

# Can quantitative and population genetics help us understand evolutionary computation?

## ABSTRACT

Even though both population and quantitative genetics, and evolutionary computation, deal with the same questions, they have developed largely independently of each other. I review key results from each field, emphasising those that apply independently of the (usually unknown) relation between genotype and phenotype. The infinitesimal model provides a simple framework for predicting the response of complex traits to selection, which in biology has proved remarkably successful. This allows one to choose the schedule of population sizes and selection intensities that will maximise the response to selection, given that the total number of individuals realised,  $C = \sum_t N_t$  is constrained. This argument shows that for an additive trait, the optimum population size and the maximum possible response are both proportional to  $\sqrt{C}$ .

## Categories and Subject Descriptors

F.2.2 [Analysis of Algorithms and Problem Complexity]: Nonnumerical Algorithms and Problems

## General Terms

Theory

## Keywords

population genetics

## 1. INTRODUCTION

Evolutionary biology began with Darwin's development of the idea of natural selection, by analogy with the artificial selection of domesticated plants and animals. The rediscovery of Mendelian genetics in 1900 led to its eventual synthesis with Darwinian selection; by the 1930s, the population genetics developed by Fisher, Haldane and Wright provided a sophisticated quantitative theory of evolution [31]. Following the elucidation of the physical basis of heredity in the

1960s, Kimura applied stochastic population genetics to the evolution of DNA sequence [18]. In parallel, quantitative genetics developed in the practical context of plant and animal breeding, providing a statistical description of the evolution of complex traits that is largely independent of the genetic details.

It was realised very early that selection provides a powerful and general algorithm for designing complex systems: indeed, this was the central problem which Darwin set out to explain. Fisher [7] presented a simple geometrical model for optimisation in multiple dimensions, which has been widely applied in recent years [28]. Wright [40] introduced the metaphor of the "adaptive landscape", and the theory of a "shifting balance", to understand how evolution might avoid being trapped at local optima. Kimura [16] estimated the amount of information that could be accumulated by selection ( $\sim 10^8$  bits since the Cambrian). Digital computers were first used to simulate evolving populations in the 1950's, and the idea that such simulations could be used to solve computational problems was taken up soon after [8]. This has led to diverse approaches to evolutionary computation: genetic algorithms, genetic programming, artificial life, and so on. Population geneticists also used computer simulation extensively, yet there has been remarkably little interaction between these parallel developments.

Why have the fields developed almost independently? Evolutionary biology is concerned with many issues that are not directly relevant to computation: tracing the actual history of species and genes, understanding the genetic basis of [phenotypic variation and the processes that maintain it, understanding the origin of new species, and so on. Yet, natural selection plays a special role as the sole cause of adaptation, and understanding how selection leads to complex adaptation is central to both evolutionary biology and evolutionary computation.

An important difference between the fields is that evolutionary biologists are resistant to ideas that evolution is in any sense optimal, leading to "progress" or to any necessary increase in complexity (though, see [35]). This is largely a reaction against naive ideas that evolution acts "for the good of the species" (see [39]). Thus, rather than asking what genetic system would be most effective, we ask what in fact evolves - which may in extreme cases even lead to extinction (e.g. [9]). Another difference is that in evolutionary computation, we are free to make up arbitrary rules - retaining the fittest individuals indefinitely, allowing sex between more than two individuals, and changing parameters at will. Nevertheless, actual algorithms tend to use the same basic

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

processes, and as we shall see, diverse implementations are often equivalent. The lack of communication between evolutionary biology and evolutionary computation may simply reflect the human tendency to form separate communities; indeed, both fields are fragmented into many subfields, grouped around different journals and conferences. The arguments set out here suggest that recombination between fields as well as subfields would be beneficial.

Evolutionary biology and evolutionary computation share the same fundamental questions, and can learn much from each other. First, how effective can selection be? Thinking on the largest scale, were  $\sim 4,000$  million years (Myr) needed for the present complexity of life to evolve? Could we have evolved in 400 Myr, or in 40 Myr? What limits the ability of selection to evolve and maintain complex organisms and complex algorithms? On a smaller scale, how should we optimise a selection scheme, given constraints on the number of organisms and their reproductive capacity, or on the number of computations? Second, does the genetic system evolve to be evolvable? In evolutionary biology, a long-standing question has been whether mutation and recombination are maintained because these processes produce the variation that is essential for adaptation; such questions are close to the parallel questions in evolutionary computation, of what are the optimal rates of mutation and cross-over, and how can they be chosen. A harder question - which has only recently received sustained attention in biology - concerns the relation between genotype and phenotype [20]. In evolutionary computation, it is clear that the way an algorithm is coded is crucial. It is extraordinary that the development and behaviour of a complex organism can be coded by relatively little information (at most  $6 \times 10^9$  bits in the human genome, for example), and that organisms are robust to random changes in this code.

## 2. A COMMON FRAMEWORK

The basic theoretical framework is common to both fields. A genotype is represented by a linear sequence that represents a DNA molecule or a bit string. In real organisms, each nucleotide can take four values  $\{A, T, G, C\}$ , but this is most often approximated by  $\{0, 1\}$ , since there are usually at most two variants per site. Blocks of sequence can be represented by multiple discrete values (alleles) or by a continuous value. Nevertheless, for most cases a bit string  $\underline{X}$  suffices.

Selection acts through the fitness of each individual; I take the fitness of an individual to be the number of offspring it produces after one discrete generation,  $W$ , or its rate of reproduction in continuous time,  $r$ . A full stochastic model would require the distribution of fitness of each genotype. However, when selection is weak, all that matters is the expected fitness; the variance in fitness determines the rate of random fluctuation, and can be taken to be the same for all genotypes. I write the expected fitness of genotype  $\underline{X}$  as  $W(\underline{X})$  or  $r(\underline{X})$ , in discrete or continuous time. In evolutionary computation, “fitness” often refers to some measure that is to be optimised, and reproduction is ~~is~~ some function of this measure. I will keep to the population genetic usage, referring to such measures as fitness components,  $\omega(\underline{X})$ . Often, we also need to consider a set of traits, whose expected value is an arbitrary function of genotype,  $\underline{z}(\underline{X})$ , and which in turn determine the fitness components,  $\omega(\underline{z})$ . To give a concrete example, a cow can be described by a set of traits

(milk yield, growth rate, lifespan, etc), and these depend in a complex way on both genotype and environment. This traits in turn determine some selection index,  $\omega$ , which the farmer uses to determine which cows will breed: for example, those below some threshold ail have zero fitness. (Note that there are complex issues concerning the way we average over variation at each level, which we ignore).

Crucially, in both the real and the virtual worlds, the relation between phenotype (i.e., the traits,  $z$ ) and genotype,  $\underline{X}$ , is **unknown**. It can be determined for any particular  $\underline{X}$  by rearing an organism, or by carrying out a computation, both being expensive. I take this realisation of the organism or algorithm to determine the cost of the breeding program. Of course, in a computer we directly observe the code, and in biology, it is now feasible to determine the entire genome of an organism. However, observing the phenotype of even a very large number of individual genotypes does not tell us which sites cause differences in phenotype, or how they interact: this fundamental problem frustrates large-scale genome-wide association studies aimed at finding the variants that cause human disease. In plant and animal breeding, it is not at all clear how much knowing the full genome sequence helps.

The basic processes of selection, mutation, recombination and random sampling are common to both fields. (Migration between distinct populations may also be important, but is not considered here). In evolutionary computation, any kind of mutation or recombination can be invented, but also in biology, there can be arbitrary errors and rearrangements in copying the DNA sequence.

Provided that the composition of a population changes slowly, evolution can be approximated as being continuous in time, with the details of each process being absorbed into a single parameter - the selection coefficient,  $s$ , the rate of random sampling fluctuations,  $1/N_e$ , the rate of mutation,  $\mu$ , and so on. As we shall see, the evolution of allele frequencies at individual loci may be slow, even when selection on the overall phenotype is strong. The outcome depends on the ratios between these rates ( $N_e s$ ,  $N_e \mu$ , etc.); this gives population genetics a remarkable generality, which carries over to describe the idiosyncrasies of genetic algorithms.

Under this continuous time approximation, the distribution of states of the population follows a diffusion equation that depends on the mean and variance of change between generations: there is a precise analogy between the stochastic evolution of a population, and the diffusion of a molecule. In physics, we may deal with an actual collection of molecules, whereas in population genetics, we must imagine an abstract probability distribution across possible states of a single population; however, the mathematical description is the same. (The diffusion approximation developed independently in physics and population genetics, both tracing back originally to a model for share prices; [5]). Kimura [18] developed the diffusion approximation and applied it to molecular evolution; it is intimately connected with the coalescent process that describes the evolution of genealogical relations between sampled genes.

Even where the diffusion equation cannot be solved, it plays a key role in justifying the use of scaled parameter combinations, and allowing extrapolation from simulations of small populations over short times out to very large populations over long times. Moreover, it provides a precise connection with statistical physics, which rests on the same

mathematical foundation.

## 2.1 Some key results

Although any model can in principle be calculated numerically, we seek general analytical results that aid our intuitive understanding. The main difficulty is in dealing with arbitrary interactions in the nonlinear relation between genotype and phenotype,  $\underline{z}(\underline{X})$ . Nevertheless, there are some general results, summarised in Table 1. Most of these results apply with arbitrary interactions; exceptions are the three results for genetic load are derived assuming no interaction, but which extend to allow some forms of interaction.

It is surprising that few results have been transferred between the fields, or independently re-derived. Perhaps the most prominent theoretical concept in evolutionary computation is Holland’s [14] idea of “schema”, which correspond in population genetics to the *average excess* of a combination of alleles; Livnat et al.’s [22] coefficients of “mixability” are closely related, and correspond to *average effects* - a regression of trait values onto the allelic values. Attention has been drawn to Wright’s “shifting balance” theory, but only in a qualitative way [27, 37]. The strongest connection between evolutionary biology and evolutionary computation has been made by Mühlenbein and Mahnig [24], who use the *breeder’s equation* to make a deterministic analysis of genetic algorithms; they emphasise the independent evolution of allele frequencies, which will be the focus of this note. Several statistical physicists have included the effect of random sampling on evolving populations; Prugel-Bennet (1997) describes a maximum-entropy approximation for trait evolution, whilst [4, 36] independently derive the stationary distribution in the limit where populations are close to fixation for a single type. These results are closely related to Wright’s [41] more general formula for the stationary distribution of allele frequencies [2, 3].

## 2.2 Evolution as hill-climbing

I will focus on how selection can most efficiently accumulate favourable alleles, viewing evolution as a hill-climbing algorithm. This needs some justification, since the central problem in optimisation is often seen quite differently: how to find the best solution, without being trapped at local peaks. These different viewpoints were at the heart of the long dispute between Fisher and Wright [32]. Fisher’s view was that although “adaptive valleys” obviously exist, populations may never cross them: there may always be ways by which fitness can increase, especially since in nature, the adaptive landscape changes with the environment. The metaphor of an adaptive landscape may be misleading when visualised in two dimensions: if there are enough degrees of freedom, progress may always be possible in some direction. If the landscape were in fact so rugged that populations are trapped within a few steps, then any search algorithm would be frustrated: values of individuals already tested must carry some information about untested individuals, otherwise search might as well be random.

It is important to understand that there are distinct versions of the adaptive landscape: it may refer to individuals or to populations, and it may refer to genotypes or to traits. The first distinction is between the fitness of an individual, considered as a function of its genotype or traits ( $W(\underline{X})$  or  $W(\underline{z})$ ), and the mean fitness of a population, considered as a function of its allele frequencies, or

trait means ( $\bar{W}(\underline{p})$  or  $\bar{W}(\underline{z})$ ); Wright used these without distinction [32]. However, it is the latter that leads to a quantitative theory: selection moves populations through allele frequency or trait space at a rate proportional to the gradient of mean fitness, and (provided that genes are well shuffled by recombination) the stationary distribution is proportional to  $\bar{W}^{2N_e}$  [41].

The second distinction is between the adaptive landscape in allele frequency space versus trait space. I will assume a smooth relation between fitness and traits, and ask how efficiently selection can push a population towards the locally fittest trait combination. However, even if there is a single peak in trait space, there may be multiple peaks in genotype space: even in the simplest case where traits are additive, many combinations of + and - alleles can give the optimal phenotype. However, populations cannot evolve between different combinations by selection alone, because mixed populations will produce sub-optimal genotypes by recombination. We shall see that despite this microscopic ruggedness of the landscape, low rates of mutation and random sampling can allow populations to evolve smoothly and predictably at the phenotypic level.

Note that in a large sexual population, if the individual-level landscape has multiple peaks, then so will the population-level version - provided that we consider allele-frequency space: if all single-step changes from the locally fittest genotype are deleterious, then none can invade from low frequency, and a fitter *combination* of alleles will be broken up by recombination. In contrast, an asexual population can move towards a higher peak provided that the mutation rate is high enough that fitter combinations, involving multiple changes, are generated. A similar smoothing occurs in a sexual population, when we consider the evolution of phenotypic traits: the mean fitness of a population ( $\bar{W}(\underline{z})$ ) is a smoothed version of the individual-level landscape ( $W(\underline{z})$ ), so that if the genetic variance is high enough, adaptive valleys may be smoothed out [19].

## 2.3 Evolution of allele frequencies

In the following, I assume that alleles are well shuffled (i.e., in *linkage equilibrium*), so that the composition of a population is described by its allele frequencies. This is a good approximation for natural populations if recombination is faster than other processes, and indeed, applies to most outcrossing sexual species. In simulations, it can be imposed by constructing individuals locus by locus, from the list of allele frequencies; this is a kind of mass meiosis, in which all the genomes in the population take part, rather than just two at a time [24]. In a breeding program, the same result could be obtained by allowing several generations of reproduction in the absence of selection.

The prevalence of sexual reproduction suggests that natural selection acts most efficiently in this limit of complete shuffling. However, it has taken a substantial theoretical effort to show exactly why this is so, and how it leads to the selection of higher rates of recombination. Essentially, recombination is favoured because it breaks up negative associations between favourable alleles (+ + -, - + + etc.), increasing the additive genetic variance that drives the response to selection. An allele that increases recombination is at an immediate disadvantage, because it breaks up favourable gene combinations, but gains a long-term advantage, be-

<hr/> Population genetics <hr/>			
Breeder's equation	$\Delta \bar{z} = \frac{V_a}{V_P} S$	$V_a$ : additive variance $V_P$ : total variance $S$ : change in $\bar{z}$ due to selection	[23]
Wright's gradient formula	$\Delta \bar{z} = V_a \frac{\partial \log(\bar{W})}{\partial \bar{z}}$		[21, 41]
Secondary theorem of natural selection	$\Delta \bar{z} = \text{cov} \left( \frac{\partial \bar{z}}{\partial \bar{W}}, z \right)$		[30, 34]
Mutation load	$\frac{\Delta \bar{W}}{\bar{W}} \sim -U$	$U$ : total rate of deleterious mutation	[10, 25]
Substitution load	$\sum_t \Delta \log(\bar{W}) \sim -\log \left( \frac{1}{p_0} \right)$	$p_0$ : initial allele frequency	[11]
Drift load	$\frac{\Delta \bar{W}}{\bar{W}} \sim -\frac{d}{2N_e}$	$d$ : #of degrees of freedom $N_e$ : effective population size	[17]
Wright's distribution	$\left( \prod_i p_i^{4N_e \mu - 1} q_i^{4N_e \nu - 1} \right) \bar{W}^{2N_e}$	$\mu, \nu$ : rate of mutation away from alleles P, Q	[41]
<hr/> Statistical physics <hr/>			
Fitness flux	$\langle e^{-N_e \Phi + \Delta \mathcal{H}} \rangle = 1$	$\Phi$ : net fitness flux $\Delta \mathcal{H}$ : change in log likelihood	[15, 26]
Maximum entropy	$\frac{\partial}{\partial t} \langle A \rangle \sim B \cdot (\alpha - \alpha^*)$		[3]
<hr/> Evolutionary computation <hr/>			
Schema theorem	$E \left[ \frac{g_{t+1}[H]}{g_t[H]} \right] \geq \frac{W_t[H]}{\bar{W}_t} (1 - \pi)$	$g_t[H]$ : frequency of H at time $t$ $\pi$ : probability that H is broken up	[14]
Drift analysis			[12]

**Table 1:** ~~Table 1~~ summarises key results from the various fields, in schematic form. The first three lines give alternative formulae for the change in trait mean,  $\Delta \bar{z}$ , due to selection and reproduction, ignoring random fluctuations. The next three lines give the net reduction in mean fitness (i.e., the *genetic load*) due to deleterious mutations at total rate  $U$ , due to substitution of an allele that was initially at frequency  $p_0$ , and due to random sampling (termed *drift* in population genetics). The final formula drawn from population genetics gives the stationary distribution of allele frequencies under mutation, selection and random sampling. The first result from statistical physics relates the change in log likelihood,  $\Delta \mathcal{H}$ , to the net *fitness flux*,  $\Phi$ , a measure of the total selection acting along an arbitrary path. The next result approximates the change in expectation of arbitrary traits,  $A$ , under selection on those traits  $\alpha$ ;  $\alpha^*$  is the selection that would maintain the current  $\langle A \rangle$ , and  $B$  is a generalised additive genetic covariance. The penultimate result sets a bound on the change in frequency of a combination of alleles (or *schema*),  $H$ , given by the product of its relative fitness and its chance of surviving recombination and mutation,  $1 - \pi$ . Finally, *drift analysis* sets bounds on the number of computations (or *runtime*) needed to achieve some goal.

cause it is associated with increased additive variance [1]. The key issue, then, is why there should tend to be negative associations among alleles. Such associations may be generated by selection, but this requires a delicate choice of parameters, for which there is little evidence. More generally, random drift interacts with directional selection to generate negative associations: in the simplest case, different favourable mutations almost always arise separately, and so must be brought together by recombination. This is known as the Hill-Robertson effect [13]; when a large number of loci are selected, it is more significant than epistatic interactions, even in large populations [29]. This justifies our focus on well-shuffled populations.

## 2.4 The infinitesimal model

Here, I focus on the *infinitesimal model*, introduced by Fisher [6]. This has been developed as the foundation of animal breeding, yet has received relatively little attention within population genetics. It allows a remarkably general analysis of the performance of genetic algorithms, without requiring detailed knowledge of the complex relation between genotype and phenotype.

At the phenotypic level, the infinitesimal model is easily defined. Unrelated parents produce offspring with trait values that are normally distributed around the mean of the parents, with fixed variance  $\frac{V_0}{2}$ . If individuals mate randomly, and all have the same fitness, then the population rapidly tends to a normal distribution with variance  $V_0$ : random mating reduces the variance twofold in each generation, but this is compensated by the release of genetic variance within families (i.e.,  $V_{t+1} = \frac{V_t}{2} + \frac{V_0}{2}$ ). Selection can produce arbitrary distortions in the trait distribution, but these rapidly dissipate, so that in the long term, only the mean changes, the variance remaining constant.

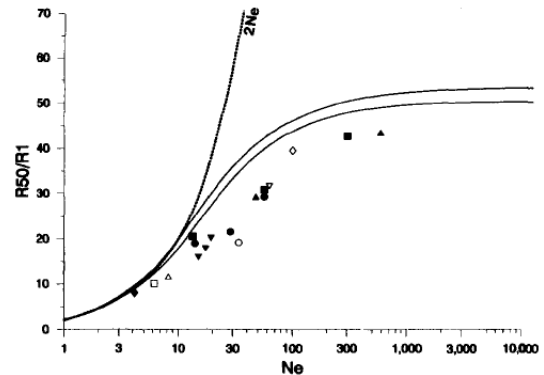
If parents are related, sharing a fraction  $F$  of their genes *identical by descent*, the within-family variance is reduced by a factor  $(1 - F)$ ; in a population of effective size  $N_e$ ,  $(1 - F)$  decreases by a factor  $(1 - \frac{1}{N_e})$  per generation. Mutation adds a constant variance  $V_m$  per generation, and so at an equilibrium between mutation, random sampling, and sexual reproduction, the trait variance equilibrates at  $N_e V_m$ .

This simple phenotypic model can be justified in the limit of a very large number of unlinked loci,  $n$ , with additive effects. The within-family variance is generated by recombination, at a rate proportional to the number of heterozygous loci, which is in turn proportional to  $(1 - F)$ . Selection on each locus is weak ( $s \sim \frac{1}{n}$ ), and so only changes the genetic variance over timescales of  $\sim n$  generations. Short-term changes in the distribution are due to correlations between loci (*linkage disequilibria*), and so dissipate rapidly in the absence of genetic linkage. Here, we assume that the population is in linkage equilibrium, so that the distribution is close to Gaussian; this is a good approximation even with quite strong selection, and can be imposed directly in a genetic algorithm.

The infinitesimal model leads immediately to a simple prediction for the total response to selection [33]. Genetic variance dissipates at a rate  $1 - \frac{1}{N_e}$ , and so the total genetic variance, summed over generations, is  $V_1 N_e$ . Since the change in mean (i.e., the *selection response*) is proportional to  $V_t$  in each generation, the total change is just  $\sum_{t=1}^{\infty} \Delta \bar{z}_t = N_e \Delta \bar{z}_1$ . Robertson [33] gave an alternative derivation of this result, based on the probability  $u(a, p)$

that an allele with effect  $a$  and with frequency  $p$  will ultimately be fixed in the population. The total change in the trait mean must be  $\sum_{t=1}^{\infty} \Delta \bar{z}_t = \sum_i a_i (u(a_i, p_i) - p_i)$ , where the sum on the right is over all loci,  $i$ . If the effect of an allele,  $a$ , is weak, then  $u = p + N_e \beta a p(1 - p) + O(a^2)$ , where  $\beta = \partial W / \partial z$  [33]. This immediately gives  $\sum_{t=1}^{\infty} \Delta \bar{z}_t = N_e \beta \sum_i a_i^2 p_i (1 - p_i) = N_e \beta V_1 = N_e \Delta \bar{z}_1$ . This shows that under the infinitesimal model, the change in the mean is due to the cumulative effect of small perturbations to the distribution of allele frequencies at each locus. This simple additive approximation predicts the selection response over 50 generations remarkably well, for a wide range of traits and organisms (Fig. 1, from [38]). This suggests that the infinitesimal model may also give a good approximation to arbitrary genetic algorithms.

**Figure 1:** ~~Figure 1.~~ The ratio between the change in mean over 50 generations, and the change in the first generation ( $R_{50}/R_1$ ), plotted against effective population size,  $N_e$ . Symbols represent the outcome of independent selection experiments in maize, mouse and *Drosophila*. The upper curve (labelled  $2N_e$ ) is Robertson's (1960) prediction for the ultimate response,  $R_{\infty} = 2N_e R_1$ , whilst the lower curve is the prediction for the response at 50 generations, from the same infinitesimal model. The middle curve is the prediction including a contribution from mutation,  $V_m = 0.001V_g$ , which makes little difference over this short timescale. The observed responses are about 10-20% below the prediction, probably reflecting the effect of selection on alleles of large effect; however, the overall relationship fits surprisingly well, given that the selected traits are unlikely to have an additive genetic basis. From Fig. 4 in [38].



Robertson's derivation suggests a generalisation to arbitrary relations between genotype and phenotype. If traits depend on a large number of loci, then strong selection on the traits may only slightly perturb the underlying distribution of allele frequencies from its neutral trajectory. Moreover, this perturbation depends only on the marginal effect of the allele - that is, its additive effect on the trait. In the short term, the mean changes at a rate proportional to the additive variance, which itself stays approximately constant. In the longer term, the additive effects change as the genetic background changes, in an unpredictable way that depends on how genes interact. However, it may still be possible to optimise the selection scheme, using purely

“local” information.

## 2.5 Optimising the selection response

How can we optimise the net response to selection, for a given total number of individuals? For simplicity, we neglect mutation, and assume that the mean changes solely by selection on the initial pool of variation. We begin by assuming *truncation selection* on an additive trait,  $z$ , choosing the best  $\theta$  of individuals out of  $N$  ( $0 < \theta < 1$ ). In order to select a fraction  $\theta$  from a normal distribution, those more than  $x$  standard deviations above the mean must be chosen;  $\theta = \frac{1}{2} \operatorname{erfc} \left[ \frac{x}{\sqrt{2}} \right]$ . The direct effect of selection is to change the mean by  $S = i(\theta)\sqrt{V}$ , where  $V$  is the variance of the trait, and where  $i = \frac{\epsilon - \frac{x^2}{2}}{\theta\sqrt{2\pi}}$ . Throughout, we assume that there is no non-genetic variance.

The problem is to choose the optimal  $N_t, \theta_t$ , given the constraint  $C = \sum_{t=1}^{\infty} N_t$ . We will see that the maximum possible response has the form  $A\sqrt{V_1 C_1}$ , where  $A$  is to be determined. With this *ansatz*, we must trade an immediate gain of  $i(\theta_1)\sqrt{V_1}$  against a future gain of  $A\sqrt{(C_1 - N_1)V_1} \left(1 - \frac{1}{N_1\theta_1}\right)$ ; note that the loss of variance depends on the number of *selected* individuals,  $N_1\theta$ , whereas the cost is counted as the total number of individuals,  $C_1 = \sum_{t=1}^{\infty} N_t$ . Assuming that  $C_t, N_t\theta_t$  are large, so that  $C_t \gg N_t, N_t\theta_t \gg 1$ . Then, the optimal scheme is  $N_1 = \sqrt{C/\theta_1}$ ,  $A = -2i'\theta_1^{3/2}$ . The same argument applies throughout the process, and so gives the full schedule of  $N_t, \theta_t$ . We see immediately that the optimal  $\theta$  is the same throughout, but that the optimal  $N_t$  decreases with the remaining number of individuals,  $\sqrt{C_t}$ . Approximating change as continuous in time, this implies that  $\partial C/\partial t = -N = -\sqrt{C/\theta}$ , so that  $C = \left(\sqrt{C_1} - \frac{t}{2\sqrt{\theta}}\right)^2$ , with selection ending at time  $t_{\max} = 2\sqrt{C_1\theta}$ . Integrating over time, we have

$$A\sqrt{V_1 C_1} = i\sqrt{V_1} \int_0^{2\sqrt{C_1\theta}} \exp\left(-\int_0^t \frac{d\tau}{2\sqrt{C_\tau/\theta}}\right) dt = \frac{2i\sqrt{V_1 C_1 \theta}}{1+\theta}.$$

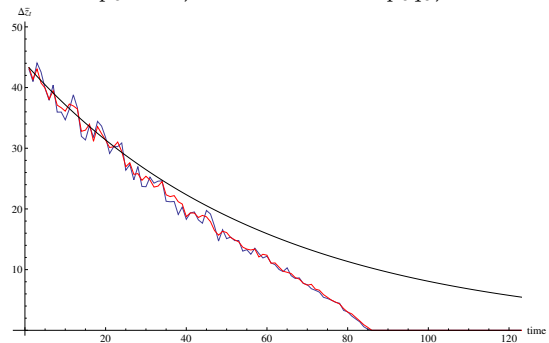
This confirms the *ansatz*, and implies  $A = \frac{2i\sqrt{\theta}}{1+\theta}$ . Combining this with the criterion for optimising  $\theta$ ,  $A = -2i'\theta^{3/2}$ , we find  $\theta=0.391$ ,  $A=0.883$ . For comparison, if we select with constant  $N, \theta$  for time  $\frac{C}{N}$ , then the response is  $i\sqrt{V_1} \int_0^{C/N} \exp\left(-\frac{t}{N\theta}\right) dt = 2N\theta i\sqrt{V_1} \left(1 - \exp\left(-\frac{T}{2N\theta}\right)\right)$ . The optimal solution is now when  $\theta=0.270$ ,  $N = 1.213\sqrt{C}$ ,  $T = \frac{\sqrt{C}}{1.213} = \frac{N}{1.471}$ , and the net response is  $0.574\sqrt{V_1 C_1}$  - rather less than with the optimal schedule, in which  $N$  falls over time, and the total response is  $0.883\sqrt{V_1 C_1}$ .

These arguments are illustrated by simulations of truncation selection on an additive trait, based on  $10^4$  loci of equal effect,  $\alpha=1$ , and starting with average allele frequency 0.5 - and hence a trait mean  $\sum_{i=1}^n \alpha_i p_i = \frac{n}{2}$ . Figure 2 shows an example with  $N = 30$  selected individuals, showing how the standard deviation initially decreases as  $\sqrt{(1-1/N)^t}$ , but then falls away more rapidly than predicted by the infinitesimal model, as selection fixes favourable alleles. In any generation, the change in mean is very close to the prediction,  $i\sqrt{V_t}$ . The area under these jagged curves gives the total response to selection, which is somewhat less than the infinitesimal prediction, shown by the area under the smooth curve. Figure 3 shows that the infinitesimal model predicts the ultimate response quite well for  $N \leq 30$ ; in larger

populations, selection is more effective at eliminating variation relative to sampling, and so the infinitesimal prediction is much too high. This must necessarily be so, because the ultimate response approaches the maximum possible,  $\frac{n}{2}$ , as  $N$  increases. Figure 4 shows the optimal choice of  $N$ , and the maximum possible response, given the constraint that the total number of individuals realised is  $C = NT$  (and using constant  $N, \theta$  for simplicity). This fits the predictions  $\hat{N} = 1.213\sqrt{CV_0}$ ,  $\hat{R} = 0.574\sqrt{C}$  well for  $N \leq 50$ ; for larger values of  $C$ ,  $\hat{N}$  is larger and  $\hat{R}$  is smaller, because selection as well as sampling eliminate variation.

*Figs. 2-4 here*

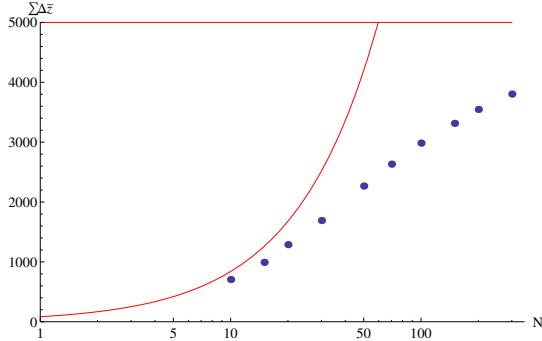
**Figure 2:** The decrease in the rate of response to truncation selection over time. The blue curve shows the change in mean in successive generations, averaged over 5 replicates. The red curve shows the predicted change,  $i[\theta]\sqrt{V_t}$ , which is proportional to the standard deviation of the trait. The upper smooth curve shows the predicted rate of decay of the standard deviation,  $\sqrt{(1-1/N)^t} \sim e^{-t/(2N)}$ . The area under the blue curve gives the total selection response, whilst the area under the smooth curve gives the prediction based on the infinitesimal model. There are  $N=30$  selected individuals, each with  $10^4$  selected loci with equal additive effects,  $\alpha=1$ . In each generation,  $N/\theta=111$  individuals are generated; the fraction selected is  $\theta=0.27$ . Initial allele frequencies are drawn from a beta distribution with mean  $p_0=0.5$ , and variance  $F p_0 q_0$ ,  $F = 0.5$ .



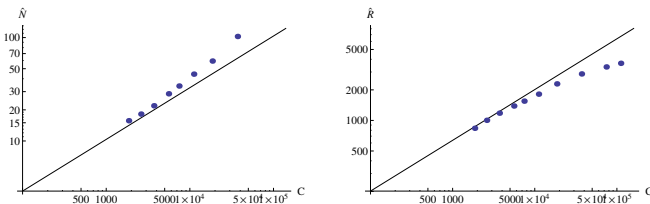
## 3. CONCLUSIONS

This note has outlined how quantitative genetics can predict the response of a population to selection on a trait, without any detailed knowledge of the genetic basis of that trait: the assumption is essentially that selection is spread over so many loci that the distribution of allele frequencies at each is hardly perturbed. In biology, this very simple model, which assumes additive effects, is remarkably successful in predicting the short-term evolution of traits that are clearly not additive (e.g. Fig. 1). Figure 3 shows that the ultimate change due to selection on an additive trait - made without making any assumption whatever about the number of loci involved, the distribution of allele frequencies, or the distribution of allelic effects - is predicted well for small population sizes, provided that the number of loci is large. It overestimates the response for larger population sizes, essentially because the population approaches the global optimum, which sets an upper bound on the response.

**Figure 3:** Points show the ultimate response to selection,  $\sum_{t=1}^{\infty} \Delta \bar{z}_t$ , plotted against the number of selected individuals,  $N$ ; other parameters are as in Fig. 2. The upper flat line shows the maximum possible change,  $\frac{n}{2} = 5000$ , and the curve shows the prediction based on the infinitesimal model,  $2N \Delta \bar{z}_1$ .



**Figure 4:** The optimal population size,  $\hat{N}$ , and the corresponding maximum possible selection response,  $\hat{R} = \left(\sum_{t=1}^T \Delta \bar{z}_t\right)_{\max}$ , plotted against the total number of individuals realised,  $C = TN/\theta$ . These values are calculated by interpolation from the simulation results summarised in Fig. 3. The straight lines show the theoretical predictions,  $\hat{N} = 1.213\sqrt{CV_0}$ ,  $\hat{R} = 0.574\sqrt{C}$ .



One might object that in evolutionary computation, one is interested in precisely this regime. However, local predictions for the response over some tens of generations, may allow optimisation of the selection scheme. In biology, artificial selection of domesticated plants and animals has been remarkable, even though it is very far from producing an ultimate “global optimal” (whatever that might mean).

There is clearly much more to be done in applying quantitative genetic methods to improving evolutionary algorithms: most obviously, in finding the range of non-additive models for which these methods are sufficient approximations. Mutation should also be incorporated, and here one must take account of the effect of mutation in degrading the mean, as well as increasing the variance. An intriguing possibility is to use the regression of mean and variance of trait values on the number of mutations in the realised individuals, to estimate the optimal mutation rate.

### 3.1 References

- [1] Barton, N. H. 1995. A general model for the evolution of recombination. *Genetical Research* 65:123-144.
- [2] Barton, N. H., and J. B. Coe. 2009. On the application of statistical physics to evolutionary biology. *Journal of Theoretical Biology* 259:317–324.
- [3] Barton, N. H., and H. P. de Vladar. 2009. Statistical mechanics and the evolution of polygenic quantitative traits. *Genetics* 181:997-1011.
- [4] Berg, J., S. Willmann, and M. Lässig. 2004. Adaptive evolution of transcription factor binding sites. *BMC Evolutionary Biology* 4:42.
- [5] Davis, M., and A. M. Etheridge. 2006. *Louis Bachelier’s Theory of Speculation: The Origins of Modern Finance*. Princeton University Press, Princeton, NJ.
- [6] Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Proceedings of the Royal Society of Edinburgh* 52:399-433.
- [7] Fisher, R. A. 1930. *The genetical theory of natural selection*. Oxford University Press, Oxford.
- [8] Gerrish, P. J., A. Colato, A. S. Perelson, and P. D. Sniegowski. 2007. Complete genetic linkage can subvert natural selection. *Proc. Natl. Acad. Sci. U. S. A.* 104:6266-6271.
- [9] Fogel, D. B. 1998. *Evolutionary computation: the fossil record*. IEEE Press, New York.
- [10] Haldane, J. B. S. 1937. The effect of variation on fitness. *American Naturalist* 71:337-349.
- [11] Haldane, J. B. S. 1957. The cost of natural selection. *J.Genet.* 55:511-524.
- [12] He, J., and X. Yao. 2004. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing* 3:21-35.
- [13] Hill, W. G., and A. Robertson. 1966. The effect of linkage on limits to artificial selection. *Genet. Res.* 8:269-294.
- [14] Holland, J. 1975. *Adaptation in natural and artificial systems*. Univ. of Michigan Press. Ann Arbor.
- [15] Jarzynski, C. 1997. Nonequilibrium equality for free energy differences. *Physical Review Letters* 78:2690-2693.
- [16] Kimura, M. 1961. Natural selection as the process of accumulating genetic information in adaptive evolution. *Genetical Research* 2:127-140.
- [17] Kimura, M., and T. Ohta. 1970. Genetic loads at a polymorphic locus maintained by frequency dependent se-

lection. *Genet.Res.* 16:145-150.

[18] Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge Univ. Press. Cambridge.

[19] Kirkpatrick, M. 1982. Quantum evolution and punctuated equilibrium in continuous genetic characters. *Amer. Nat.* 119:833-848.

[20] Kirschner, M., and J. Gerhart. 1998. Evolvability. *Proceedings of the National Academy of Sciences (U.S.A.)* 95:8420-8427.

[21] Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30:314-334.

[22] Livnat, A., C. Papadimitriou, N. Pippenger, and M. W. Feldman. 2010. Sex, mixability, and modularity. *Proceedings of the National Academy of Sciences (USA)* 107:1452-1457.

[23] Lush, J. 1937. *Animal Breeding Plans*. Iowa State College Press, Ames, Iowa.

[24] Mühlenbein, H., and T. Mahnig. 2001. Evolutionary computation and beyond. Pp. 123-188 in Y. Uesaka, P. Kanerva, and H. Asoh, eds. *Foundations of real-world intelligence*. CSLI Publications, Stanford, CA.

[25] Muller, H. J. 1950. Our load of mutations. *Amer. J. Hum. Genet.* 2:111-176.

[26] Mustonen, V., and M. Lassig. 2010. Fitness flux and ubiquity of adaptive evolution. *Proceedings of the National Academy of Sciences (U.S.A.)* 107:4248-4253.

[27] Oppacher, F., and M. Wineberg. 2000. Reconstructing the Shifting Balance Theory in a Genetic Algorithm: Taking Sewall Wright Seriously. Pp. 219-226. *Congress on Evolutionary Computation*, San Diego CA.

[28] Orr, H. A. 1998. The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* 52:935-949.

[29] Otto, S. P., and N. H. Barton. 2001. Selection for recombination in small populations. *Evolution* 55:1921-1931.

[30] Price, G. R. 1970. Selection and covariance. *Nature* 227:520-521.

[31] Provine, W. 1971. *The origins of theoretical population genetics*. Univ. of Chicago Press. Chicago.

[32] Provine, W. 1986. *Sewall Wright and evolutionary biology*. Univ. of Chicago Press. Chicago.

[33] Robertson, A. 1960. A theory of limits in artificial selection. *Proceedings of the Royal Society (London)* B 153:234-249.

[34] Robertson, A. 1966. A mathematical model of the culling process in dairy cattle. *Animal Production* 8:95-108.

[35] Ruse, M. 1996. *Monad to Man: the concept of progress in evolutionary biology*. Harvard University Press, Harvard.

[36] Sella, G., and A. E. Hirsh. 2005. The application of statistical physics to evolutionary biology. *Proceedings of the National Academy of Sciences (U.S.A.)* 102:9541-9546; correction 102:14475-14475.

[37] Toquenaga, Y., and M. J. Wade. 1996. Sewall Wright meets artificial life: the origin and maintenance of evolutionary novelty. *Trends in Ecology and Evolution* 11:478-482.

[38] Weber, K. E., and L. T. Diggins. 1990. Increased selection response in larger populations. II Selection for ethanol vapour tolerance in *D melanogaster* at two population sizes. *Genetics* 125:585-597.

[39] Williams, G. C. 1966. *Adaptation and natural selection*. Princeton University Press, Princeton, New Jersey.

[40] Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97-159.

[41] Wright, S. 1937. The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences (U.S.A.)* 23:307-320.