|A^{(t-1-\tau)}(b(\tau)-b^*)\right| \\
&\leq \sum_{\tau=0}^{T_b-1} \left|A^{(t-1-\tau)}\right| \cdot \left|(b(\tau)-b^*)\right| + \sum_{\tau=T_b}^{t-1} \left|A^{(t-1-\tau)}\right| \cdot \left|(b(\tau)-b^*)\right| \\
&\leq \sum_{\tau=0}^{T_b-1} \left|A^{(t-1-\tau)}\right| B + \sum_{\tau=T_b}^{t-1} \left|A^{(t-1-\tau)}\right| \epsilon_b \\
&= B \sum_{n=t-T_b}^{t-1} |A^n| + \epsilon_b \sum_{n=0}^{t-T_b-1} |A^n| \\
&\leq B \sum_{n=t-T_b}^{t-1} |A|^n + \epsilon_b \sum_{n=0}^{t-T_b-1} |A|^n \\
&= \left[B\left(|A|^{t-T_b} - |A|^t\right) + \epsilon_b\left(1-|A|^{t-T_b}\right)\right] \times (1-|A|)^{-1} \\
&\leq \left[B\left(|A|^{t-T_b} - |A|^t\right) + \epsilon_b\right] \times (1-|A|)^{-1}.
\end{aligned}$$

If we choose $\epsilon_b = \frac{\epsilon_a}{2}(1-|A|)$, calculate the corresponding T_b and then choose T_a such that $B\left(|A|^{T_a-T_b} - |A|^{T_a}\right) < \epsilon_b$, we have that given any $\epsilon_a > 0$, there exists a T_a such that $\forall t > T_a \implies \left|\sum_{\tau=0}^{t-1} A^{(t-1-\tau)}b(\tau) - \sum_{\tau=0}^{t-1} A^{\tau}b^*\right| < \epsilon_a$. That is, the limit for $t \to \infty$ of the second term in Equation 18 exists and is $(I-A)^{-1}b^*$. As a result, also $\lim_{t\to\infty} x(t) = x^*$ exists. So, a fixed point for Equation 10 exists, is unique, is non-negative and is given by

$$x^* = (I-A)^{-1}b^*. \tag{19}$$

□

A.4 Theorem 4

Proof. Since the fitness landscape is flat, $p(H,t) = \Phi(H,t)$ for any schema. Also, because the population is infinite, $E[\Phi(H,t+1)] = \Phi(H,t+1)$. Then, for a duplication-free

GRD we can rewrite the schema evolution equations as

$$\Phi(h,t+1) = \sum_{r\in\mathcal{R}_\ell^\ell} p_c(m,v)\Phi\left(\bigotimes_{k=1}^{|I_r|}\left(*^{v_{i_k}-v_{i_{k-1}}-1}h_{i_k}\right)*^{\ell-v_{i_{|I_r|}}},t\right) \Phi\left(\bigotimes_{k=1}^{|I_{\bar{r}}|}\left(*^{v_{j_k}-v_{j_{k-1}}-1}h_{j_k}\right)*^{\ell-v_{j_{|I_{\bar{r}}|}}},t\right), \quad (20)$$

where i_k and j_k are elements of the sets $I_r = \{i_1, i_2, \cdots, i_{|I_r|}\}$ and $I_{\bar{r}} = \{j_1, j_2, \cdots, j_{|I_{\bar{r}}|}\}$, respectively, which are assumed to be ordered as indicated in Part I.

We can prove that Equation 13 is a fixed point for this equation by substituting the right-hand side of Equation 13 into the right-hand side of this equation and then showing that the resulting expression for $\Phi(h_1, ..., h_\ell, t+1)$ has exactly the same form as the right-hand side of Equation 13.

Let us start by splitting each I_r into disjoint subsets I_{rn} for $n \in Q(p_c)$ where subset I_{rn} includes the elements of I_r from component n. That is $I_{rn} = I_r \cap n$. Then at the fixed point

$$\Phi\left(\bigotimes_{k=1}^{|I_r|}\left(*^{v_{i_k}-v_{i_{k-1}}-1}h_{i_k}\right)*^{\ell-v_{i_{|I_r|}}},t\right) = \prod_{n\in Q(p_c)}\prod_{i\in I_{rn}} c(n,h_i).$$

A similar result holds for $I_{\bar{r}}$ and the last term of Equation 20.

So, from the substitution of the fixed point in Equation 20 we obtain

$$\Phi(h,t+1) = \sum_{r\in\mathcal{R}_\ell^\ell} p_c(r)\prod_{n\in Q(p_c)}\prod_{i\in I_{rn}} c(n,h_i)\prod_{n\in Q(p_c)}\prod_{j\in I_{\bar{r}n}} c(n,h_j)$$

Because I_r and $I_{\bar{r}}$ are disjoint and their union is $\{1, \cdots, \ell\}$, for all $n \in Q(p_c)$ we have $I_{rn} \cup I_{\bar{r}n} = n$ and, so,

$$\begin{aligned}\Phi(h,t+1) &= \sum_{r\in\mathcal{R}_\ell^\ell} p_c(r)\prod_{n\in Q(p_c)}\prod_{i\in n} c(n,h_i)\\ &= \prod_{n\in Q(p_c)}\prod_{i\in n} c(n,h_i)\underbrace{\sum_{r\in\mathcal{R}_\ell^\ell} p_c(r)}_{=1}\\ &= \prod_{n\in Q(p_c)}\prod_{i\in n} c(n,h_i),\end{aligned}$$

which proves that Equation 13 is a fixed point for the distribution of strings and, more generally, schemata. □

References

Bäck, T., Fogel, D. B., and Michalewicz, T., editors (2000). *Evolutionary Computation 1: Basic Algorithms and Operators*. Institute of Physics Publishing.

Baker, J. E. (1987). Reducing bias and inefficiency in the selection algorithm. In Grefenstette, J. J., editor, *Genetic algorithms and their applications: Proceedings of the Second*

International Conference on Genetic Algorithms, pages 14–21, Hillsdale, NJ. Lawrence Erlbaum Assoc.

Bürger, R. (2000). *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, Chichester, UK.

Davis, T. E. and Principe, J. C. (1993). A Markov chain framework for the simple genetic algorithm. *Evolutionary Computation*, 1(3):269–288.

Geiringer, H. (1944). On the probability theory of linkage in Mendelian heredity. *Annals of Mathematical Statistics*, 15(1):25–57.

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

Langdon, W. B. and Poli, R. (2002). *Foundations of Genetic Programming*. Springer-Verlag.

Liekens, A. M. L., ten Eikelde, H. M. M., and Hilbers, P. A. J. (2003). Modelling and simulating diploid simple genetic algorithms. In Jong, K. D., Poli, R., and Rowe, J., editors, *Proceedings of the Foundations of Genetic Algorithm (FOGA-VII) Workshop*, pages 151–167, Malaga, Spain, (4–6 September, 2002). Morgan Kaufmann.

McPhee, N. F. and Poli, R. (2002). Using schema theory to explore interactions of multiple operators. In Langdon, W. B., Cantú-Paz, E., Mathias, K., Roy, R., Davis, D., Poli, R., Balakrishnan, K., Honavar, V., Rudolph, G., Wegener, J., Bull, L., Potter, M. A., Schultz, A. C., Miller, J. F., Burke, E., and Jonoska, N., editors, *GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 853–860, New York. Morgan Kaufmann Publishers.

McPhee, N. F., Poli, R., and Rowe, J. E. (2001). A schema theory analysis of mutation size biases in genetic programming with linear representations. In *Proceedings of the 2001 Congress on Evolutionary Computation CEC2001*, pages 1078–1085, Seoul, Korea. IEEE Press.

Mitavskiy, B. and Rowe, J. (2006). An extension of geiringer's theorem for a wide class of evolutionary search algorithms. *Evolutionary Computation*, 14(1):87–118.

Mitavskiy, B. and Rowe, J. E. (2005). A schema-based version of Geiringer's theorem for nonlinear genetic programming with homologous crossover. In Wright, A. H., Vose, M. D., De Jong, K. A., and Schmitt, L. M., editors, *Foundations of Genetic Algorithms 8*, volume 3469 of *Lecture Notes in Computer Science*, pages 156–175. Springer-Verlag, Berlin Heidelberg.

Poli, R., Rowe, J. E., Stephens, C. R., and Wright, A. H. (2002a). Allele diffusion in linear genetic programming and variable-length genetic algorithms with subtree crossover. In Foster, J. A., Lutton, E., Miller, J., Ryan, C., and Tettamanzi, A. G. B., editors, *Genetic Programming, Proceedings of the 5th European Conference, EuroGP 2002*, volume 2278 of *LNCS*, pages 212–227, Kinsale, Ireland. Springer-Verlag.

Poli, R., Stephens, C. R., Wright, A. H., and Rowe, J. E. (2002b). On the search biases of homologuous crossover in linear genetic programming and variable-length genetic algorithms. In Langdon, W. B., Cantú-Paz, E., Mathias, K., Roy, R., Davis, D., Poli, R., Balakrishnan, K., Honavar, V., Rudolph, G., Wegener, J., Bull, L., Potter,

M. A., Schultz, A. C., Miller, J. F., Burke, E., and Jonoska, N., editors, *GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 868–876, New York. Morgan Kaufmann.

Poli, R., Stephens, C. R., Wright, A. H., and Rowe, J. E. (2003). A schema-theory-based extension of Geiringer's theorem for linear GP and variable-length GAs under homologous crossover. In Jong, K. D., Poli, R., and Rowe, J., editors, *Proceedings of the Foundations of Genetic Algorithm (FOGA-VII) Workshop*, pages 45–62, Torremolinos, (4–6 September, 2002). Morgan Kaufmann.

Ridley, M. (1993). *Evolution*. Blackwell Scientific Publications, Boston.

Rogers, A. and Prügel-Bennett, A. (1999). Genetic drift in genetic algorithm selection schemes. *IEEE-EC*, 3(4):298.

Stephens, C. R. (2001). Some exact results from a coarse grained formulation of genetic dynamics. In Spector, L., Goodman, E. D., Wu, A., Langdon, W. B., Voigt, H.-M., Gen, M., Sen, S., Dorigo, M., Pezeshk, S., Garzon, M. H., and Burke, E., editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, pages 631–638, San Francisco, California. Morgan Kaufmann.

Stephens, C. R. and Poli, R. (2005a). Coarse grained dynamics for generalized recombination. Technical Report CSM-432, Department of Computer Science, University of Essex.

Stephens, C. R. and Poli, R. (2005b). Coarse graining in an evolutionary algorithm with recombination, duplication and inversion. *In The 2005 IEEE Congress on Evolutionary Computation*, Vol. 2, pages 1683–1690. IEEE Press.

Stephens, C. R. and Poli, R. (2005c). Coarse graining in an evolutionary algorithm with recombination, duplication and inversion. Technical Report CSM-427, Department of Computer Science, University of Essex.

Vose, M. D. (1999). *The simple genetic algorithm: Foundations and theory*. MIT Press, Cambridge, MA.