The Equilibrium and Transient Behavior of Mutation and Recombination

William M. Spears
AI Center - Code 5515
Naval Research Laboratory
Washington, D.C. 20375
spears@aic.nrl.navy.mil

Abstract

This paper investigates the limiting distributions for mutation and recombination. The paper shows a tight link between standard schema theories of recombination and the speed at which recombination operators drive a population to equilibrium. A similar analysis is performed for mutation. Finally the paper characterizes how a population undergoing recombination and mutation evolves.

1 INTRODUCTION

In a previous paper Booker (1992) showed how the theory of "recombination distributions" can be used to analyze evolutionary algorithms (EAs). First, Booker re-examined Geiringer's Theorem (Geiringer 1944), which describes the equilibrium distribution of an arbitrary population that is undergoing recombination. Booker suggested that "the most important difference among recombination operators is the rate at which they converge to equilibrium". Second, Booker used recombination distributions to re-examine analyses of schema dynamics. In this paper we show that the two themes are tightly linked, in that traditional schema analyses such as schema disruption and construction (Spears 2000) yield important information concerning the speed at which recombination operators drive the population to equilibrium. Rather than focus solely on the dynamics near equilibrium, however, we also examine the transient behavior that occurs before equilibrium is reached.

This paper also investigates the equilibrium distribution of a population undergoing only mutation, and demonstrates precisely (with a closed-form solution) how the mutation rate μ affects the rate at which this distribution is reached. Again, we will focus both on the

transient and the equilibrium dynamics. Finally, this paper characterizes how a population of chromosomes evolves under recombination and mutation. We discuss mutation first.

2 THE LIMITING DISTRIBUTION FOR MUTATION

This section will investigate the limiting distribution of a population of chromosomes undergoing mutation, and will quantify how the mutation rate μ affects the rate at which the equilibrium is approached. Mutation will work on alphabets of cardinality C in the following fashion. An allele is picked for mutation with probability μ. Then that allele is changed to one of the other $C-1$ alleles, uniformly randomly.

Theorem 1 *Let S be any string of L alleles: $(a_1, ..., a_L)$. If a population is mutated repeatedly (without selection or recombination) then:*

$$\lim_{t\to\infty} p_S(t) \;=\; \prod_{i=1}^{L} \frac{1}{C}$$

where $p_S(t)$ is the expected proportion of string S in the population at time t and C is the cardinality of the alphabet.

Theorem 1 states that a population undergoing only mutation approaches a "uniform" equilibrium distribution in which all possible alleles are uniformly likely at all loci. Thus all strings will become equally likely in the limit. Clearly, since the mutation rate μ does not appear, it does not affect the equilibrium distribution that is reached. Also, the initial population will not affect the equilibrium distribution. However, both the mutation rate and the initial population may affect the *transient* behavior, namely the *rate* at which the distribution is approached. This will be explored further in the next two subsections.

2.1 A MARKOV CHAIN MODEL OF MUTATION

To explore the (non-)effect that the mutation rate and the initial population have on the equilibrium distribution, the dynamics of a finite population of strings being mutated will be modeled as follows. Consider a population of P individuals of length L, with cardinality C. Since Geiringer's Theorem for recombination (Geiringer 1944) (discussed in the next section) focuses on loci, the emphasis will be on the L loci. However, since each locus will be perturbed independently and identically by mutation, it is sufficient to consider only one locus. Furthermore, since each of the alleles in the alphabet are treated the same way by mutation, it is sufficient to focus on only one allele (all other alleles will behave identically).

Let the alphabet be denoted as $\mathcal{A}$ and $\alpha \in \mathcal{A}$ be one of the particular alleles. Let $\overline{\alpha}$ denote all the other alleles. Then define a state to be the number of α's at some locus and a time step to be one generation in which all individuals have been considered for mutation. More formally, let S_t be a random variable that gives the number of α's at some locus at time t. S_t can take on any of the $P+1$ integer values from 0 to P at any time step t. Since this process is memory-less, the transitions between states can be modeled with a Markov chain. The probability of transitioning from state i to state j in one time step will be denoted as $P(S_t = j \mid S_{t-1} = i) \equiv p_{i,j}$. Thus, transitioning from i to j means moving from a state with $S_{t-1} = i$ α's and $(P-i)$ $\overline{\alpha}$'s to a state with $S_t = j$ α's and $(P-j)$ $\overline{\alpha}$'s.

When $0.0 < \mu < 1.0$ all $p_{i,j}$ entries are non-zero and the Markov chain is ergodic. Thus there is a steady-state distribution describing the probability of being in each state after a long period of time. By the definition of steady-state distribution, it can not depend on the initial state of the system, hence the initial population will have no effect on the long-term behavior of the system. The steady-state distribution reached by this Markov chain model can be thought of as a sequence of P Bernoulli trials with success probability $1/C$. Thus the steady-state distribution can be described by the binomial distribution, giving the probability π_i of being in state i (i.e., the probability that i α's appear at a locus after a long period of time):

$$\lim_{t\to\infty} P(S_t = i) \;\equiv\; \pi_i \;=\; \binom{P}{i}\left(\frac{1}{C}\right)^i\left(1 - \frac{1}{C}\right)^{P-i}$$

Note that the steady-state distribution does not depend on the mutation rate μ or the initial population, although it does depend on the cardinality C. Now Theorem 1 states that the equilibrium distribution is one in which all possible alleles are equally likely. This can be proven by showing that the expected number of α's at any locus of the population (at steady state) is:

$$\lim_{t\to\infty} E[S_t] \;=\; \sum_{i=0}^{P}\binom{P}{i} i \left(\frac{1}{C}\right)^i\left(1 - \frac{1}{C}\right)^{P-i} \;=\; \frac{P}{C}$$

The Markov chain model will also yield the transient behavior of the system, if we fully specify the one-step probability transition values $p_{i,j}$. First, suppose $j \geq i$. This means we are increasing (or not changing) the number of α's. To accomplish the transition requires that $j - i$ more $\overline{\alpha}$'s are mutated to α's than α's are mutated to $\overline{\alpha}$'s. The transition probabilities are:

$$p_{i,j} = \sum_{x=0}^{min\{i,P-j\}} \binom{i}{x}\binom{P-i}{x+j-i}\mu^x\left(\frac{\mu}{C-1}\right)^{x+j-i}(1-\mu)^{i-x}\left(1-\frac{\mu}{C-1}\right)^{P-j-x}$$

Let x be the number of α's that are mutated to $\overline{\alpha}$'s. Since there are i α's in the current state, this means that $i - x$ α's are *not* mutated to $\overline{\alpha}$'s. This occurs with probability $\mu^x(1-\mu)^{i-x}$. Also, since x α's are mutated to $\overline{\alpha}$'s then $x + j - i$ $\overline{\alpha}$'s must be mutated to α's. Since there are $P - i$ $\overline{\alpha}$'s in the current state, this means that $P - i - x - j + i = P - x - j$ $\overline{\alpha}$'s are *not* mutated to α's. This occurs with probability $(\mu/(C-1))^{x+j-i}(1-\mu/(C-1))^{P-x-j}$. The combinatorials yield the number of ways to choose x α's out of the i α's, and the number of ways to choose $x + j - i$ $\overline{\alpha}$'s out of the $P - i$ $\overline{\alpha}$'s. Clearly, it isn't possible to mutate more than i α's. Thus $x \leq i$. Also, since it isn't possible to mutate more than $P - i$ $\overline{\alpha}$'s, $x + j - i \leq P - i$, which indicates that $x \leq P - j$. The minimum of i and $P - j$ bounds the summation correctly.

Similarly, if $i \geq j$, we are decreasing (or not changing) the number of α's. Thus one needs to mutate $i - j$ more α's to $\overline{\alpha}$'s than $\overline{\alpha}$'s to α's. The transition probabilities are:

$$p_{i,j} = \sum_{x=0}^{min\{P-i,j\}} \binom{i}{x+i-j}\binom{P-i}{x}\mu^{x+i-j}\left(\frac{\mu}{C-1}\right)^{x}(1-\mu)^{j-x}\left(1-\frac{\mu}{C-1}\right)^{P-i-x}$$

The explanation is almost identical to before. Let x be the number of $\overline{\alpha}$'s that are mutated to α's. Since there are $P-i$ $\overline{\alpha}$'s in the current state, this means that $P-i-x$ $\overline{\alpha}$'s are *not* mutated to α's. This occurs with probability $(\mu/(C-1))^x(1-\mu/(C-1))^{P-i-x}$. Also, since x $\overline{\alpha}$'s are mutated to α's then $x+i-j$ α's must be mutated to $\overline{\alpha}$'s. Since there are i α's in the current state, this means that $i-x-i+j=j-x$ α's are *not* mutated to $\overline{\alpha}$'s. This occurs with probability $\mu^{x+i-j}(1-\mu)^{j-x}$. The combinatorials yield the number of ways to choose x $\overline{\alpha}$'s out of the $P-i$ $\overline{\alpha}$'s, and the number of ways to choose $x+i-j$ α's out of the i α's. Clearly, it isn't possible to mutate more than $P-i$ $\overline{\alpha}$'s. Thus $x \leq P-i$. Also, since it isn't possible to mutate more than i α's, $x+i-j \leq i$, which indicates that $x \leq j$. The minimum of $P-i$ and j bounds the summation correctly.

In general, these equations are not symmetric ($p_{i,j} \neq p_{j,i}$), since there is a distinct tendency to move towards states with a $1/C$ mixture of α's (the limiting distribution). We will not make further use of these equations in this paper, but they are included for completeness.

2.2 THE RATE OF APPROACHING THE LIMITING DISTRIBUTION

The previous subsection showed that the mutation rate μ and the initial population have no effect on the limiting distribution that is reached by a population undergoing only mutation. However, these factors do influence the transient behavior, namely, the *rate* at which that limiting distribution is approached. This issue is investigated in this subsection. Rather than use the Markov chain model, however, an alternative approach will be taken.

In order to model the rate at which the process approaches the limiting distribution, consider an analogy with radioactive decay. In radioactive decay, nuclei disintegrate and thus change state. In the world of binary strings ($C=2$) this would be analogous to having a sea of 1's mutate to 0's, or with arbitrary C this would be analogous to having a sea of α's mutate to $\overline{\alpha}$'s. In radioactive decay, nuclei can not change state back from $\overline{\alpha}$'s to α's. However, for mutation, states can continually change from α to $\overline{\alpha}$ and vice versa. This can be modeled as follows. Let $p_\alpha(t)$ be the expected proportion of α's at time t. Then the expected time evolution of the system, which is a classic birth-death process (Feller 1968), can be described by a differential equation:[1]

$$\frac{dp_\alpha(t)}{dt} = -\mu\, p_\alpha(t) + \left(\frac{\mu}{C-1}\right)(1-p_\alpha(t)) = \left(\frac{\mu}{C-1}\right)(1-C\, p_\alpha(t))$$

The term $\mu\, p_\alpha(t)$ represents a loss (death), which occurs if α is mutated. The other term is a gain (birth), which occurs if an $\overline{\alpha}$ is successfully mutated to an α. At steady state the differential equation must be equal to 0, and this is satisfied by $p_\alpha(t)=1/C$, as would be expected.

The general solution to the differential equation was found to be:

$$p_\alpha(t) = \frac{1}{C} + \left(p_\alpha(0) - \frac{1}{C}\right) e^{\frac{-C\mu t}{C-1}}$$

[1]Since the system is discrete in time, *difference* equations would seem more appropriate (e.g., for $C=2$ see Equation (44) of Beyer (1998) with $p_\alpha(t)=P_M$ and $p_\alpha(t-1)=P_R$). However, in this case differential equations are easier to work with and are adequate approximations to the behavior explored in this paper.

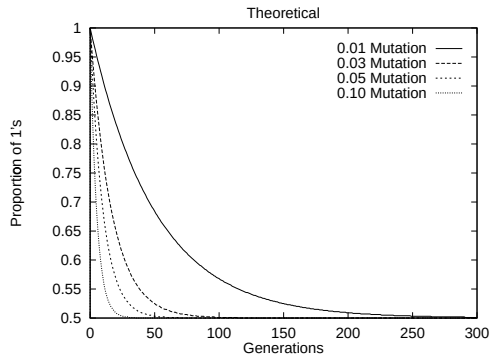

Figure 1: Decay rate for mutation when $C = 2$.

where $-C\mu/(C-1)$ plays a role analogous to the decay rate in radioactive decay. This solution indicates a number of important points. First, as expected, although μ does not change the limiting distribution, it does affect how fast it is approached. Also, the cardinality C also affects that rate (as well as the limiting distribution itself). Finally, different initial conditions will also affect the rate at which the limiting distribution is approached, but will not affect the limiting distribution itself. For example, if $p_\alpha(0) = 1/C$ then $p_\alpha(t) = 1/C$ for all t, as would be expected.

Assume that binary strings are being used ($C = 2$) and $\alpha = 1$. Also assume the population is initially seeded only with 1's. Then the solution to the differential equation is:

$$p_1(t) \;=\; \frac{e^{-2\mu t} \;+\; 1}{2} \tag{1}$$

which is very similar to the equation derived from physics for radioactive decay.

Figure 1 shows the decay curves derived via Equation 1 for different mutation rates. Although μ has no effect on the limiting distribution, increasing μ clearly increases the rate at which that distribution is approached. Although this result is quite intuitively obvious, the key point is that we can now make quantitative statements as to how the initial conditions and the mutation rate affect the speed of approaching equilibrium.

3 THE LIMITING DISTRIBUTION FOR RECOMBINATION

Geiringer's Theorem (Geiringer 1944) describes the equilibrium distribution of an arbitrary population that is undergoing recombination, but no selection or mutation. To understand Geiringer's Theorem, consider a population of ten strings of length four. In the initial population, five of the strings are "AAAA" while the other five are "BBBB". If these strings are recombined repeatedly, eventually 2^4 strings will become equally likely in the population. In equilibrium, the probability of a particular string will approach the product of the initial probabilities of the individual alleles – thus asserting a condition of independence between alleles. Geiringer's Theorem can be stated as follows:

Theorem 2 *Let S be any string of L alleles: $(a_1, ..., a_L)$. If a population is recombined*

repeatedly (without selection or mutation) then:

$$\lim_{t\to\infty} p_S(t) \;=\; \prod_{i=1}^{L} p_{a_i}(0)$$

where $p_S(t)$ is the expected proportion of string S in the population at time t and $p_{a_i}(0)$ is the proportion of allele a at locus (position) i in the initial population.

Thus, the probability of string S is simply the product of the proportions of the individual alleles in the initial ($t = 0$) population. The equilibrium distribution illustrated in Theorem 2 is referred to as "Robbins' equilibrium" (Robbins 1918). Theorem 2 holds for all standard recombination operators, such as n-point recombination and P_0 uniform recombination.[2] It also holds for arbitrary cardinality alphabets. The key point is that recombination operators do not change the distribution of alleles at any locus; they merely shuffle those alleles at each locus.

3.1 OVERVIEW OF MARGINAL RECOMBINATION DISTRIBUTIONS

According to Booker (1992) and Christiansen (1989), the population dynamics of a population undergoing recombination (but no selection or mutation) is governed by *marginal recombination distributions.* To briefly summarize, $\mathcal{R}_A(B)$ is "the marginal probability of the recombination event in which one parent transmits the loci $B \subseteq A$ and the other parent transmits the loci in $A \backslash B$" (Booker 1992). A and B are sets and $A \backslash B$ represents set difference. For example, suppose one parent is xyz and the other is XYZ. Since there are three loci, $A = \{1, 2, 3\}$. Let $B = \{1, 2\}$ and $A \backslash B = \{3\}$. This means that the two alleles xy are transmitted from the first parent, while the third allele Z is transmitted from the second parent, producing an offspring xyZ. The marginal distribution is defined by the probability terms $\mathcal{R}_A(B)$, $B \subseteq A$. Clearly $\sum_{B \subseteq A} \mathcal{R}_A(B) = 1$ and under Mendelian segregation, $\mathcal{R}_A(B) = \mathcal{R}_A(A \backslash B)$. In terms of the more traditional schema analysis, the set A designates the defining loci of a schema. Thus, the terms $\mathcal{R}_A(A) = \mathcal{R}_A(\emptyset)$ refer to the *survival* of the schema at the defining loci specified by A.

3.2 THE RATE AT WHICH ROBBINS' EQUILIBRIUM IS APPROACHED

As stated earlier, Booker (1992) has suggested that the rate at which the population approaches Robbins' equilibrium is the significant distinguishing characterization of different recombination operators. According to Booker, "a useful quantity for studying this property is the coefficient of *linkage disequilibrium*, which measures the deviation of current chromosome frequencies from their equilibrium levels". Such an analysis has been performed by Christiansen (1989), but given its roots in mathematical genetics the analysis is not explicitly tied to more conventional analyses in the EA community. The intuitive hypothesis is that those recombination operators that are more disruptive should drive the population to equilibrium more quickly (see Mühlenbein (1998) for empirical evidence to support this hypothesis). Christiansen (1989) provides theoretical support for this hypothesis by stating that the eigenvalues for convergence are given by the $\mathcal{R}_A(A)$ terms in the marginal distributions. The smaller $\mathcal{R}_A(A)$ is, the more quickly equilibrium is reached, *in the limit.* Since

[2] P_0 is the probability of swapping alleles. See Stephens et al. (1998) for a recent related proof of Geiringer's Theorem, stemming from exact evolution equations.

disruption is the opposite of survival, the direct implication is that equilibrium is reached more quickly when a recombination operator is more disruptive.

One very important caveat, however, is that this theoretical analysis holds only in the limit of large time, or when the population is near equilibrium. As GA practitioners we are far more interested in the short-term *transient* behavior of the population dynamics. Although equilibrium behavior can be studied by use of the marginal probabilities $\mathcal{R}_A(A)$, studying the transient behavior requires *all* of the marginals $\mathcal{R}_A(B), B \subseteq A$. The primary goal of this section is to tie the marginal probabilities to the more traditional schema analyses, in order to analyze the *complete* (transient and equilibrium) behavior of a population undergoing recombination. The focus will be on recombination operators that are commonly used in the GA community: n-point recombination and P_0 uniform recombination. Several related questions will be addressed. For example, lowering P_0 from 0.5 makes P_0 uniform recombination less disruptive ($\mathcal{R}_A(A)$ increases). How do the remainder of the marginals change? Can we compare n-point recombination and P_0 uniform recombination in terms of the population dynamics? Finally, what can we say about the transient dynamics? Although these questions can often only be answered in restricted situations the picture that emerges is that traditional schema analyses such as schema disruption *and* construction (Spears and De Jong 1998) do in fact yield important information concerning the dynamics of a population undergoing recombination.

3.3 THE FRAMEWORK

The framework used in this section consists of a set of differential equations that describe the *expected* time evolution of the strings in a population of finite size (equivalently this can be considered to be the evolution of an infinite-size population). The treatment will hold for hyperplanes (schemata) as well, so the term "hyperplane" and "string" can be used interchangeably.

Consider having a population of strings. Each generation, pairs of strings (parents) are repeatedly chosen uniformly randomly for recombination, producing offspring for the next generation. Let S_h, S_i, and S_j be strings of length L (alternatively, they can be considered to be hyperplanes of order L). Let $p_{S_i}(t)$ be the proportion of string S_i at time t. The time evolution of S_i will again involve terms of loss (death) and gain (birth). A loss will occur if parent S_i is recombined with another parent such that neither offspring is S_i. A gain will occur if two parents that are not S_i are recombined to produce S_i. Thus the following differential equation can be written for each string S_i:

$$\frac{dp_{S_i}(t)}{dt} = -loss_{S_i}(t) + gain_{S_i}(t)$$

The losses can occur if S_i is recombined with another string S_j such that S_i and S_j differ by $\Delta(S_i, S_j) \equiv k$ alleles, where k ranges from two to L. For example the string "AAAA" can (potentially) be lost if recombined with "AABB" (where $k = 2$). If S_i and S_j differ by one or zero alleles, there will be no change in the proportion of string S_i. In general, the expected loss for string S_i at time t is:

$$loss_{S_i}(t) = \sum_{S_j} p_{S_i}(t)\, p_{S_j}(t) P_{\mathrm{d}}(H_k) \quad where\ 2 \leq \Delta(S_i, S_j) \equiv k \leq L \tag{2}$$

The product $p_{S_i}(t)\ p_{S_j}(t)$ is the probability that S_i will be recombined with S_j, and $P_{\mathrm{d}}(H_k)$ is the probability that neither offspring will be S_i. Equivalently, $P_{\mathrm{d}}(H_k)$ refers to the probability of disrupting the kth-order hyperplane H_k defined by the k different alleles. This is identical to the probability of disruption as defined by De Jong and Spears (1992).

Gains can occur if two strings S_h and S_j of length L can be recombined to construct S_i. It is assumed that neither S_h or S_j is the same as S_i at all defining positions (because then there would be no gain) and that either S_h or S_j has the correct allele for S_i at every locus. Suppose that S_h and S_j differ at $\Delta(S_h, S_j) \equiv k$ alleles. Once again k must range from two to L. For example, the string "AAAA" can (potentially) be constructed from the two strings "AABB" and "ABAA" (where $k = 3$). If S_h and S_j differ by one or zero alleles, then either S_h or S_j is equivalent to S_i and there is no true construction (or gain).

Of the k differing alleles, m are at string S_h and $n = k - m$ are at string S_j. Thus what is happening is that two non-overlapping, lower-order building blocks H_m and H_n are being constructed to form H_k (and thus the string S_i). In general, the expected gain for string S_i at time t is:

$$gain_{S_i}(t) \;=\; \sum_{S_h, S_j} p_{S_h}(t)\ p_{S_j}(t)\ P_{\mathrm{c}}(H_k \mid H_m \wedge H_n) \quad where\ 2 \leq \Delta(S_h, S_j) \equiv k \leq L \quad (3)$$

The product $p_{S_h}(t)\ p_{S_j}(t)$ is the probability that S_h will be recombined with S_j, and $P_{\mathrm{c}}(H_k \mid H_m \wedge H_n)$ is the probability that an offspring will be S_i. Equivalently, $P_{\mathrm{c}}(H_k \mid H_m \wedge H_n)$ is the probability of constructing the kth-order hyperplane H_k (and hence string S_i) from the two strings S_h and S_j that contain the non-overlapping, lower-order building blocks H_m and H_n. This is identical to the probability of construction as defined by Spears and De Jong (1998).

If the cardinality of the alphabet is C then there are C^L different strings. This results in a system of C^L simultaneous first-order differential equations. What is important to note is the explicit connection between Equations 2 – 3 and the more traditional schema theory for recombination, as exemplified by the probability of disruption $P_{\mathrm{d}}(H_k)$ and the probability of construction $P_{\mathrm{c}}(H_k \mid H_m \wedge H_n)$.

3.4 TRADITIONAL SCHEMA THEORY AND THE MARGINALS

We are now in a position to also explain the link between the traditional schema theory and the marginal probabilities. Consider the prior example, where the recombination of the parents xyz and XYZ produced an offspring xyZ. The other offspring produced is XYz. This occurs if $B = \{1, 2\}$ and $A \backslash B = \{3\}$ or in the complementary situation where $B = \{3\}$ and $A \backslash B = \{1, 2\}$. Hence, as pointed out earlier, $\mathcal{R}_A(B) = \mathcal{R}_A(A \backslash B)$. However, this is identical to the situation in which the third-order hyperplane $H_3 = xyZ$ is constructed from the first-order hyperplane $H_1 = \#\#Z$ and the second-order hyperplane $H_2 = xy\#$.[3] If we use $|\cdot|$ to denote the cardinality of a set, we can explicitly tie the probability of construction with the marginal probabilities:

$$P_{\mathrm{c}}(H_{|A|} \mid H_{|B|} \wedge H_{|A \backslash B|}) \;=\; \mathcal{R}_A(B) \;+\; \mathcal{R}_A(A \backslash B) \;=\; 2\mathcal{R}_A(B) \quad (4)$$

[3]This assumes that the two lower-order hyperplanes differ at all 3 defining positions. In general, they must differ at all k. Using the notation in Spears (2000), $P_{\mathrm{eq}} = 0$.

where a hyperplane of order $k = |A|$ is being constructed from lower-order hyperplanes of order $m = |B|$ and order $n = |A \backslash B|$. Interestingly, Equation 4 allows us to correct an error in Booker (1992), where he computes the marginal distribution for P_0 uniform recombination. The marginal distribution should be:

$$\mathcal{R}_A(B) \;=\; [\; {P_0}^{|B|}(1-P_0)^{|A\backslash B|} \;+\; {P_0}^{|A\backslash B|}(1-P_0)^{|B|} \;] \;/\; 2$$

Survival is a special form of construction in which one of the lower-order hyperplanes has zero order. Since disruption is the opposite of survival we can write:

$$P_{\mathrm{d}}(H_{|A|}) \;\equiv\; 1 \;-\; P_{\mathrm{c}}(H_{|A|} \mid H_{|\emptyset|} \wedge H_{|A\backslash\emptyset|}) \;=\; 1 \;-\; 2\mathcal{R}_A(A)$$

As mentioned earlier, according to Christiansen (1989), the smaller $\mathcal{R}_A(A)$ is, the more quickly equilibrium is reached *in the limit*. We can now connect this result to the concepts of disruption and construction in the framework given by Equations 2 – 3. If some hyperplane is above the equilibrium proportion then the loss terms will be more important, as they drive the hyperplane down to equilibrium. A decrease in $\mathcal{R}_A(A)$ indicates that a recombination operator is more disruptive. This increases the loss terms and drives the hyperplane down towards equilibrium more quickly. Likewise, if some hyperplane is below the equilibrium proportion then the gain terms will be more important, as they drive the hyperplane up towards equilibrium. Since marginals must sum to one, if $\mathcal{R}_A(A)$ decreases some (or all) of the other marginals will increase to compensate. Thus, *on the average* a more disruptive recombination operator will increase the $P_{\mathrm{c}}(H_k \mid H_m \wedge H_n)$ terms and hence drive that hyperplane to equilibrium more quickly.

However, as stated before, the caveat lies in the phrase "in the limit". Although a reduction in $\mathcal{R}_A(A)$ increases the other marginals on the average, there are situations where some marginals increase while others decrease. This can cause quite interesting transient behavior. The major contributor to this phenomena appears to be the order of the hyperplane, which we investigate in the following subsections.

3.5 SECOND-ORDER HYPERPLANES

We first consider the case of second-order hyperplanes, which are easiest to analyze. Consider the situation where the cardinality of the alphabet $C = 2$. In this situation there are four hyperplanes of interest: (#0#0#, #0#1#, #1#0#, #1#1#).[4] Then the four differential equations describing the expected time evolution of these hyperplanes are:

$$\begin{aligned}
\frac{dp_{00}(t)}{dt} &= -p_{00}(t)\; p_{11}(t)\; P_{\mathrm{d}}(H_2) \;+\; p_{01}(t)\; p_{10}(t)\; P_{\mathrm{c}}(H_2 \mid H_1 \wedge H_1) \\
\frac{dp_{01}(t)}{dt} &= -p_{01}(t)\; p_{10}(t)\; P_{\mathrm{d}}(H_2) \;+\; p_{00}(t)\; p_{11}(t)\; P_{\mathrm{c}}(H_2 \mid H_1 \wedge H_1) \\
\frac{dp_{10}(t)}{dt} &= -p_{01}(t)\; p_{10}(t)\; P_{\mathrm{d}}(H_2) \;+\; p_{00}(t)\; p_{11}(t)\; P_{\mathrm{c}}(H_2 \mid H_1 \wedge H_1) \\
\frac{dp_{11}(t)}{dt} &= -p_{00}(t)\; p_{11}(t)\; P_{\mathrm{d}}(H_2) \;+\; p_{01}(t)\; p_{10}(t)\; P_{\mathrm{c}}(H_2 \mid H_1 \wedge H_1)
\end{aligned}$$

[4] These four hyperplanes have been chosen arbitrarily for illustrative purposes. Also, we assume a binary-string representation, although that isn't necessary.

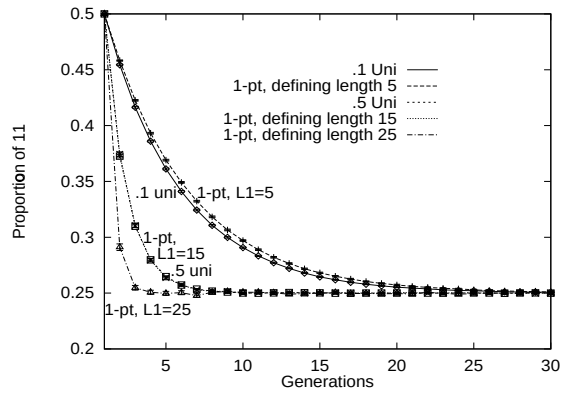

Figure 2: The rate of approaching Robbins' equilibrium for $H_2 = \#1\#1\#$, when $L = 30$.

Thus for this special case the loss and gain terms are controlled fully by one computation of disruption and one computation of construction. If two recombination operators have precisely the same disruption and construction behavior on second-order hyperplanes, the system of differential equations will be the same, and the time evolution of the system will be the same. This is true regardless of the initial conditions of the system.

For example, consider one-point recombination and P_0 uniform recombination. Suppose the defining length of the second-order hyperplane is L_1. Then, $P_{\mathrm{d}}(H_2) = L_1/L$ for one-point recombination, and $P_{\mathrm{d}}(H_2) = 2P_0(1-P_0)$ for uniform recombination. The computations for $P_{\mathrm{c}}(H_2 \mid H_1 \wedge H_1)$ equal $P_{\mathrm{d}}(H_2)$ for one-point and uniform recombination. Thus, one-point recombination should act the same as uniform recombination when the defining length $L_1 = 2LP_0(1-P_0)$.

To illustrate this, an experiment was performed in which a population of binary strings was initialized so that 50% of the strings were all 1's, while 50% were all 0's. The strings were of length $L = 30$ and were repeatedly recombined, generation by generation, while the percentage of the second-order hyperplane $\#1\#1\#$ was monitored. When Robbins' equilibrium is reached the percentage of any of the four hyperplanes should be 25%. The experiment was run with 0.1 and 0.5 uniform recombination. Under those settings of P_0, the theory indicates that one-point recombination should perform identically when the second-order hyperplanes have defining length 5.4 and 15, respectively. Since an actual defining length must be an integer, the hyperplanes of defining length 5 and 15 were monitored.

Figure 2 graphs the results.[5] As expected, the results show a perfect match when comparing the evolution of H_2 under 0.5 uniform recombination and one-point recombination when $L_1 = 15$ (the two curves coincide almost exactly on the graph). The agreement is almost perfect when comparing 0.1 uniform recombination and one-point recombination when $L_1 = 5$, and the small amount of error is due to the fact that the defining length had to be rounded to an integer. As an added comparison, the second-order hyperplanes of defining length 25 were also monitored. In this situation one-point recombination should drive these

[5] These results (and all subsequent results) are averaged over 1000 independent runs, with the same starting conditions. The population size was 50 and selection was uniform. 95% confidence intervals are also shown.

hyperplanes to equilibrium even faster than 0.5 uniform recombination (because one-point recombination is more disruptive in this situation). The graph confirms this observation.

It is important to note that the above analysis holds even for arbitrary cardinality alphabets C, although it was demonstrated for $C = 2$. The system of differential equations would have more equations and terms as C increases, but the computations would still only involve one computation of $P_{\mathrm{d}}(H_2)$ and $P_{\mathrm{c}}(H_2 \mid H_1 \wedge H_1)$, and those computations would be precisely the same. To see this, consider having $C = 3$, with an alphabet of $\{0, 1, 2\}$. Then #0#0# can be disrupted if recombined with #1#1#, #1#2#, #2#1#, or #2#2#. The probability of disruption is the same as it was above. Similarly it can be shown that the probability of construction is the same as it was above.

3.5.1 The Rate of Decay for Second-Order Hyperplanes

It is interesting to note that, as with mutation, the decay of the curves in Figure 2 appears to be exponential in nature. This turns out to be true. To prove this, let us reconsider the differential equation describing the change in the expected proportion of the second-order hyperplane #1#1# at time t:

$$\frac{dp_{11}(t)}{dt} = -p_{00}(t)\, p_{11}(t)\, P_{\mathrm{d}}(H_2) + p_{01}(t)\, p_{10}(t)\, P_{\mathrm{c}}(H_2 \mid H_1 \wedge H_1)$$

In this case we know that $P_{\mathrm{d}}(H_2) = P_{\mathrm{c}}(H_2 \mid H_1 \wedge H_1)$, so the equation can be simplified to:

$$\frac{dp_{11}(t)}{dt} = P_{\mathrm{d}}(H_2)[\, p_{01}(t)\, p_{10}(t) - p_{00}(t)\, p_{11}(t)\,] \tag{5}$$

For the experiment leading to Figure 2 we also know that $p_{01}(t) = p_{10}(t)$ and $p_{00}(t) = p_{11}(t)$:

$$\frac{dp_{11}(t)}{dt} = P_{\mathrm{d}}(H_2)[\, p_{01}(t)\, p_{01}(t) - p_{11}(t)\, p_{11}(t)\,]$$

However, since $p_{00}(t) + p_{01}(t) + p_{10}(t) + p_{11}(t) = 1$ it is easy to show that $p_{01}(t) = \frac{1}{2} - p_{11}(t)$. With some simplification this leads to:

$$\frac{dp_{11}(t)}{dt} = P_{\mathrm{d}}(H_2)[\, \frac{1}{4} - p_{11}(t)\,]$$

Given the initial condition that $p_{11}(0) = \frac{1}{2}$ the solution to the differential equation is:

$$p_{11}(t) = p_{00}(t) = \frac{1 + e^{-P_{\mathrm{d}}(H_2)t}}{4}$$

Similarly, it is easy to show that the proportions of the #0#1# and #1#0# hyperplanes grow as follows (given the initial conditions that $p_{01}(0) = p_{10}(0) = 0$):

$$p_{01}(t) = p_{10}(t) = \frac{1 - e^{-P_{\mathrm{d}}(H_2)t}}{4}$$

3.5.2 The Rate of Approaching Equilibrium

It is also possible to derive a more general result concerning the rate at which equilibrium is approached. To see this, consider Equation 5 again:

$$\frac{dp_{11}(t)}{dt} = P_{\mathrm{d}}(H_2)[\, p_{01}(t)\; p_{10}(t) \; - \; p_{00}(t)\; p_{11}(t)\,]$$

Let $\delta(t) = p_{01}(t)\; p_{10}(t) - p_{00}(t)\; p_{11}(t)$. Then it is very clear that:

$$\frac{dp_{11}(t)}{dt} = \frac{dp_{00}(t)}{dt} = -\frac{dp_{01}(t)}{dt} = -\frac{dp_{10}(t)}{dt} = \delta(t)P_{\mathrm{d}}(H_2)$$

Since $\delta(t)$ goes to zero as the proportions of all the second-order hyperplanes approach equilibrium, we can consider $\delta(t)$ to be a measure of *linkage disequilibrium* for second-order hyperplanes. We can now write:

$$\frac{d\delta(t)}{dt} = p_{01}(t)\frac{dp_{10}(t)}{dt} + p_{10}(t)\frac{dp_{01}(t)}{dt} - p_{00}(t)\frac{dp_{11}(t)}{dt} - p_{11}(t)\frac{dp_{00}(t)}{dt}$$

This simplifies to:

$$\frac{d\delta(t)}{dt} = -\frac{dp_{11}(t)}{dt} = -\delta(t)P_{\mathrm{d}}(H_2)$$

The solution to this is simply:

$$\delta(t) = -\delta(0)e^{-P_{\mathrm{d}}(H_2)t} \tag{6}$$

where $\delta(0) = p_{01}(0)\; p_{10}(0) - p_{00}(0)\; p_{11}(0)$.

3.5.3 Summary of Second-Order Hyperplane Results

These results explicitly link traditional schema theory with the rate at which Robbins' equilibrium is approached, for second-order hyperplanes. We have shown that for second-order hyperplanes, the time evolution of two different recombination operators can be compared simply by comparing their disruptive and constructive behavior. Two recombination operators will drive a second-order hyperplane to Robbins' equilibrium at the same rate if their disruptive and constructive behavior are the same. It was also shown that $\delta(t)$, a measure of linkage disequilibrium for second-order hyperplanes, exponentially decays towards zero with the probability of disruption $P_{\mathrm{d}}(H_2)$ being the rate of decay.

Unfortunately, similar results are harder to determine for higher-order hyperplanes, although some interesting results can be shown for P_0 uniform recombination, as shown in the following subsections.

3.6 UNIFORM RECOMBINATION AND LOW-ORDER HYPERPLANES

For P_0 uniform recombination the loss and gain terms of Equations 2 – 3 are especially easy to compute. As stated earlier, losses can occur if an Lth-order hyperplane S_i is recombined

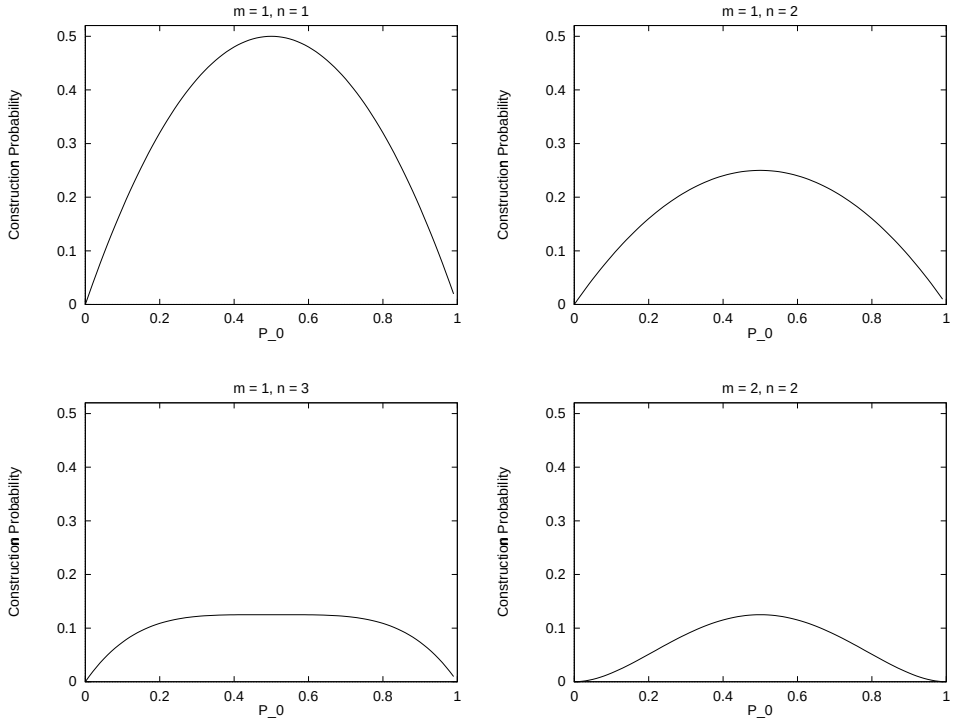

Figure 3: The probability of construction $P_c(H_k \mid H_m \wedge H_n)$ for P_0 uniform recombination on second-order ($m = 1$, $n = 1$), third-order ($m = 1$, $n = 2$), and fourth-order hyperplanes ($m = 1$, $n = 3$ and $m = 2$, $n = 2$). The probability of construction monotonically decreases as P_0 is decreased/increased from 0.5.

with an Lth-order hyperplane S_j such that S_i and S_j differ by k alleles, where k ranges from two to L. But, according to De Jong and Spears (1992) this occurs with probability:

$$P_d(H_k) \;=\; 1 \;-\; {P_0}^k \;-\; (1 - P_0)^k \qquad 2 \leq k \leq L$$

It can be shown that this is a unimodal function with a maximum at 0.5. Thus, the key point is that when the time evolution of the population undergoing recombination is expressed with C^L differential equations, the effect of increasing or decreasing P_0 from 0.5 reduces *all* of the loss terms in the differential equations, regardless of the order of the hyperplane. This slows the rate at which the equilibrium is approached.

Gains will occur if two hyperplanes S_h and S_j of order L can be recombined to construct S_i. Again, suppose that S_h and S_j differ at k alleles, where k ranges from two to L. Of the k differing alleles, m are at hyperplane S_h and $n = k - m$ are at hyperplane S_j. Then the probability of construction is:

$$P_c(H_k \mid H_m \wedge H_n) \;=\; {P_0}^m (1 - P_0)^n \;+\; {P_0}^n (1 - P_0)^m \qquad 2 \leq k \leq L,\; 0 < m < k$$

This has zero slope at $P_0 = 0.5$. The question now is under what conditions of n and m

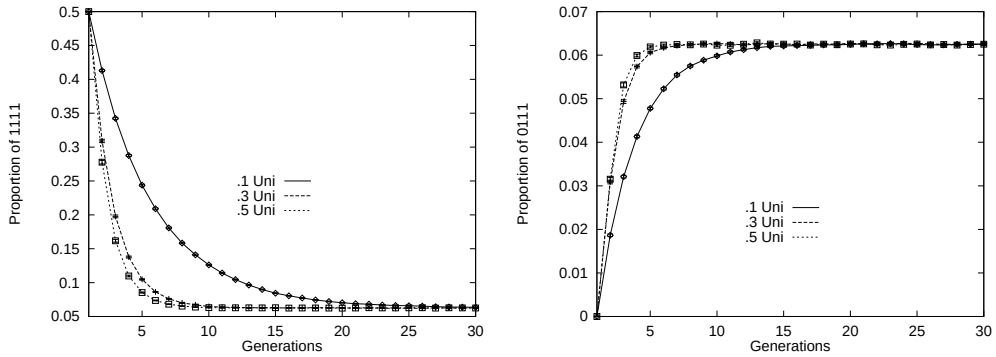

Figure 4: The rate of approaching Robbins' equilibrium for the fourth-order hyperplanes $H_4 = \#1\#1\#1\#1\#$ (left) and $H_4 = \#0\#1\#1\#1\#$ (right).

will $P_0 = 0.5$ represent a global (and the only) maximum. It is easy to show by counter-example that $P_0 = 0.5$ is not a global maximum for arbitrary m and n (e.g., $m = 1$ and $n = 4$). However, there are various cases where $P_0 = 0.5$ is a global maximum – namely when $m = 1$ and $n = 1$, $m = 1$ and $n = 2$, $m = 1$ and $n = 3$, and when $m = 2$ and $n = 2$ (and the symmetric cases where m and n are interchanged). See Figure 3 for graphs of the probability of construction, as P_0 changes.

Since we are interested in kth-order hyperplanes (where $k = m + n$), we have shown that for low-order hyperplanes ($k < 5$), construction is at a maximum when $P_0 = 0.5$ and construction decreases as P_0 decreases or increases from 0.5.

Thus, consider the time evolution of the hyperplanes in a population that are undergoing recombination, as modeled with the above differential equations. What we have shown is that if the hyperplanes have low order ($k < 5$), the effect of increasing or decreasing P_0 from 0.5 reduces *all* of the gain terms in the differential equations. Put in terms of the marginal probabilities we have shown that, for P_0 uniform recombination on low-order hyperplanes ($k < 5$), increasing $\mathcal{R}_A(A)$ (moving P_0 from 0.5) will decrease *all* of the other marginals $\mathcal{R}_A(B)$, $\emptyset \subset B \subset A$. Given this, we can expect that reducing or increasing P_0 from 0.5 should monotonically decrease the rate at which the equilibrium is approached, even during the transient behavior of the system.

To illustrate this, an experiment was performed in which a population of binary strings was initialized so that 50% of the strings were all 1's, while 50% were all 0's. The strings were of length $L = 30$ and were repeatedly recombined, generation by generation, while the percentages of the fourth-order hyperplanes #1#1#1#1# and #0#1#1#1# were monitored. When Robbins' equilibrium is reached the percentage of any of the fourth-order hyperplanes should be 6.25%. The experiment was run with uniform recombination, with P_0 ranging from 0.1 to 0.5 (higher values were ignored due to symmetry).

Figure 4 graphs the results. One can see that as P_0 increases to 0.5, the rate at which Robbins' equilibrium is approached also increases, as expected. This holds even throughout the transient dynamics of the system.

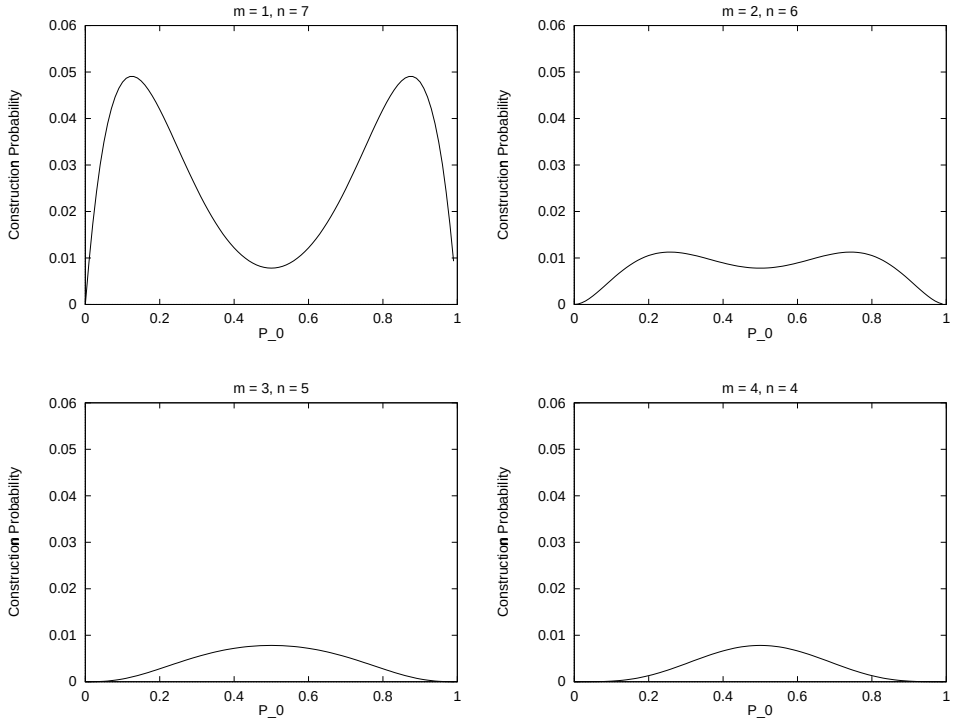

Figure 5: The probability of construction $P_c(H_k \mid H_m \wedge H_n)$ for P_0 uniform recombination, on eighth-order hyperplanes. The probability of construction does not always monotonically decrease as P_0 is decreased/increased from 0.5.

3.7 UNIFORM RECOMBINATION AND HIGH-ORDER HYPERPLANES

It is natural to wonder how this extends to higher-order hyperplanes. Unfortunately, as pointed out above, there will be situations (of m and n) where $P_0 = 0.5$ does not represent a global maximum for construction. However, it is easy to prove that when $m = n$ once again construction decreases as P_0 decreases or increases from 0.5. It appears as if this also holds for those situations where m and n are roughly equal (i.e., both are roughly $k/2$), but eventually fails when m and n are sufficiently different (either m or n is close to 1). Figure 5 illustrates this for eighth-order hyperplanes. Although construction is maximized at $P_0 = 0.5$ when $m = 4$ and $n = 4$, this certainly isn't true when $m = 1$ and $n = 7$. In fact, in that case construction is maximized when P_0 is roughly 0.13.

Put in terms of the marginal probabilities we have shown that, for P_0 uniform recombination on higher-order hyperplanes ($k > 4$), increasing $\mathcal{R}_A(A)$ (moving P_0 from 0.5) will decrease *some* (but not all) of the other marginals $\mathcal{R}_A(B)$, $\emptyset \subset B \subset A$. Given this, we can expect that reducing or increasing P_0 from 0.5 should not necessarily monotonically decrease the rate at which the equilibrium is approached, during the transient behavior of the system.

To illustrate this, an experiment was performed in which a population of binary strings was initialized so that 50% of the strings were all 1's, while 50% were all 0's. The

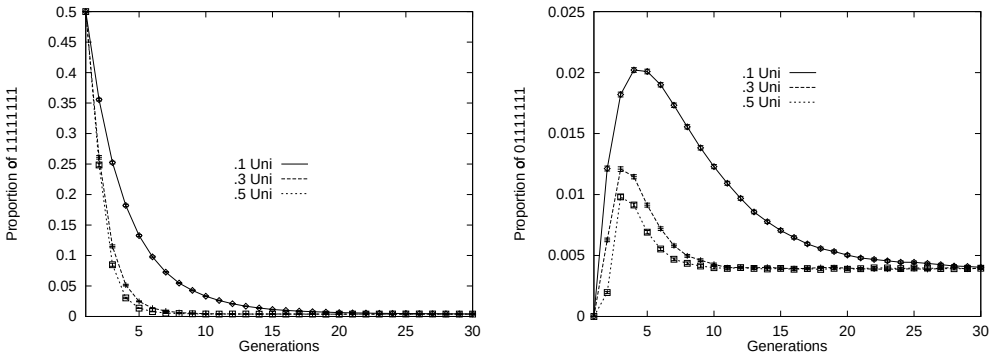

Figure 6: The rate of approaching Robbins' equilibrium for the eighth-order hyperplanes $H_8 = \#1\#1\#1\#1\#1\#1\#1\#$ (left) and $H_8 = \#0\#1\#1\#1\#1\#1\#1\#$ (right).

strings were of length $L = 30$ and were repeatedly recombined, generation by generation, while the percentages of the eighth-order hyperplanes $\#1\#1\#1\#1\#1\#1\#1\#$ and $\#0\#1\#1\#1\#1\#1\#1\#$ were monitored. When Robbins' equilibrium is reached the percentage of any of the eighth-order hyperplanes should be approximately 0.39%. The experiment was run with uniform recombination, with P_0 ranging from 0.1 to 0.5 (higher values were ignored due to symmetry).

Figure 6 graphs the results, which are quite striking. Although the proportion of the hyperplane $\#1\#1\#1\#1\#1\#1\#1\#$ decays smoothly towards its equilibrium proportion, this is certainly not true for the hyperplane $\#0\#1\#1\#1\#1\#1\#1\#$. Although $P_0 = 0.5$ uniform recombination does provide the fastest convergence in the limit of large time, as would be expected, it is also clear that $P_0 = 0.1$ provides much larger changes in the proportions during the early transient behavior. In fact, for all values of P_0 the change in the proportion of this hyperplane is so large that it temporarily overshoots the equilibrium proportion!

In summary, for higher-order hyperplanes, one can see that as P_0 increases to 0.5, the rate at which Robbins' equilibrium is approached also increases, in the limit. However, this does not necessarily hold throughout the transient dynamics of the system. In fact, we have shown an example in which a *less* disruptive recombination operator provides more substantive changes in the early transient behavior.

4 THE LIMITING DISTRIBUTION FOR MUTATION AND RECOMBINATION

The previous sections have considered mutation and recombination in isolation. A population undergoing recombination approaches Robbins' equilibrium, while a population undergoing mutation approaches a uniform equilibrium. What happens when both mutation and recombination act on a population? The answer is very simple. In general, Robbins' equilibrium is not the same as the uniform equilibrium; hence the population can not approach both distributions in the long term. In fact, in the long term, the uniform equilibrium prevails and we can state a similar theorem for mutation and recombination.

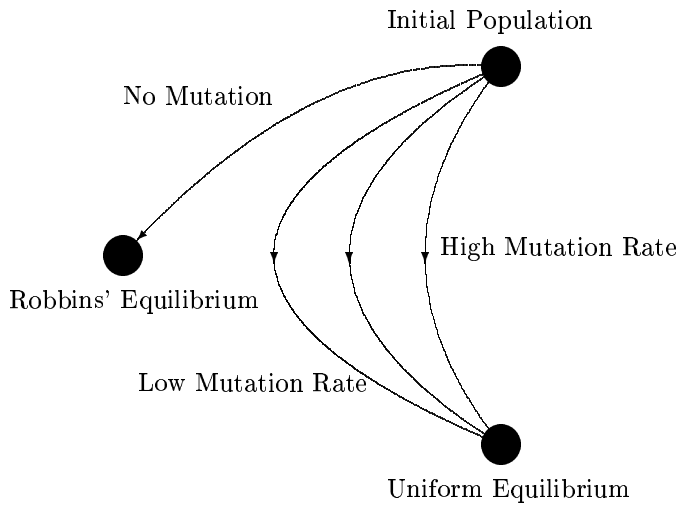

Figure 7: Pictorial representation of the action of mutation and recombination on the initial population

Theorem 3 *Let S be any string of L alleles: $(a_1, ..., a_L)$. If a population is mutated and recombined repeatedly (without selection) then:*

$$\lim_{t \to \infty} p_S(t) \;=\; \prod_{i=1}^{L} \frac{1}{C}$$

where $p_S(t)$ is the expected proportion of string S in the population at time t and C is the cardinality of the alphabet.

This is intuitively obvious. Recombination can not change the distribution of alleles at any locus – it merely shuffles alleles. Mutation, however, actually changes that distribution. Thus, the picture that arises is that a population that undergoes recombination and mutation attempts to approach a Robbins' equilibrium that is itself approaching the uniform equilibrium. Put another way, Robbins' equilibrium depends on the distribution of alleles in the initial population. This distribution is continually changed by mutation, until the uniform equilibrium distribution is reached. In that particular situation Robbins' equilibrium is the same as the uniform equilibrium distribution. Thus the effect of mutation is to move Robbins' equilibrium to the uniform equilibrium distribution. The speed of that movement will depend on the mutation rate μ (the greater that μ is the faster the movement). This is displayed pictorially in Figure 7.

5 SUMMARY

This paper investigated the limiting distributions of recombination and mutation, focusing not only on the dynamics near equilibrium, but also on the transient dynamics before equilibrium is reached. A population undergoing mutation approaches a uniform equilibrium in which every string is equally likely. The mutation rate μ and the initial population have no

effect on that limiting distribution, but they do affect the transient behavior. The transient behavior was examined via a differential equation model of this process (which is analogous to radioactive decay in physics). This allowed us to make quantitative statements as to how the initial population, the cardinality C of the alphabet, and the mutation rate μ affect the speed at which the equilibrium is approached.

We then investigated recombination. A population undergoing only recombination will approach Robbins' equilibrium. Geiringer's Theorem indicates that this equilibrium distribution depends only on the distribution of alleles in the initial population. The form of recombination and the cardinality are irrelevant. The paper then attempted to characterize the transient behavior of the system, by developing a differential equation model of the population. Using this, it is possible to show that the probability of disruption (P_{d}) and the probability of construction (P_{c}) of schemata are crucial to the time evolution of the system. These probabilities can be obtained from traditional schema analyses. We also provide the connection between the traditional schema analyses and an alternative framework based on marginal recombination distributions $\mathcal{R}_A(B)$ (Booker 1992). Survival (the opposite of disruption) is given by $\mathcal{R}_A(A)$ while construction is given by the remaining marginals $\mathcal{R}_A(B)$, $\emptyset \subset B \subset A$.

The analysis supports the theoretical result by Christiansen (1989) that, in the limit, more disruptive recombination operators (higher values of P_{d} or lower values of $\mathcal{R}_A(A)$) drive the population to equilibrium more quickly. However, we also show that the transient behavior can be subtle and can not be captured this simply. Instead the transient behavior depends on the whole probability distribution $\mathcal{R}_A(B)$, $B \subseteq A$ (and hence on the values of P_{c}).

The major contributor to interesting transient behavior appears to be the order of the hyperplane. We first examined second-order hyperplanes. By comparing one-point recombination and P_0 uniform recombination directly on second-order hyperplanes, we were able to derive a relationship showing when one-point recombination and uniform recombination both drive hyperplanes towards equilibrium at the same speed. We were also able to show that the linkage disequilibrium for second-order hyperplanes exponentially decays towards zero, with the probability of disruption P_{d} being the rate of decay. In these situations a more disruptive recombination operator drives hyperplanes towards equilibrium more quickly, even during the transient dynamics.

We then examined P_0 uniform recombination on hyperplanes of order $k > 2$. When $k < 5$ it is possible to show that when recombination becomes less disruptive ($\mathcal{R}_A(A)$ increases), *all* of the remaining marginals $\mathcal{R}_A(B)$ ($\emptyset \subset B \subset A$) decrease. Due to this, once again a more disruptive recombination operator drives hyperplanes towards equilibrium more quickly, even during the transient dynamics. However, when $k > 4$ the situation becomes much more interesting. In these situations some remaining marginals will decrease while others increase. This leads to behavior in which *less* disruptive recombination operators can in fact provide *larger* changes in hyperplane proportions, during the transient phase. These results are important because, due to the action of selection on a real GA population, the transient behavior of a population undergoing recombination is all that really matters.

Finally, we investigated the joint behavior of a population undergoing both mutation and recombination. We showed that, in a sense, the behavior of mutation takes priority, in that mutation actually moves Robbins' equilibrium until it is the same as the uniform equilibrium (i.e., all strings being equally likely).

References

Beyer, H.-G. (1998). On the dynamics of EAs without selection. In W. Banzhaf and C. Reeves (Eds.), *Foundations of Genetic Algorithms*, Volume 5, pp. 5–26. Morgan Kaufmann.

Booker, L. (1992). Recombination distributions for genetic algorithms. In D. Whitley (Ed.), *Foundations of Genetic Algorithms*, Volume 2, pp. 29–44. Morgan Kaufmann.

Christiansen, F. (1989). The effect of population subdivision on multiple loci without selection. In M. Feldman (Ed.), *Math. Evol. Theory*, pp. 71–85. Princeton University Press.

De Jong, K. and W. Spears (1992). A formal analysis of the role of multi-point crossover in genetic algorithms. *Annals of Mathematics and Artificial Intelligence 5*(1), 1–26.

Feller, W. C. (1968). *An Introduction to Probability Theory and its Applications*, Volume 1. Wiley.

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

Mühlenbein, H. (1998). The equation for response to selection and its use for prediction. *Evolutionary Computation 5*(3), 303–346.

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

Spears, W. (2000). *Evolutionary Algorithms: The Role of Mutation and Recombination.* Springer-Verlag.

Spears, W. and K. De Jong (1998). Dining with GAs: Operator lunch theorems. In *Foundations of Genetic Algorithms*, Volume 5. Morgan Kaufmann.

Stephens, C., H. Waelbroeck, and R. Aguirre (1998). Schemata as building blocks: Does size matter? In W. Banzhaf and C. Reeves (Eds.), *Foundations of Genetic Algorithms*, Volume 5, pp. 117–133. Morgan Kaufmann.