Understanding the Biases of Generalised Recombination

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

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

Department of Computer Science
University of Essex
Technical Report CSM-433
ISSN: 1744-8050
July 2005

Abstract

In this paper 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. The analysis of the model reveals that the notion of schema emerges naturally from the model's equations. The study provides a variety of fixed points for the case where recombination is used alone, which generalise 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

An important objective in evolutionary computation (EC) is to exactly model classes of evolutionary algorithms (EAs) and, further, to be able to draw inferences from these models that enhance theoretical understanding and, hopefully, aid "practitioners" in finding more competent EAs. Early models for genetic algorithms (GAs), proposed by Holland, Goldberg, Whitley and others in the seventies and eighties were either approximate or not easily scalable (Holland, 1975; Goldberg, 1989; Whitley, 1992; Whitley, 1994). Exact probabilistic models have been developed, such as the dynamical systems model of Vose and collaborators (Vose, 1999; Rowe et al., 2002). More recently, an alternative exact approach, based on the notion of coarse graining of GA dynamics, has been proposed (Stephens and Waelbroeck, 1999). In physics, coarse graining methods are used to obtain more concise and easily understandable, but sometimes approximate, models of complex systems involving large numbers of microscopic degrees of freedom. The variables used to describe a system in a coarse grained model are often called *effective degrees of freedom*. Frequently, these represent macroscopic quantities which summarise the original microscopic degrees of freedom. Coarse grained models are often naturally suggested by the system under investigation itself. When applying coarse graining to EAs with crossover, schemata emerge naturally as the appropriate effective degrees of freedom. This is why the notion of schema has recently led to a

spate of both new exact theoretical results (Stephens and Waelbroeck, 1999; Stephens et al., 1999; Stephens and Vargas, 2000; Poli, 2001; Poli and McPhee, 2003a; Poli and McPhee, 2003b) and practical recipes for implementation (McPhee and Poli, 2002; Poli, 2003). Also, schema models have revealed that tight links between GAs and GP exist (Langdon and Poli, 2002).[1]

Exact models for EAs are vital, in that they offer the fundamental mathematical foundation upon which approximate models should derive their validity. Based on this exact foundation, theory should possess the twin properties of providing intutition and understanding at the qualitative level; while allowing for predictions for measurable quantities at the quantitative level. Comparison with empirical studies should be feasible at both these levels. Up to now, the vast majority of theoretical work in EAs, at least for classical fixed-length binary and real-valued representations, has been centred on the "canonical" GA, or on Evolutionary Strategies, with selection, mutation and "homologous" recombination. In the latter, a position (locus) in the offspring can *only* be filled using gene values (alleles) from *the same position* in one of the parents. In nature, though, there are many more ways of combining parental genetic material into an offspring than just homologous crossover, some of which have been used in EAs. Gene duplication, for example, has been studied in biology (Clark, 1994), as well as in the context of GAs (Sawai and Adachi, 2000) and GP (Koza, 1995), while inversion, albeit of a more sophisticated kind than that considered here, was one of the operators used by Holland (Holland, 1975) in the original formulation of the GA.

In this paper, we begin with an exact, microscopic probabilistic model for the evolution of a population of fixed length, ℓ, strings undergoing selection and generalised recombination, the latter accounting for *any* redistribution of the parental genes to the offspring, including as special cases, among others – fixed-length versions of gene duplication and deletion, as well as inversion and homologous crossover.

With this more general form of recombination, there are potentially many more ways of reshuffling genetic material between parents and offspring when compared to homologous crossover ($(2\ell)^\ell$ versus 2^ℓ). Interestingly, however, as in the case of homologous crossover, a coarse graining naturally appears, revealing that the notion of schemata as building blocks emerges from the model's equations, irrespective of how genetic material is redistributed.

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 generalized recombination, as with homologous recombination, is to destroy correlations, thereby implying that the asymptotic dynamics is governed by that of the order one 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

[1] Microscopic and macroscopic (coarse grained) approaches are not in competition. Indeed, it is possible to prove that one can construct Vose-like Markov chain models for EAs by using exact schema equations and, vice versa, one can track schema frequencies by coarse graining the dynamical systems model's equations (Poli et al., 2001; Poli et al., 2004). So, EC theory is now effectively unified to a considerable degree (Stephens and Poli, 2004).

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.

The paper is organised as follows: In Section 2 we introduce some basic definitions relating to the notion of schema, and summarise and further characterise the notion of generalised recombination that we previously introduced in (Poli and Stephens, 2005a; Poli and Stephens, 2005b; Stephens and Poli, 2005a). In Section 3 we write down the microscopic evolution equations for strings and schemata that form the starting point for our study, using the standard interpretation of schemata as sets (an extensive characterisation of these equations in terms of projection operations is given in (Stephens and Poli, 2005a)) and illustrating the features of these equations with simple examples. We explain the hierarchical nature of the schema evolution equations in Section 4 and show how the theory presented here generalises past work in Section 5. We study the biases of generalised recombination by finding the fixed points of the evolution equations for order-1 schemata in Section 6 for different classes of recombination distributions and under the standard assumption of infinite populations. As mentioned, 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 7. The infinite-population results 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 8. We discuss the expected behaviour of an evolutionary system under generalised recombination and selection in Section 9. There we also consider how finite-population effects, such as drift, can alter the dynamics of the system. In Section 10 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 extend 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 11). 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 12 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 13. Finally, to improve the exposition of this material, some proofs are relegated to Appendix A, for the benefit of theoreticians and more mathematically oriented readers.

2 Background

In this section we present some background material that is necessary to understand the contributions of this paper. We start with some definitions related to the notion of schema, and then provide a description of generalised recombination and some of its properties.

2.1 Schemata

In this paper we restrict our attention to a search space of fixed-length strings of length ℓ, where string elements (or alleles) take values from a generic alphabet Ω of any fixed cardinality. If by $\times$ we indicate the Cartesian product operator, the search space is $\overbrace{\Omega \times \Omega \times \cdots \times \Omega}^{\ell \text{ times}}$. Following standard convention we will denote this with Ω^ℓ.

A *schema* is a subset of the search space. Syntactically we will represent schemata as strings in the standard notation based on the "don't care" symbol $*$. So, for example, the schema $h = *a*bc$ represents all strings in the search space whose second, fourth and fifth characters are a, b and c, respectively. The *order* of a schema is the number of non-$*$ (or defining) symbols in it. So, the order of h is 3. The *defining length* of a schema is the distance between the furthest defining symbols. The defining length of h is 3. The schema $***\cdots*$ represents the whole search space.

In the following we will need to be able to express and perform operations on schemata both as sets and as strings with don't care symbols. So, here we introduce some notation that will facilitate this.

In order to represent schemata in which one or more symbols are repeated a certain number of times, we use the standard computer science notation x^y to indicate pattern x repeated y times. So, for example, we may represent the schema $**\cdots*$ ($*$ repeated ℓ times) as $*^\ell$, and we may represent the schema $***11111*$ as $*^3 1^5 *$. In some cases we will need to specify schemata where a pattern of alleles and don't care symbols is repeated a certain number of times. If the pattern is constant, we can do this by using brackets and the power notation just introduced. For example, $*^2(10)^5*$ represents the schema $**1010101010*$. However, we will encounter cases where the pattern varies as it is repeated. In these cases we will make use of the *concatenation operator* $\bigotimes_i$. This operator is for concatenation what $\sum_i$ is for sums and $\prod_i$ for products. For example, the schema $*1**1***1****1$ could be represented as $\bigotimes_{i=1}^{4}(*^i 1)$. In addition, we will use the convention that x^0 is the *empty sequence* whatever the sequence x (i.e. x^0 can be safely edited out from any sequence of characters).

Let us now consider some useful operations we can perform on schemata seen as sets. Given any schema of order greater than 1, we can always represent it as the intersection of some other schemata of lower order. For example, the schema $h = *a*bc$ can be seen as the intersection between the schemata $*a***$, $***b*$ and $****c$. So, we can write $h = *a*** \cap ***b* \cap ****c$. Because of this property, schemata of order 1 are particularly important, and, so, we introduce a special notation to represent them. We denote a schema of order 1 with its single defining symbol a at position s as H_s^a. Note this is simply a shorthand notation for $*^{s-1}a*^{\ell-s}$ that we can use whenever the length, ℓ, of the strings in the population is implied. So, if $\ell = 3$ we have $H_2^1 = *1*$ while $H_3^0 = **0$. Also, if $\ell = 5$, with this notation the schema $h = *a*bc$ can be rewritten as $h = H_2^a \cap H_4^b \cap H_5^c$.[2] More generally, a schema h of order n with defining symbols $a_1, a_2, \cdots, a_n$ at positions $l_1, l_2, \cdots, l_n$ can be represented as the following set intersection

$$h = \bigcap_{i=1}^{n} H_{l_i}^{a_i}.$$

The advantage of the set intersection notation is that it does not require the sequence l_i to be sorted. However, if the sequence l_i is ordered so that $l_1 \leq l_2 \leq \cdots \leq l_n$, this same set can be represented using the string $h = *^{l_1-1}a_1*^{l_2-l_1-1}a_2*^{l_3-l_2-1}a_3 \cdots *^{l_n-l_{n-1}}a_n*^{\ell-l_n}$ or, much more concisely, as

$$h = \bigotimes_{i=1}^{n}(*^{l_i-l_{i-1}-1}a_i)*^{\ell-l_n},$$

where we conventionally extended the l_i sequence by setting $l_0 = 0$ for notational convenience.

On some occasions it will be advantageous to use a concise notation for schemata of order higher than 1. Extending the order-1 schema notation H_s^a, we will denote with $H_{l_1,l_2,\cdots,l_n}^{a_1,a_2,\cdots,a_n}$ an order $n \leq \ell$ schema with defining symbols $a_1, a_2, \cdots, a_n$ at positions $l_1, l_2, \cdots, l_n$. The interpretation is, of course, the one provided in the previous two equations. For example, a generic order-2 schema with alleles a and b at positions s and u (with $s \neq u$), respectively, is represented as $H_{s,u}^{a,b} = H_s^a \cap H_u^b$. Naturally, due to the commutativity of set intersection, there are symmetries in this representation, e.g. $H_{s,v}^{a,b} = H_{v,s}^{b,a}$. Also, clearly $H_{s,u}^{a,b}$ can also be represented using the string $*^{s-1}a*^{u-s-1}b*^{\ell-u}$ if $s < u$. If instead $s > u$, then $H_{s,u}^{a,b}$ is represented by $*^{u-1}b*^{s-u-1}a*^{\ell-s}$. As another example, order 3 schemata with exactly three alleles specified can be represented as $H_{s,u,q}^{a,b,c} = H_s^a \cap H_u^b \cap H_q^c$ where $s \neq u \neq q$.

As we have seen, all schemata of order 2 or above can be represented as intersections of order-1 schemata. Later in the paper we will compute intersections between order-1 schemata, and, so, we might wonder if the reverse of the afore-mentioned property holds. That is: are all order-1 schema intersections always representable as higher order schemata? The answer is "no": the

[2] Because the set-intersection operation commutes, we can also write $h = H_5^c \cap H_2^a \cap H_4^b$, etc.

intersection of any number of schemata is either a schema or it is the empty set $\emptyset$. For example, $*1*** \cap ***1* \cap *0*** = \emptyset$. A sufficient (but not necessary) condition for the result of $\bigcap_{i=1}^{n} H_{l_i}^{a_i}$ being guaranteed to be a schema of order n if that the schemata $H_{l_i}^{a_i}$ are all from different order-1 schema partitions.[3]

Although in the following we will mainly use schemata to coarse grain over a search space of fixed-length strings of a given length, we can also use them to coarse grain over other spaces that will be introduced later in the paper to model a GA using generalised recombination. In particular, whenever we have a space $\mathcal{V}^\ell$ of vectors $v = (v_1, \cdots, v_\ell)$ with elements $v_i \in \mathcal{V}$, where $\mathcal{V}$ is a generic set, we can use the schema notation to represent subsets of $\mathcal{V}^\ell$ where one or more coordinates take any value. For example, if $\mathcal{V} = \{1, 2, 3, 4\}$ and $\ell = 7$, the schema $***3*2*$ represents the set $\{v \in \mathcal{V}^\ell : v_4 = 3 \text{ and } v_6 = 2\}$. This will be very useful to coarse grain over crossover events, as will be shown in the following section.

Finally, a very important property of schemata in relation to probability distributions is the following. If a probability distribution $\Pr(v)$ is associated to the elements v of a space $\mathcal{V}^\ell$ (with $\sum_{v \in \mathcal{V}^\ell} \Pr(v) = 1$), be it the search space or some other space, and we consider the elements of the space as mutually exclusive events, then given a subset $S \subseteq \mathcal{V}^\ell$, the probability that one of the events $v \in S$ takes place is, clearly, $\Pr(S) = \sum_{v \in S} \Pr(v)$. Obviously, this applies also to schemata since they are special subsets. So, in the next sections every time we have a term of the form $\sum_{v \in h} \Pr(v)$ for some schema h and probability distribution $\Pr(v)$ we will replace that with $\Pr(h)$. Note that effectively $\Pr(h)$ is a marginal of $\Pr(v)$. This is the fundamental mechanism by which schemata allow coarse graining over the genetic dynamics of strings and over recombination events.

2.2 Generalised Recombination

In homologous recombination the loci in the offspring can by filled only by using alleles coming from corresponding loci in one of the parents. *Generalised recombination* is a form of recombination which allows the offspring's alleles to come from *any* locus in *any* of the parents.

In order to model mathematically a GA with this form of recombination we need a good notation to represent all possible recombination events. Homologous recombination events are often modelled using the notion of crossover mask. A crossover mask is a binary string of the same length as strings in the population. A 1 at a given position in the mask indicates that the allele at the corresponding position in the offspring should be taken from that locus in the first parent. A 0 in a given position of the mask indicates that the corresponding offspring's allele should be taken from the second parent. By restricting which masks are allowed and properly fixing the probability of choosing crossover masks (a distribution known as the recombination distribution), one can model any homologous crossover operator.

Crossover masks and recombination distributions are sufficient to model a crossover operator when only alleles at the same locus can be exchanged, i.e. homologous crossover. However, if we want to cope with other ways of redistributing genetic material, such as inversion, gene duplication, gene deletion, and, more generally, unequal crossing over, we need a model that allows for the possibility that the allele in one particular locus of the offspring comes from a different locus of a parent. This new level of generality can be represented mathematically in several equivalent ways (see (Poli and Stephens, 2005b; Poli and Stephens, 2005a; Stephens and Poli, 2005a)). Here, however, we focus only on one that naturally extends the notion of crossover masks and recombination distribution.

[3] A set of sets $\{S_1, \cdots, S_n\}$ forms a partition of a set S if $S = \bigcup_i S_i$ and $\forall i \neq j$ we have that $S_i \cap S_j = \emptyset$. The search space can be partitioned using families of schemata, called *schema partitions* in the literature. The schemata in a schema partition are all of the same order and have don't care symbols at exactly the same loci. The whole set is obtained by setting defining symbols in the remaining loci in all possible ways. For example, for $\ell = 5$ the schemata $**0*0$, $**0*1$, $**1*0$ and $**1*1$ form a partition of Ω^ℓ. Schema partitions are often represented with strings of symbols from the alphabet $\{*, \#\}$, where the $\#$'s indicate the position of the defining characters in the schemata in a schema partition. In this notation the partition $\{**0*0, **0*1, **1*0, **1*1\}$ is represented as $**\#*\#$.

For strings of length ℓ, to represent a possible recombination event we use a *Generalised Crossover Mask* (GCM) $r = (m, v)$ 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 $\mathcal{N}_\ell = \{1, \cdots, \ell\}$. So, $m \in \{0,1\} \otimes \{0,1\} \otimes \cdots \otimes \{0,1\} = \{0,1\}^\ell$ and $v \in \mathcal{N}_\ell \otimes \mathcal{N}_\ell \otimes \cdots \otimes \mathcal{N}_\ell = \mathcal{N}_\ell^\ell$ whereby we can see that GCMs live in the space $\mathcal{R}_\ell^\ell = \{0,1\}^\ell \otimes \mathcal{N}_\ell^\ell$, and so the total number of GCMs is $(2\ell)^\ell$, many more than the 2^ℓ possible masks for homologous recombination.

The semantics of the GCM representation is very simple. The elements in m specify which parent contributes the alleles that fill each locus in the offspring, while the elements of v tell us which particular alleles in a parent will be transferred to the offspring. So, $m_i = 1$ means locus i will be filled with an allele from parent 1, $m_i = 0$ means parent 2 will contribute the allele instead. If the corresponding entry $v_i = j$ then locus i will be filled with the allele currently in position j in a parent. More formally we can express the offspring $h = h_1 \cdots h_\ell$, produced by parents $a = a_1 \cdots a_\ell$ and $b = b_1 \cdots b_\ell$, with GCM $r = (m, v)$ as

$$h_i = m_i a_{v_i} + (1 - m_i) b_{v_i}, \tag{1}$$

where a_{v_i} is the allele from the first parent picked out by the GCM r, and similarly for b_{v_i} from the second.[4] Naturally, for our model of generalised recombination to be complete we need to specify the probability $p_c(r)$ of choosing any particular GCM r. This is a generalisation of the notion of recombination distribution – the *Generalised Recombination Distribution* (GRD).

As a first example of how the representation of a GCM works consider the following example using standard one-point crossover for $\ell = 3$. The associated traditional crossover masks are 100 and 110 each invoked with probability $\frac{1}{2}$. These are equivalent to the GCMs $r_1 = (100, (1,2,3))$ and $r_2 = (110, (1,2,3))$. As a second example that illustrates the larger variety of ways in which parental genetic material can be distributed among the offspring, consider the case of $\ell = 2$, where the $(2 \times 2)^2 = 16$ GCMs are

(00,(1,1))	(00,(1,2))	(00,(2,1))	(00,(2,2))
(01,(1,1))	(01,(1,2))	(01,(2,1))	(01,(2,2))
(10,(1,1))	(10,(1,2))	(10,(2,1))	(10,(2,2))
(11,(1,1))	(11,(1,2))	(11,(2,1))	(11,(2,2))

If the associated GRD is such that each is invoked with probability $p_c(r) = \frac{1}{16}$, this would represent a recombination operator where each locus in the offspring is filled with a randomly chosen allele from the parents. Clearly this operator could not be represented with ordinary crossover masks. As a final example, the following GRD represents a single-parent inversion operator in the case of a three-locus system:

$$p_c(111, (2,1,3)) = p_c(111, (1,3,2)) = p_c(111, (3,2,1)) = \frac{1}{3}$$

In the EA theory for homologous crossovers, the operation of bit-wise negation of a crossover mask is often used. Here we extend this notion to GCMs as follows. The *negation of a GCM* $r = (m, v)$ is the mask $\bar{r} = (\overline{m}, v)$, where $\overline{m}$ is the bit-wise negation of the bit-string m.

Since the GRD is a probability distribution over the space $\mathcal{R}_\ell^\ell$, we can use schemata (over this space) to represent entire classes of recombination events. Such schemata can be represented using GCMs in which some of the symbols have been replaced with $*$'s. The don't care symbols can be in the m component of a GCM, in the v component or both. The probability that a recombination event matching a *schema* R be the case is then given by $p_c(R)$ (see end of Section 2.1). So, for example, if $\ell = 5$, we can represent all events where crossover fills the first element of the offspring with the fourth element of the first parent with the schema $(1****, 4****)$, where we omitted the commas between the elements of v for conciseness. The associated probability is $p_c(1****, 4****) = \sum_{r \in (1****, 4****)} p_c(r)$.

[4] In this notation, homologous crossover events can be represented with GCMs of the form $r = (m, (1, 2, \cdots, \ell))$ where, effectively, m is the only element that can vary and, so, r can be seen as a traditional crossover mask.

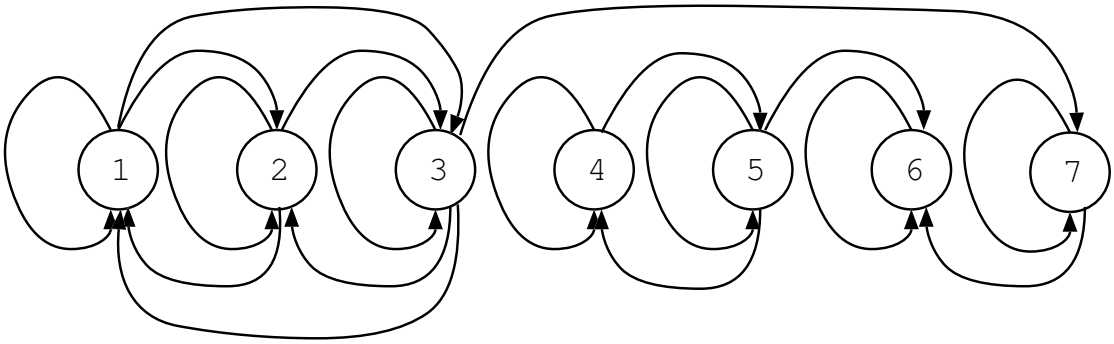

Figure 1: Example of order-1 mixing graph for $\ell = 7$.

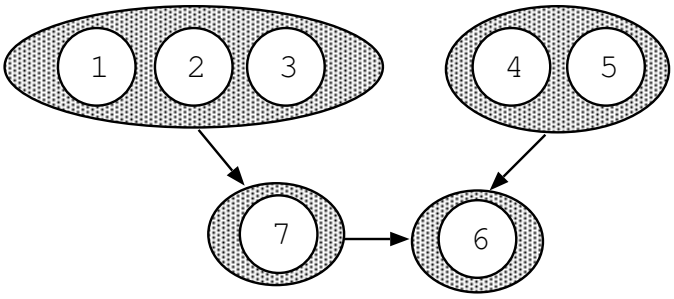

Figure 2: The order-1 recombination clique graph for the mixing graph in Figure 1.

2.3 Mixing graphs and recombination cliques

An important concept when considering redistribution of genetic material as determined by the GRD is: in which direction can one have a flow of genes? As qualitatively different behaviours are exhibited by genetic systems with different GRDs, to understand which features are important we model the effects of the GRD through a set of *mixing graphs*. The first one of these is the *order-1 mixing graph*, the nodes of which represent different loci. The arcs are *directed* and represent causal relationships between loci. Thus, we will connect locus i with an arrow from locus j if the frequency of alleles in locus i can be influenced by the allele frequency of locus j. Figure 1 shows an example of an order-1 mixing graph for a 7-locus representation.

The network of causal influences is completely determined by the GRD. The elements c_{ij} of the connection matrix C for the order-1 mixing graph are given by

$$c_{ij} = \delta\big(p_c(*^{\ell}, *^{i-1} j *^{\ell-i}) > 0\big)$$

where $p_c(*^{\ell}, *^{i-1} j *^{\ell-i})$ is a schema-based coarse graining of the GRD and $\delta(x) = 1$ if x is true, while $\delta(x) = 0$ otherwise. If there is a directed path between each pair of nodes in the order-1 mixing graph (the mixing graph is strongly connected), we define the recombination to be *order-1 mixing*.

Imagine a population of strings and focus attention on a particular allele, a, at a particular locus, i, of a particular string, $a_1 \cdots a_\ell$. An order-1 mixing generalised crossover allows for the migration of allele a to different loci in different strings. So, generalised crossover promotes a process of "diffusion" of alleles from one locus to other loci. That is, unlike the case of homologous crossover, in general, generalised recombination does not keep the alleles in their original position, i.e. allele a might migrate to loci different from i. Because of this, in repeated applications of crossover, a copy of the allele can be placed back into the original string $a_1 \cdots a_\ell$ (which may now have a different allele composition) but at a different locus, effectively creating a sort of gene duplication (indeed unequal crossing over seems to be a mechanism for gene duplication

in nature (Ridley, 1993)). Put another way, crossover is trying to spread each allele as thinly as possible over every locus available in the population. On the other hand, for homologous crossovers, the connection matrix is diagonal and so each node in the graph is isolated (having only a self-connection).

Naturally, many qualitatively different intermediate situations are also possible. In all intermediate cases we can divide the order-1 mixing graph into two or more *order-1 recombination cliques*. These are characterised by the fact that all pairs of nodes in a clique are mutually accessible by traversing only nodes and arcs in the clique, while none of the nodes in a clique is mutually accessible from any other node outside the clique. In Figure 1, loci 1–3 form a recombination clique, nodes 4 and 5 form another, and nodes 6 and 7 form two single-node cliques. Formally, recombination cliques are the strong components of the order-1 recombination graph. So, each locus belongs to one and only one clique. Also, the cliques themselves form a directed acyclic graph (component graph) that we will call the *order-1 recombination clique graph*. This has one node for each recombination clique and an arc between two nodes if there is an edge between the corresponding cliques. Figure 2 shows an example of an order-1 recombination clique graph.

We can define, in a very similar way, the notion of higher order mixing graphs, higher order recombination cliques, higher order clique graphs and higher-order-mixing GRDs. For example, the order-2 mixing graph associated to a GRD includes $\binom{\ell}{2}$ nodes (labelled with all possible pairs of loci), and has connection matrix C with elements $c_{ij,mn}$ given by

$$c_{ij,mn} = \delta\big(p_c(*^{\ell}, *^{i-1}m*^{j-i-1}n*^{\ell-j}) > 0\big),$$

for $i < j$, while the elements $c_{ij,mn}$ for $i > j$ can be computed in a similar fashion.[5] If there is a directed path between each pair of nodes in the order-2 mixing graph (the mixing graph is strongly connected), we define the recombination to be *order-2 mixing*. Again, as for the case of the order-1 connection matrix, this property depends only on a set of relevant schema-based marginals of the GRD. Figure 3 shows an example of an order-2 mixing graph for a 4-locus representation. As one can easily verify, the graph is strongly connected and, so, the corresponding GRD is order-2 mixing.

We define a recombination distribution which is order-1 through to order-ℓ mixing a *fully mixing recombination distribution*.

3 Evolution equations

In (Poli and Stephens, 2005b; Poli and Stephens, 2005a) we derived evolution equations for strings and schemata for a GA using selection and generalised recombination. In (Stephens and Poli, 2005a) we extended the string equations by adding bit-flip mutation and considering also the case of variable length strings. Here we summarise the main results providing only proof fragments wherever these are important for what follows.

3.1 Evolution equations for strings

Let us consider a generational evolutionary system with selection and generalised recombination exploring a search space of fixed-length strings Ω^{ℓ}. For simplicity let us consider the case where crossover is performed with 100% probability (as we will see below we can do this is without loss of generality). Under these assumptions the expected frequency of a string $h = h_1 \cdots h_{\ell} \in \Omega^{\ell}$ is given by

$$E[\Phi(h, t+1)] \quad = \quad \sum_{a \in \Omega^{\ell}} p(a,t) \sum_{b \in \Omega^{\ell}} p(b,t) \sum_{r \in \mathcal{R}^{\ell}_{\ell}} p_c(r) \gamma(a, b, r \rightarrow h)$$

where $E[\cdot]$ is the expectation operator, $\Phi(h, t+1)$ is the proportion of strings of type h in the population at generation $t+1$, $p(a,t)$ is the probability of selecting a string of type a as a parent

[5] We use double indices to identify rows and columns. So, ij represents a row and mn a column of C.

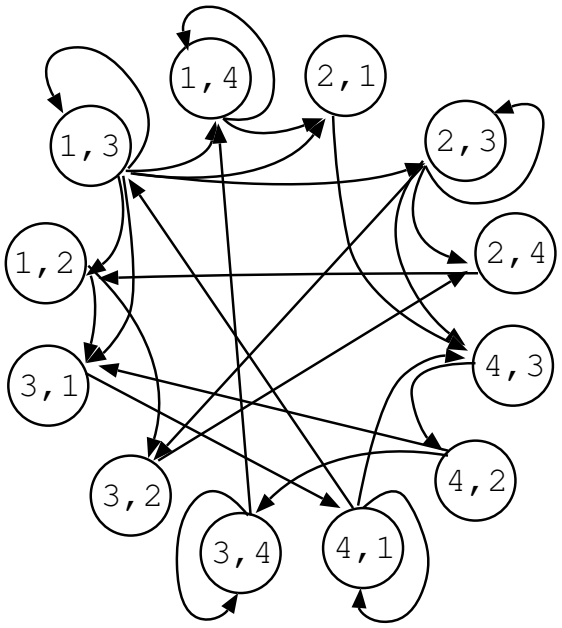

Figure 3: Example of order-2 mixing graph for $\ell = 4$. The graph has only one order-2 recombination clique and the corresponding GRD is order-2 mixing.

from the population at generation t, $\mathcal{R}_\ell^\ell$ is the set of all possible GCMs, $p_c(r)$ is the GRD and $\gamma(a, b, r \to h)$ is the conditional probability that the offspring h is formed given the parents a and b and a GCM r. This takes the value 1 if h is created from a and b using the GCM r and 0 otherwise. Note that the string summations cover the entire search space Ω^ℓ rather than just the strings in the population because the selection probability for any string not currently in the population is zero.

The offspring $h = h_1 \cdots h_\ell$, produced by parents $a = a_1 \cdots a_\ell$ and $b = b_1 \cdots b_\ell$, with GCM $r = (m, v)$ can be computed using Equation 1. From this, it follows that

$$\gamma(a, b, r \to h) = \prod_{i \in I_r} \delta(h_i = a_{v_i}) \prod_{j \in I_{\bar{r}}} \delta(h_j = b_{v_j})$$

where $I_r = \{i : m_i = 1\}$ represents the genes picked out from the first parent by r that go to form part of the offspring h, and $I_{\bar{r}} = \{i : m_i = 0\}$ is the complementary set picked out from the second parent via the negation of GCM r (see Section 2.2). As the full genetic composition of h has to come from the parents together we have $I_r \cup I_{\bar{r}} = \{1, 2, \cdots, \ell\}$. By substituting this result into the evolution equation for h and reordering terms, we obtain

$$E[\Phi(h, t+1)] \quad = \quad \sum_{r \in \mathcal{R}_\ell^\ell} p_c(r) \sum_{a \in \Omega^\ell} p(a, t) \prod_{i \in I_r} \delta(h_i = a_{v_i}) \sum_{b \in \Omega^\ell} p(b, t) \prod_{j \in I_{\bar{r}}} \delta(h_j = b_{v_j}).$$

The effect of terms of the form $\prod_{i \in I_r} \delta(h_i = a_{v_i})$ in this equation is simply to limit the summations to subsets of Ω^ℓ. These subsets are intersections of order one schemata of the form introduced in Section 2.1. More precisely, for each GCM r these are

$$\Gamma(h, I_r) = \bigcap_{i \in I_r} H_{v_i}^{h_i} \qquad \text{and} \qquad \Gamma(h, I_{\bar{r}}) = \bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}$$

with the conventions $\bigcap_{i \in \{j\}} H_{v_i}^{h_i} = H_{v_j}^{h_j}$ and $\bigcap_{i \in \emptyset} H_{v_i}^{h_i} = *^\ell$. We call these *building blocks* for the string h. This leads directly to the following

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

$$E[\Phi(h,t+1)] \quad = \quad \sum_{r\in\mathcal{R}_\ell^\ell} p_c(r)p(\Gamma(h,I_r),t)p(\Gamma(h,I_{\bar{r}}),t). \tag{2}$$

Thus, as in the case of homologous crossover, we see that evolution proceeds by building a string from its component building block schemata. Of course, to make further progress, one would then need to have the equations that govern these schemata. We will do this in the next section. Before we do that, however, we would like to discuss the differences between strings and schemata for describing the evolution. Firstly, note that there are an exponentially large number of ways of reshuffling genetic material from parents to offspring. To emphasise once again, there are $(2\ell)^\ell$ GCMs irrespective of whether strings or schemata are used to describe the dynamics. Of course, it may well be that only a small subset of these masks have non-zero probability. For instance, for homologous crossover there are 2^ℓ possible masks. However, for one-point crossover only $\ell-1$ of these masks have non-zero probability. For a given GCM, there remains the question of how many combinations of strings or schemata can lead to a particular offspring. This is where the advantage of building block schemata plays a crucial role as for a given GCM there is uniquely only *one* relevant pair of schemata and, therefore, correspondingly only one term in the r.h.s. of Equation 2. For strings, however, there are an exponential number of terms to consider. Also, even if there are an exponential number of GCMs, as we will show later, we can study schema equations formally (i.e., for any ℓ and without having to compute the actual terms in the equations) to infer general properties of genetic systems.

3.2 Coarse-grained evolution equations

For homologous crossover, one of the most remarkable features of the coarse grained exact schema equations is their form invariance under a further coarse graining (Stephens and Waelbroeck, 1999), i.e. that the functional form of the equations for a building block schema is identical to that of the equations for the strings themselves. This means that building blocks for a string are composed, in their turn, by other more coarse grained (lower order) building blocks, which in their turn etc., the whole hierarchy terminating at the 1-schemata. It is precisely the existence of this form invariance and the hierarchical nature of the relationship between the different building blocks that has led to so many new results using the coarse grained formulation. We are thus led to consider whether for generalised recombination the same features appear which can then be further exploited to gain a better theoretical understanding and derive new practical results.

Thus, we begin by considering what happens when we coarse grain such that $h_1\cdots h_\ell \to \bigcup_{h_s} h_1\cdots h_s\cdots h_\ell = h_1\cdots h_{s-1}{*}h_{s+1}\cdots h_\ell$. Thus

$$\begin{aligned} E[\Phi(h_1\cdots h_{s-1}{*}h_{s+1}\cdots h_\ell,t+1)] &= \sum_{h_s} E[\Phi(h_1\cdots h_s\cdots h_\ell,t+1)] \\ &= \sum_{h_s}\sum_{r\in\mathcal{R}_\ell^\ell} p_c(r)p\left(\Gamma(h,I_r),t\right)p\left(\Gamma(h,I_{\bar{r}}),t\right). \end{aligned}$$

If we use the notation $expr/y \leftarrow z$ to mean "replace y with z in expression $expr$", clearly $E[\Phi(h_1\cdots h_{s-1}{*}h_{s+1}\cdots h_\ell,t+1)] = E[\Phi(h\Big/h_s \leftarrow *,t+1)]$. Because in each crossover event allele h_s in an offspring comes from one parent or the other (depending on the value of the s-th bit in the m component of the corresponding GCM) but not both at the same time, it then follows that

$$E[\Phi(h\Big/h_s \leftarrow *,t+1)] \quad = \quad \sum_{r\in\mathcal{R}_\ell^\ell} p_c(r)p\left(\Gamma(h,I_r)\Big/h_s \leftarrow *,t\right)p\left(\Gamma(h,I_{\bar{r}})\Big/h_s \leftarrow *,t\right) \tag{3}$$

(see (Poli and Stephens, 2005b; Poli and Stephens, 2005a) for a proof). This result can be iterated to coarse grain over any number of variables (Poli and Stephens, 2005b; Poli and Stephens, 2005a), leading to the following

Theorem 2 (Schema evolution equation). *Equation 2 is applicable to both strings and schemata of any order.*

Interestingly, we can collect some terms in Equation 3. To see that, let us assume, without loss of generality, that $m_s = 1$. In other words, $s \in I_r$. Therefore

$$\Big(\Gamma(h, I_r)\Big/ h_s \leftarrow *\Big) = \left(\bigcap_{i \in I_r} H_{v_i}^{h_i}\Big/ h_s \leftarrow *\right) = \left(\bigcap_{i \in I_r \setminus \{s\}} H_{v_i}^{h_i}\right) \cap \overbrace{\Big(H_{v_s}^{h_s}\Big/ h_s \leftarrow *\Big)}^{=*^\ell} = \bigcap_{i \in I_r \setminus \{s\}} H_{v_i}^{h_i}$$

where the operator $\setminus$ represents set subtraction and $\{s\}$ is the singleton set including only the integer s.

A similar result holds for $s \in I_{\bar{r}}$. This means that neither m_s nor v_s appear explicitly in any of the terms in Equation 3, except in the factor $p_c(r)$. So, we can rewrite the equation as:

$$\begin{aligned} E[\Phi(h_1 \cdots h_{s-1} * h_{s+1} \cdots h_\ell, t+1)] \;=\; & \sum p_c(m_1 \cdots m_{s-1} * m_{s+1} \cdots m_\ell, v_1 \cdots v_{s-1} * v_{s+1} \cdots v_\ell) \\ & p\Big(\Gamma(h, I_r)\Big/ h_s \leftarrow *, t\Big)\, p\Big(\Gamma(h, I_{\bar{r}})\Big/ h_s \leftarrow *, t\Big) \end{aligned} \tag{4}$$

where the summation ranges over all GCMs $r = (m_1 \cdots m_{s-1} m_{s+1} \cdots m_\ell, v_1 \cdots v_{s-1} v_{s+1} \cdots v_\ell) \in \mathcal{R}_\ell^{\ell-1}$. Naturally this result generalises to any number of "don't care" symbols, leading to the following

Theorem 3. *For a schema h with d don't care symbols at positions $l_1, \cdots, l_d$, the summation in Equation 2 can be turned into a summation over $(m', v') \in \mathcal{R}_\ell^{\ell-d}$ provided the recombination distribution p_c is replaced with the marginal distribution p'_c obtained by summing $p_c(m, v)$ over all m_{l_i} and v_{l_i} for $1 \le i \le d$.*

3.3 A more explicit notation

As we saw in Section 2.1, the intersection of order one schemata can be either a schema or the empty set. When, for a given GCM $r = (m, v) \in \mathcal{R}_\ell^\ell$, v is a permutation of the vector $(1, 2, \cdots, \ell)$, then $\Gamma(h, I_r) = \bigcap_{i \in I_r} H_{v_i}^{h_i}$ is guaranteed to be an ordinary schema and the same applies to $\Gamma(h, I_{\bar{r}})$. In order to be able to express exactly which schemata these are we need to order the sets $I_r = \{i_1, i_2, \cdots, i_{|I_r|}\}$ and $I_{\bar{r}} = \{j_1, j_2, \cdots, j_{|I_{\bar{r}}|}\}$ on the basis of the corresponding entries in the vector v. That is, the elements i_k of I_r are ordered in such a way that $v_{i_k} \le v_{i_{k+1}}$ for any k, and likewise for $I_{\bar{r}}$. For example, if $\ell = 4$ and $r = (m, v) = (1101, (4, 3, 4, 1))$, then I_r is obtained as follows. As before, we first collect the indices of the elements of m that are 1 in a set (in this example, $\{1, 2, 4\}$). Then we sort the elements of this set based on the values of the corresponding elements in v. So, because $v_4 \le v_2 \le v_1$, $I_r = \{4, 2, 1\}$. Naturally, $I_{\bar{r}} = \{3\}$.

With this ordering, when v is a permutation, then $v_{i_k} < v_{i_{k+1}}$ for all k. Therefore we can use the concatenation operator introduced in Section 2.1 to express our building block schemata

$$\Gamma(h, I_r) = \bigotimes_{k=1}^{|I_r|} \left(*^{v_{i_k} - v_{i_{k-1}} - 1} h_{i_k}\right) *^{\ell - v_{i_{|I_r|}}}$$

where we adopt the conventions of Section 2.1 that $v_{i_0} = 0$ and $*^0$ is the empty sequence. For example, if $r = (m, v)$ with $m = 1101$ and $v = (4, 3, 4, 1)$ as above, and $h = 1001$, then $I_r = \{4, 2, 1\}$

and

$$\begin{aligned}\Gamma(h, I_r) &= (*^{v_{i_1}-1}h_{i_1})(*^{v_{i_2}-v_{i_1}-1}h_{i_2})(*^{v_{i_3}-v_{i_2}-1}h_{i_3})*^{4-v_{i_3}} \\ &= (*^{v_4-1}h_4)(*^{v_2-v_4-1}h_2)(*^{v_1-v_2-1}h_1)*^{4-v_1} \\ &= (*^{1-1}h_4)(*^{3-1-1}h_2)(*^{4-3-1}h_1)*^{4-4} \\ &= h_4 * h_2 h_1\end{aligned}$$

We can interpret $\Gamma(h, I_r)$ as a schema also when $v_{i_k} = v_{i_{k-1}}$ for some k, as long as $h_{i_k} = h_{i_{k-1}}$. If this is not the case, then $\Gamma(h, I_r)$ is the empty set $\emptyset$ (naturally $p(\emptyset, t) = 0$). Therefore, in general we can write

$$p\left(\Gamma(h, I_r), t\right) = p\Big(\bigotimes_{\substack{1 \le k \le |I_r| \\ i_k \neq i_{k-1}}} \left(*^{v_{i_k}-v_{i_{k-1}}-1}h_{i_k}\right) *^{\ell - v_{i_{|I_r|}}}, t\Big) \times \prod_{\substack{1 \le k \le |I_r| \\ i_k = i_{k-1}}} \delta(h_{i_k} = h_{i_{k-1}})$$

3.4 Symmetry induced by independent parent selection

Collecting terms with common factors in string and schema evolution equations reveals a symmetry with respect to the order in which parents are selected. Interestingly, this is captured by the notion of negation for GCMs introduced in Section 2.2.

Let us make the effects induced by the symmetry in parent selection explicit. In Equation 2, for any given GCM $r = (m, v)$ we have a term of the form

$$p_c(r)p(\Gamma(h, I_r), t)p(\Gamma(h, I_{\bar{r}}), t).$$

Obviously, we also have a term corresponding to the negation $\bar{r}$ of such a mask. This is

$$p_c(\bar{r})p(\Gamma(h, I_{\bar{r}}), t)p(\Gamma(h, I_{\bar{\bar{r}}}), t) = p_c(\bar{r})p(\Gamma(h, I_{\bar{r}}), t)p(\Gamma(h, I_r), t),$$

since $\bar{\bar{r}} = r$. Therefore, the two terms involve the same selection probabilities, although with different coefficients, namely $p_c(r)$ and $p_c(\bar{r})$. So, we can rewrite Equation 2 as

$$E[\Phi(h, t+1)] = \sum_{r \in \widetilde{\mathcal{R}}_\ell^\ell} \left(p_c(r) + p_c(\bar{r})\right) p(\Gamma(h, I_r), t)p(\Gamma(h, I_{\bar{r}}), t) \tag{5}$$

where $\widetilde{\mathcal{R}}_\ell^\ell$ is any set representing half of the space of possible GCMs such that $\forall r \in \widetilde{\mathcal{R}}_\ell^\ell \implies \bar{r} \notin \widetilde{\mathcal{R}}_\ell^\ell$ (for example, $\widetilde{\mathcal{R}}_\ell^\ell$ could be the set of all GCM (m, v) with $m_1 = 1$, i.e. $\widetilde{\mathcal{R}}_\ell^\ell = (1*^{\ell-1}, *^\ell)$).

3.5 The case $p_{xo} < 1$

Let us re-consider our earlier assumption that crossover is performed with 100% probability, i.e. that no offspring are created by selection followed by cloning (a.k.a. reproduction) alone. What would happen if recombination was applied with probability p_{xo} and reproduction with probability $(1 - p_{xo})$? The evolution equations for a generic schema h would transform into

$$E[\Phi(h, t+1)] = (1 - p_{xo})p(h, t) + p_{xo} \sum_{r \in \mathcal{R}_\ell^\ell} p_c(r)p(\Gamma(h, I_r), t)p(\Gamma(h, I_{\bar{r}}), t). \tag{6}$$

This might seem rather different from Equation 2, but this is not the case. Because $\mathcal{R}_\ell^\ell$ includes the trivial GCM $r' = (11\cdots1, (1, 2, 3, \cdots, \ell))$ and because $I_{r'} = \mathcal{N}_\ell$ and $\bar{I}_{r'} = \emptyset$, then, since $\Gamma(h, \emptyset) = *^\ell$, $p(*^\ell, t) = 1$ and $\Gamma(h, \mathcal{N}_\ell) = h$, the action of the GCM r is simply to produce an exact copy of the first parent. That is, when the mask r' is used, crossover behaves exactly like reproduction. So, we can model crossover and reproduction together by using Equation 2 but with a modified recombination distribution $p'_c(r) = p_{xo}p_c(r) + (1 - p_{xo})\delta(r = r')$. So, any results obtained with Equation 2 (that is for $p_{xo} = 1$) can trivially be generalised to the case $p_{xo} < 1$.

3.6 Infinite population assumption

Infinitely large populations are a mathematical idealisation that we will use extensively in the rest of the paper. This is an essential assumption if we wish to separate the intrinsic operator biases from the bias due to genetic drift. Also, from a practical point of view, the infinite-population assumption makes life easier whenever one wants to iterate schema and string evolution equations to evaluate the behaviour of an EA over multiple time steps – a technique we will use to assess the interactions between selection and recombination in Section 10. It is then natural to ask what sort of errors one should expect to see when going from finite to infinite populations. Interestingly, it is possible to assess this from the schema evolution equations themselves (see (Poli et al., 1998)).

Let us start by considering that for a finite population, schema (and string) frequencies are always of the form

$$\Phi(h,t) = \frac{m(h,t)}{M}$$

where $m(h,t)$ is the number of strings in the population which match the schema h and M is the population size. If we know the state of the population at time t, the number of strings matching a schema at the next generation, $m(h,t+1)$, is a stochastic variable which is binomially distributed (Poli et al., 1998). Its success probability $\pi_h(t)$ is given by the r.h.s. of Equation 2, that is

$$\pi_h(t) = \sum_{r \in \mathcal{R}_\ell^\ell} p_c(r) p(\Gamma(h, I_r), t) p(\Gamma(h, I_{\bar{r}}), t),$$

and so, we can re-interpret Equation 2 as a statement about the mean of $m(h,t+1)$, i.e.

$$E[m(h,t+1)] = M\pi_h(t).$$

Once we know the success probability for a binomially distributed variable, we can calculate all the moments of the distribution and, indeed, the distribution itself. In the specific case of the variable $m(h,t+1)$, we have

$$\Pr\{m(h,t+1) = k\} = \binom{M}{k} \pi_h(t)^k (1-\pi_h(t))^{M-k}.$$

Of particular interest for our infinite-vs-finite population discussion is the variance of $m(h,t+1)$, which is

$$\mathrm{Var}[m(h,t+1)] = M\pi_h(t)(1-\pi_h(t)).$$

From this we can easily compute the standard deviation of $\Phi(h,t+1)$, obtaining

$$\mathrm{StdDev}[\Phi(h,t+1)] = \sqrt{\frac{\pi_h(t)(1-\pi_h(t))}{M}}.$$

So, *the deviations we should expect in each generation between the infinite population model predictions and the finite-population behaviour are of the order* $O\left(\frac{1}{\sqrt{M}}\right)$. These deviations can be seen as "noise", and, so, naturally, their importance depends on the amplitude of the corresponding "signal", $\pi_h(t)$ (see discussion on signal-to-noise ratios in (Poli et al., 1998)).

In addition, it is possible to use exact schema evolution equations to exactly study finite population evolution. Indeed, it is easy to construct a Markov chain model for generalised recombination. This can be done by applying Equation 2 to calculate the expected frequencies of all strings in Ω^ℓ at the next generation. For a given population, these can then be used as the success probabilities for a multinomial distribution which, for any possible population, provides the probability of it being the successor of the current population. So, by iterating this process for all possible "current" populations, we obtain the entries of the transition matrix of the Markov chain (we used this technique is (Poli et al., 2004) for an example). Then, iterating the chain gives an exact probabilistic characterisation of the behaviour of a GA with finite populations.

The main obstacle to this procedure is that, as a stochastic process, a GA has an enormous number of possible states (populations). For the case of binary strings of length ℓ, a GA with a population of M individuals can be in any of $N = (M + 2^\ell - 1)!/M!(2^\ell - 1)!$ different states (Nix and Vose, 1992). So, a Markov chain for a GA requires an immense $(N \times N)$ transition matrix, implying computations that are much worse than exponential. This is why we prefer to investigate finite-population effects using actual runs. We will do this in Section 11.

3.7 Examples

As a first example, let us write the evolution equations for a generic string of length $\ell = 2$ from Equation 2

$$
\begin{aligned}
& E[\Phi(ab, t+1)] \qquad (7)\\
&= p_c(11,(1,1))p(a* \cap b*,t) + p_c(11,(1,2))p(ab,t) + p_c(10,(1,1))p(a*,t)p(b*,t)\\
&+ p_c(10,(1,2))p(a*,t)p(*b,t) + p_c(11,(2,1))p(ba,t) + p_c(11,(2,2))p(*a \cap *b,t)\\
&+ p_c(10,(2,1))p(*a,t)p(b*,t) + p_c(10,(2,2))p(*a,t)p(*b,t) + p_c(01,(1,1))p(b*,t)p(a*,t)\\
&+ p_c(01,(1,2))p(*b,t)p(a*,t) + p_c(00,(1,1))p(a* \cap b*,t) + p_c(00,(1,2))p(ab,t)\\
&+ p_c(01,(2,1))p(b*,t)p(*a,t) + p_c(01,(2,2))p(*b,t)p(*a,t) + p_c(00,(2,1))p(ba,t)\\
&+ p_c(00,(2,2))p(*a \cap *b,t).
\end{aligned}
$$

Collecting terms with common factors in this equation clearly shows the symmetry with respect to the order in which parents are selected mentioned in Section 3.4. For example, in the terms $p_c(11,(1,1))p(a* \cap b*,t)$ and $p_c(00,(1,1))p(a* \cap b*,t)$ we have that $(11,(1,1)) = (\overline{00},(1,1)) = \overline{(00,(1,1))}$ (by definition of negation for GCMs).

If one replaces a and b with some values from Ω, all schema intersections of the form $*a \cap *b$ turn into either the order-1 schema $*a$ (if $a = b$) or into the empty set $\emptyset$, for which $p(\emptyset,t) = 0$, and so it is possible to further simplify the equation. For example, if $a = b = 1$ and all GCMs have equal probability ($p_c(r) = 1/16$), we obtain

$$
E[\Phi(11,t+1)] = \frac{1}{8}\left[p(1*,t)^2 + p(1*,t) + 2p(1*,t)p(*1,t) + p(*1,t)^2 + 2p(11,t) + p(*1,t)\right]
$$

Notice that in order to solve for the dynamics of the strings in Equation 7 we need to have a solution for the building blocks $a*$, $*a$, $b*$ and $*b$. Notice, also, that expected frequency of ab is a linear function of the selection probabilities of that string and its permutation ba and a linear-quadratic function of the selection probabilities of lower order schemata (building blocks). That is, if we term the latter $\mathfrak{b}_{ab}(t)$, we can rewrite:

$$
\begin{aligned}
& E[\Phi(ab,t+1)]\\
&= (p_c(11,(1,2)) + p_c(00,(1,2)))p(ab,t) + (p_c(11,(2,1)) + p_c(00,(2,1)))p(ba,t) + \mathfrak{b}_{ab}(t),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathfrak{b}_{ab}(t) &= p_c(11,(1,1))p(a* \cap b*,t) + p_c(10,(1,1))p(a*,t)p(b*,t) + p_c(10,(1,2))p(a*,t)p(*b,t)\\
&+ p_c(11,(2,2))p(*a \cap *b,t) + p_c(10,(2,1))p(*a,t)p(b*,t) + p_c(10,(2,2))p(*a,t)p(*b,t)\\
&+ p_c(01,(1,1))p(b*,t)p(a*,t) + p_c(01,(1,2))p(*b,t)p(a*,t) + p_c(00,(1,1))p(a* \cap b*,t)\\
&+ p_c(01,(2,1))p(b*,t)p(*a,t) + p_c(01,(2,2))p(*b,t)p(*a,t) + p_c(00,(2,2))p(*a \cap *b,t)
\end{aligned}
$$

and acts as a "source" for creating the string ab from lower order objects.

As an example, the evolution equation for the schema $a*$ (a building block for ab) can be

obtaining by simply replacing b with $*$ in Equation 7, obtaining

$$
\begin{aligned}
&E[\Phi(a*, t+1)] \\
&\quad = \; p_c(11,(1,1))p(a* \cap **, t) + p_c(11,(1,2))p(a*, t) + p_c(10,(1,1))p(a*, t)p(**, t) \\
&\quad + \; p_c(10,(1,2))p(a*, t)p(**, t) + p_c(11,(2,1))p(*a, t) + p_c(11,(2,2))p(*a \cap **, t) \\
&\quad + \; p_c(10,(2,1))p(*a, t)p(**, t) + p_c(10,(2,2))p(*a, t)p(**, t) + p_c(01,(1,1))p(**, t)p(a*, t) \\
&\quad + \; p_c(01,(1,2))p(**, t)p(a*, t) + p_c(00,(1,1))p(a* \cap **, t) + p_c(00,(1,2))p(a*, t) \\
&\quad + \; p_c(01,(2,1))p(**, t)p(*a, t) + p_c(01,(2,2))p(**, t)p(*a, t) + p_c(00,(2,1))p(*a, t) \\
&\quad + \; p_c(00,(2,2))p(*a \cap **, t) \\
&\quad = \; \big(p_c(11,(1,1)) + p_c(11,(1,2)) + p_c(10,(1,1)) + p_c(10,(1,2)) + \\
&\qquad\; p_c(01,(1,1)) + p_c(01,(1,2)) + p_c(00,(1,1)) + p_c(00,(1,2))\big)\, p(a*, t) \\
&\quad + \; \big(p_c(11,(2,1)) + p_c(11,(2,2)) + p_c(10,(2,1)) + p_c(10,(2,2)) + \\
&\qquad\; p_c(01,(2,1)) + p_c(01,(2,2)) + p_c(00,(2,1)) + p_c(00,(2,2))\big)\, p(*a, t).
\end{aligned}
$$

That is

$$
\begin{aligned}
E[\Phi(a*, t+1)] &= \big(p_c(**,(1,1)) + p_c(**,(1,2))\big)p(a*, t) + \big(p_c(**,(2,1)) + p_c(**,(2,2))\big)p(*a, t) \\
&= p_c(**,(1,*))p(a*, t) + p_c(**,(2,*))p(*a, t),
\end{aligned}
$$

as predicted from Equation 4. Note that there is now no corresponding $\mathfrak{b}(t)$ for order-1 schemata as they cannot be created from any more elementary object.

A much deeper analysis of the $\ell = 2$ case is provided in (Stephens and Poli, 2005b), where a complete, exact solution, is derived, showing how the dynamical behaviour is radically different from that of homologous crossover. Even in such a simple case, new qualitatively different behaviours are observed. For example, inversion is shown to potentially introduce oscillations in the dynamics, while gene duplication leads to an asymmetry between homogeneous and heterogeneous strings. Also, all non-homologous operators lead to allele "diffusion" along the chromosome.

The general form of the evolution equations for $\ell = 3$ for the generic string abc is much bigger than for $\ell = 2$ (see explicit expansion in (Poli and Stephens, 2005b; Poli and Stephens, 2005a)), this including 216 terms – a number that, although quite big, is only $\frac{1}{64}$ of the number of terms one would get in the absence of coarse graining. Interestingly, also in this case, the expected frequency of abc is a linear function of the selection probabilities of that string and all its permutations and a linear-quadratic function of the selection probabilities of lower order schemata (building blocks). That is:

$$
\begin{aligned}
&E[\Phi(abc, t+1)] \\
&\quad = \; (p_c(111,(1,2,3)) + p_c(000,(1,2,3)))p(abc, t) + (p_c(111,(1,3,2)) + p_c(000,(1,3,2)))p(acb, t) \\
&\quad + \; (p_c(111,(2,1,3)) + p_c(000,(2,1,3)))p(bac, t) + (p_c(111,(2,3,1)) + p_c(000,(2,3,1)))p(cab, t) \\
&\quad + \; (p_c(111,(3,1,2)) + p_c(000,(3,1,2)))p(bca, t) + (p_c(111,(3,2,1)) + p_c(000,(3,2,1)))p(cba, t) \\
&\quad + \; \mathfrak{b}_{abc}(t)
\end{aligned}
$$

Again, the symmetry in the selection process induces a symmetry in the equations whereby terms relating to a mask and its negation always collect.

Naturally, in order to solve for the string dynamics we need to have the dynamics of the building blocks that determine the driving term $\mathfrak{b}_{abc}(t)$. Fortunately, again, the equations for the building blocks can be derived by trivial syntactic manipulations on the string equations (see (Poli and Stephens, 2005b; Poli and Stephens, 2005a; Stephens and Poli, 2005a)).

4 Hierarchical nature of schema evolution equations

In the examples in Section 3.7 for $\ell = 2$ and $\ell = 3$, string and schema evolution equations have right-hand sides with the same structure, i.e. with a linear part which depends on the selection probabilities of schemata of the same order as the schema on the left-hand side of the equation, and a non-linear forcing term which depends on lower-order schemata. The only exception to this is order one objects, in which case there is no forcing term. So, as in the case of normal recombination a hierarchical structure emerges, where objects at a higher level of the hierarchy depend on the evolution of objects at a lower level. Strings are at the highest level of the hierarchy and order one schemata the lowest. As we will show in a moment this hierarchical organisation is a general property of schema evolution equations. To show this we will need to analyse the terms in Equation 2 in more detail.

Let us consider a generic schema h and let us define $D(h)$ to be the set of loci where h has defining symbols, that is $D(h) = \{i | h_i \neq *\}$. Then, we can split the set of GCMs $\mathcal{R}_\ell^\ell$ into three disjoint sets:

$$\mathcal{R}_1(h) = \{r | D(h) \subseteq I_r\}, \quad \mathcal{R}_2(h) = \{r | D(h) \subseteq I_{\bar{r}}\}, \quad \text{and} \quad \mathcal{R}_3(h) = \mathcal{R}_\ell^\ell \setminus \mathcal{R}_1(h) \setminus \mathcal{R}_2(h),$$

where we could alternatively have expressed $\mathcal{R}_2(h) = \{r | D(h) \cap I_r = \emptyset\}$ or, extending the notion of negation from GCMs to sets of GCMs, as $\mathcal{R}_2(h) = \overline{\mathcal{R}}_1(h) = \{r | \bar{r} \in \mathcal{R}_1(h)\}$. Then we can rewrite Equation 2 as

$$\begin{aligned}
&E[\Phi(h, t+1)] \\
&= \sum_{r \in \mathcal{R}_1(h)} p_c(r) p(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t) p(\bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}, t) + \sum_{r \in \mathcal{R}_2(h)} p_c(r) p(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t) p(\bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}, t) \\
&+ \sum_{r \in \mathcal{R}_3(h)} p_c(r) p(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t) p(\bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}, t) \\
&= \sum_{r \in \mathcal{R}_1(h)} p_c(r) p(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t) p(*^\ell, t) + \sum_{r \in \mathcal{R}_2(h)} p_c(r) p(*^\ell, t) p(\bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}, t) \\
&+ \sum_{r \in \mathcal{R}_3(h)} p_c(r) p(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t) p(\bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}, t) \\
&= \sum_{r \in \mathcal{R}_1(h)} p_c(r) p(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t) + \sum_{r \in \overline{\mathcal{R}}_1(h)} p_c(r) p(\bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}, t) \\
&+ \sum_{r \in \mathcal{R}_3(h)} p_c(r) p(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t) p(\bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}, t) \\
&= \sum_{r \in \mathcal{R}_1(h)} (p_c(r) + p_c(\bar{r})) p(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t) + \sum_{r \in \mathcal{R}_3(h)} p_c(r) p(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t) p(\bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}, t) \qquad (8)
\end{aligned}$$

Let us further split $\mathcal{R}_1(h)$ into

$$\mathcal{R}_1'(h) = \{r \in \mathcal{R}_1(h) : \forall i, j \in D(h), i \neq j \implies v_i \neq v_j\} \quad \text{and} \quad \mathcal{R}_1''(h) = \mathcal{R}_1(h) \setminus \mathcal{R}_1'(h).$$

Also, let us also assume that the elements of I_r are ordered based on the values of the elements of v (see Section 3.3). Then, for $r \in \mathcal{R}_1'(h)$ we have $\bigcap_{i \in I_r} H_{v_i}^{h_i} = \bigotimes_{k=1}^{|I_r|} \left(*^{v_{i_k} - v_{i_{k-1}} - 1} h_{i_k} \right) *^{\ell - v_{i_{|I_r|}}}$ and so we can transform Equation 8 into the following

Theorem 4 (Schema hierarchy). *The evolution of a string or schema h under selection and generalised recombination is governed by the following equation*

$$E[\Phi(h, t+1)] = \sum_{r \in \mathcal{R}_1'(h)} (p_c(r) + p_c(\bar{r})) p \left(\bigotimes_{k=1}^{|I_r|} \left(*^{v_{i_k} - v_{i_{k-1}} - 1} h_{i_k} \right) *^{\ell - v_{i_{|I_r|}}}, t \right) + \mathfrak{b}_h(t) \qquad (9)$$

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

$$\begin{aligned}\mathfrak{b}_h(t) &= \sum_{r\in\mathcal{R}_1''(h)} (p_c(r)+p_c(\bar{r}))p(\bigcap_{i\in I_r} H_{v_i}^{h_i},t) \\ &+ \sum_{r\in\mathcal{R}_3(h)} p_c(r)p(\bigcap_{i\in I_r} H_{v_i}^{h_i},t)p(\bigcap_{i\in I_{\bar{r}}} H_{v_i}^{h_i},t)\end{aligned} \tag{10}$$

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

The order of the schemata $\bigotimes_{k=1}^{|I_r|}\left(*^{v_{i_k}-v_{i_{k-1}}-1}h_{i_k}\right)*^{\ell-v_{i_{|I_r|}}}$ is the same as h because, for $r\in\mathcal{R}_1'$, we have that $D(h)\subseteq I_r$ and $h_i=*$ for any $i\in I_r\setminus D(h)$. Also, the order of the schemata in $\mathfrak{b}_h(t)$ is lower than h for the following reasons: $r\in\mathcal{R}_1''(h) \implies \exists i\neq j : v_i=v_j$, and so $\bigcap_{i\in I_r} H_{v_i}^{h_i}$ is either $\emptyset$ or is a schema of lower order than h. Also, $r\in\mathcal{R}_3(h) \implies \exists i,j\in D(h), i\notin I_r, j\notin I_{\bar{r}}$, and so $\bigcap_{i\in I_r} H_{v_i}^{h_i}$ and $\bigcap_{i\in I_{\bar{r}}} H_{v_i}^{h_i}$ are either $\emptyset$ or schemata of lower order than h.

Theorem 4 generalises the properties we observed in Section 3.7 for the cases $\ell=2$ and $\ell=3$. In particular, for order-1 schemata, the right-hand side of Equation 10 collapses to 0. This is because they cannot be composed of lower order objects.[6] These objects, therefore, evolve independently but contribute to all higher-order schemata. So, *order one schemata act as pacemakers for a genetic system evolving under generalised recombination* just as they do for the case of homologous crossover. For these reasons we will analyse the evolution equations for such a case in more detail in Section 6. Before we do that, however, we want to show how the result in Theorem 4 relate to previously published work on schema theory.

5 Relationship with previous schema theory

As we have mentioned in Section 2.2 generalised recombination is more general than the notion of homologous crossover. Therefore, all general results obtained in this paper extend previous results including the exact schema theory for one-point crossover in (Stephens and Waelbroeck, 1999) and the exact schema theory for general homologous crossover in (Stephens, 2001).

As an illustrative example, in this section we will first derive a generalisation of Holland's schema theorem for the case of generalised recombination, and then we will show how Holland's original theorem emerges from a specialisation of this.

Theorem 5 (Generalised Holland's schema theorem). *A lower bound for the expected frequency a string or schema h at the next generation under selection, reproduction and generalised recombination applied with probability p_{xo} is given by the following equation*

$$\begin{aligned}E[\Phi(h,t+1)] &\geq (1-p_{xo})p(h,t) \\ &+ p_{xo}\sum_{r\in\mathcal{R}_1(h)} (p_c(r)+p_c(\bar{r}))p(\bigcap_{i\in D(h)} H_{v_i}^{h_i},t) \\ &+ p_{xo}\sum_{r\in\mathcal{R}_3(h)} p_c(r)p(\bigcap_{i\in D(h)} H_{v_i}^{h_i},t)p(\bigcap_{i\in D(h)} H_{v_i}^{h_i},t)\end{aligned} \tag{11}$$

(See Appendix A.1 for a proof.)

It is easy to see that this equation turns into Holland's schema theorem (to be more precise, into Whitley's version of it (Whitley, 1994)). For this purpose, we apply the previous theorem to the case of a homologous crossover operator.

[6] Formally one can verify this by observing that, for order-1 schemata, the set $D(h)$ contains only one element, and so all $r\in\mathcal{R}_\ell^\ell$ can only either be in $\mathcal{R}_1(h)$ or $\mathcal{R}_2(h)$. So, $\mathcal{R}_3(h)=\emptyset$. In addition, also $\mathcal{R}_1''(h)=\emptyset$.

Corollary 1. *A lower bound for the expected frequency of a string or schema, h, at the next generation under selection, reproduction and homologous recombination applied with probability p_{xo}, is given by the following equation*

$$E[\Phi(h,t+1)] \quad \geq \quad p(h,t)\left[1-p_{xo}\left(\sum_{r\in\mathcal{R}_3(h)} p_c(r)\right)\cdot(1-p(h,t))\right]. \tag{12}$$

(See Appendix A.2 for a proof.)

Equation 12 is a more recognisable result. Indeed, under one-point crossover, the sum over crossover masks $\sum_{r\in\mathcal{R}_3(h)} p_c(r)$ turns into the traditional $\frac{\mathcal{L}(h)}{\ell-1}$, where $\mathcal{L}(h)$ is the defining length (see Section 2.1) of the schema h. So, Equation 12 *is* a version of Holland's schema theorem (see (Whitley, 1994)).

6 Dynamics of order-1 schemata

As we have shown that order-1 schemata play a privileged role in recombination, let us focus schemata H_s^a where only one allele is specified. By coarse-graining on the recombination distribution as illustrated in Section 2.2, the schema evolution equations for these schemata transform 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., 2002; 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

$$\vec{\Phi}^a(t+1)=A\vec{\Phi}^a(t) \tag{13}$$

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 one schemata in the infinite-population flat-landscape case is given by

$$\vec{\Phi}^a(t)=A^t\vec{\Phi}^a(0) \tag{14}$$

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 with $\Phi^*(H_s^a)$.

6.1 Fixed points

Let us look for fixed points for the dynamical system defined by Equation 13. 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 for 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, finding analytically other fixed points may not be simple. Also, determining whether a fixed point is a global attractor for the system is non-trivial.[7] 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.

6.1.1 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 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.)

6.1.2 Order-1 mixing recombination

Theorem 6. *In a GA with an infinite population, with a flat fitness landscape and using an order-1 mixing GRD $p_c(r)$, 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 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.

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

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 this 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 assume that λ_i are the eigenvalues of A and v_i are the corresponding right eigenvectors (normalised so as to be unit vectors). Also, let u_i be the left eigenvectors of A. Then we can express the initial conditions $\vec{\Phi}^a(0)$ as a linear combination of eigenvectors, 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 $i = m$,

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

So, 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. □

This result is compatible with the family of fixed points $\Phi^*(H^a_s) = c(a)$ 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 2. *Under the conditions of the previous theorem, if the GRD is duplication-free, order-1 schema frequencies converge to the following fixed point*

$$\Phi^*(H^a_s) = \frac{1}{\ell} \sum_i \Phi(H^a_i, 0).$$

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 $\alpha_i = \sum_k p_c(*^\ell, *^{k-1} i *^{\ell-k})$ (see Section 6.1.1) are compatible with this type of GRD. Let us assume that $\alpha_i < 1$ for some i. Because the A matrix is row stochastic then the sum of all the elements in the matrix must be ℓ. So, matrix elements not in column i must add up to $\ell - \alpha_i > \ell - 1$. Therefore, 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, then 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 (in addition to being row stochastic), i.e. $\sum_s a_{sk} = \sum_s p_c(*^\ell, *^{s-1}k*^{\ell-s}) = 1$, then the 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

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

□

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 left eigenvectors coincide with the 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).

6.1.3 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

$$\Phi^*(H_s^a) = \Phi(H_s^a, 0).$$

6.1.4 Fully disconnected recombination cliques

Let $Q(p_c)$ be the set of recombination cliques 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 clique 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 clique (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 cliques, but where the cliques themselves are completely disconnected. In other words, we consider the case where the recombination clique DAG 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 $\vec{\Phi}_n^a$ 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:

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

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

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

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

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

If each clique is order-1 mixing then the results presented in Section 6.1.2 can trivially be generalised to fully disconnected cliques. In particular, if there is no form of duplication, then we have the following attractors for the cliques:

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

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

7 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 4 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 9 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 Section 2.1, we use the notation $H_{l_1,l_2,\cdots,l_n}^{a_1,a_2,\cdots,a_n}$ 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 $h = H_{l_1,l_2,\cdots,l_n}^{a_1,a_2,\cdots,a_n}$, Equation 9 can be rewritten as

$$\Phi(h,t+1) \quad = \quad \sum_{r\in\mathcal{R}_1'(h)} (p_c(r) + p_c(\bar{r}))p\left(H_{v_1,v_2,\cdots,v_n}^{a_1,a_2,\cdots,a_n}, t\right) + \mathfrak{b}_h(t). \tag{15}$$

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

$$\begin{aligned}\Phi(h,t+1) \quad &= \quad \sum_{v_1\neq v_2\neq\cdots\neq v_n=1}^{\ell} \left[p_c(H_{l_1,l_2,\cdots,l_n}^{0,0,\cdots 0}, H_{l_1,l_2,\cdots,l_n}^{v_1,v_2,\cdots,v_n}) + p_c(H_{l_1,l_2,\cdots,l_n}^{1,1,\cdots 1}, H_{l_1,l_2,\cdots,l_n}^{v_1,v_2,\cdots,v_n})\right] \\ &\times \quad p\left(H_{v_1,v_2,\cdots,v_n}^{a_1,a_2,\cdots,a_n}, t\right) + \mathfrak{b}_h(t)\end{aligned} \tag{16}$$

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 16 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. So, 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 16. In vector notation the system of equations turns into the following dynamical system

$$\vec{\Phi}^{a_1,a_2,\cdots,a_n}(t+1) = A^{a_1,a_2,\cdots,a_n}\vec{\Phi}^{a_1,a_2,\cdots,a_n}(t) + \vec{\mathfrak{b}}^{a_1,a_2,\cdots,a_n}(t), \tag{17}$$

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 $p_c(H_{l_1,l_2,\cdots,l_n}^{0,0,\cdots 0}, H_{l_1,l_2,\cdots,l_n}^{v_1,v_2,\cdots,v_n}) + p_c(H_{l_1,l_2,\cdots,l_n}^{1,1,\cdots 1}, H_{l_1,l_2,\cdots,l_n}^{v_1,v_2,\cdots,v_n})$. 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.

7.1 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 4 and using the more explicit notation introduced in Section 3.3, we can write the schema evolution equations for these schemata as:

$$\begin{aligned} E[\Phi(H^{a,b}_{s,u}, t+1)] &= \sum_{v_s \neq v_u = 1}^{\ell} \left(p_c(H^{0,0}_{s,u}, H^{v_s,v_u}_{s,u}) + p_c(H^{1,1}_{s,u}, H^{v_s,v_u}_{s,u})\right) p\left(H^{a,b}_{v_s,v_u}, t\right) \\ &+ \delta(a=b) \sum_{v_s=1}^{\ell} \left(p_c(H^{0,0}_{s,u}, H^{v_s,v_s}_{s,u}) + p_c(H^{1,1}_{s,u}, H^{v_s,v_s}_{s,u})\right) p\left(H^{a}_{v_s}, t\right) \qquad (18) \\ &+ \sum_{v_s,v_u=1}^{\ell} \left(\sum_{m_s \neq m_u \in \{0,1\}} p_c(H^{m_s,m_u}_{s,u}, H^{v_s,v_u}_{s,u}) \right) p\left(H^{a}_{v_s}, t\right) p\left(H^{b}_{v_u}, t\right). \end{aligned}$$

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 $a = 1, b = 0$.

7.2 Fixed points

In order to say something more about the properties of the dynamical systems in Equation 17 we need to introduce the following general result (see Section A.3 for a proof):

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

$$x(t+1) = Ax(t) + b(t) \qquad (19)$$

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 following 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 10 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, then, by Lemma 1, there is a unique, non-negative fixed point

$$\vec{\Phi}^{a_1,a_2,\cdots,a_n *} = (I - A^{a_1,a_2,\cdots,a_n})^{-1}\ \vec{\mathfrak{b}}^{a_1,a_2,\cdots,a_n *}. \qquad (20)$$

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 20 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 20 is really a *family of fixed-points*, different fixed-points in the family being reached depending on the initial conditions.

8 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 18 obtaining:

$$\begin{aligned}\Phi(H^{a,b}_{s,u}, t+1) &= \sum_{v_s \neq v_u = 1}^{\ell} \left(p_c(H^{0,0}_{s,u}, H^{v_s,v_u}_{s,u}) + p_c(H^{1,1}_{s,u}, H^{v_s,v_u}_{s,u})\right) \Phi\left(H^{a,b}_{v_s,v_u}, t\right) \\ &+ \sum_{v_s \neq v_u = 1}^{\ell} \left(\sum_{m_s \neq m_u \in \{0,1\}} p_c(H^{m_s,m_u}_{s,u}, H^{v_s,v_u}_{s,u}) \right) \Phi\left(H^{a}_{v_s}, t\right) \Phi\left(H^{b}_{v_u}, t\right)\end{aligned} \tag{21}$$

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 20)?

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 21 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 21 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 Section A.4 for a proof):

Theorem 7 (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*

$$\Phi^*(h_1 \cdots h_\ell) = \prod_{q \in Q(p_c)} \prod_{i \in q} c(q, h_i) \tag{22}$$

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 one schemata described in Section 6.

It is interesting to rewrite Equation 22 in a slightly different form. If $\nu(h, n, a)$ represents the number of times symbol a appears in one of the loci in clique n of the string or schema h, and Ω

represents our alphabet, then

$$\Phi^*(h) = \prod_{n \in Q(p_c)} \prod_{a \in \Omega} (c(n,a))^{\nu(h,n,a)}. \tag{23}$$

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.[8]

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 of a complete eigenvalue/eigenvector analysis.

Theorem 8 (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

$$\Phi^*(h_1 \cdots h_\ell) = \prod_{i=1}^{\ell} c(h_i) \tag{24}$$

with

$$c(a) = \frac{1}{\ell} \sum_i \Phi(H_i^a, 0) \tag{25}$$

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 clique 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 2 gives us the answer. □

9 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.

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

9.1 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 5. 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 (because with these operators, when one crosses an individual with itself, one gets that individual back again). This is illustrated in Figure 4. Finally, we know that point 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.[9] The crossover bias, however, should generally be expected to be different. In Section 8 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, 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 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 near the Geiringer manifold in virtually all cases.

In fact, we should generally expect to see trajectories like the ones illustrated in Figure 6. That is, we expect that there would be a strong initial crossover bias and that that would be 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 crossover's reaction. 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".[10]

[9] 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.

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

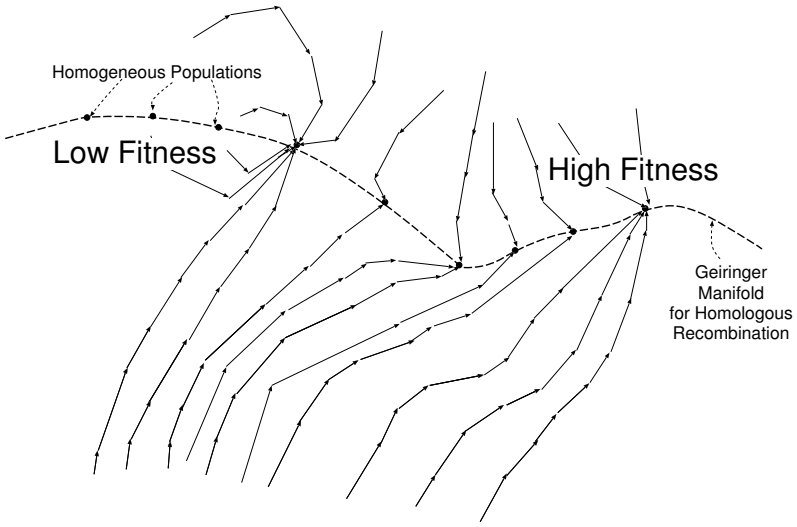

Figure 4: Selection bias. The plane represent the space of all possible populations. Circles represent homogeneous populations. Arrows represent the possible trajectories of a GA initialised in different areas of the population space. The system tends to converge towards high fitness populations, but this may or may not happen (depending on its initial state).

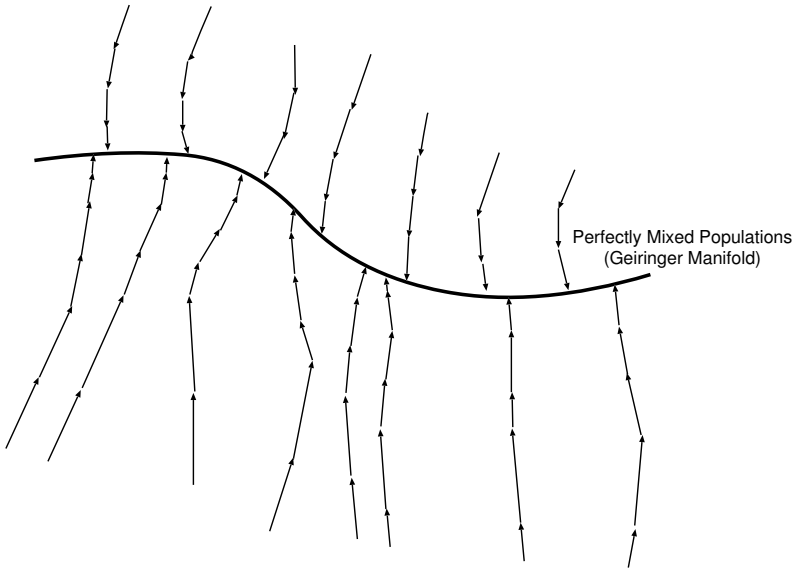

Figure 5: Crossover bias. When crossover alone acts on the population, the system moves towards the Geiringer manifold. The trajectory of the system and the fixed-point reached on the manifold depend on the initial conditions.

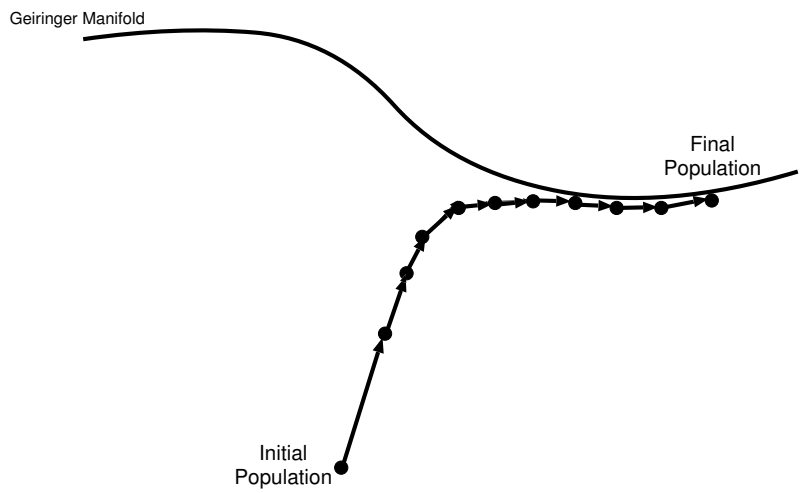

Figure 6: Typical interaction between the selection bias and the bias of generalised recombination. Crossover tries to push the population towards the Geiringer manifold (the sub-space of all perfectly mixed populations). Selection tries to drive the population towards highly-fit homogeneous populations. 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.

In the absence of selection, the manifold is described by the factorisation in Equation 24. 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

$$\Phi(h_1 \cdots h_\ell, t) = \prod_{i=1}^{\ell} c(h_i, t), \tag{26}$$

where, unlike Equation 25, 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 $c(*,t) = 1$). Equation 26 is a generalisation of what was found in the case of homologous crossover and selection in (Stephens, 2001).

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?

9.2 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

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 stays 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.

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.

9.2.1 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 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's 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 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 11.

9.2.2 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 11, 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).[11]

[11] 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.

9.2.3 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 6.1.1) 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 to what is stated in Theorem 8 which holds for infinite populations. This is because the fixed-points for order one schemata provided in Section 6 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 25) 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 11.

10 "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 7 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 clique, $\mathcal{N}_\ell = \{1, \cdots, \ell\}$, 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 7, 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 the fixed point provided in Equation 22. 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 22 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 $\Phi^*(110) = \Phi^*(011) =$

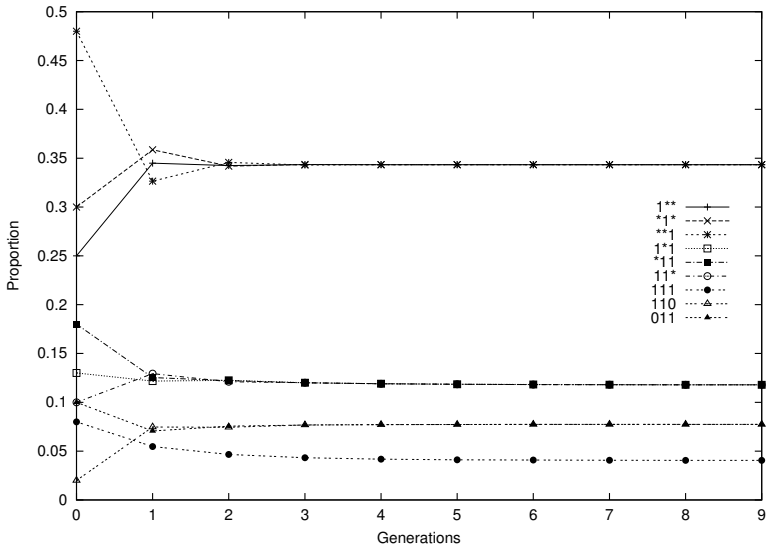

Figure 7: Infinite-population dynamics of strings and schemata for $\ell = 3$ and a duplication-free, order-1 mixing, random recombination distribution.

$c(\mathcal{N}_\ell, 1)^2 c(\mathcal{N}_\ell, 0) = \frac{103 \times 197^2}{300^3} \approx 0.077$ 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 2, i.e. it is one for which

$$c(n, a) = \frac{1}{|n|} \sum_{i \in n} \Phi(H_i^a, 0). \tag{27}$$

Since in this particular example we only have one clique,

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

In order to understand the behaviour of systems where crossover is used in conjuction with selection we have 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 8 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. That is, after a transient, the population moves along the Geiringer manifold for generalised crossover.

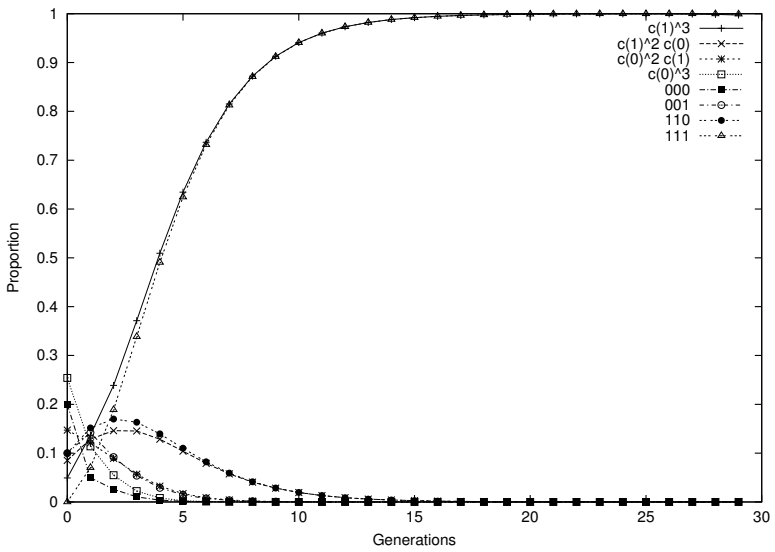

Figure 8: Infinite-population dynamics of a representative sample of strings in a 3-bit GA solving the one-max problem in the presence of generalised of recombination and selection. The predicted trajectory along the Geiringer manifold for the strings is also shown.

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

Figure 9 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 more slowly than in the one-max problem, the similarity between the dynamics of the two is apparent.[12] 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.

11 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

[12] 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 $\prod_{i=1}^{\ell} c(h_i, t)$.

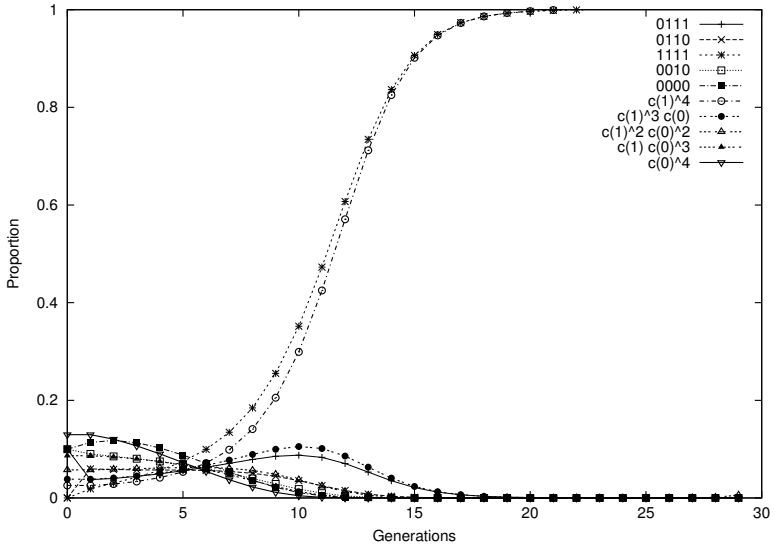

Figure 9: Infinite-population dynamics of a representative sample of strings in a 4-bit GA solving the needle-in-a-haystack problem in the presence of generalised of recombination and selection. The predicted trajectory along the Geiringer manifold for the strings is also shown.

steps per generation. Selection was implemented using stochastic universal sampling.

Figure 10 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.[13] 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 homogeneuos 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 10.

The situation is qualitatively similar in the case of populations of size $M = 10$. In this case, however, as shown in Figure 11 the movement towards homogeneous populations is slower since larger populations present less drift. This is confirmed in Figure 12 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 initialised uniformly at random (i.e., 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 7. Figures 13, 14 and 15 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,

[13] 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.

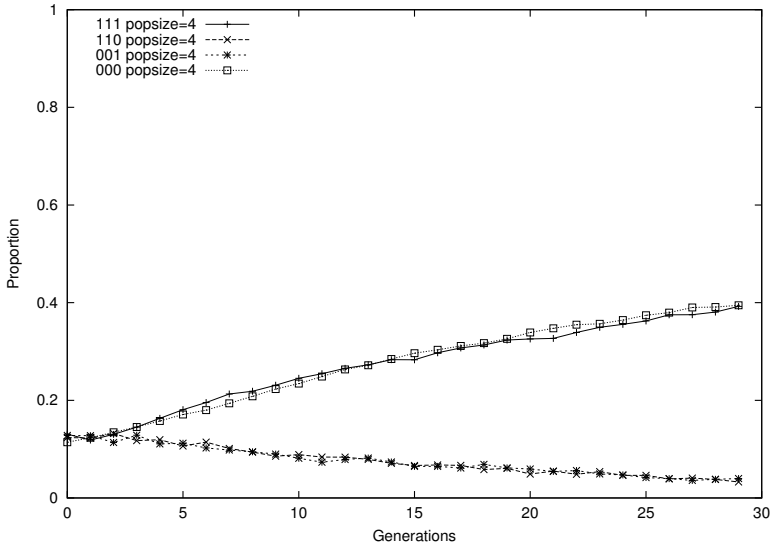

Figure 10: Finite-population dynamics of a population of $M = 4$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen uniformly at random.

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 15 and 7). 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 poportions 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 16, 17 and 18. In this case the similarity between the predictions of the infinite-population model (Figure 8) and actual runs is striking, the behaviour of actual runs having being captured by the infinite population model. Indeed, in the case of $M = 100$ the plots in Figure 18 are almost indistinguishable from the corresponding plots in Figure 8.

The results for the 4-bit needle-in-a-haystack problem are shown in Figures 19, 20 and 21. In this case the similarity between the predictions of the theory (Figure 9) 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 which is only twice the fitness of the "hay" (the rest of the search space).[14]

In all cases, for the first few generations, where the crossover bias is strongest, the fit between infinite population model and actual runs is very good.

12 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

[14]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).

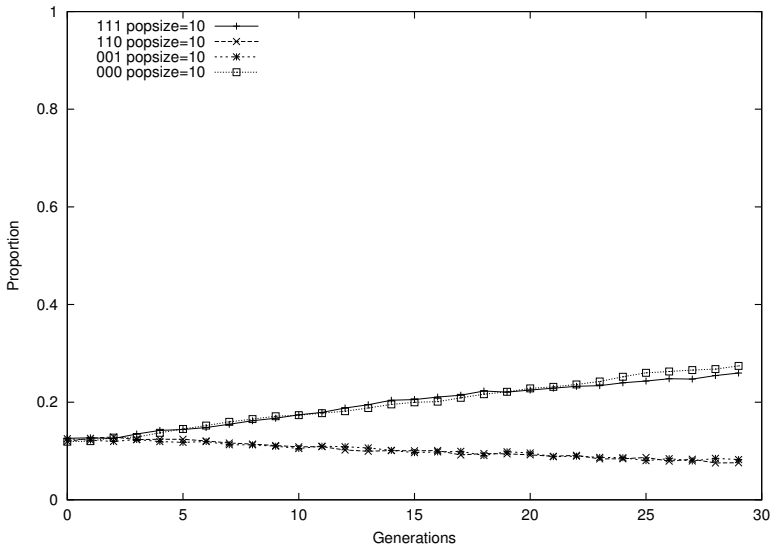

Figure 11: Finite-population dynamics of a population of $M = 10$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen uniformly at random.

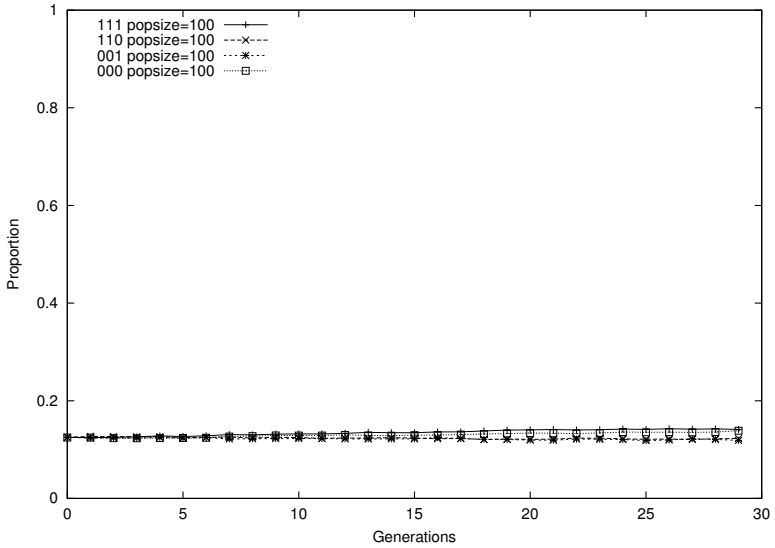

Figure 12: Finite-population dynamics of a population of $M = 100$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen uniformly at random.

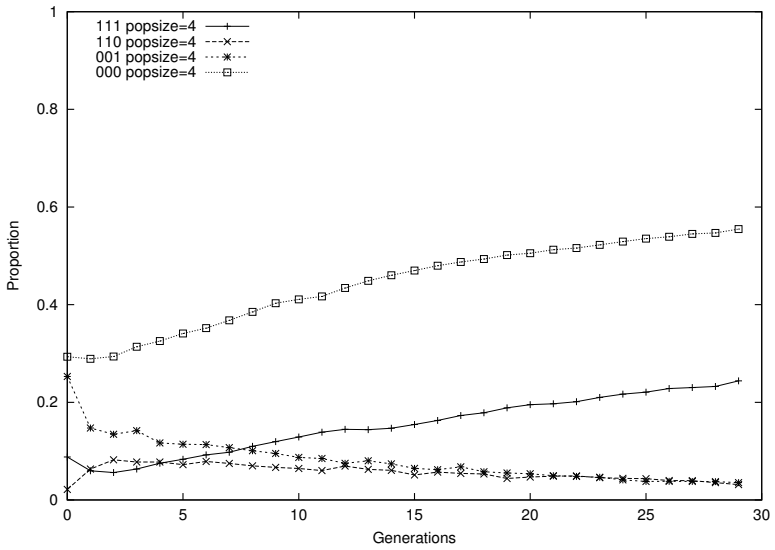

Figure 13: Finite-population dynamics of a population of $M = 4$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

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 2. 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.

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.

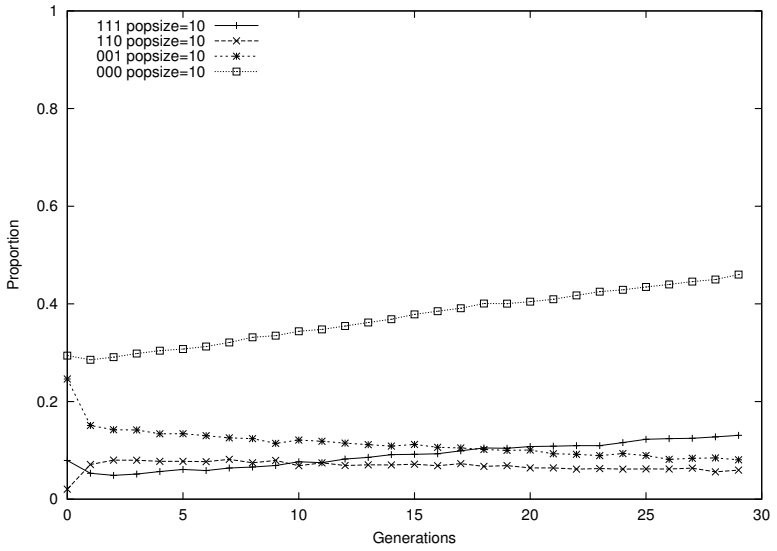

Figure 14: Finite-population dynamics of a population of $M = 10$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

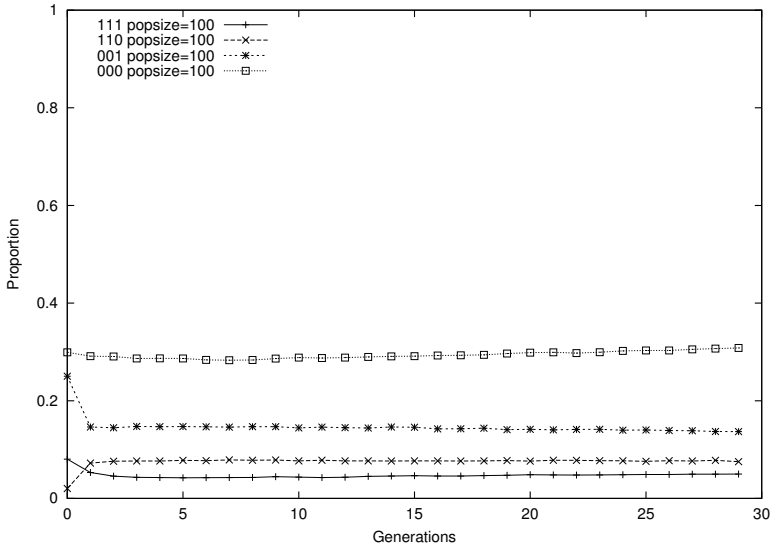

Figure 15: Finite-population dynamics of a population of $M = 100$ strings of length $\ell = 3$ evolving on a flat landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

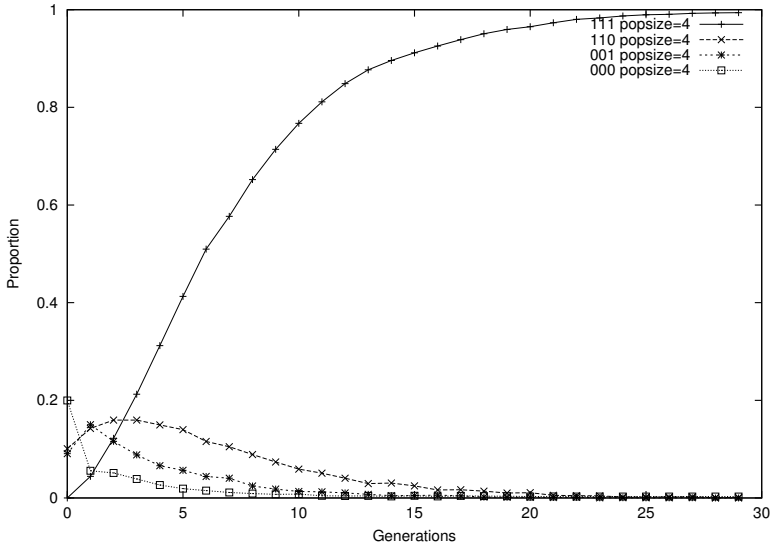

Figure 16: Finite-population dynamics of a population of $M = 4$ strings of length $\ell = 3$ evolving on the one-max landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

- 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 is 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 cliques: 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

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

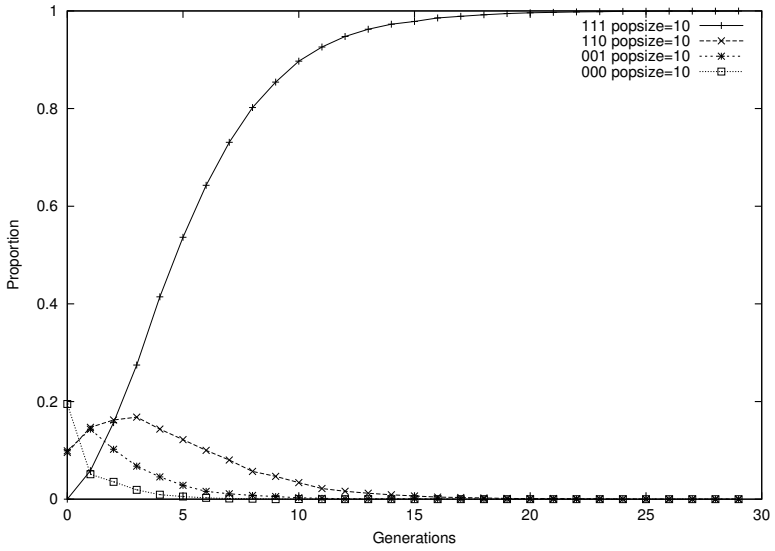

Figure 17: Finite-population dynamics of a population of $M = 10$ strings of length $\ell = 3$ evolving on the one-max landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

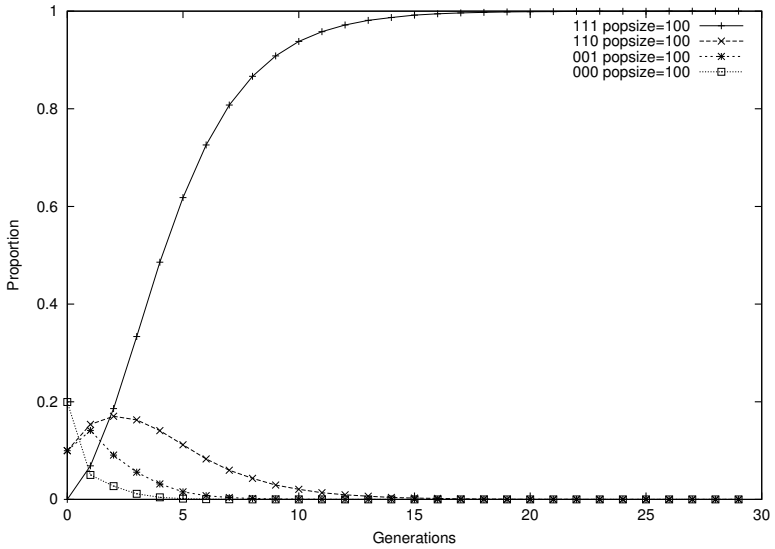

Figure 18: Finite-population dynamics of a population of $M = 100$ strings of length $\ell = 3$ evolving on the one-max landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

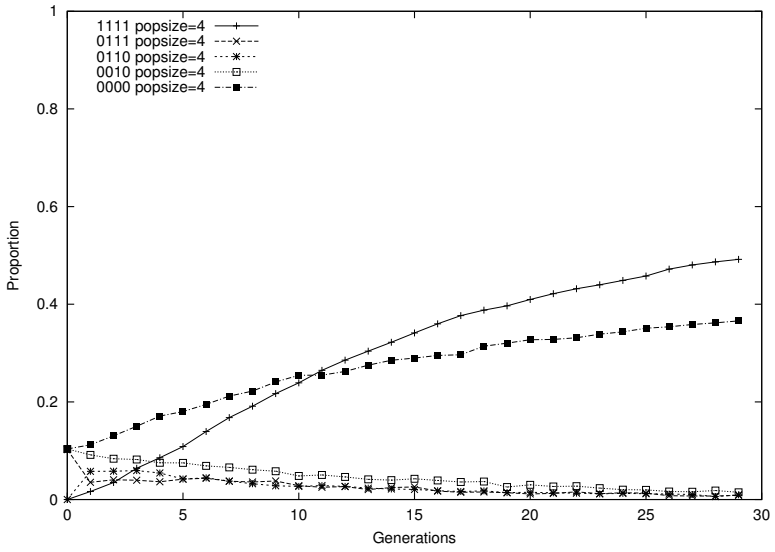

Figure 19: Finite-population dynamics of a population of $M = 4$ strings of length $\ell = 4$ evolving on a needle-in-a-haystack landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

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

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

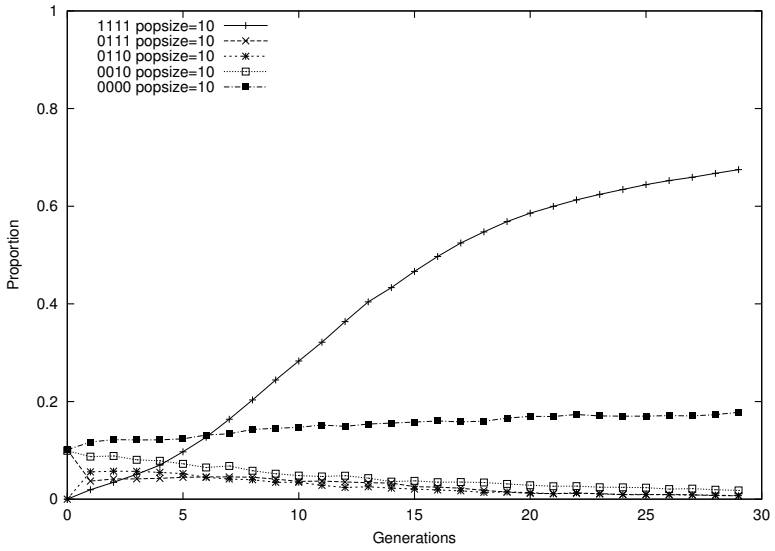

Figure 20: Finite-population dynamics of a population of $M = 10$ strings of length $\ell = 4$ evolving on a needle-in-a-haystack landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

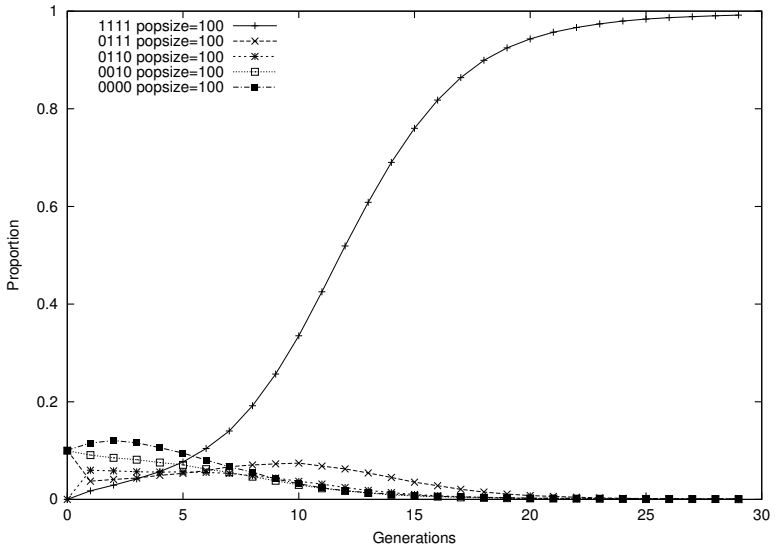

Figure 21: Finite-population dynamics of a population of $M = 100$ strings of length $\ell = 4$ evolving on a needle-in-a-haystack landscape (means of 1 000 runs). Strings in the initial population were chosen randomly but with unequal proportions (see text).

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

This way of proceeding may look rather artificial: typically in genetics the focus is on homozygote frequencies. In our model we use a finer level of detail. However, we can calculate homozygote frequencies, $\Phi(h_1 \cdots h_\ell, t)$, by using marginals of the distribution $\Phi(h'_1 \cdots h'_\ell h''_1 \cdots h''_\ell, t)$. 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 $a = a'_1 a'_2 a'_3 a'_4 a''_1 a''_2 a''_3 a''_4$ and $b = b'_1 b'_2 b'_3 b'_4 b''_1 b''_2 b''_3 b''_4$ the GCM $r = (11110000, (1, 2, 2, 3, 1, 2, 3, 4))$ would produce the offspring $a'_1 a'_2 a'_2 a'_3 b'_1 b'_2 b'_3 b'_4$ which would correspond to an 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.

13 Discussion and Conclusions

In this 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. This model includes as a special case previous models, such as the exact schema theory in (Stephens and Waelbroeck, 1999; Stephens, 2001).

As mentioned in the introduction, 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 summarize 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 $|\Omega|$ 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 explicit 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 it 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 future research 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 have already 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.[15] 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

[15] In this sense generalised recombination has some features of mutation offering search possibilities that are not available to homologous recombination.

(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 instructions for a register based CPU 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 for an operator (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 (see (McPhee and Poli, 2002) for an example).

Acknowledgements

CRS and RP thank Bill Langdon for helpful comments and EPSRC for financial support (grant number GR/T24616/01). CRS also thanks DGAPA of the UNAM for a Sabbatical Fellowship and Conacyt project 30422-E.

A Proofs

A.1 Theorem 5

Proof. Remebering that $\mathcal{R}_1(h) = \{r | D(h) \subseteq I_r\}$, for $r \in \mathcal{R}_1(h)$ we have that

$$\bigcap_{i \in I_r} H_{v_i}^{h_i} = \bigcap_{i \in D(h)} H_{v_i}^{h_i}$$

because $h_i = *$ for any $i \in I_r \setminus D(h)$. Also, recalling that $\mathcal{R}_3(h) = \mathcal{R}_\ell^\ell \setminus \mathcal{R}_1(h) \setminus \mathcal{R}_2(h)$ with $\mathcal{R}_2(h) = \{r | D(h) \subseteq I_{\bar{r}}\}$, for $r \in \mathcal{R}_3(h)$ we have that

$$\bigcap_{i \in D(h)} H_{v_i}^{h_i} \subset \bigcap_{i \in I_r} H_{v_i}^{h_i}$$

and

$$\bigcap_{i \in D(h)} H_{v_i}^{h_i} \subset \bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}$$

because the set $D(h)$ is always split between I_r and $I_{\bar{r}}$. So, from Equation 8 we obtain

$$\begin{aligned}
E[\Phi(h, t+1)] &= \sum_{r \in \mathcal{R}_1(h)} (p_c(r) + p_c(\bar{r})) p(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t) && (28) \\
&+ \sum_{r \in \mathcal{R}_3(h)} p_c(r) p(\bigcap_{i \in I_r} H_{v_i}^{h_i}, t) p(\bigcap_{i \in I_{\bar{r}}} H_{v_i}^{h_i}, t) \\
&\geq \sum_{r \in \mathcal{R}_1(h)} (p_c(r) + p_c(\bar{r})) p(\bigcap_{i \in D(h)} H_{v_i}^{h_i}, t) && (29) \\
&+ \sum_{r \in \mathcal{R}_3(h)} p_c(r) p(\bigcap_{i \in D(h)} H_{v_i}^{h_i}, t) p(\bigcap_{i \in D(h)} H_{v_i}^{h_i}, t)
\end{aligned}$$

By properly manipulating the recombination distribution, as indicated at the end of Section 3.1, we can then apply this result to the case in which crossover is invoke with probability p_{xo}. □

A.2 Corollary 1

Proof. For homologous crossover all elements of the vector v must satisfy $v_i = i$. So,

$$\bigcap_{i \in D(h)} H_{v_i}^{h_i} = \bigcap_{i \in D(h)} H_i^{h_i} = h$$

and Equation 11 turns into

$$\begin{aligned}
E[\Phi(h,t+1)] &\geq (1-p_{xo})p(h,t) \\
&+ p_{xo}p(h,t) \sum_{r \in \mathcal{R}_1(h)} (p_c(r) + p_c(\bar{r})) + p_{xo}p(h,t)^2 \sum_{r \in \mathcal{R}_3(h)} p_c(r) \\
&= p(h,t)\left[1 - p_{xo}\left(1 - \underbrace{\sum_{r \in \mathcal{R}_1(h)} (p_c(r) + p_c(\bar{r}))}_{=1-\sum_{r \in \mathcal{R}_3(h)} p_c(r)} - p(h,t) \sum_{r \in \mathcal{R}_3(h)} p_c(r)\right)\right] \\
&= p(h,t)\left[1 - p_{xo}\left(\sum_{r \in \mathcal{R}_3(h)} p_c(r)\right) \cdot (1 - p(h,t))\right]. \qquad (30)
\end{aligned}$$

□

A.3 Lemma 1

Proof. The explicit solution of Equation 19 is

$$x(t) = A^t x(0) + \sum_{\tau=0}^{t-1} A^{(t-1-\tau)} b(\tau) \qquad (31)$$

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 31. 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 31 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 19 exists, is unique, is non-negative and is given by

$$x^* = (I - A)^{-1}b^*. \tag{32}$$

□

A.4 Theorem 7

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

$$\Phi(h,t+1) = \sum_{r\in\mathcal{R}_\ell^\ell} p_c(m,v)\Phi\left(\bigotimes_{k=1}^{|I_r|}\left(*^{v_{i_k}-v_{i_{k-1}}-1}h_{i_k}\right)*^{\ell-v_{i_{|I_r|}}},t\right) \Phi\left(\bigotimes_{k=1}^{|I_{\bar{r}}|}\left(*^{v_{j_k}-v_{j_{k-1}}-1}h_{j_k}\right)*^{\ell-v_{j_{|I_{\bar{r}}|}}},t\right), \quad (33)$$

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 Section 3.3.

We can prove that Equation 22 is a fixed point for this equation by substituting the right-hand side of Equation 22 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 22.

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 clique 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 33.

So, from the substitution of the fixed point in Equation 33 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 22 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.

Clark, A. (1994). Invasion and maintenance of a gene duplication. *Proceedings of the National Academy of Sciences*, 91:2950–2954.

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. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, Massachusetts.

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

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

Koza, J. R. (1995). Gene duplication to enable genetic programming to concurrently evolve both the architecture and work-performing steps of a computer program. In *IJCAI-95 Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, volume 1, pages 734–740, Montreal, Quebec, Canada. Morgan Kaufmann.

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*, Torremolinos, (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, COEX, World Trade Center, 159 Samseong-dong, Gangnam-gu, Seoul, Korea. IEEE Press.

Nix, A. E. and Vose, M. D. (1992). Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence*, 5:79–88.

Poli, R. (2001). Exact schema theory for genetic programming and variable-length genetic algorithms with one-point crossover. *Genetic Programming and Evolvable Machines*, 2(2):123–163.

Poli, R. (2003). A simple but theoretically-motivated method to control bloat in genetic programming. In Ryan, C., Soule, T., Keijzer, M., Tsang, E., Poli, R., and Costa, E., editors, *Genetic Programming, Proceedings of the 6th European Conference, EuroGP 2003*, LNCS, pages 211–223, Essex, UK. Springer-Verlag.

Poli, R., Langdon, W. B., and O'Reilly, U.-M. (1998). Analysis of schema variance and short term extinction likelihoods. In Koza, J. R., Banzhaf, W., Chellapilla, K., Deb, K., Dorigo, M., Fogel, D. B., Garzon, M. H., Goldberg, D. E., Iba, H., and Riolo, R., editors, *Genetic Programming 1998: Proceedings of the Third Annual Conference*, pages 284–292, University of Wisconsin, Madison, Wisconsin, USA. Morgan Kaufmann.

Poli, R. and McPhee, N. F. (2003a). General schema theory for genetic programming with subtree-swapping crossover: Part I. *Evolutionary Computation*, 11(1):53–66.

Poli, R. and McPhee, N. F. (2003b). General schema theory for genetic programming with subtree-swapping crossover: Part II. *Evolutionary Computation*, 11(2):169–206.

Poli, R., McPhee, N. F., and Rowe, J. E. (2004). Exact schema theory and markov chain models for genetic programming and variable-length genetic algorithms with homologous crossover. *Genetic Programming and Evolvable Machines*, 5(1):31–70.

Poli, R., Rowe, J. E., and McPhee, N. F. (2001). Markov chain models for GP and variable-length GAs with homologous crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, San Francisco, California, USA. Morgan Kaufmann.

Poli, R. and Stephens, C. R. (2005a). Theoretical analysis of generalised recombination. Technical Report CSM-426, Department of Computer Science, University of Essex.

Poli, R. and Stephens, C. R. (2005b). Theoretical analysis of generalised recombination. In *CEC-2005*. Accepted.

Poli, R., Stephens, C. R., Wright, A. H., and Rowe, J. E. (2002). 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 Publishers.

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.

Rowe, J. E., Vose, M. D., and Wright, A. H. (2002). Group properties of crossover and mutation. *Evolutionary Computation*, 10(2):151–184.

Sawai, H. and Adachi, S. (2000). A comparative study of gene-duplicated GAs based on pfGA and SSGA. In Whitley, D., Goldberg, D., Cantu-Paz, E., Spector, L., Parmee, I., and Beyer, H.-G., editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2000)*, pages 74–81, Las Vegas, Nevada, USA. Morgan Kaufmann.

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, USA. Morgan Kaufmann.

Stephens, C. R. and Poli, R. (2004). EC theory – "in theory": Towards a unification of evolutionary computation theory. In Menon, A., editor, *Frontiers of Evolutionary Computation*, pages 129–156. Kluwer, Boston, MA.

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 *CEC-2005*. Accepted.

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.

Stephens, C. R. and Vargas, J. M. (2000). Effective fitness as an alternative paradigm for evolutionary computation I: General formalism. *Genetic Programming and Evolvable Machines*, 1(4):363–378.

Stephens, C. R. and Waelbroeck, H. (1999). Schemata evolution and building blocks. *Evolutionary Computation*, 7(2):109–124.

Stephens, C. R., Waelbroeck, H., and Aguirre, R. (1999). Schemata as building blocks: Does size matter? In Banzhaf, W. and Reeves, C., editors, *Foundations of Genetic Algorithms 5*, pages 117–133. Morgan Kaufmann, San Francisco, CA.

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

Whitley, D. (1992). An executable model of a simple genetic algorithm. In Whitley, D., editor, *Foundations of Genetic Algorithms Workshop (FOGA-92)*, Vail, Colorado.

Whitley, L. D. (1994). A Genetic Algorithm Tutorial. *Statistics and Computing*, 4:65–85.