Dining with GAs: Operator Lunch Theorems

William M. Spears
AI Center - Code 5514
Naval Research Laboratory
Washington, D.C. 20375
spears@aic.nrl.navy.mil
and
Kenneth A. De Jong
Computer Science Department
George Mason University
Fairfax, VA 22030
kdejong@gmu.edu

Abstract

There has been considerable discussion of the pros/cons of recombination and mutation operators in the context of Holland's schema theory. In this paper we define a common framework for extending and relating previous "disruption" and "construction" analyses for both recombination and mutation. This results in several insights into the properties of recombination and mutation, including a No Free Lunch theorem for recombination operators as well as the lack of such a theorem for mutation.

1 Introduction

The motivation for the original schema analysis of Holland (1975) was to compute the expected number of instances of hyperplanes at time $t+1$, given their number at time t. To use the notation of Goldberg (1987), let $m_t(H)$ be the number of individuals in hyperplane H at time t. Then let $f_t(H)$ be the observed average fitness of the hyperplane at time t and let $\bar{f}_t$ be the observed average fitness of the population at time t. Then the *expected* number of individuals in H at time $t+1$ is given by the schema theorem:

$$E[m_{t+1}(H)] \;\geq\; m_t(H)\frac{f_t(H)}{\bar{f}_t}P_{survival}(H)$$

where $P_{survival}(H)$ is the probability that the hyperplane will not be disrupted by *either* mutation or recombination (i.e., it survives). The inequality refers to the fact that not only may a hyperplane H survive, it also may be *constructed* from other hyperplanes. See Holland (1975) or Goldberg (1987) for further discussions of this equation.

In order to improve this result to a precise equality, detailed aspects of the makeup of the entire population must be modeled (Goldberg 1989; Whitley 1992; Vose 1995), resulting in exact but difficult to analyze characterizations of GA behavior. In this paper we extend earlier work in which a less precise but useful characterization of population makeup is adopted, resulting in more tractable models of GA behavior (Spears and De Jong 1991; De Jong and Spears 1992). This allows us to define a framework for extending and relating previous "disruption" and "construction" analyses for both recombination and mutation. This results in several insights into the properties of recombination and mutation, including a No Free Lunch theorem (Wolpert 1992; Shaffer 1994; Rao, Gordon, and Spears 1995; Culberson 1996; English 1997; Wolpert and Macready 1997) for recombination operators as well as the lack of such a theorem for mutation.

2 Framework

The schema theorem articulates how the *expected* makeup of the next generation $E[m_{t+1}]$ is a function of the current generation m_t, selection, and the reproductive operators. If cloning is the only reproductive operator, $E[m_{t+1}]$ is completely determined by m_t and selection. Of interest then is characterizing the perturbation effects due to reproductive operators such as recombination and mutation. Following the standard schema analysis, we focus on a particular kth order hyperplane H_k and ask how recombination and mutation perturb the expected number of offspring in H_k in the next generation.

We do so in the following manner. First, to allow a direct comparison of the effects of recombination and mutation, we treat them both as reproductive operators that take two parents as input and produce two children as output. In the case of mutation, this is equivalent to two independent applications of the standard one-parent mutation operator, but has the advantage that it allows us to compare the expected number of offspring residing in H_k as a result of a single application of either operator.

More formally, let B_k be a random variable describing the number of offspring residing in H_k as a result of a single application of any two-parent reproductive operator. Then the expected number of offspring in H_k can be computed as follows:

$$E[B_k] \;=\; \sum_{b\in\{0,1,2\}} b \times P(B_k = b) \;\;=\;\; P(B_k = 1) \;+\; 2P(B_k = 2) \qquad (1)$$

2.1 Survival analysis

Historically, the schema theorem has emphasized the disruptive aspects of reproductive operators by assuming that at least one of the parents is a member of H_k and calculating

the likelihood that neither of the children will be instances of H_k. The complement of this is "survival" in which at least one of the offspring are in H_k.

More formally, let A_k be a random variable which describes the number of parents that are instances of H_k. For survival analysis, then, A_k can take on values 1 and 2. Again, let B_k be the random variable describing the number of offspring that are instances of H_k. B_k can take on values 0, 1, and 2. We can then express the probability of survival as:

$$\begin{aligned} P_s(H_k) &= P(B_k = 1 \ \vee \ 2 \mid A_k = 1 \ \vee \ 2) \\ &= P(B_k = 1 \mid A_k = 1 \ \vee \ 2) \ + \ P(B_k = 2 \mid A_k = 1 \ \vee \ 2) \end{aligned} \tag{2}$$

For survivability, the general formula given by equation 1 for $E[B_k]$ (the expected number of offspring in H_k) specializes to:

$$E_s[B_k] \ = \ \sum_{b \in \{0,1,2\}} b \times P(B_k = b \mid A_k = 1 \ \vee \ 2) \tag{3}$$

The subscript s is a reminder that the situation is one in which the *survival* of an individual in H_k is at stake. By expanding the summation and using equation 2 we have:

$$\begin{aligned} E_s[B_k] &= P(B_k = 1 \mid A_k = 1 \ \vee \ 2) \ + \ 2P(B_k = 2 \mid A_k = 1 \ \vee \ 2) \\ &= P_s(H_k) \ + \ P(B_k = 2 \mid A_k = 1 \ \vee \ 2) \end{aligned} \tag{4}$$

So, we can see that the expected number of offspring that will reside in H_k is determined by the traditional survival analysis $P_s(H_k)$ and an additional term giving the probability that both offspring will be in H_k. The particulars are, of course, operator specific and will be explored in more detail in later sections.

2.2 Construction analysis

The constructive view of a two-parent operator is that members of H_k are built up from instances of lower-order hyperplanes (Syswerda 1989). Following Syswerda, we consider the creation of a kth-order hyperplane H_k from two lower-order hyperplanes H_m and H_n of order m and n respectively, with H_m and H_n non-overlapping and $k = m + n$.

Of the k alleles in H_k, m are supplied by one parent, while n are supplied by the other. Since they are non-overlapping, there are 2^k ways the k alleles necessary to construct H_k can be distributed between the two parents. We refer to each of these ways as a *situation*, which is denoted by $0 \leq \mathcal{S} < 2^k$. The binary representation of $\mathcal{S}$ indicates which parent has each of the k alleles. Thus, $\mathcal{S}$ uniquely identifies H_m and H_n. For example, $\mathcal{S} = 0$ corresponds with $H_m = H_k$ and $\mathcal{S} = 2^k - 1$ corresponds with $H_n = H_k$. Then, for each situation we denote the probability of constructing a member of H_k from H_m and H_n as $P_c(H_k \mid \mathcal{S})$.

Once again consider A_k to be the random variable that describes the number of parents that are instances of H_k. For survival analysis it is always assumed that at least one parent

is an instance of H_k. However, for construction it is also possible for neither parent to be in H_k, so for construction A_k can take on values 0, 1, and 2. Again, B_k is the random variable describing the number of offspring that are instances of H_k. B_k can take on values 0, 1, and 2. This allows us to more formally express the probability that H_k will be constructed from a given situation of H_m and H_n:

$$\begin{aligned} P_c(H_k \mid \mathcal{S}) &= P(B_k = 1 \ \vee \ 2 \mid (A_k = 0 \ \vee \ 1 \ \vee \ 2) \ \wedge \ \mathcal{S}) \\ &= P(B_k = 1 \ \vee \ 2 \mid \mathcal{S}) \\ &= P(B_k = 1 \mid \mathcal{S}) \ + \ P(B_k = 2 \mid \mathcal{S}) \end{aligned} \tag{5}$$

As before, the general formula given by equation 1 (for computing the expected number of offspring in H_k) can be specialized for construction:

$$E_c[B_k \mid \mathcal{S}] \ = \ \sum_{b \in \{0,1,2\}} b \times P(B_k = b \mid \mathcal{S})$$

The subscript c is a reminder that the situation is one in which the *construction* of an individual in H_k is at stake. By expanding the summation and using equation 5, we have:

$$\begin{aligned} E_c[B_k \mid \mathcal{S}] &= P(B_k = 1 \mid \mathcal{S}) \ + \ 2P(B_k = 2 \mid \mathcal{S}) \\ &= P_c(H_k \mid \mathcal{S}) \ + \ P(B_k = 2 \mid \mathcal{S}) \end{aligned}$$

So, we can see that the expected number of offspring that will be in H_k is determined by the traditional construction analysis $P_c(H_k \mid \mathcal{S})$ and an additional term giving the probability that both offspring will be in H_k. Finally we can compute the average $E_c[B_k]$ over all constructive situations. Of the 2^k situations, all are constructive except for the two extreme cases in which all alleles are on one of the parents ($\mathcal{S} = 0$ and $\mathcal{S} = 2^k - 1$). These two cases represent survival, not construction. Hence,

$$\begin{aligned} E_c[B_k] &= \frac{1}{2^k - 2} \sum_{\mathcal{S}=1}^{2^k-2} E_c[B_k \mid \mathcal{S}] \\ &= \frac{1}{2^k - 2} \sum_{\mathcal{S}=1}^{2^k-2} \Big[P_c(H_k \mid \mathcal{S}) \ + \ P(B_k = 2 \mid \mathcal{S}) \Big] \\ &= \frac{1}{2^k - 2} \sum_{\mathcal{S}=1}^{2^k-2} P_c(H_k \mid \mathcal{S}) \ + \ \frac{1}{2^k - 2} \sum_{\mathcal{S}=1}^{2^k-2} P(B_k = 2 \mid \mathcal{S}) \end{aligned} \tag{6}$$

Note that the first term is just the probability of constructing a member of H_k averaged over all constructive situations. If we denote this average by $P_c(H_k)$, i.e.,

$$P_c(H_k) \ = \ \frac{1}{2^k - 2} \sum_{\mathcal{S}=1}^{2^k-2} P_c(H_k \mid \mathcal{S})$$

then the previous equation simplifies to:

$$E_c[B_k] \;=\; P_c(H_k) \;+\; \frac{1}{2^k-2}\sum_{\mathcal{S}=1}^{2^k-2} P(B_k = 2 \mid \mathcal{S}) \tag{7}$$

Further simplifications are operator specific and will be explored in detail in later sections.

2.3 Survival + Construction analysis

The formalism of the previous section provides the basis to express the combined effects of both survival and construction by simply including the two survival terms that were removed ($H_m = H_k$ and $H_n = H_k$). Thus the expected number of offspring in H_k is given by:

$$\begin{aligned}
E_{c,s}[B_k] &= \frac{1}{2^k}\sum_{\mathcal{S}=0}^{2^k-1} E_{c,s}[B_k \mid \mathcal{S}] \\
&= \frac{1}{2^k}\sum_{\mathcal{S}=0}^{2^k-1} \Big[P_{c,s}(H_k \mid \mathcal{S}) \;+\; P(B_k = 2 \mid \mathcal{S})\Big] \\
&= \frac{1}{2^k}\sum_{\mathcal{S}=0}^{2^k-1} P_{c,s}(H_k \mid \mathcal{S}) \;+\; \frac{1}{2^k}\sum_{\mathcal{S}=0}^{2^k-1} P(B_k = 2 \mid \mathcal{S})
\end{aligned}$$

The subscript c,s is a reminder that the situation is one in which the construction or survival of an individual in H_k is at stake. As before, we note that the first term is just the probability of obtaining a member of H_k from either construction or survival, averaged over all situations. If we denote this average by $P_{c,s}(H_k)$, then the previous equation simplifies to:

$$E_{c,s}[B_k] \;=\; P_{c,s}(H_k) \;+\; \frac{1}{2^k}\sum_{\mathcal{S}=0}^{2^k-1} P(B_k = 2 \mid \mathcal{S}) \tag{8}$$

$P_{c,s}(H_k)$ can be further expanded to explicitly identify the contributions of survival and construction. Recall that the two survival cases are when $H_m = H_k$ ($\mathcal{S} = 0$) and $H_n = H_k$ ($\mathcal{S} = 2^k - 1$). Hence,

$$\begin{aligned}
P_{c,s}(H_k) &= \frac{1}{2^k}\sum_{\mathcal{S}=0}^{2^k-1} P_{c,s}(H_k \mid \mathcal{S}) \\
&= \frac{1}{2^k}\Big[P_{c,s}(H_k \mid \mathcal{S}=0) \;+\; P_{c,s}(H_k \mid \mathcal{S}=2^k-1) \;+\; \sum_{\mathcal{S}=1}^{2^k-2} P_{c,s}(H_k \mid \mathcal{S})\Big] \\
&= \frac{1}{2^k}\Big[2P_s(H_k) \;+\; (2^k-2)P_c(H_k)\Big] \qquad (9)
\end{aligned}$$

This general framework can now be applied to specific two-parent operators. We do so for recombination and mutation in the following sections.

3 Dining with Recombination

3.1 Survival under recombination

Equation 4 provides the general expression for the expected number of offspring B_k in H_k via survival, namely:

$$E_s[B_k] \;=\; P_s(H_k) \;+\; P(B_k = 2 \mid A_k = 1 \;\vee\; 2)$$

For the specific case of recombination, the term $P(B_k = 2 \mid A_k = 1 \;\vee\; 2)$ can be simplified by noting that the only way both offspring can reside in H_k is if both parents do. Hence,

$$\begin{aligned}
P(B_k = 2 \mid A_k = 1 \;\vee\; 2) \quad &= \quad P(B_k = 2 \;\wedge\; A_k = 1) \;+\; P(B_k = 2 \;\wedge\; A_k = 2) \\
&= \quad P(B_k = 2 \mid A_k = 1)\; P(A_k = 1) \;+ \\
&\quad\;\; P(B_k = 2 \mid A_k = 2)\; P(A_k = 2) \\
&= \quad 0 \times P(A_k = 1) \;+\; 1 \times P(A_k = 2) \\
&= \quad P(A_k = 2)
\end{aligned}$$

It remains then to derive $P(A_k = 2)$. In general, it is difficult to derive precise expressions for such probabilities at a particular point in time because they vary from generation to generation in complex, non-linear, and interacting ways. We can, however, for visualization purposes obtain considerable insight into the effects of population homogeneity by making the simplifying assumption that at a particular point in time, the probability that both parents have the same allele at a particular defining position d is given by $P_{eq}(d) = P_{eq}$. That is, the probabilities at each defining position are independent and identical.

Adopting the $P_{eq}(d) = P_{eq}$ assumption here allows us to express the probability that both parents are in H_k simply as $P(A_k = 2) \;=\; {P_{eq}}^k$. Hence,

$$E_s[B_k] \;=\; P_s(H_k) \;+\; {P_{eq}}^k$$

It is easily seen from this formulation how the expected number of survivors can vary from 0 to 2 as a function of the disruptiveness of the recombination operator and the homogeneity of the population. The actual form this expected value takes is easily visualized by simply adding the constant ${P_{eq}}^k$ to typical survival probability curves for standard N-point recombination and parameterized P_0 uniform recombination operators (see De Jong and Spears (1992) for more details). In the special case of $P_{eq} = 0$ the graphs are identical (see Figure 1).[1] If $P_{eq} > 0$ all the curves in Figure 1 are translated upward – their basic form remains the same.

[1] L is the length of the individual.

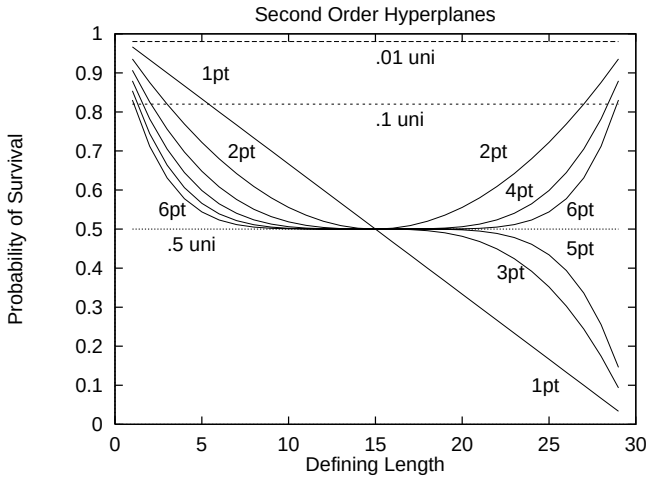

Figure 1: $P_s(H_k)$ for H_2 when $L = 30$ and $P_{eq} = 0.0$, for N-point and P_0 uniform recombination.

3.2 Construction via recombination

Equation 7 gives the general expression for the expected number of offspring B_k in H_k via construction, namely:

$$E_c[B_k] \;=\; P_c(H_k) \;+\; \frac{1}{2^k - 2} \sum_{\mathcal{S}=1}^{2^k-2} P(B_k = 2 \mid \mathcal{S})$$

Again, the only way two recombination offspring can reside in H_k is if both parents do. Hence, $P(B_k = 2 \mid \mathcal{S}) \;=\; P(A_k = 2)$ and

$$\begin{aligned} E_c[B_k] \quad &= \quad P_c(H_k) \;+\; \frac{1}{2^k - 2} \sum_{\mathcal{S}=1}^{2^k-2} P(A_k = 2) \\ &= \quad P_c(H_k) \;+\; P(A_k = 2) \end{aligned}$$

If, as before, we adopt the $P_{eq}(d) = P_{eq}$ assumption for visualization purposes, then we have:

$$E_c[B_k] \;=\; P_c(H_k) \;+\; {P_{eq}}^k$$

It is easily seen from this formulation how the expected number of constructed members of H_k can vary from 0 to 2 as a function of the constructiveness of the recombination operator and the homogeneity of the population. The actual form this expected value takes is easily visualized by simply adding the constant ${P_{eq}}^k$ to typical construction probability curves for standard N-point recombination and parameterized P_0 uniform recombination operators

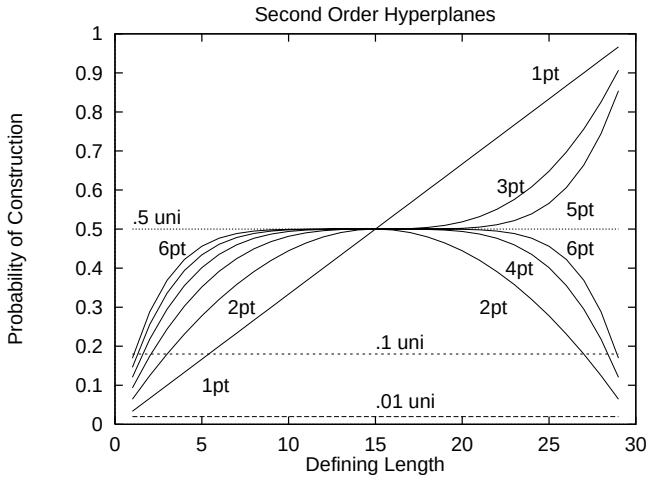

Figure 2: $P_c(H_k)$ of H_2 when $L = 30$ and $P_{eq} = 0.0$, for N-point and P_0 uniform recombination.

(see Spears and De Jong (1991) for more details). In the special case of $P_{eq} = 0$ the graphs are identical (see Figure 2). If $P_{eq} > 0$ all the curves in Figure 2 are translated upward – their basic form remains the same.

In comparing the survival and construction graphs for recombination, one is immediately struck by their complementary features: more disruptive operators (low survivability) have high constructive potential and vice versa. This suggests there is a No Free Lunch Theorem lurking in the background. We explore this possibility in the next section.

3.3 Survival + Construction using recombination

Equation 8 provides a general expression for the expected number of offspring B_k in H_k (via survival or construction), namely:

$$E_{c,s}[B_k] \;=\; P_{c,s}(H_k) \;+\; \frac{1}{2^k}\sum_{\mathcal{S}=0}^{2^k-1} P(B_k = 2 \mid \mathcal{S})$$

As we have seen before, with recombination B_k can only be 2 when A_k is 2. Thus,

$$P(B_k = 2 \mid \mathcal{S}) \;=\; P(A_k = 2)$$

Using this fact along with equation 9 we have:

$$\begin{aligned} E_{c,s}[B_k] \quad &= \quad P_{c,s}(H_k) \;+\; P(A_k = 2) \\ &= \quad \frac{1}{2^k}\Big[2P_s(H_k) \;+\; (2^k - 2)P_c(H_k)\Big] \;+\; P(A_k = 2) \end{aligned} \tag{10}$$

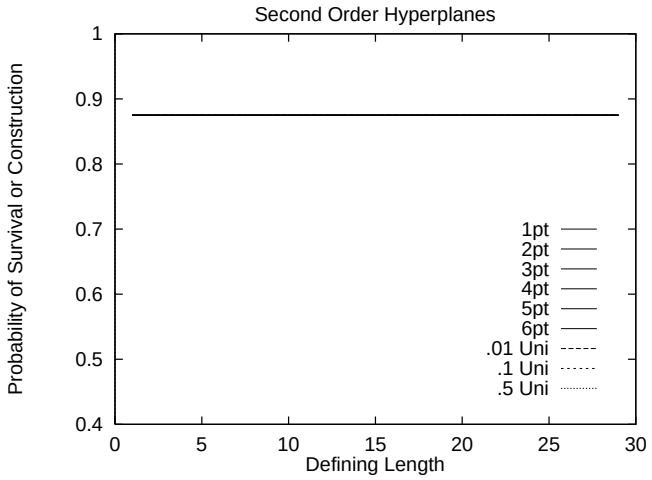

Figure 3: $P_{c,s}(H_k)$ of H_2 when $L = 30$ and $P_{eq} = 0.5$. The results are the same regardless of recombination operator.

If we make our $P_{eq}(d) = P_{eq}$ assumption for visualization purposes, then $P(A_k = 2) = {P_{eq}}^k$ as before and:

$$E_{c,s}[B_k] = \frac{1}{2^k}\Big[2P_s(H_k) + (2^k - 2)P_c(H_k)\Big] + {P_{eq}}^k$$

If we plot $P_{c,s}(H_k)$ (or $E_{c,s}[B_k]$) for all the standard N-point and P_0 uniform recombination operators, we obtain the rather surprising result illustrated in Figure 3. *All* recombination operators have the same graphs. $P_{c,s}(H_k)$ and $E_{c,s}[B_k]$ are only affected by H_k and population homogeneity, but *not* by the form of recombination. Or, in more familiar terms, *there appears to be no free lunch with recombination*! Increasing survivability results in an offsetting reduction in constructability, and vice versa.

The key question is whether this "no free lunch" observation is due to our $P_{eq}(d) = P_{eq}$ assumption for visualization purposes, or whether this result is independent of any such assumptions. The rather surprising answer is that it is true in the more general case. We provide the proof for this in the remainder of this section.

For H_k there are 2^k possible recombination events, denoted by $\mathcal{R}$, where $0 \leq \mathcal{R} < 2^k - 1$. Each recombination event $\mathcal{R}$ can be represented by a bit mask of length k (i.e., the binary representation of $\mathcal{R}$), where a '1' at position j indicates that recombination swapped the alleles at position j between the two parents, and a '0' means that recombination did not swap the alleles at position j. All N-point recombination events and all parameterized uniform recombination events can be described with these bit masks.

Consider a breakdown of $P_{c,s}(H_k)$ over all situations $\mathcal{S}$ and recombination events $\mathcal{R}$ as follows:

$$P_{c,s}(H_k) = \frac{1}{2^k}\sum_{\mathcal{S}} P_{c,s}(H_k \mid \mathcal{S})$$

$$P_{c,s}(H_k) \;=\; \frac{1}{2^k} \sum_{\mathcal{R}} \sum_{\mathcal{S}} P(\mathcal{R})\; P_{c,s}(H_k \mid \mathcal{S} \;\wedge\; \mathcal{R})$$

$$P_{c,s}(H_k) \;=\; \frac{1}{2^k} \sum_{\mathcal{R}} \Big[P(\mathcal{R}) \sum_{\mathcal{S}} P_{c,s}(H_k \mid \mathcal{S} \;\wedge\; \mathcal{R}) \Big]$$

The following important fact illustrates a tight relationship between situations and recombination events:

$$P_{c,s}(H_k \mid \mathcal{S} \;\wedge\; \mathcal{R}) \;=\; P_{c,s}(H_k \mid \mathcal{S} \oplus z \;\wedge\; \mathcal{R} \oplus z) \quad \forall\, z \;\; \forall\, \mathcal{S} \;\; \forall\, \mathcal{R} \tag{11}$$

The variable z represents any integer and the $\oplus$ operator represents addition modulo 2^k. The point is that since there are 2^k situations and 2^k recombination events, nothing changes if both the situation and the recombination event are changed the same way. For example, suppose one considers the situation $\mathcal{S} = 0$ and recombination event $\mathcal{R} = 0$. The situation $\mathcal{S} = 0$ indicates that all of the alleles for hyperplane H_k are in the first parent. The recombination event $\mathcal{R} = 0$ indicates that no alleles are exchanged during recombination. Now also consider situation $\mathcal{S} = 1$ and recombination event $\mathcal{R} = 1$. In this case the second parent contains one of the desired alleles. However, since $\mathcal{R} = 1$ will in fact exchange that allele, the offspring will be the same as that produced from situation $\mathcal{S} = 0$ and recombination event $\mathcal{R} = 0$. This example is easily generalized to yield equation 11. Thus:

$$P_{c,s}(H_k) \;=\; \frac{1}{2^k} \sum_{\mathcal{R}} \Big[P(\mathcal{R}) \sum_{\mathcal{S}} P_{c,s}(H_k \mid \mathcal{S} \oplus z \;\wedge\; \mathcal{R} \oplus z) \Big] \quad \forall\, z$$

If we let $z = -\mathcal{R}$ (modulo 2^k), we can rephrase the inner sum in terms of one recombination event only:

$$P_{c,s}(H_k) \;=\; \frac{1}{2^k} \sum_{\mathcal{R}} \Big[P(\mathcal{R}) \sum_{\mathcal{S}} P_{c,s}(H_k \mid \mathcal{S} \ominus \mathcal{R} \;\wedge\; \mathcal{R} = 0) \Big]$$

where the $\ominus$ operator represents subtraction modulo 2^k. Since the inner summation is summing over all situations (they are just shifted by $\mathcal{R}$), this is equivalent to:

$$P_{c,s}(H_k) \;=\; \frac{1}{2^k} \sum_{\mathcal{R}} \Big[P(\mathcal{R}) \sum_{\mathcal{S}} P_{c,s}(H_k \mid \mathcal{S} \;\wedge\; \mathcal{R} = 0) \Big]$$

This inner summation can now be separated from the events $\mathcal{R}$:

$$P_{c,s}(H_k) \;=\; \frac{1}{2^k} \Big[\sum_{\mathcal{S}} P_{c,s}(H_k \mid \mathcal{S} \;\wedge\; \mathcal{R} = 0) \Big] \Big[\sum_{\mathcal{R}} P(\mathcal{R}) \Big]$$

Now, the probability of all recombination events must sum to 1.0, so:

$$P_{c,s}(H_k) \;=\; \frac{1}{2^k}\Big[\sum_{\mathcal{S}} P_{c,s}(H_k \mid \mathcal{S} \;\wedge\; \mathcal{R}=0)\Big] \tag{12}$$

Clearly this does not depend on the form of recombination, since the probability of recombination events is absent. What this says is that $P_{c,s}(H_k)$ is the same, regardless of the form of recombination. Only the population homogeneity will change the value of $P_{c,s}(H_k)$. This is true also for $E_{c,s}[B_k]$ (see equation 10).

This particular No Free Lunch theorem is similar in spirit to the previous results for concept learning (Wolpert 1992; Shaffer 1994; Rao, Gordon, and Spears 1995), search, and optimization (Wolpert and Macready 1997). Roughly speaking, the results in concept learning depend upon the assumption that concept learning problems are uniformly likely. Similarly, the results in search and optimization depend upon the assumption that the functions to be searched are uniformly likely (for further details see the cited papers). Although our results do not depend on assumptions about functions or problems per se, they do depend on the assumption that all situations $\mathcal{S}$ are uniformly likely. It is an open issue as to whether our results will generalize to non-uniform distributions of situations.

4 Dining with Mutation

To provide insight into the relative roles of recombination and mutation, we develop a similar analysis for mutation in this section. As noted earlier we do so by also viewing mutation as a two-parent operator in order to directly compare the expected number of offspring B_k in H_k produced via mutation with that of recombination.

We assume mutation works on alphabets of cardinality C in the following fashion.[2] An allele is picked for mutation with probability μ. Then that allele is changed to one of the other $C-1$ alleles, uniformly randomly.[3]

4.1 Survival under mutation

Equation 3 provides the general expression for the expected number of offspring B_k in H_k via survival, namely:

$$E_s[B_k] \;=\; \sum_{b\in\{0,1,2\}} b \times P(B_k = b \mid A_k = 1 \;\vee\; 2) \tag{13}$$

Without loss of generality, assume that the first parent is in H_k, while the second parent is arbitrary. To compute $E_s[B_k]$ it turns out to be more convenient to focus on the similarity of the two parents, as opposed to concentrating on A_k explicitly. This is done by letting Q be a random variable that describes the set of alleles (at the defining positions) in the second parent that do not match H_k. Then we can write $E_s[B_k]$ as follows:

[2] The analysis for recombination holds for arbitrary cardinality alphabets as well.

[3] This form of mutation is reasonable for discrete representations, however, it should be modified for real-valued representations.

$$E_s[B_k] \;=\; \sum_Q P(Q)\Big[\; P(B_k = 1 \mid Q) \;+\; 2P(B_k = 2 \mid Q)\;\Big]$$

Now consider the derivation of $P(B_k = 2 \mid Q)$. In order to have both offspring be in H_k (i.e., $B_k = 2$), the k alleles in the first parent (associated with the hyperplane H_k) must not be mutated, since the first parent is already in H_k. However, the $|Q|$ differing alleles in the second parent must be mutated, while the remaining $k - |Q|$ alleles in the second parent must not be mutated, in order to place the second offspring in H_k.[4] For a general alphabet of cardinality C, if an allele is mutated, there is a $1/(C-1)$ probability of mutating it to the desired allele. Thus, the probability of placing *both* offspring in H_k is simply computed as:

$$P(B_k = 2 \mid Q) \;=\; (1-\mu)^k \;\left[\; \left(\frac{\mu}{C-1}\right)^{|Q|} (1-\mu)^{k-|Q|} \;\right]$$

where the probability of not mutating the k alleles of the first parent is $(1-\mu)^k$, and the remainder of the expression is the probability of mutating the second parent into the hyperplane H_k.

It is now possible to compute the probability that *only one* offspring will be in H_k. Clearly that will occur if the first parent is kept in H_k while the second parent is not mutated into H_k, or if the first parent is mutated out of H_k while the second parent is mutated into H_k. This can be simply computed by using the components of the previous equation:

$$\begin{aligned} P(B_k = 1 \mid Q) \quad = \quad & (1-\mu)^k \;\left[\; 1 \;-\; \left(\frac{\mu}{C-1}\right)^{|Q|} (1-\mu)^{k-|Q|} \;\right] \;+ \\ & \left[\; 1 \;-\; (1-\mu)^k \;\right] \left[\; \left(\frac{\mu}{C-1}\right)^{|Q|} (1-\mu)^{k-|Q|} \;\right] \end{aligned}$$

With some simplification $E_s[B_k]$ can now be expressed for μ mutation:

$$E_s[B_k] \;=\; \sum_Q P(Q)\Big[\; (1-\mu)^k \;+\; \left(\frac{\mu}{C-1}\right)^{|Q|} (1-\mu)^{k-|Q|} \;\Big] \tag{14}$$

It will be noted that $P(Q)$ depends on the population homogeneity. If we make our $P_{eq}(d) = P_{eq}$ assumption for visualization purposes, then $P(Q)$ is simply:

$$P(Q) \;=\; (1-P_{eq})^{|Q|}\; {P_{eq}}^{k-|Q|} \tag{15}$$

Figure 4 illustrates $E_s[B_k]$ when $C = 2$, for mutation rates ranging from 0.0 to 1.0, for H_2, while P_{eq} ranges from 0.0 to 1.0. Inspection of Figure 4 for plausible estimates of population

[4] $|Q|$ is the cardinality of the set Q.

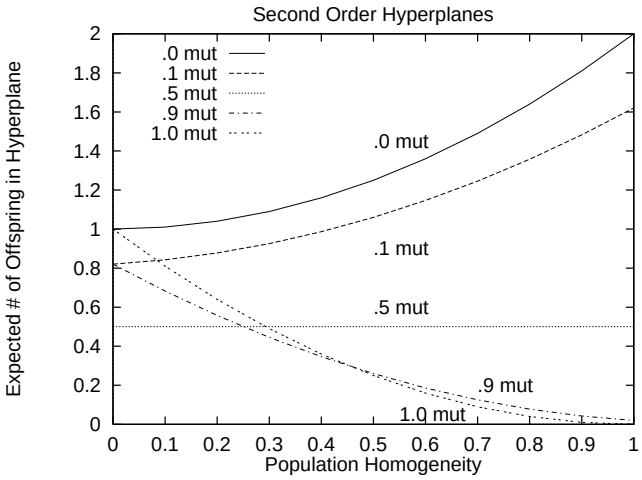

Figure 4: $E_s[B_k]$ of H_2 for mutation when $C = 2$.

homogeneity (where $P_{eq} > 1/C$) indicates that the maximum survivability (highest $E_s[B_k]$) occurs when $\mu = 0.0$, while the minimum survivability (lowest $E_s[B_k]$) occurs when $\mu = 1.0$, as would be expected. Note also that $E_s[B_k]$ is unaffected by P_{eq} when $\mu = 0.5$. This makes sense, since when $C = 2$ and the mutation rate is 0.5, both parents are just randomly reinitialized.

4.2 Construction via mutation

By equation 6 we can write (for mutation):

$$E_c[B_k] \;=\; \frac{1}{2^k - 2} \sum_{\mathcal{S}=1}^{2^k-2} E_c[B_k \mid \mathcal{S}]$$

where the situations range over the $2^k - 2$ possible constructive ways in which the k alleles can be distributed across the two parents. What remains, then, is to derive an expression for $E_c[B_k \mid \mathcal{S}]$.

Without loss of generality, assume that the first parent is in H_n, while the second parent is in H_m. To compute $E_c[B_k \mid \mathcal{S}]$ it will be convenient to let Q be a random variable that describes the set of alleles (at the defining positions) in the second parent that do not match H_n. Similarly, let R be a random variable that describes the set of alleles (at the defining positions) in the first parent that do not match H_m. Then we can write $E_c[B_k \mid \mathcal{S}]$ as follows:

$$E_c[B_k \mid \mathcal{S}] \;=\; \sum_{Q} \sum_{R} P(Q \;\wedge\; R) \Big[\; P(B_k = 1 \mid Q \;\wedge\; R) \;+\; 2P(B_k = 2 \mid Q \;\wedge\; R) \;\Big]$$

Now consider the derivation of $P(B_k = 2 \mid Q \;\wedge\; R)$. In order to have *both* offspring be in

H_k (i.e., $B_k = 2$), the n alleles in the first parent (associated with the hyperplane H_n) must not be mutated. Also, of the remaining m alleles in the first parent, $|R|$ must be mutated (while $m - |R|$ are not). Finally, the m alleles in the second parent (associated with the hyperplane H_m) must not be mutated. Of the remaining n alleles in the second parent, $|Q|$ must be mutated (while $n - |Q|$ are not). For a general alphabet of cardinality C, if an allele is mutated, there is a $1/(C-1)$ probability of mutating it to the desired allele. Thus, the probability of placing *both* offspring in H_k is simply computed as:

$$P(B_k = 2 \mid Q \ \wedge \ R) \ = \\ \left[\left(\frac{\mu}{C-1}\right)^{|R|} (1-\mu)^{m-|R|} \ (1-\mu)^n \right] \left[\left(\frac{\mu}{C-1}\right)^{|Q|} (1-\mu)^{n-|Q|} \ (1-\mu)^m \right]$$

The first term expresses the probability of placing the first offspring in H_k. The probability of not mutating the n correct alleles of the first parent is $(1-\mu)^n$. Also, since $|R|$ of the remaining m alleles are incorrect, $|R|$ must be mutated to the correct allele while $m - |R|$ are not mutated. The second term expresses the probability of placing the second offspring in H_k.

It is now possible to compute the probability that *only one* offspring will be in H_k. Clearly that will occur if the first offspring is placed in H_k while the second offspring is not, or if the second offspring is placed in H_k while the first offspring is not. This can be easily computed by using the components of the previous equation:

$$P(B_k = 1 \mid Q \ \wedge \ R) \ = \\ \left[\left(\frac{\mu}{C-1}\right)^{|R|} (1-\mu)^{m-|R|} \ (1-\mu)^n \right] \left[1 \ - \ \left(\frac{\mu}{C-1}\right)^{|Q|} (1-\mu)^{n-|Q|} \ (1-\mu)^m \right] \ + \\ \left[1 \ - \ \left(\frac{\mu}{C-1}\right)^{|R|} (1-\mu)^{m-|R|} \ (1-\mu)^n \right] \left[\left(\frac{\mu}{C-1}\right)^{|Q|} (1-\mu)^{n-|Q|} \ (1-\mu)^m \right]$$

Thus, with some simplification:

$$E_c[B_k \mid \mathcal{S}] \ = \\ \sum_Q \sum_R P(Q \ \wedge \ R) \left[\left(\frac{\mu}{C-1}\right)^{|R|} (1-\mu)^{k-|R|} \ + \ \left(\frac{\mu}{C-1}\right)^{|Q|} (1-\mu)^{k-|Q|} \right] \tag{16}$$

It will be noted that $P(Q \ \wedge \ R)$ depends on the population homogeneity. If we make our $P_{eq}(d) = P_{eq}$ assumption for visualization purposes, then $P(Q \ \wedge \ R)$ is simply:

$$P(Q \ \wedge \ R) \ = \ (1 - P_{eq})^{|Q|+|R|} \ {P_{eq}}^{k-|Q|-|R|} \tag{17}$$

Figure 5 illustrates $E_c[B_k]$ when the cardinality of the alphabet $C = 2$, for mutation rates ranging from 0.0 to 1.0, for H_2, while P_{eq} ranges from 0.0 to 1.0. Inspection of Figure 5

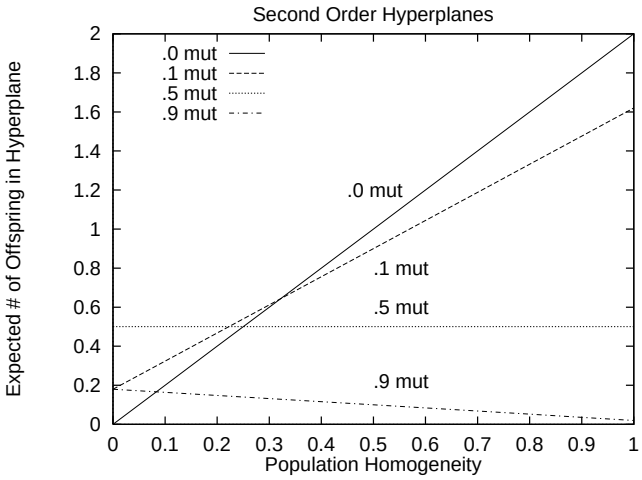

Figure 5: $E_c[B_k]$ of H_2 for mutation when $C = 2$.

where $P_{eq} > 1/C$ indicates that the maximum construction (highest $E_c[B_k]$) occurs when $\mu = 0.0$, while the minimum construction (lowest $E_c[B_k]$) occurs when $\mu = 1.0$, as would be expected. Note also that $E_c[B_k]$ is unaffected by P_{eq} when $\mu = 0.5$. This makes sense, since when $C = 2$ and the mutation rate is 0.5, both parents are just randomly reinitialized.

4.3 Survival + Construction using mutation

In an earlier section we saw that more disruptive recombination operators achieve higher levels of construction. However, this is not the case for mutation. Although high levels of mutation are the most disruptive (low values of $E_s[B_k]$), they also achieve the worst levels of construction (lowest values of $E_c[B_k]$). Thus, in general, a No Free Lunch theorem with respect to the disruptive and constructive aspects of mutation does not hold. The implications of this particular difference between mutation and recombination are not yet clear.

5 Summary and Conclusions

There has been considerable discussion of the pros/cons of recombination and mutation operators in the context of Holland's schema theory. In this paper we defined a common framework for extending and relating previous "disruption" and "construction" analyses for both recombination and mutation.

The framework indicated that more disruptive recombination operators achieve higher levels of construction. This led to a No Free Lunch theorem for recombination operators, with respect to survivability (the opposite of disruption) and construction. The theorem makes the assumption that all situations $\mathcal{S}$ are equally probable. This seems reasonable, since if one is given no prior information concerning the problem the evolutionary algorithm is to see, there appears to be no reason to assume that any particular situation will be more

likely than another.[5] Thus, given no prior information, it appears problematic to assume that any particular recombination operator will yield desirable behavior.

On the other hand, the more disruptive mutation rates yield lower levels of construction. Thus, there is no general No Free Lunch for mutation, with respect to disruption and construction. The implications of this particular result are not yet clear.

The framework introduced in this paper can also be used to help characterize the roles of recombination and mutation. These results are beyond the scope of this paper, but the interested reader can find the details in Spears (1998). We provide a brief synopsis here. It turns out that mutation can achieve any level of disruption that recombination can achieve, but can also achieve higher levels of disruption than recombination. Thus one role of mutation appears to be disruption. On the other hand, recombination can achieve higher levels of construction than mutation, indicating that one role of recombination does in fact appear to be construction. Moreover, the constructive advantage of recombination (over mutation) is maximized when the lower-order hyperplanes H_m and H_n are of roughly the same order. Finally, there is one specific case where mutation obeys the above No Free Lunch theorem. It turns out that when the cardinality $C = 2$ and the population is maximally diverse ($P_{eq} = 0$), μ mutation acts just like P_0 uniform recombination, thus allowing the No Free Lunch theorem for recombination to carry over to mutation.

References

Culberson, J. (1996). On the futility of blind search. Technical Report TR-18, University of Alberta.

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.

English, T. (1997). Information is conserved in optimization. Technical report, Texas Tech University.

Goldberg, D. (1987). Simple genetic algorithms and the minimal, deceptive problem. In L. Davis (Ed.), *Genetic Algorithms and Simulated Annealing*. Morgan Kaufmann.

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

Holland, J. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press.

Rao, B., D. Gordon, and W. Spears (1995). For every generalization action is there an equal and opposite reaction? (Analyzing the conservation law for generalization performance). In *Machine Learning Conference*, Volume 12, pp. 471–479.

Shaffer, C. (1994). A conservation law for generalization performance. In *Machine Learning Conference*, Volume 11, pp. 259–265.

Spears, W. (1998). *The role of mutation and recombination in evolutionary algorithms*. Ph. D. thesis, George Mason University, Fairfax, VA.

[5] Although selection may cause certain situations to be more likely than others (for a given problem), this will vary from problem to problem. Thus, given no prior information about the problems being searched we have no choice but to assume a uniform distribution over $\mathcal{S}$.

Spears, W. and K. De Jong (1991). On the virtues of parameterized uniform crossover. In R. K. Belew and L. B. Booker (Eds.), *International Conference on Genetic Algorithms*, Volume 4, pp. 230–236. Morgan Kaufmann.

Syswerda, G. (1989). Uniform crossover in genetic algorithms. In *International Conference on Genetic Algorithms*, Volume 3, pp. 2–9. Morgan Kaufmann.

Vose, M. (1995). Modeling simple genetic algorithms. *Evolutionary Computation 3*(4), 453–472.

Whitley, D. (1992). An executable model of a simple genetic algorithm. In D. Whitley (Ed.), *Foundations of Genetic Algorithms*, Volume 2, pp. 45–62. Morgan Kaufmann.

Wolpert, D. (1992). On the connection between in-sample testing and generalization error. *Complex Systems 6*, 47–94.

Wolpert, D. and W. Macready (1997). No free lunch theorems for optimization. *IEEE Trans. on Evolutionary Computation 1*(1), 67–82.