Genetic Set Recombination

Nicholas J. Radcliffe

`njr@epcc.ed.ac.uk`

Edinburgh Parallel Computing Centre

University of Edinburgh

King's Buildings

EH9 3JZ

Scotland

Abstract

The application of genetic algorithms to optimisation problems for which the solution is a set or multiset (bag) is considered. A previous extension of schema analysis, known as forma analysis, is further developed and used to construct principled representations and operators for problems in this class. The extensions to forma analysis include the introduction of genes whose values cannot be assigned independently and a method for mediating between desirable but sometimes incompatible properties of recombination operators.

1 Introduction

This paper is concerned with optimisation problems for which the solution is a set or multiset (bag). Examples include selecting an investment portfolio, choosing a connectivity for a neural network and finding the best sites for a network of retail outlets given a choice of possible locations. Both the case in which the size of the set or multiset is fixed and the case in which it is subject to optimisation are considered.

The approach taken is based on *forma analysis,* an extension to schema analysis (Holland, 1975) developed previously (Radcliffe, 1991a, 1991b). Section 2 presents a brief review of forma analysis, but the reader unfamiliar with this approach may find it helpful to read Radcliffe (1991b), which provides a gentler introduction. Various formae (generalised

schemata) for sets and multisets are introduced in sections 4–6, and suitable recombination operators for their manipulation are constructed. The key problems with set and multiset optimisation arise when their size is constrained. In this case, the natural formae are said to be *non-separable,* and the construction of satisfactory recombination operators is especially hard. This difficulty is tackled in section 5.2 by introducing genes whose values cannot always be assigned independently. This is achieved through the formalism of *non-orthogonal bases.* A new recombination operator, called *random assorting recombination,* is introduced to deal with this case in section 7.

2   Forma Analysis: Summary and Definitions

This section reviews forma analysis as developed in Radcliffe (1991a, 1991b). In later sections this will be used to analyse recombination of sets and multisets.

2.1   Equivalence Relations and Formae

Let $\mathcal{S}$ be a search space and let Ψ be a set of equivalence relations over $\mathcal{S}$. Then the equivalence classes of all the equivalence relations in Ψ are referred to as *formae,* which are generalisations of schemata. The set of all formae induced by a set of equivalence relations Ψ will be written $\Xi(\Psi)$. Formae satisfy the "schema theorem" (Holland, 1975) provided that suitable disruption coefficients are chosen (Radcliffe, 1991a, Vose & Liepins, 1991). Disruption is discussed further in section 7.3.

It will be assumed throughout this paper that there are enough equivalence relations in Ψ to ensure that specifying all of the equivalence classes to which a solution in $\mathcal{S}$ belongs suffices to identify that solution uniquely.

2.2   Respect

A recombination operator X can be conveniently described by a function

$$X : \mathcal{S} \times \mathcal{S} \times \mathcal{K}_X \longrightarrow \mathcal{S} \tag{1}$$

which takes two parent solutions A and B from the search space, $\mathcal{S}$ together with a control parameter $k \in \mathcal{K}_X$, and produces a child in $\mathcal{S}$. The control parameter, k, determines which of the (typically) many possible children is produced, so that for one point crossover k would be the cross point and for uniform crossover it would be a binary mask. A recombination operator X is said to *respect* the set $\Xi(\Psi)$ of formae induced by the equivalence relations in Ψ if and only if

$$\forall \xi \in \Xi \ \forall A \in \xi \ \forall B \in \xi \ \forall k \in \mathcal{K}_X : \ X(A, B, k) \in \xi. \tag{2}$$

Thus respect requires that all children produced by recombination alone are members of all formae to which both their parents belong. For example, suppose there were equivalence relations for hair and eye colour in Ψ. Then if both parents had blue eyes and brown hair, respect would require that recombination only produce children with blue eyes and brown hair.

2.3   Similarity Set

It is convenient to introduce the notion of the *similarity set* of two parents, which is— loosely—the set of all solutions which share the characteristics that the two parent solutions
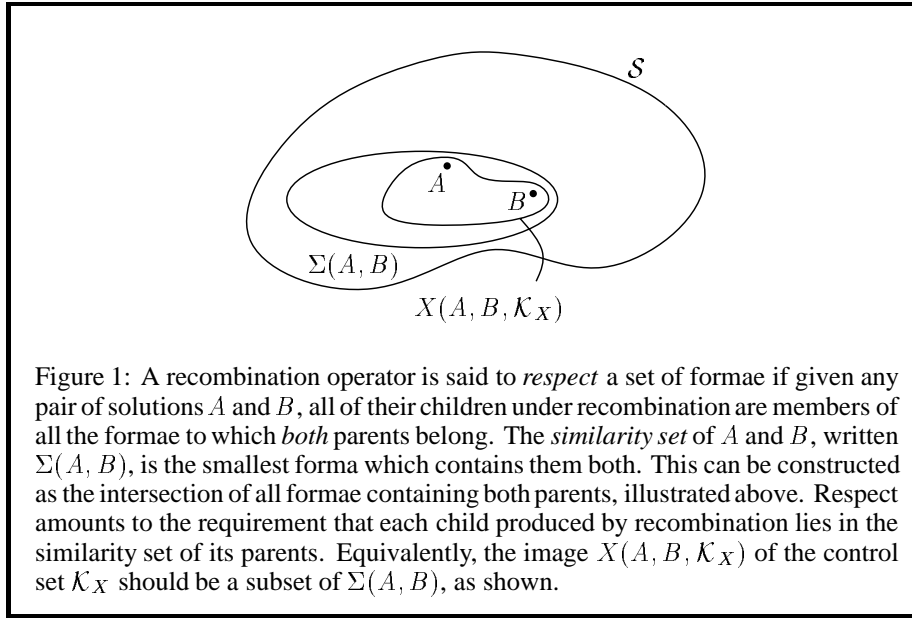
Figure 1: A recombination operator is said to *respect* a set of formae if given any pair of solutions A and B, all of their children under recombination are members of all the formae to which *both* parents belong. The *similarity set* of A and B, written $\Sigma(A, B)$, is the smallest forma which contains them both. This can be constructed as the intersection of all formae containing both parents, illustrated above. Respect amounts to the requirement that each child produced by recombination lies in the similarity set of its parents. Equivalently, the image $X(A, B, \mathcal{K}_X)$ of the control set $\mathcal{K}_X$ should be a subset of $\Sigma(A, B)$, as shown.

share. Formally, the similarity set of A and B, written $\Sigma(A, B)$, is defined by

$$\Sigma(A, B) = \bigcap \left\{ \xi \in \Xi(\Psi) \mid A, B \in \xi \right\} . \tag{3}$$

This is the intersection of all formae to which both parents belong—the smallest forma to which they both belong provided that this intersection is itself an equivalence class induced by some equivalence relation in Ψ.

Respect can be seen to be identical to the requirement that every child lie in the similarity set of its parents, as is illustrated in figure 1. This allows respect to be re-expressed as

$$\forall A, B \in \mathcal{S}: \ X(A, B, \mathcal{K}_X) \subset \Sigma(A, B). \tag{4}$$

2.4 Assortment

A recombination operator X is said *properly to assort* the formae in $\Xi(\Psi)$ if and only if

$$\forall \xi_1, \xi_2 \in \Xi(\Psi) \ (\xi_1 \cap \xi_2 \neq \varnothing) \ \forall A_1 \in \xi_1 \ \forall A_2 \in \xi_2 \ \exists k \in \mathcal{K}_X: \ X(A_1, A_2, k) \in \xi_1 \cap \xi_2. \tag{5}$$

Thus proper assortment requires that, given parents $A_1 \in \xi_1$ and $A_2 \in \xi_2$, a single recombination can generate a child in $\xi_1 \cap \xi_2$ provided this intersection is non-empty. Continuing the same example used to illustrate respect, if one parent has blue eyes and the other has brown hair, then recombination must allow the construction of a child with blue eyes and brown hair, provided that these characteristics are compatible.

2.5 Separability and Random Respectful Recombination

A set $\Xi(\Psi)$ of formae is said to be *separable* if it is capable of being simultaneously respected and properly assorted. The *random respectful recombination* operator (R^3)

makes a uniform random selection from the similarity set of the two parents. This operator respects and properly assorts the formae if they are separable. To see this, recall that respect is the requirement that children lie in the similarity set of their parents (equation 4). R^3 guarantees this, and generation of any solution outside the similarity set would by definition violate respect. So if respect and assortment are to be compatible conditions, any solutions which the latter requires be capable of production must fall within the similarity set. Since every solution in the similarity set is generated with non-zero probability by the R^3 operator, it must properly assort as well as respect whenever these conditions are compatible.

2.6 Complete Orthogonal Basis

For present purposes an equivalence relation $\sim$ in Ψ is conveniently expressed as a binary function

$$\psi : \mathcal{S} \times \mathcal{S} \longrightarrow \{\,0,1\,\} \tag{6}$$

which returns 1 if its arguments are equivalent and 0 if they are not:

$$\psi(A,B) = \begin{cases} 1, & \text{if } A \sim B, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

The intersection of two equivalence relations $\psi_1, \psi_2 \in \Psi$ can then be defined by

$$(\psi_1 \cap \psi_2)(A,B) = \begin{cases} 1, & \text{if } \psi_1(A,B) = \psi_2(A,B) = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Given this, a subset $E \subset \Psi$ is said to constitute a *complete orthogonal basis* for Ψ provided that

- (*Completeness*) Each relation $\psi \in \Psi$ can be constructed as the intersection of some subset of the basic relations:

$$\forall \psi \in \Psi \; \exists \, E_\psi \subset E : \bigcap E_\psi = \psi. \tag{9}$$

- (*Orthogonality*) Given any subset of the equivalence relations in E, the intersection of each possible combination of equivalence classes (basic formae) induced by these equivalence relations should be non-empty. Formally, let Ξ_ψ be the set of formae induced by the equivalence relation ψ and given a subset F of the equivalence relations in Ψ, let

$$\Xi_F = \prod_{\psi \in F} \Xi_\psi, \tag{10}$$

the space of vectors of equivalence classes induced by the various relations in F. Then orthogonality requires that

$$\forall F \subset E \; \forall (\xi_1, \xi_2, \dots, \xi_{|F|}) \in \Xi_F : \bigcap_{i=1}^{|F|} \xi_i \neq \emptyset \; . \tag{11}$$

This means, in effect, that the basic equivalence class (basic forma) can be chosen independently for each basic equivalence relation in E without introducing incompatibilities. (It should be noted that the definition of orthogonality given here is different from that given in Radcliffe (1991b). The earlier definition, which required only pairwise compatibility between basic equivalence relations, is not strict enough for the purposes of this paper, and is altogether a less satisfactory definition.) If equation 11 is satisfied only for sets F up to some size N then the basis will be said to be *orthogonal to order* N.

2.7 Genes, Alleles and Gene Transmission

A recombination operator is said to be *strictly transmitting* if it is the case that for each basic equivalence relation $\psi \in \Psi$ the children produced by recombination always lie in the same basic forma as one or other of the parents. (Equivalently, every child should be equivalent to at least one of its parents under each basic equivalence relation.) Strict gene transmission trivially implies respect. Basic equivalence relations can be identified with *genes,* and basic equivalence classes (basic formae) with *alleles.* Thus if eye colour were one of the basic equivalence relations, eye colour would constitute a "gene" with alleles "blue", "green" and "brown".

2.8 Example: Schemata

In later sections formae which differ significantly from schemata will be introduced, but for illustration the characteristics of one-point crossover, uniform crossover[1] and random respectful recombination (R^3) are compared for the case of k-ary (base k) schemata in table 1.

Schemata can be identified as equivalence classes induced by certain equivalence relations as follows. Consider the set of equivalence relations $\{\psi_i\}$, defined by

$$\psi_i(A, B) = \begin{cases} 1, & \text{if } A_i = B_i, \\ 0, & \text{otherwise,} \end{cases} \tag{12}$$

which relate two solutions if they have the same value (allele) for the ith gene. The ith such equivalence relation can conveniently be denoted

$$\psi_i = \square\square \cdots \square\blacksquare\square\square \cdots \square \tag{13}$$

where $\square$ is the familiar "don't care" character used to describe schemata, and the "care" symbol $\blacksquare$ occurs at the ith position. The set of such equivalence relations over all positions i induce all first order schemata (basic formae, alleles) and form a complete orthogonal basis for equivalence relations constructed by the arbitrary intersection of members of the basis. Such intersected equivalence relations induce all higher order schemata.

One-point, two-point and uniform crossover can immediately be seen to transmit genes, because each of a child's genes comes from one or other parent. By virtue of their strict transmission, these operators also plainly respect schemata. While uniform crossover also properly assorts schemata, because a child can be constructed with any combination of its parents' genes in a single recombination, one-point crossover does not, because $10\,10$ cannot be constructed from parents 0000 and 1111. It does, however, weakly assort in the sense that repeated incestuous recombination will allow an arbitrary admixture of parental genes.

R^3 is guaranteed to respect and properly assort schemata since schemata can be seen to be separable by virtue of their separation by uniform crossover. R^3 can be implemented by initially copying across all genes common to the two parents into the child and then filling remaining positions with random alleles. Except in the case $k = 2$, this random allocation of alleles to genes which take on different values in the parents prevents random respectful recombination from strictly transmitting genes.

---

[1] with parameter half, i.e. with each bit in the mask equally likely to be one or zero.

Table 1: Summary of the characteristics of various recombination operators.

Operator	Respect	Assortment	Gene Transmission
one-point	●	weak	●
two-point	●	weak	●
uniform	●	proper	●
R^3	●	proper	$k = 2$ only

3 Sets and Multiset Recombination

Recall that the distinction between a set and a multiset is that duplication of elements is not significant in sets, so that

$$\{ a, a, b \} \equiv \{ a, b \}, \tag{14}$$

whereas in multisets (also known as "bags") an element may appear more than once

$$\{\!\{ a, a, b \}\!\} \neq \{\!\{ a, b \}\!\}. \tag{15}$$

(The notation $\{\!\{ \cdots \}\!\}$ is used in this paper to indicate a multiset.) A number of different set and multiset optimisation problems may be distinguished. In general there will be a "universal set", $\mathcal{E}$, from which elements are drawn. The aim is to construct a set or multiset consisting of elements drawn from this universal set so as to optimise some property of the resulting set or multiset. Examples could include finding locations for bottle banks so as to maximise recycling in some area, selecting members of a committee to make a environmental impact assessment or choosing connections in a neural network to minimise its average learning time to some acceptable error (Radcliffe, 1992). Whitley (1987) has studied the use of genetic search over restricted poker hands using a multiset formulation; this is discussed in Radcliffe (1990).

4 Recombining Fixed-Size Sets

Given a universal set $\mathcal{E}$, the search space for sets of fixed size N is

$$\mathcal{S} = \big\{ A \subset \mathbb{P}(\mathcal{E}) \,\big|\, |A| = N \big\}, \tag{16}$$

(figure 2), where $\mathbb{P}(\mathcal{E})$ is the *power set* (set of all subsets) of $\mathcal{E}$:

$$\mathbb{P}(\mathcal{E}) = \{ B \subset \mathcal{E} \}. \tag{17}$$

4.1 Equivalence Relations and Formae

Given a universal set $\mathcal{E}$, with $a \in \mathcal{E}$, and solutions $A, B \in \mathcal{S}$, let

$$\psi_{\{a\}}(A, B) = \begin{cases} 1, & \text{if } a \in A \cap B \text{ or } a \notin A \cup B, \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$
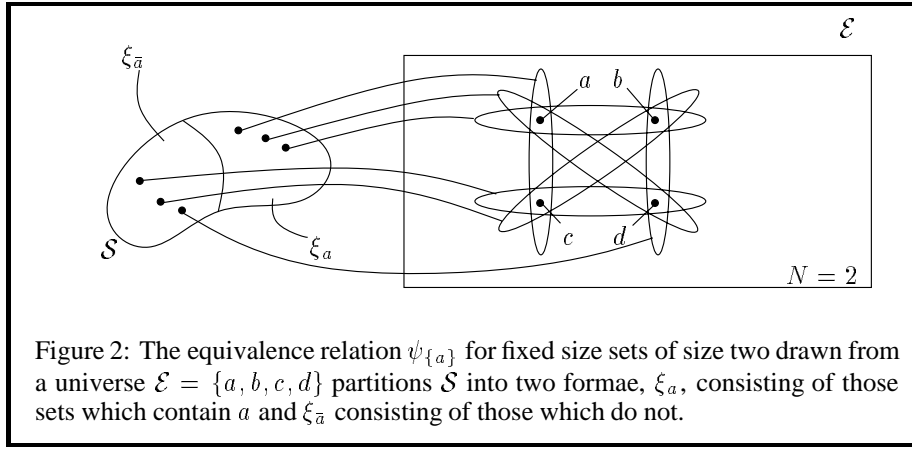
Figure 2: The equivalence relation $\psi_{\{a\}}$ for fixed size sets of size two drawn from a universe $\mathcal{E} = \{a, b, c, d\}$ partitions $\mathcal{S}$ into two formae, ξ_a, consisting of those sets which contain a and $\xi_{\bar{a}}$ consisting of those which do not.

This equivalence relation induces two equivalence classes, one comprising the solutions containing the element a and another comprising those which do not (figure 2). There is clearly an equivalence relation $\psi_{\{x\}}$ of the form described by equation 18 for *each* $x \in \mathcal{E}$. Moreover, these are intuitively natural candidates for a basis for a set Ψ of equivalence relations which might generate all formae specifying the presence or absence of any subset of the elements in $\mathcal{E}$. As will now be demonstrated, if the rule for intersection of equivalence relations described by equation 8 is followed, the set

$$E = \left\{ \psi_{\{x\}} \mid x \in \mathcal{E} \right\} \tag{19}$$

forms a basis, orthogonal to some order K, for a set of equivalence relations Ψ which induces a useful set of formae. Ψ can be defined by

$$\Psi = \left\{ \psi \mid \exists\, E_\psi \in E : \ \psi = \bigcap E_\psi \right\}. \tag{20}$$

To see this, consider the intersection of $\psi_{\{a\}}$ and $\psi_{\{b\}}$, which will be denoted $\psi_{\{a,b\}}$. According to the definition of intersection for equivalence relations (equation 8)

$$(\psi_{\{a\}} \cap \psi_{\{b\}})(A, B) = \begin{cases} 1, & \text{if } \psi_{\{a\}}(A, B) = \psi_{\{b\}}(A, B) = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

This equivalence relation, illustrated in figure 3, induces four equivalence classes, which might conveniently be written

$$\begin{aligned}
\xi_{ab} &= \{A \in \mathcal{S} \mid a \in A, b \in A\}, \\
\xi_{a\bar{b}} &= \{A \in \mathcal{S} \mid a \in A, b \notin A\}, \\
\xi_{\bar{a}b} &= \{A \in \mathcal{S} \mid a \notin A, b \in A\}, \\
\xi_{\bar{a}\bar{b}} &= \{A \in \mathcal{S} \mid a \notin A, b \notin A\}.
\end{aligned} \tag{22}$$

The generalisation of this is rather obvious. A general equivalence relation, $\psi \in \Psi$, has an associated *description set,* conveniently written $\langle \psi \rangle$, which is a subset of the universal set $\mathcal{E}$. Members of the search space (themselves subsets of $\mathcal{E}$) are then equivalent under ψ
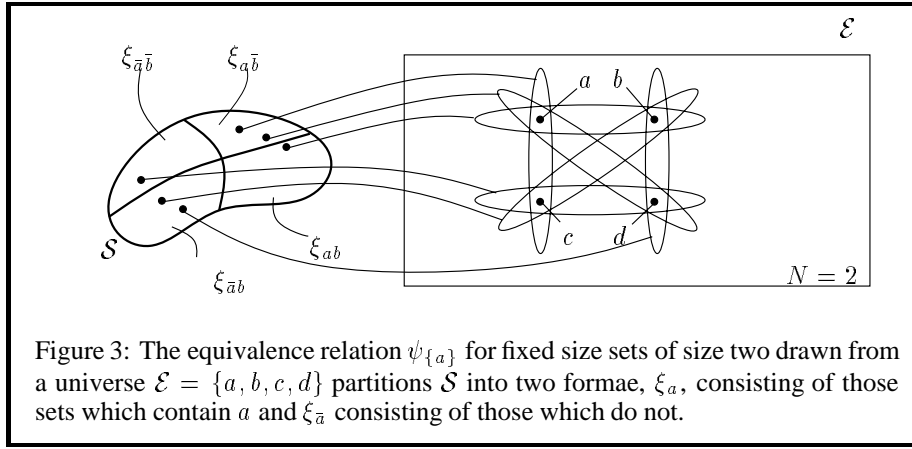
Figure 3: The equivalence relation $\psi_{\{a\}}$ for fixed size sets of size two drawn from a universe $\mathcal{E} = \{a, b, c, d\}$ partitions $\mathcal{S}$ into two formae, ξ_a, consisting of those sets which contain a and $\xi_{\bar{a}}$ consisting of those which do not.

precisely if they contain the same subset of the members of the description set $\langle\psi\rangle$ (figure 3). Formally,

$$\psi(A, B) = \begin{cases} 1, & \text{if } \langle\psi\rangle \cap A = \langle\psi\rangle \cap B, \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

It is clear that E (defined in equation 19) does indeed form a basis for Ψ, but that this basis is not fully orthogonal. Completeness follows because

$$\forall \psi \in \Psi : \; \psi = \bigcap_{x \in \langle\psi\rangle} \psi_{\{x\}}, \tag{24}$$

so that every relation can be expressed as an intersection of the basic relations defined in equation 19, as required for completeness (equation 9). Orthogonality to order K follows provided that for up to K basic equivalence relations it is possible independently to choose whether or not a set should include the element labelling the basic equivalence relation. K is given by

$$K = \max\{N, |\mathcal{E}| - N\}. \tag{25}$$

To see this, notice that if $|\mathcal{E}| \geq 2N$ then $K = N$, because after picking N elements to include it would be impossible to include an $N + 1$th. Similarly, if $|\mathcal{E}| < 2N$ then once $|\mathcal{E}| - N$ elements had been chosen for exclusion from the set it would be impossible to exclude any more. The lack of full orthogonality will prove to be problematical when recombination operators are constructed.

A general forma ξ induced by an equivalence relation $\psi \in \Psi$ can then be characterised by a partition of the description set $\langle\psi\rangle$. It thus becomes convenient to describe a forma by a 2-tuple

$$\langle\xi\rangle = (\xi^+, \xi^-) \tag{26}$$

where

$$\xi^+ \cap \xi^- = \emptyset \tag{27}$$

and

$$\xi^+ \cup \xi^- = \langle\psi\rangle \tag{28}$$

with the interpretation

$$A \in \xi \iff \left(A \cap \xi^+ = \xi^+ \text{ and } A \cap \xi^- = \emptyset \right). \qquad (29)$$

This says that a set A is a member of the forma ξ if and only if it contains all those elements in ξ^+ and none of those in ξ^-. Having introduced this formalism, it is possible to identify the similarity set of two solutions with respect to the formae $\Xi(\Psi)$ induced by Ψ. This will allow the random respectful recombination operator R^3 to be constructed.

Recall that the similarity set of two solutions is defined by

$$\Sigma(A, B) = \bigcap \{\xi \in \Xi(\Psi) \mid A, B \in \xi\}. \qquad (30)$$

Clearly A and B will share membership of a forma ξ if and only if

$$A \cap \xi^+ = B \cap \xi^+ = \xi^+ \qquad (31)$$

and

$$A \cap \xi^- = B \cap \xi^- = \emptyset. \qquad (32)$$

Equation 31 can be satisfied if and only if

$$\xi^+ \subset A \cap B \qquad (33)$$

and equation 32 can be satisfied if and only if

$$\xi^- \subset \mathcal{E} - (A \cup B), \qquad (34)$$

where the minus sign denotes set subtraction. It is thus clear that when performing an intersection of all formae satisfying these conditions the similarity set as specified in equation 30 must be described by

$$\langle \Sigma(A, B) \rangle = \left(A \cap B, \ \mathcal{E} - (A \cup B) \right). \qquad (35)$$

The R^3 operator makes a random (uniform) selection from this similarity set. For example, with the universe

$$\mathcal{E} = \{a, b, c, d, e, f\} \qquad (36)$$

and $N = 3$,

$$\langle \Sigma(\{a, b, c\}, \{a, d, e\}) \rangle = \left(\{a, b, c\} \cap \{a, d, e\}, \ \mathcal{E} - (\{a, b, c\} \cup \{a, d, e\}) \right)$$

$$= \left(\{a\}, \{f\} \right). \qquad (37)$$

This describes the forma containing those sets which contain a and exclude f:

$$\Sigma(\{a, b, c\}, \{a, d, e\}) = \left\{ \{a, b, c\}, \{a, b, d\}, \{a, b, e, \}, \{a, c, d\}, \{a, c, e\}, \{a, d, e, \} \right\}. \qquad (38)$$

Thus, R^3 for these formae can be understood as an operator which

1. copies all the elements which are common to the two parents into the child;

2. fills the remaining places in the child with a random selection of the unused elements from the two parents.

A child C of A and B thus has the natural properties

$$A \cap B \subset C \subset A \cup B. \tag{39}$$

It is clear, therefore, that in this case R^3 strictly transmits genes, a gene being labelled by an element of $\mathcal{E}$ and an allele corresponding to the presence or absence of that element (equation 18). Notice, however, that the formae are not separable, with the consequence that neither R^3 nor any other respectful operator can assort them. To see this, simply observe that $\{a, b, c\}$ is a member of the forma ξ_{bc} and $\{a, d, e\}$ is a member of the forma ξ_d but that R^3 cannot produce a member in the intersection $\xi_{bc} \cap \xi_d = \xi_{bcd}$ of these formae because respect restricts the choice of children to those in the similarity set given in equation 38. This arises directly from the restriction to fixed-size sets. A way of trading-off respect and assortment is discussed in section 7.

An alternative way of viewing this operator is to imagine a conventional linear chromosome in which every position represents an element from the universal set, and to imagine an operator like uniform crossover, but constrained so that the total number of 1's in the child is constant and equal to N, the fixed size of the set.

5 Recombination of Fixed-Size Multisets

5.1 Equivalence Relations and Formae

The extension of the previous case from sets to multisets is in essence simple, but involves one complication. The basic idea will be that rather than specify whether or not certain elements are in the multiset, a forma will specify the *multiplicities* of some elements. Again, assume that $\mathcal{E}$ is a universal set from which all elements are to be drawn, but that elements may now be taken more than once. Then let $\mathbb{P}_m(\mathcal{E})$ be the *multipower set* of $\mathcal{E}$, that is, the set of all multisets whose elements are drawn from $\mathcal{E}$. Then the *multiplicity* function

$$m : \mathcal{E} \times \mathbb{P}_m(\mathcal{E}) \longrightarrow \mathbb{Z}^+ \cup \{0\} \tag{40}$$

is defined so that $m(x, A)$ is the number of copies of x in the multiset A.

A forma for multisets could either specify exact multiplicities for certain elements or could give bounds on their multiplicities. Since the former is a special case of the latter, where the bounds are maximally tight, the more general case will be examined.

A forma is now conveniently described by a set of 3-tuples of the form $(x, N_x^\downarrow, N_x^\uparrow)$ each of which is understood to specify that the multiplicity $m(x, A)$ of the element x in the multiset A lies in the inclusive range $N_x^\downarrow$ to $N_x^\uparrow$. For example, a forma ξ with the description set

$$\langle \xi \rangle = \{(a, 0, 0), (b, 1, 3)\} \tag{41}$$

contains all those multisets over $\mathcal{E}$ of size N which contain no copies of a and contain between one and three copies of b (figure 4). Such formae are closely related to the "range formae" discussed in Radcliffe (1991a).

5.2 Non-Orthogonality

There are a number of sets of equivalence relations which could be constructed to generate these formae. An obvious starting point is equivalence relations which induce formae
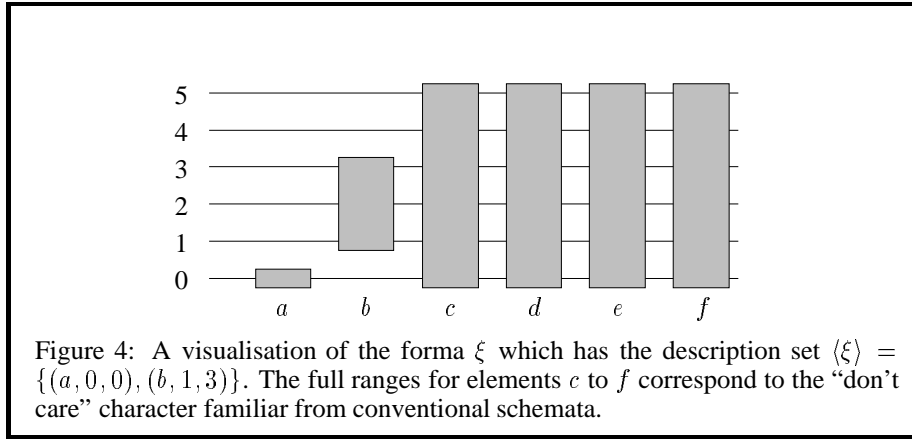
Figure 4: A visualisation of the forma ξ which has the description set $\langle \xi \rangle = \{(a, 0, 0), (b, 1, 3)\}$. The full ranges for elements c to f correspond to the "don't care" character familiar from conventional schemata.

defined with respect to a single element x from $\mathcal{E}$. The equivalence relation ψ which induces the forma described by $\langle \xi \rangle = \{(x, N_x^{\downarrow}, N_x^{\uparrow})\}$ would have the same description set

$$\langle \psi \rangle = \left\{(x, N_x^{\downarrow}, N_x^{\uparrow})\right\} \tag{42}$$

and would be defined by

$$\psi(A, B) = \begin{cases} 1, & \text{if } (m(x, A) < N_x^{\downarrow} \text{ and } m(x, B) < N_x^{\downarrow}) \\ & \text{or } m(x, A), m(x, B) \in [N_x^{\downarrow}, N_x^{\uparrow}] \\ & \text{or } (m(x, A) > N_x^{\uparrow} \text{ and } m(x, B) > N_x^{\uparrow}), \\ 0, & \text{otherwise,} \end{cases} \tag{43}$$

where

$$[N_x^{\downarrow}, N_x^{\uparrow}] = \left\{ n \in \mathbb{Z} \mid N_x^{\downarrow} \le n \le N_x^{\uparrow} \right\}. \tag{44}$$

As was the intention, formae can now specify a range of multiplicities for any element and a single equivalence relation will be seen to suffice to define up to three ranges simultaneously. The natural candidates for a basis are the equivalence relations which divide the range of multiplicities for a single element into a lower portion and an upper portion, as shown in figure 5,

$$E = \left\{ \psi \in \Psi \mid \langle \psi \rangle = \{(x, 0, N_x^{\uparrow})\}, x \in \mathcal{E}, N_x^{\uparrow} \in [0, N^{\star}] \right\} \tag{45}$$

where $N^{\star}$ is the maximum allowed multiplicity for an element. These equivalence relations can easily be seen to be complete, for any equivalence relation with a description set $\{(x, N_x^{\downarrow}, N_x^{\uparrow})\}$ can be constructed as an intersection of the relations with description sets $\{(x, 0, N_x^{\downarrow} - 1)\}$ and $\{(x, 0, N_x^{\uparrow})\}$ (figure 6). Equivalence relations defined with respect to more than one member of $\mathcal{E}$ can then be constructed trivially by intersection.

It is easy, however, to see that the relations in E do not satisfy the condition of orthogonality specified in equation 11. To verify this, simply note that if a multiset is a member of the forma with description set $\{(x, 0, 1)\}$ (induced by ψ_1, as labelled in figure 5) it cannot also be a member of the forma with the description set $\{(x, 4, N^{\star})\}$ (induced by ψ_3) as would be required if E were orthogonal (equation 11). This is because no multiset can have both fewer than two and more than three copies of the element x. Thus a multiset cannot

Figure 5: The set of equivalence relations with description sets of the form $\{(x, 0, N)\}$ divide the range $[0, N^\star]$ into a lower and and upper portion as shown: the equivalence relations may be thought of as simple dividing lines at integer-plus-half values.
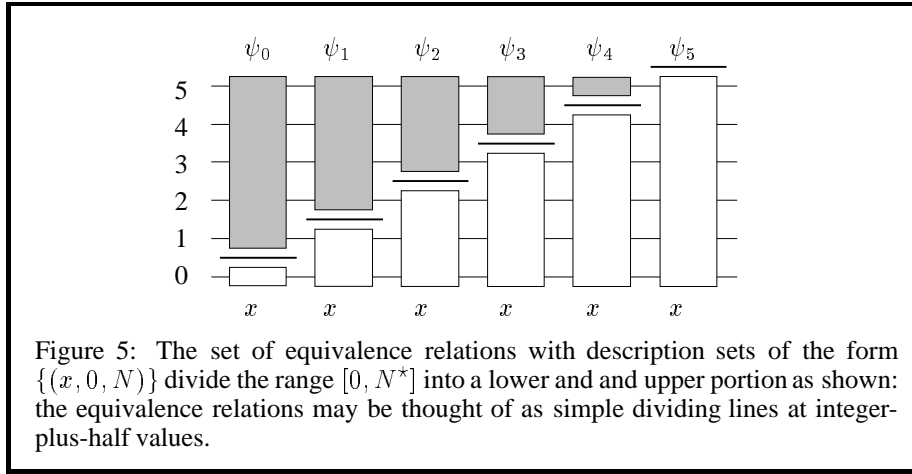
be constructed with independent choice of its forma membership of for the equivalence relations ψ_1 and ψ_3, as shown in figure 5, violating orthogonality.

Rather than abandon this potential basis, it is instructive to return to the analogy with linear algebra which led to the original formulation of the conditions on a basis, namely completeness and orthogonality. In linear algebra there is a weaker notion than orthogonality known as *linear independence*: a set of vectors is said to be linearly independent if no one of them can be expressed as a linear combination of the others. Following this analogy, the following definition suggests itself:
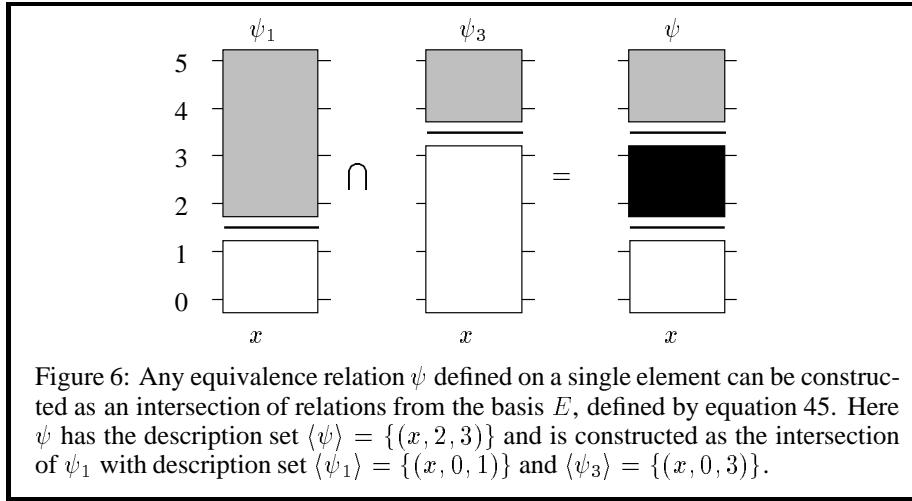
- (*Independence*) A set E of equivalence relations will be said to be *independent* if no one of the relations $\psi \in E$ can be expressed as the intersection of some subset of the others, i.e.

$$\forall \psi \in E \; \nexists \, E_\psi \subset E - \{\psi\} : \bigcap E_\psi = \psi. \tag{46}$$

The set E defined in equation 45 satisfies this condition of independence,[2] as well as completeness.

The purpose of introducing the notion of a complete orthogonal basis for a set of equivalence relations was to generalise the notion of a gene and allow a principle of strict gene transmission to be extended to more general formae. It will be demonstrated below that the weaker notion of a non-orthogonal *independent basis* suffices for the definition of genes, and thus is adequate for the original purpose. Using the same definition of genes and alleles for non-orthogonal bases as for orthogonal bases, (i.e. genes are the basic equivalence relations and alleles are the basic equivalence classes) it is now possible to construct the *random transmitting recombination* operator (RTR) induced by the basis E for Ψ, described by equation 45.

---

[2] A rather minor point which should nevertheless be made in passing is that the formae now being considered violate closure as discussed in Radcliffe (1991a, 1991b): this turns out to be unimportant.

Figure 6: Any equivalence relation ψ defined on a single element can be constructed as an intersection of relations from the basis E, defined by equation 45. Here ψ has the description set $\langle\psi\rangle = \{(x,2,3)\}$ and is constructed as the intersection of ψ_1 with description set $\langle\psi_1\rangle = \{(x,0,1)\}$ and $\langle\psi_3\rangle = \{(x,0,3)\}$.

5.3 Random Transmitting Recombination

The random transmitting recombination operator[3] (RTR) can be defined in a way similar to random respectful recombination (R^3), the difference being that instead of selecting from the entire similarity set of the two parents, RTR selects from that subset of solutions which share every gene with at least one parent. This subset, which for parents A and B is written $\Gamma(A,B)$, is called their *dynastic potential,* and is defined by

$$\Gamma(A,B) = \{C \in \Sigma(A,B) \mid \forall\psi \in E : \psi(C,A) = 1 \text{ or } \psi(C,B) = 1\}. \tag{47}$$

The RTR operator picks each element in the dynastic potential of the parents with equal probability, and both strictly transmits genes (which trivially implies respect) and properly assorts formae, provided that these conditions are compatible. (In the special case of k-ary schemata, RTR is identical to uniform crossover with parameter half.)

5.4 Application to Fixed Size Multisets

The formalism developed above can now be applied to the problem of recombining fixed-size multisets. The similarity set of two solutions (now multisets) is the forma with the description set

$$\langle\Sigma(A,B)\rangle = \left\{(x,N_x^{\downarrow},N_x^{\uparrow}) \mid N_x^{\downarrow} = \min\{m(x,A),m(x,B)\}, \right. \tag{48}$$
$$\left. N_x^{\uparrow} = \max\{m(x,A),m(x,B)\}\right\}$$

This similarity set contains all those multisets of the given fixed size N which have at least as many copies of each element as the parent with fewer copies, and no more than the number held by the parent with more. For example, if the chosen fixed size for the multisets is five, and the universal set $\mathcal{E}$ is given by equation 36, then given

$$A = \{a,a,a,b,c\}, \tag{49}$$

---

[3] previously called *inheritance crossover*

and
$$B = \{\!| a, b, b, c, d |\!\}, \tag{50}$$

the similarity set $\Sigma(A, B)$ is described by

$$\langle \Sigma(A, B) \rangle = \{(a, 1, 3), (b, 1, 2), (c, 1, 1), (d, 0, 1), (e, 0, 0), (f, 0, 0)\}. \tag{51}$$

The similarity set itself thus contains those multisets containing $\{\!| a, b, c |\!\}$ together with exactly two elements from $\{\!| a, a, b, d |\!\}$. For fixed size multisets it happens that, given the equivalence relations discussed, the dynastic potential of any parents A and B is identical to their similarity set. To see this, consider any basic equivalence relation ψ with the description set

$$\langle \psi \rangle = \{(x, 0, n)\}. \tag{52}$$

This has two equivalence classes, described by

$$\langle \xi_\downarrow \rangle = \{(x, 0, n)\} \tag{53}$$

and

$$\langle \xi_\uparrow \rangle = \{(x, n + 1, N^\star)\}. \tag{54}$$

If both parents belong to the same basic forma, then their similarity set is clearly a subset of this forma, so gene transmission imposes no extra requirement. If, however, they belong to different basic formae, then since there are only two of these, the requirement to lie in their union is no restriction at all, because

$$\xi_\downarrow \cup \xi_\uparrow = \mathcal{S}. \tag{55}$$

Thus dynastic potentials for these equivalence relations are indeed identical to similarity sets and so it can be seen that strict gene transmission is in this case no stronger a requirement than respect. In this special case, RTR reduces to R^3. This is not, of course, true for general formae, as is shown by the difference for the k-ary schemata discussed in section 2.8.

Notice that the restriction to multisets of fixed size, as was the case with sets, ensures non-separability, so that RTR and R^3 are unable to assort the formae discussed. This can be seen by from the same example as was used in section 2.1 to demonstrate that formae for fixed size sets are non-separable, re-interpreting the sets as multisets. This weakness is discussed further in section 7.

6   Recombining Variable-Size Multisets

Variable-size multisets can be dealt with simply by relaxing the constraint of fixed size as discussed in the previous section. The formae then arrived at are separable and random transmitting recombination (which is in this special case again identical to R^3) not only properly assorts and respects the formae, but also strictly transmits genes.

In summary, R^3/RTR for variable-size multisets simply inserts a number of copies of each element from the universal set which is bounded by the number of copies in the two parents, and in doing so strictly transmits genes and properly assorts the formae induced by the equivalence relations generated by the independent basis of equation 45.

7 Assorting Non-Separable Formae

When there are non-separable formae, such as those discussed in this paper, the question arises as to which among respect and proper assortment should be given priority. R^3 and RTR both give priority to respect, but arguably proper assortment, which embodies the exploratory power of the search, is more important.

The present section shows how it is possible to trade off the degree of violation of respect with the "thoroughness" of assortment. (Here thoroughness may be taken to mean the likelihood of generating the solutions required by assortment.) Recall that proper assortment only requires that the probability of generating a child in the intersection of a given pair of formae to which the parents belong be finite. (Equation 5 says that there must be some control parameter which allows generation of a solution in the intersection, but does not specify any required density of such parameters.) It is therefore technically possible to guarantee proper assortment by defining an operator which respects the given formae with a very high probability $1 - \varepsilon$, but with low probability ε randomly selects a solution which violates respect. (Indeed, mutation could be viewed as performing this rôle in combination with a respectful recombination operator.) More narrowly, the choice of solution outside the similarity set (the set of children allowed by respect) could be restricted to those which are required to be capable of being generated by assortment. The parameter ε can be viewed as controlling the degree to which respect is violated. While any non-zero value for ε technically guarantees assortment, clearly the larger the value, the more thorough will be the assortment.

7.1 Random Assorting Recombination

The following *random assorting recombination* operator (RAR) uses these ideas to ensure proper assortment by sacrificing respect (and, by implication, gene transmission) in a controlled way. This operator takes a positive integer parameter w (for "weight") which is like an inverse of the parameter ε discussed above.

1. Place w copies of each allele present in both parents in a bag (multiset).
2. Add to the bag one copy of each allele present in only one parent.
3. Repeatedly draw alleles from the bag without replacement. Whenever it is possible to add the allele to the child being formed, do so; otherwise discard the allele. Continue until the bag is empty or the child is fully specified, i.e. until a basic forma (allele) has been chosen for every one of the basic equivalence relations (genes).
4. If the child is not fully specified at the end of this process, assign alleles to any remaining genes at random, from among the remaining legal values.

A number of observations should be made about this operator.

- The operator is general, and can be applied to any problem in which genes are properly specified in the sense used in this paper, provided that their number is finite.
- If the formae are induced by a set of equivalence relations with an orthogonal basis, this operator will separate them, and reduces to RTR. To see this, simply observe that the only circumstances in which it would not be possible to add an allele drawn from the bag to the child would be those in which a different allele for the same gene had already be chosen from the other parent.

- If the formae are separable, RAR will separate them. To see this, observe that if a gene is common to both parents, only one allele will be placed in the bag for that gene (albeit w times). In the initial phase of RAR's operation, only combinations of alleles which assortment requires will ever be included in the child. If, therefore, assortment is compatible with respect, this cannot compromise the ability to include the shared alleles in the child, as required by respect.

- The action of the operator is only unusual, therefore, when the formae are non-separable. In other cases, the value of w is irrelevant. With non-separable formae, the higher the parameter w is set, the greater will be the degree of respect which RAR achieves and the less thorough will be the assortment. This is because the higher w is set, the more likely is it that alleles common to the two parents will be drawn early, so that incompatibilities between respect and assortment will be more likely to be concluded in favour of respect. General guidance as to the appropriate value for w probably requires experimental evidence, though $w = 2$ has aesthetic appeal because filling the bag then amounts merely taking every allele from each parent.

- Finally, the account of the RAR given describes its theoretical application. In most circumstances it will be possible to find implementations which are very much more efficient than that described, but which replicate its behaviour identically.

7.2 Example of Random Transmitting Recombination

Consider again the example used in section 4.1 to show that fixed-size sets are non-separable. Recall that $\mathcal{E} = \{\, a, b, c, d, e, f \,\}$, the fixed size is 3 and the parents are $A = \{\, a, b, c \,\}$ and $B = \{\, a, b, d \,\}$. The full genetic description of A is now the singleton forma

$$\xi_{abc\bar{d}\bar{e}\bar{f}} = \{A\}. \tag{56}$$

Similarly, B is given by

$$\xi_{ab\bar{c}d\bar{e}\bar{f}} = \{B\}. \tag{57}$$

Taking $w = 3$, the bag will initially be filled as follows:

$$\{a, a, a, b, b, b, c, \bar{c}, d, \bar{d}, \bar{e}, \bar{e}, \bar{e}, \bar{f}, \bar{f}, \bar{f}\}. \tag{58}$$

Clearly drawing out a solution containing $\{\, c, d \,\}$ is now possible in a number of ways, but is not particularly likely because a and b are each three times as likely to get drawn out at each stage as are c and d. Notice that, just as there is a small possibility of being unable to include a and b, there is a possibility of having to include, say, f, which is present in neither parent. This would happen, for example, if the first three alleles drawn happened to be $\bar{c}, \bar{d}$ and $\bar{e}$. In this case, when the $\bar{f}$ allele was chosen it would be discarded because it is impossible for a solution to this problem to omit c, d, e and f, and in the final stage an e allele would necessarily be introduced. This slightly counterintuitive behaviour is required by assortment, and again becomes ever more unlikely as w is increased.

7.3 Forma Disruption

Depending partly on the weight used, random assorting recombination can be a fairly disruptive operator. As with other recombination operators, the amount disruption can be reduced by biasing the operator to take genes preferentially from one parent (Syswerda, 1989, Spears & DeJong, 1991). This may well prove to be sensible in the current context. It may further be appropriate to introduce linkage and reordering operators, though these are out of vogue. Such extensions will not, however, be discussed in detail in this paper.

8 Summary

This paper has shown how forma analysis can be used to construct operators for principled recombination of sets and multisets. A single operator, defined in generic terms and called *random assorting recombination* (RAR), suffices for this. Previous work had introduced the notion of a complete orthogonal basis for a set of equivalence relations and shown that when such a basis exists a linear chromosome can be constructed which uniform crossover can manipulate effectively in the sense that it transmits and properly assorts genes. This paper has demonstrated that in cases where no orthogonal basis exists it may still be possible to construct a non-orthogonal basis. Such a basis suffices for the definition of genes, but these cannot be independently assigned because some combinations of alleles will be illegal. Moreover, in some cases it will be possible to transmit and properly assort genes forming a non-orthogonal basis, but conventional operators will be unable to do this. The RAR operator is guaranteed to transmit and properly assort genes whenever this is possible. The paper has further shown that when respect and assortment are incompatible, it is possible to parameterise the degree of violation of respect and to use this to control the thoroughness of assortment. The random assorting recombination operator provides a convenient and general mechanism for doing this.

References

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

(Radcliffe, 1990) Nicholas J. Radcliffe. *Genetic Neural Networks on MIMD Computers*. PhD thesis, University of Edinburgh, 1990.

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

(Radcliffe, 1991b) Nicholas J. Radcliffe. Forma analysis and random respectful recombination. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 222–229. Morgan Kaufmann (San Mateo), 1991.

(Radcliffe, 1992) Nicholas J. Radcliffe. Genetic set recombination and its application to neural network topology optimisation. *Neural Computing and Applications*, 1(1), 1992.

(Spears and De Jong, 1991) William M. Spears and Kenneth A. De Jong. On the virtues of parameterised uniform crossover. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 230–236. Morgan Kaufmann (San Mateo), 1991.

(Syswerda, 1989) Gilbert Syswerda. Uniform crossover in genetic algorithms. In *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann (San Mateo), 1989.

(Vose and Liepins, 1991) Michael D. Vose and Gunar E. Liepins. Schema disruption. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 237–243. Morgan Kaufmann (San Mateo), 1991.

(Whitley, 1987) Darrell Whitley. Using reproductive evaluation to improve genetic search and heuristic discovery. In *Proceedings of the Second International Conference on Genetic Algorithms*. Lawrence Erlbaum Associates (Hillsdale), 1987.