Epistasis in Genetic Algorithms: An Experimental Design Perspective *

Colin R. Reeves and Christine C. Wright
School of Mathematical and Information Sciences
Coventry University
UK
Email: CRReeves@coventry.ac.uk

Abstract

In an earlier paper we examined the relationship between genetic algorithms (GAs) and traditional methods of *experimental design*. This was motivated by an investigation into the problems caused by epistasis in the implementation and application of GAs to optimization problems. We showed how this viewpoint enables us to gain further insights into the determination of epistatic effects, and into the value of different forms of encoding a problem for a GA solution. We also demonstrated the equivalence of this approach to *Walsh transform* analysis.

In this paper we consider further the question of whether the epistasis metric actually gives a good prediction of the ease or difficulty of solution of a given problem by a GA. Our original analysis assumed, as does the rest of the related literature, knowledge of the complete solution space. In practice, we only ever sample a fraction of all possible solutions, and this raises significant questions which are the subject of the second part of this paper. In order to analyse these questions, we introduce the concept of *alias sets*, and conclude by discussing some implications for the traditional understanding of how GAs work.

1 Introduction

In an earlier paper [1], we introduced the experimental design (ED) decomposition model as a useful perspective for the analysis of genetic algorithms (GAs). However, the necessarily expository nature of that paper meant that we were not able fully to explore the

---

*Published in L.J.Eshelman (Ed.) (1995) *Proceedings of the 6th International Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo, CA.

value of the proposed approach to the measurement of epistasis, and we intend to return to this subject in greater depth in the current article.

In order to make this paper self-contained for readers who have not seen [1], we repeat the basic ED model here. For those needing more general information on the field of experimental design, a very comprehensive introduction can be found in Hinkelmann and Kempthorne [2].

We assume that we have populations of binary strings $\{S\}$ of length l, and that the fitness of string S is denoted by $v(S)$. We use the term *Universe* to denote the set of all possible 2^l strings, and reserve the use of the term *population* for the sense in which it is commonly used in the GA community.

The idea of assuming an underlying linear model (defined on the bits) for the fitness of a string is implicit in several studies of GAs. Davidor [3, 4] for example, did so in his attempt to define measures of epistasis—a study which we dealt with in some detail in the first paper and will return to again here. Assuming no epistasis, we can write such a model as

$$v(S) = \text{constant} + \sum_{i=1}^{l} (\text{effect of allele at gene } i),$$

while at the other extreme, we can express the full epistatic model as

$$
\begin{aligned}
v(S) =\ & \text{constant} + \sum_{i=1}^{l} (\text{effect of allele at gene } i) \\
& + \sum_{i=1}^{l-1} \sum_{j=i+1}^{l} (\text{interaction of alleles at genes } i \text{ and } j) \\
& + \dots \\
& + (\text{interaction of alleles at genes } 1,\dots,l)
\end{aligned}
$$

In conventional experimental design, the above model would actually be written in parametric form, and would also allow for the possibility of random error.

For example, the model for a string of 3 bits could be written as follows:

$$
\begin{aligned}
v_{pqrs} =~& \mu + \alpha_p + \beta_q + \gamma_r + (\alpha\beta)_{pq} + (\alpha\gamma)_{pr} + (\beta\gamma)_{qr} \\
& + (\alpha\beta\gamma)_{pqr} + \varepsilon_{pqrs}
\end{aligned}
$$

where v_{pqrs} is the fitness of the string (p, q, r), and the subscript s denotes the replication number (i.e. the sth occurrence of the string). If there is no intrinsic random error, we can of course drop the final term (and of course the subscript s). The parameters on the right-hand side are as follows:

μ	average fitness
α_p	effect of allele p at gene 1
β_q	effect of allele q at gene 2
γ_r	effect of allele r at gene 3
$(\alpha\beta)_{pq}$	joint effect of allele p at gene 1 and allele q at gene 2
$(\alpha\gamma)_{pr}$	joint effect of allele p at gene 1 and allele r at gene 3
$(\beta\gamma)_{qr}$	joint effect of allele q at gene 2 and allele r at gene 3
$(\alpha\beta\gamma)_{pqr}$	joint effect of allele p at gene 1, allele q at gene 2 and allele r at gene 3
ε_{pqrs}	random error for replication s of string (p, q, r)

Davidor assumes zero random error, which is reasonable in many, although not all, applications of GAs, and we shall follow suit.

In the first paper we also followed Davidor in assuming that we knew the fitness of every one of the whole universe of strings, in order to present the basic ED approach in as uncomplicated a manner as possible. However, in the real situation, the various quantities proposed by Davidor for obtaining an epistasis metric are only *estimates* of parameters. In fact, not only are these measures compromised, but (as we shall show) so are the estimates of schema fitness—quantities which are fundamental in the traditional understanding of how GAs work.

1.1 An example

To motivate the arguments, suppose we have a 3-bit string, and the fitness of every string in the Universe is known. There are of course $2^3 = 8$ strings , and therefore 8 fitness values, but the experimental design model above has 27 parameters. It is thus essential to impose some side conditions if these parameters are to be estimated; the usual ones are the obvious constraints that at every order of interaction, the parameters sum to zero for each subscript. This results in an additional 19 independent relationships and thus allows the 'solution' of the above model—in the sense that all the parameter values can be determined if we have observed every one of the 8 possible strings—the

Universe. For example, we find that

$$
\begin{aligned}
\mu &= v_{***} \\
\mu + \alpha_p &= v_{p**} \quad \text{for } p = 0, 1 \\
\mu + \beta_q &= v_{*q*} \quad \text{for } q = 0, 1 \\
\mu + \gamma_r &= v_{**r} \quad \text{for } r = 0, 1
\end{aligned}
$$

where the notation v_{p**}, for instance, means averaging over subscripts q and r. These effects are exactly equivalent to Davidor's 'excess allele values', while his 'excess genic values' are found by summing α_p, β_q and γ_r for each possible combination of p, q, r. Finally, his 'string genic value' is clearly

$$
\mu + \alpha_p + \beta_q + \gamma_r.
$$

The difference between the actual value and the genic value, $\epsilon(S)$, is therefore simply the sum of all the interaction terms; putting it another way, zero epistasis is seen to be equivalent to having no interactions in the model.

In [1] we showed how this information can be obtained by the well-known statistical method of 'Analysis of Variance' (Anova), whereby the variability of the fitness values (measured by sums of squared deviations from mean fitness, and denoted by SS) is partitioned into orthogonal components from identifiable sources. Associated with these SS, are the *degrees of freedom*—the number of independent elements in the associated SS.

It is well-known (and easy to prove) that

Total SS = Main effects SS + Interaction SS

and these values Davidor simply divides by a constant to obtain his 'variances'. Thus for the Universe, it is hardly surprising to find that

Total 'variance' = Genic 'variance'+Epistasis 'variance'.

However, when Davidor examined the case of a sample, this result *appeared* no longer to be true; in particular, some of the 'variances' turned out to be negative. Later we shall show why this occurs, and how his analysis would have to be modified in order to retain the additivity of the variances.

2 The Measurement of Epistasis

Given the partitioning of the variances in the above way, an obvious metric for the degree of epistasis in a problem is to express the Interaction SS as a percentage of the Total SS. The four 3-bit functions Davidor used were analysed in [1], and using this metric epistasis varied from 0% for the case of a linear function to 93% for a deceptive function. The first question that arises is naturally whether this metric is useful and meaningful in the sense that it relates to the likely degree of difficulty for solving the problem by means of a GA.

In order to explore this question further, we carried out some experiments using NK-landscapes as described by Manderick *et al.* [5]. Varying the parameter K for these functions has the effect of changing the amount of epistasis, and in [5] it is suggested that a GA will find problems with high values of K more difficult than low values. Table 1 shows the results of some experiments for $N = 9$ and $K = 1, 2, 3, 4$ (10 experiments for each K), where the interaction SS has been measured as a percentage of the total SS.

Table 1: Values of the epistasis metric for 10 $9K$-landscapes

K	Mean	Standard deviation
1	40.2	9.3
2	62.5	13.5
3	77.6	8.6
4	85.5	6.4

As a contrasting example, Grefenstette's 'easy deceptive' function [6, p.81] measured only 7.2% using the epistasis metric. From these results it would seem a reasonable deduction that using this metric leads to a means of detecting epistatic problems, and therefore a method of determining the degree of difficulty facing a GA.

However, before leaping to this conclusion, we need to examine the nature of interaction effects more carefully. Figure 1 gives a pictorial representation of interaction in the case of two genes A and B.

What this illustrates is as follows: in the upper diagram, the best allele for each gene is 1, and while there is epistasis, in that the joint effect of having the alleles of both A and B set at 1 exceeds the sum of the individual main effects, its influence is benign since the interaction reinforces the main effects. However, in the lower diagram, the interaction has a malign influence: the best allele for both A and B is 1, but overall it is better to set gene A at 0. From a traditional GA perspective, there is clearly an element of deception about the second case: the schema average v_{1*} exceeds v_{0*}, but $v_{01} > v_{11}$.

In terms of the actual values of the effects, the first situation corresponds to the case where the interaction effect $(\alpha\beta)_{01}$ has the same sign as the main effect α_0, and the second to the case where the signs are different. In fact it is easily seen from the conditions in the second case ($v_{1*} > v_{0*}$ and $v_{01} > v_{11}$) that

$$(\alpha\beta)_{01} > -\alpha_0.$$

It is natural to ask therefore, whether the existence of effects of different sign is an important influence on the epistasis of a particular problem. The answer is: maybe!

First, we should realize that the magnitude of the interaction (relative to its associated main effects) is also
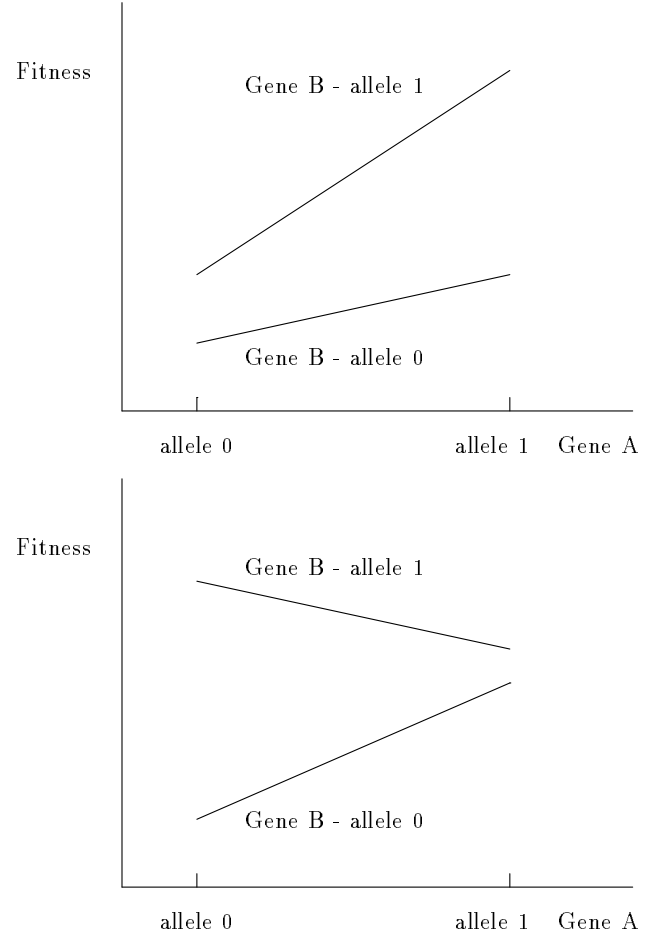

Figure 1: Benign and malign interactions

important. In Figure 2 the effects are still of opposite sign, but there will be little difficulty in finding the best combination because the best allele for gene A is now 0; in terms of schema averages, $v_{1*} < v_{0*}$.

Secondly, the *order* of the interaction terms is also relevant. The side conditions on the effects result in the constraint that for the case of 2-gene interactions,

$$(\alpha\beta)_{11} = (\alpha\beta)_{00} = -(\alpha\beta)_{01} = -(\alpha\beta)_{10}.$$

However, for a 3-gene interaction $(\alpha\beta\gamma)_{pqr}$ say, the relationships for different values of p, q, r are

$$(\alpha\beta\gamma)_{111} = (\alpha\beta\gamma)_{100} = (\alpha\beta\gamma)_{010} = (\alpha\beta\gamma)_{001} =$$
$$-(\alpha\beta\gamma)_{110} = -(\alpha\beta\gamma)_{101} = -(\alpha\beta\gamma)_{011} = -(\alpha\beta\gamma)_{000}.$$

On applying these relationships to a particular situation, it becomes clear that it may be the combined effect of all interactions up to a particular order that determines the difficulty of solving a particular problem. For example, in a 3-bit problem where we have $v_{0**} < v_{1**}$, but $v_{011} > v_{111}$, the conditions can be shown to reduce to

$$(\alpha\beta)_{01} + (\alpha\gamma)_{01} + (\alpha\beta\gamma)_{011} > -\alpha_0$$
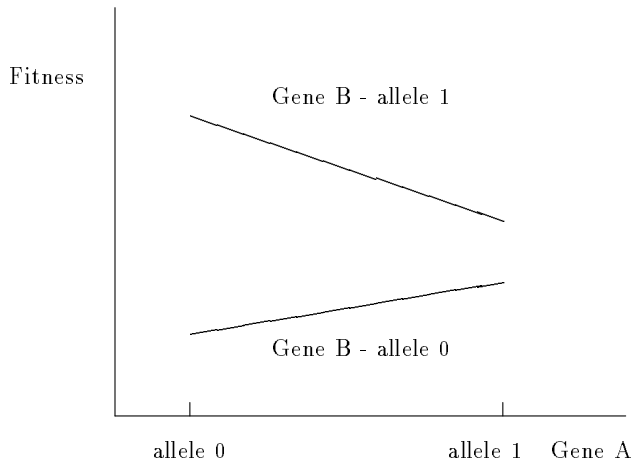
Figure 2: Relatively small interaction

whereas in the case $v_{001} > v_{111}$, they become

$$(\alpha\gamma)_{01} + (\beta\gamma)_{01} > -(\alpha_0 + \beta_0).$$

Both cases are epistatic to some degree, but in the second case the 3-gene interaction (no matter how large it is) is irrelevant.

However, if this analysis is followed through to the calculation of the Sums of Squares, it becomes clear that after cancelling some factors out, we are left with terms in the *square* of the effect, so that positive interaction effects cannot be distinguished from negative ones. For example, in the case of a 3-bit binary string, the SS due to the first-order interaction term between genes 1 and 2 is (see [1] for details)

$$\sum_p \sum_q \sum_r (v_{pq*} - v_{p**} - v_{*q*} + v_{***})^2$$

which in virtue of the side conditions reduces to

$$2\sum_p \sum_q (\alpha\beta)_{pq}^2.$$

Thus Davidor's variance metric will have the same value for functions that are actually quite different in terms of their difficulty of solution.

In summary, we see that the existence of large interaction Sums of Squares may be an important indicator of epistasis in some cases but not in others. The problem is that we cannot tell the difference simply from the SS, and we need the auxiliary information about the magnitude and sign of the effects to get a clearer picture of the difficulty of a particular problem.

In [1] we showed that Goldberg's 3-bit deceptive function [7] is characterized by a large (negative) 3-gene interaction, and by conditions that imply combinations of interactions must have a net effect greater than the main effects. This can be extended to longer strings, and in general it can be shown that a class of hard

deceptive functions may be generated by a large high-order interaction term of an appropriate sign, as in the analysis by Homaifar *et al.* [8] for example. (In [8] Walsh functions are used; that these are equivalent to ED is shown in [1].) Other hard but non-deceptive functions such as that described by Grefenstette [6] can also be shown to have large interaction terms, so it would seem that if we could establish the existence or otherwise of substantial high-order interaction effects, it would certainly be a useful indicator of problem difficulty.

3 Making Sense of Partial Information

The next question to consider is how this auxiliary information can be obtained in practice. A GA population of strings is really only a sample of the Universe of chromosomes, but this fact has not so far been taken into account in examining the measurement of epistasis.

Suppose, using Davidor's 3-bit examples again, that we actually observed the strings in table 2. These are *half-fractions* of the Universe; the first one (F_1) is balanced but F_2 and F_3 are not—we can see that there are 2 occurrences of each allele in the first case, but in the second case only allele 0 is instantiated at gene 3, while F_3 has a different frequency of occurrences of alleles 0 and 1 in both gene 1 and 3.

Table 2: Some half-fractions of the Universe

fitness value	String			
	Universe	F_1	F_2	F_3
v_1	0 0 0	0 0 0	0 0 0	
v_2	0 0 1	0 0 1		0 0 1
v_3	0 1 0		0 1 0	0 1 0
v_4	0 1 1			0 1 1
v_5	1 0 0		1 0 0	
v_6	1 0 1			1 0 1
v_7	1 1 0	1 1 0	1 1 0	
v_8	1 1 1	1 1 1		

We introduce here the experimental design concept of a *contrast*, usually denoted by upper case Roman letters. For example, the contrast

$$A = \alpha_1 - \alpha_0$$

(where α_p is as previously defined), expresses the average fitness value when allele 1 is instantiated at gene 1, compared to the instantiation of allele 0. (Since $\alpha_1 + \alpha_0 = 0$, the contrast A is readily seen to be just twice the value of the main effect α_1.) In general, any linear combination $\sum c_i v_i$ of the fitness values (with $\{c_i\}$ a set of constants) is a contrast, but only a few of them have any sensible meaning (see [2] for further details).

In terms of the vector of fitness values **v** in the Universe, the contrasts which relate to the main effects

are

$$A = \frac{1}{4}(-1, -1, -1, -1, 1, 1, 1, 1)\mathbf{v}$$

$$B = \frac{1}{4}(-1, -1, 1, 1, -1, -1, 1, 1)\mathbf{v}$$

$$C = \frac{1}{4}(-1, 1, -1, 1, -1, 1, -1, 1)\mathbf{v}.$$

Similarly, we can define contrasts relating to the interaction effects, so that

$$AB = \frac{1}{4}(1, 1, -1, -1, -1, -1, 1, 1)\mathbf{v}$$

expresses the average fitness value for cases where the instantiated alleles at genes 1 and 2 are the same, compared to those where they are different. The other contrasts are as follows:

$$AC = \frac{1}{4}(1, -1, 1, -1, -1, 1, -1, 1)\mathbf{v}$$

$$BC = \frac{1}{4}(1, -1, -1, 1, 1, -1, -1, 1)\mathbf{v}$$

$$ABC = \frac{1}{4}(-1, 1, 1, -1, 1, -1, -1, 1)\mathbf{v}.$$

This particular set of 7 contrasts constitute an orthogonal partitioning of the information available in the 8 fitness values. In what follows, we will refer to this set, with the addition of the mean

$$\mu = \frac{1}{8}(1, 1, 1, 1, 1, 1, 1, 1)\mathbf{v}$$

as *basic contrasts*.

When the Universe is known, all these can be computed and the existence (or otherwise) of epistasis identified. (In fact, this is usually still possible even when there is some random error in the fitness evaluation function.) However, when only a fraction is available, some interesting problems arise.

Suppose we have the first fraction F_1 as shown above, so that in terms of the fitness values, we know only (v_1, v_2, v_7, v_8). Using only the *available* information to calculate estimates of the basic contrasts, it would appear natural to estimate the contrasts as follows:

$$\widehat{A} = \frac{1}{2}(-1, -1, 0, 0, 0, 0, 1, 1)\mathbf{v}$$

$$\widehat{B} = \frac{1}{2}(-1, -1, 0, 0, 0, 0, 1, 1)\mathbf{v}$$

$$\widehat{C} = \frac{1}{2}(-1, 1, 0, 0, 0, 0, -1, 1)\mathbf{v}.$$

$$\widehat{AB} = \frac{1}{4}(1, 1, 0, 0, 0, 0, 1, 1)\mathbf{v}$$

$$\widehat{AC} = \frac{1}{2}(1, -1, 0, 0, 0, 0, -1, 1)\mathbf{v}$$

$$\widehat{BC} = \frac{1}{2}(1, -1, 0, 0, 0, 0, -1, 1)\mathbf{v}$$

$$\widehat{ABC} = \frac{1}{2}(-1, 1, 0, 0, 0, 0, -1, 1)\mathbf{v}.$$

From this it is clear that

$$\widehat{A} = \widehat{B}, \widehat{C} = \widehat{ABC}, \widehat{AC} = \widehat{BC}, \text{ and } \widehat{AB} = \widehat{\mu}.$$

In terms of the basic contrasts, we see that (for example)

$$(A + B) = \frac{1}{2}(-1, -1, 0, 0, 0, 0, 1, 1)\mathbf{v}$$

so that

$$\widehat{A} = \widehat{B} = (A + B).$$

In other words, we cannot estimate A and B separately from F_1, but only their sum. The other orthogonal combinations of the basic contrasts that can be computed from this fraction are

$$(C + ABC), (AC + BC) \text{ and } (AB + \mu).$$

These pairs of indistinguishable contrasts are known as *alias sets*, and each set is associated with 1 degree of freedom, corresponding to the 3 degrees of freedom associated with having 4 fitness values. The contrast which is aliased with μ (AB in the above instance) is known as the *defining contrast* for this half-fraction. By choosing a different defining contrast it is possible to generate a different half-fraction.

It may be impossible to discern from this fraction whether there are any epistasis effects, since, for instance, ABC—the 3-gene interaction term—is indistinguishable from C. Thus, if an Anova table indicates a significant source of variation due to the orthogonal contrast $(C + ABC)$, we cannot tell whether this is because the fitness is a linear function with a high contribution from gene 3, or whether it is because the function is epistatic with a 3-gene interaction. Conversely, if the table suggests that there is no variation due to this contrast, it might simply be because the effect of ABC is in the opposite direction to C.

Table 3: Anova results on F_1 for Davidor's functions

	f_1		f_2	
Source	df	SS	df	SS
$(A + B)$	1	36.00	1	196.00
$(C + ABC)$	1	1.00	1	196.00
$(AC + BC)$	1	0.00	1	196.00
Total	3	37.00	3	588.00
	f_3		f_4	
Source	df	SS	df	SS
$(A + B)$	1	100.00	1	4.00
$(C + ABC)$	1	56.25	1	9.00
$(AC + BC)$	1	49.00	1	25.00
Total	3	205.25	3	38.00

As an example, we present in table 3 the Analysis of Variance for Davidor's functions f_1, f_2, f_3, f_4. It might still seem reasonable to conclude that f_1 is *not* epistatic, and that f_4 *is*, but it would be necessary to bear in mind the possibility that C and ABC, and AC

and BC, have cancelled each other out in the case of f_1, and that A and B have in the case of f_4. It is hard to make any positive statement about f_2 and f_3.

We can now see that Davidor's attempt to estimate epistasis variance from a fraction is fatally flawed, precisely because of these alias sets. As we noted earlier, using the fraction F_1 for example, negative 'variances' were obtained in [4, Table 8]. This occurs because, for instance, the 'genic values' for *both* genes 1 and 2 (i.e. the contrasts A and B) are included in the 'variance' calculation, when it is not actually possible to compute them both simultaneously from this fraction.

Other fractions will give rise to different aliasing structures; for instance if we use the fraction F_2, we obtain the Anova table shown in Table 4, where the estimable combinations are

$$(A - AC); (B - BC); (AB - ABC); (-C + \mu).$$

Table 4: Anova results on F_2 for Davidor's functions

Source	f_1 df	f_1 SS	f_2 df	f_2 SS
$(A - AC)$	1	16.00	1	0.00
$(B - BC)$	1	4.00	1	0.00
$(AB - ABC)$	1	0.00	1	0.00
Total	3	20.00	3	0.00
Source	f_3 df	f_3 SS	f_4 df	f_4 SS
$(A - AC)$	1	4.00	1	20.25
$(B - BC)$	1	1.00	1	6.25
$(AB - ABC)$	1	0.00	1	0.25
Total	3	5.00	3	26.75

Here the confusion is yet greater: it is impossible even to make a reasonable guess as to whether any epistasis exists or not, since all the main effects are aliased with interactions. In the case of f_2 no variation has been detected at all. This is not surprising since f_2 has a solitary spike at 111 (not one of the points sampled), but the same result could have been obtained for quite different reasons.

This fraction was at least balanced across genes 1 and 2; in the case of fraction F_3, the confusion is worse. In cases like this, the estimates of contrasts need to be defined carefully if they are to be unbiased. For example, a 'natural' estimate of A would be

$$\widehat{A} = \frac{1}{3}\{0, -1, -1, -1, 0, 3, 0, 0\}\mathbf{v}.$$

However, these estimates are now complex linear combinations of the basic contrasts, and it becomes almost impossible to judge the degree of epistasis from an ANOVA table. In principle, it is possible to find expressions for the estimated contrasts in terms of the basic contrasts by expressing both as linear combinations of the fitness values v_i and equating coefficients.

For example, an expression for $\widehat{A}$ is

$$\widehat{A} = \frac{1}{3}\{3A - 2B - AB + C + 2AC - BC - 2ABC\}.$$

The results for other contrasts all display a similarly complicated structure. Nor is the situation necessarily any better if we increase the size of the population. If we take a 3/4 fraction such as $(v_1, v_2, v_3, v_5, v_6, v_7)$ and $\widehat{A} = \frac{1}{3}(-1, -1, -1, 0, 1, 1, 1, 0)\mathbf{v}$, we find

$$\widehat{A} = \frac{1}{3}(3A - AB - AC - ABC).$$

In such situations it becomes impossible to determine whether an apparent effect is really caused by the gene (or interaction of genes) with which it is ostensibly related.

Thus, any epistasis measure that relies on a decomposition into main and interaction effects must be treated with great caution. In the first place, as we have shown in section 2, Davidor's 'epistasis variance' cannot be interpreted as necessarily indicating the level of difficulty without the auxiliary information of the sign and magnitude of the effects. This was observed when we attempted to measure the epistasis of the same NK-landscapes investigated in section 2 using different orthogonal fractions of the Universe. For larger values of K, the epistasis metric was fairly consistently large, but for $K = 1$ the value of the metric varied considerably from one fraction to another, even when a half-fraction was used. There would be a considerable danger of concluding that a relatively easy problem with no interactions at higher than 2-gene order was actually quite difficult.

Secondly, as we have seen in this section, without knowing the Universe, the auxiliary information cannot be obtained. Those NK-landscapes that gave a high value on the variance metric could not be confirmed as difficult because the aliasing prevented the proper identification of the effects. Since we have shown in [1] that the ED approach is equivalent to the Walsh transform analysis popularized by Goldberg [7, 9], we should point out that this conclusion applies equally to methods based on Walsh functions as to the approach of Davidor which has been given prominence in this paper.

4 Further Implications

The implications for epistasis measurement are clear, but this also has some relevance to ideas of schema processing or hyperplane sampling. The 'traditional' GA interpretation uses the idea of hyperplane competitions to explain its operation, as in [10], for instance. It should be evident from the above that a hyperplane competition is simply a contrast—e.g. the competition between the hyperplanes $(1**)$ and $(0**)$ can be expressed by the contrast $A = \alpha_1 - \alpha_0$.

When we have partial information, our estimate of this contrast is aliased with many others. As an example, the hyperplane $(1 * *)$ could 'win' a competition between 3-bit chromosomes (i.e. $\widehat{A} > 0$) simply because in a particular population A is aliased with BC and there is a large interaction between genes 2 and 3. In practice GA populations are not chosen in any systematic way, so that there will not even be a clear-cut choice within disjoint alias sets, but rather a 'mess' of alternative explanations for the phenomenon observed.

The latter is the main reason why ED uses balanced orthogonal fractions which enable the aliasing to be controlled. Typically an *a priori* decision is made to ignore certain interactions (usually high-order ones), using problem-specific information or otherwise. Then a fraction of the Universe is chosen in a way that ensures that factors (in GA terminology, genes) which are potentially important in explaining the data are aliased, not with each other, but with the 'unimportant' interactions. Often, analysis of this first fraction leads to certain hypotheses which can be tested by the evaluation of different fractions, without forgetting the data that have already been collected. This sequential procedure may quickly lead to a position where the levels (in GA terminology, alleles) of the important main factors can be fixed with some confidence, so that further experiments would be needed only to refine the values of some of the less important ones.

This is reminiscent of the standard explanation of what a GA does, although a GA proceeds without any human interpretation and intervention. It is of course this latter characteristic which is one of the main attractions of genetic search. Experimental design methods are not usually applied to problems with large numbers of factors (genes) and/or levels (alleles), although simple designs based on Hadamard matrices are known for large problems, precisely because of the considerable human input that is needed.

However, the implication of the above analysis is that GAs may distribute their search effort in an inefficient way. Firstly, by using random populations, a genetic search is bound to create complicated alias sets. In contrast, ED uses any prior knowledge that is available to create 'clean' sets which can extract the maximum useful information from the data.

Secondly, a GA forgets the results of previous trials as it progresses, at the same time as the population becomes less and less diverse. The effect of this is further to complicate the structure of the alias sets; in the limit, if the population converges to multiple copies of a single string, all effects are aliased with each other, and the population contains no useful information at all about the problem as a whole. Again, the ED approach is different; by making use of all the information gathered over the course of the experiment, it can identify the important effects fairly rapidly and effi-

ciently.

5 Conclusions

We have used the experimental design framework to analyse in detail Davidor's suggested epistasis metric, which we have shown to be equivalent to Analysis of Variance (Anova) for the Universe. We have demonstrated that while it may give some guidance as to the likely difficulty of a given problem, we cannot draw unequivocal conclusions from it, *even when the whole Universe is used*, unless we also compute the actual ED effects themselves. As to do the latter requires more computation than simply to enumerate the Universe, we need to consider the question of whether anything useful can be deduced from a sample (a GA 'population'). This requires an understanding of the underlying alias structures induced by a particular choice of population; the lack of this understanding explains the apparent failure of Davidor's metric in [4]. We then showed that even if populations are chosen in a controlled way in order to produce disjoint alias sets, it becomes very difficult to determine whether any epistasis could be present (using the Anova table) or to estimate the influence of a single gene, unless further assumptions can be made as to the likely maximum order of interaction effects.

We also discussed some of the implications of this analysis for the traditional interpretation of GA operations, and made some comparisons between the GA and ED approaches. In conclusion, we would remark that the ideal algorithm would be one which was able to automate the decisions made in a typical ED, so that at any stage the 'best' point to evaluate next could be chosen in the light of all the information available. However, as pointed out above, such automation is difficult. The GA could be viewed as a step in this direction, and despite its inefficiencies, the weight of empirical evidence suggests that it often does fairly well.

A possible explanation for this is that many of the problems to which GAs have been applied are not all that epistatic, or that the epistasis is of the benign 're-inforcing' variety. For instance, Das and Whitley [11] argue that many problems can be 'solved' by sequentially solving the order-1 hyperplane competitions, i.e. estimating the main effects one at a time by randomly sampling from the Universe, and suggest that a GA may well be implicitly using such a strategy. If this is true, in light of our analysis, we conjecture that the GA recognizes the 'true' contrast in a particular alias set (or equivalently, the 'true' main effect) because most of the others are negligible in comparison. We have also observed empirically, that as larger samples are taken, the coefficient of the 'true' basic contrast in the alias set often becomes larger relative to the coefficients of the others, so that even where the others are not neg-

ligible the 'true' one receives sufficient weight to be recognized.

We would point out that in such situations, the results of such a strategy could be achieved far more efficiently by using an orthogonal fractional design to estimate the main effects simultaneously; for instance, we have 'solved' Das and Whitley's quadratic problem [11, p.168-9] using a 1/16 orthogonal fraction—requiring far fewer fitness evaluations than their sequential approach. Of course in many applications, fitness evaluation is relatively cheap, but for cases where this is not so, ED could be usefully employed in order to focus the genetic search. In further work, we hope to explore such possibilities in the context of a comparison of experimental design methods and GAs in solving some real engineering design problems.

References

[1] C.R.Reeves and C.C.Wright (1995) An experimental design perspective on genetic algorithms. In D.Whitley and M.Vose (Eds.) (1995) *Foundations of Genetic Algorithms 3*, Morgan Kaufmann, San Mateo, CA.

[2] K.Hinkelmann and O.Kempthorne (1994) *The Design and Analysis of Experiments*. John Wiley & Sons, New York.

[3] Y.Davidor (1990) Epistasis variance: suitability of a representation to genetic algorithms. *Complex Systems*, **4**, 369-383.

[4] Y.Davidor (1991) Epistasis variance: a viewpoint on GA-hardness. In [12], 23-35.

[5] B.Manderick, M.de Weger and P.Spiessens (1991) The genetic algorithm and the structure of the fitness landscape. In [13], 143-150.

[6] J.J.Grefenstette (1993) Deception considered harmful. In [14], 75-91.

[7] D.E.Goldberg (1989) Genetic algorithms and Walsh functions: part II, deception and its analysis. *Complex Systems*, **3**, 153-171.

[8] A.Homaifar, X.Qi and J.Fost (1991) Analysis and design of a general GA deceptive problem. In [13], 196-203.

[9] D.E.Goldberg (1989) Genetic algorithms and Walsh functions: part I, a gentle introduction. *Complex Systems*, **3**, 129-152.

[10] D.Whitley (1991) Fundamental principles of deception in genetic search. In [12], 221-241.

[11] R.Das and D.Whitley (1991) The only challenging problems are deceptive: global search ny solving order-1 hyperplanes. In [13], 166-173.

[12] G.J.E.Rawlins (Ed.) (1991) *Foundations of Genetic Algorithms*. Morgan Kaufmann, San Mateo, CA.

[13] R.K.Belew and L.B.Booker (Eds.) (1991) *Proceedings of 4th International Conference on Genetic Algorithms*. Morgan Kaufmann, San Mateo, CA.

[14] D.Whitley (Ed.) (1993) *Foundations of Genetic Algorithms 2*, Morgan Kaufmann, San Mateo, CA.