Characterising Fitness Landscapes Through Evolvability

Tom Smith[*1,2], **Phil Husbands**[1,3] **and Michael O'Shea**[1,2]
[1]Centre for Computational Neuroscience and Robotics (CCNR)
[2]School of Biological Sciences
[3]School of Cognitive and Computing Sciences
University of Sussex, Brighton, UK
[*]`toms@cogs.susx.ac.uk`

Abstract

The majority of measures developed to predict the difficulty of finding good solutions to a given problem through search focus on epistasis, while other features which affect problem difficulty such as landscape neutrality, may be ignored completely. In this paper we develop a set of measures based on evolvability statistics of the fitness landscape surrounding sampled solutions. Averaging the measures over a sample of equal fitness solutions allows us to build up a fitness portrait of the fitness landscape, which we show can be used to predict the difficulty of finding good solutions in a set of tunably rugged and tunably neutral landscapes.

Keywords: Evolvability, Fitness Landscape, Search Space, Neutral Evolution, NK System

1 Introduction

The majority of measures developed to predict the difficulty of finding good solutions to a given problem through search focus on epistasis, while other features which affect problem difficulty such as landscape neutrality, may be ignored completely. In this paper we present a novel technique based on local characteristics of the fitness landscape surrounding a solution, averaged over samples of equal fitness solutions. We show that such a technique predicts the difficulty of searching in both tunably rugged and tunably neutral fitness landscapes.

A second feature of most problem-hardness measures is that a single global metric is used to predict the difficulty of searching over the entire fitness landscape. The technique presented in this paper develop a set of continuous metrics that vary with solution fitness. This approach allows the identification of fitness levels at which direct search optimisation techniques are likely to have difficulty finding fitter solutions.

Many problems to which stochastic search techniques such as evolutionary computation are typically applied, present such highly skewed distributions of solution fitnesses that random sampling (even when some imposed distribution is applied to the sample) are vanishingly unlikely to represent fitnesses above a given level, even when such fitnesses are easily found through direct search optimisation. If we are to predict the difficulty of searching in such spaces, we must develop problem difficulty measures that work with *online* sampling (in this paper, we will use the term online sample to refer to samples collected during some search process, as opposed to samples collected through random sampling). In this paper, we show that the problem difficulty metrics presented work with samples of solutions collected both through random sampling techniques, and through online sampling of the best solution so far found during simple hill-climbing optimisation.

The paper proceeds as follows: Section 2 outlines the concepts of fitness landscapes and neutrality, and goes on to describe the notion of problem difficulty as determined by the structure of the fitness landscape. Section 3 introduces the notion of solution evolvability as defined by local characteristics of the fitness landscape surrounding the solution, and derives and applies the problem difficulty metrics used in the remainder of the paper. Section 4 describes the tunably rugged and tunably neutral terraced NK landscapes used as test problems in this work. Sections 5 and 6 apply the metrics derived in section 3 to the test landscapes, and show that they predict the difficulty of searching in the tunably rugged and tunably neutral landscapes. Finally, section 7 shows that the metrics are robust when applied to online samples collected during simple hill-climbing, and the paper closes with discussion.

2 Fitness landscapes and neutrality

This section introduces two of the main concepts used in the paper. The *fitness landscape* (section 2.1), first introduced by Wright (1932), describes the search space as a multi-dimensional landscape defined by the genotype-to-fitness mapping through which evolution moves. The classical idea of searching this landscape for good genotypes focuses on the difficulty of climbing up to the globally optimal fitness solution, and avoiding locally optimal solutions. Here we argue that in difficult search problems, much of the time spent optimising may be spent in non-adaptive *neutral* evolution (section 2.2). Thus measures aimed at predicting the difficulty of search, must take account of the nature of neutrality in the space. Section 2.3 describes how the difficulty of finding good solutions is determined by the structure of the fitness landscape, and section 2.4 outlines different methods for sampling the fitness landscape structure.

2.1 Fitness landscapes

Wright (1932) introduced the *fitness landscape* as a non-mathematical aid to visualise the action during evolution of selection and variation (in this paper we will use the term evolution to refer to both natural biological evolution, and the artificial evolution class of stochastic search processes which operate through some form of "generate-and-test" algorithm, e.g. genetic algorithms (Holland, 1992), genetic programming (Koza, 1992), evolutionary strategies (Rechenberg, 1973) and evolutionary programming (Fogel et al., 1966)). The description views the space in which evolution takes place as a landscape, with one dimension per genotype loci and an extra dimension, or height, representing the phenotype, or fitness, of that particular genotype[1]. The search space defined by a two-loci representation can thus be viewed as a three-dimensional fitness landscape (figure 1), with each point corresponding to a single genotype and fitness. Applying a mutation operator to a particular genotype A typically produces a cluster of offspring genotypes lying close to A in the landscape, while recombination of two different genotypes A, B typically produces offspring genotypes lying somewhere between A and B in the landscape. Evolution can thus be viewed as the movement of the population, represented by a set of points (genotypes), towards higher (fitter) areas of the landscape.

This view of the search space leads naturally to the identification of the major problems with which evolution will have to cope; ruggedness and modality (Kauffman, 1993; Naudts and Kallel, 2000). Highly *epistatic* problems where fitness is dependent on multiple inter-gene interactions will produce a rugged landscape, in which the direction to good solutions is obscured by local noise. Similarly, a high degree of *modality*, i.e. large numbers of local optima, will be seen as large numbers of hill-tops in the landscape with no neighbours of higher fitness. The majority of measures derived to predict the difficulty of searching in a given space are based around these features of ruggedness and modality (see e.g. Weinberger, 1990; Hordijk, 1996; Jones and Forrest, 1995; Naudts and Kallel, 2000).

---

[1] Wright defined two forms of fitness landscapes. The first version, used in this work, defines each point on the landscape as representing a single genotype with height corresponding to genotype fitness. The second version has each landscape point representing an entire population, with the values along each dimension representing the allele frequency over the population, and the height corresponding to the mean population fitness. The two approaches may show markedly different properties (Coyne et al., 1997).

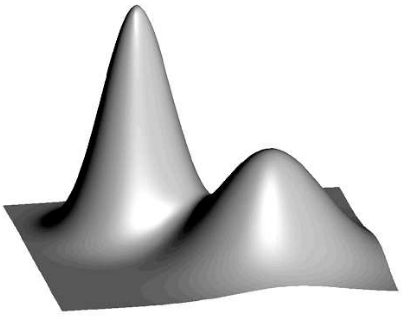


Figure 1: A two-dimensional model fitness landscape, with one globally-optimal and one locally-optimal peak. From a starting point, typically chosen at random, the search process tries to find good solutions. The process typically creates a new set of solutions through the application of genetic operators to the current solution(s), evaluating whether the new set is better than the current solutions. Evolving populations will tend to get stuck at the locally-optimal peak due to its large basin of attraction, and from there will only find the global optimum with difficulty.

A more exact picture, especially when dealing with solutions represented by discrete-valued genotypes, is the connected graph (Stadler, 1996). Solution vertices, or nodes, are connected directly through the action of the genetic operators. The graph may show the space in a very different way to the fitness landscape: mutation operators acting on more than one loci, and other operators such as recombination, may not 'see' fitness landscape hill-tops as local optima at all. However, local optima can clearly exist in the graph, occurring as graph nodes from which all connected nodes are of lower fitness. This definition may produce local optima with respect to genetic operators other than simply mutation, for example some solutions may be local optima with respect to recombination operators.

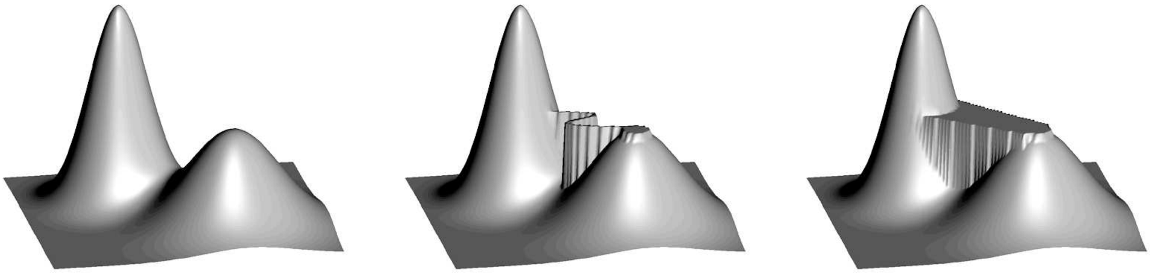
The graph definition of the search space highlights the dangers in the simple visualisable picture afforded to us by the fitness landscape description: our intuitive view may not apply in higher dimensional spaces. Fisher, for example, argued that local optima may not exist in a large class of high-dimensional spaces; the probability that a solution is optimal in every single dimension simultaneously is negligible (quoted in Provine, 1986, p. 274). However, it should be stressed that many problems clearly do show local optimality, e.g. the travelling salesman problem (Lawler et al., 1985). The next section introduces the idea of search space *neutrality*, one possible way in which some high-dimension spaces may differ radically from our intuitive viewpoint.

2.2 Fitness landscape neutrality

In the neutral theory, it is argued that evolving populations may spend relatively large periods of time undergoing non-adaptive neutral mutation (Kimura, 1983), staying at a constant height in the fitness landscape. The evolutionary timescale may be dominated by long periods of neutral epochs (van Nimwegen et al., 1999), interspersed with short periods of rapid fitness increase, i.e. *punctuated equilibrium* (Eldredge and Gould, 1972; Gould and Eldredge, 1977; Elena et al., 1996). During these neutral epochs, the population will move in the space through random drift (note that this is a separate process to Wright's idea of genetic drift due to finite population size). Despite the undirected nature of the population movement, neutrality can be of use in escaping from (nearly) locally optimal solutions: figure 2 shows three model landscapes illustrating the possible advantages of neutrality.

Neutral mutation in a fitness landscape will occur as random drift between solutions of equal fitness connected by mutation; such neutral evolution cannot be distinguished from a population stuck in a local optimum by looking at fitness. Instead, the underlying dynamics of the population must be investigated. Two key features have been predicted that distinguish the behaviour of the population during these periods of neutral drift from the behaviour of a population stuck in a local optimum. The first key feature is that of *movement*, i.e. is the population moving significantly in genotype space. The second key feature is that of *constant innovation*, i.e. the number of previously unencountered phenotypes seen over time is constant. Thus any phenotype should be accessible from a large enough neutral network; neutral drift can eventually find a higher fitness genotype and jump up to a higher neutral network.

These key features have been analysed in many theoretical landscapes. Landscapes produced by RNA



(a) Unconnected peaks (b) Single neutral pathway (c) Broad neutral plateau

Figure 2: Three two-dimensional model fitness landscapes showing the possible advantage of neutrality in a simple landscape with one globally-optimal and one (nearly) locally-optimal peak. **(a)** shows the two peaks as unconnected; populations evolving to the locally optimal peak will have difficulty moving to the global optimum. **(b)** shows the two peaks connected by a single neutral pathway; a population on the sub-optimal peak will eventually find the pathway. **(c)** shows the two peaks connected by a broad plateau; the population will move easily from the sub-optimal peak to the global optimum.

secondary structure folding algorithms show that neutral walks (a neutral variant on the random walk, where successive genotypes are generated by application of the mutation operator) have the property of constant innovation (Huynen et al., 1996; Huynen, 1996). Exhaustive analysis of some RNA landscapes has even mapped out the extent of the neutral networks, and calculated the transition probabilities between the different networks (Forst et al., 1995; Grüner et al., 1996). Barnett (1998) introduces the NKp landscape, a tunably neutral variant on Kauffman's NK systems (Kauffman, 1993), and through calculation of population diffusion coefficients shows highly neutral landscapes produce much more population movement during periods of no fitness change when compared with such periods in landscapes of zero neutrality. An important result from the NKp landscape work is that changing the amount of neutrality in the landscape has no effect on the ruggedness correlation function; correlation lengths do not predict the change in evolutionary dynamics seen for landscapes of different neutrality (Barnett, 1998). Newman and Engelhardt (1998) investigate a similar tunably neutral variant of the NK system, finding that increased neutrality allows high fitness solutions to be found more easily through search. This is backed up by Shackleton et al. (2000) who find that adding neutrality through redundancy can help improve the level of fitness found through artificial evolution.

Neutrality has also been shown in real-world problem landscapes: In experiments on evolving tone recognition circuits, populations were seen to move in genotype space during periods where fitness did not increase (Harvey and Thompson, 1996). Experiments on evolution of digital circuits under two conditions - allowing neutral mutations and not allowing such changes - have also shown the importance of neutrality to the search process. Experiments where neutral changes were allowed consistently produced two-bit multiplier solutions of higher fitness than experiments without neutral mutation (Vassilev and Miller, 2000). Also, Thompson (2001) reports on the presence of neutral plateaus in evolving speech recognition circuits, and shows through exhaustive mutation of genotypes at the start of the plateau that no transitions exist to a higher fitness level; the neutral evolution phase is necessary.

The degree of neutrality in a system is clearly a factor in the population dynamics during evolution. However, many measures aimed at predicting the difficulty of search in a particular space do not account for this factor, e.g. the work on NKp landscapes showing that the autocorrelation function does not change with the level of neutrality (Barnett, 1998). In the next section, we describe typical approaches to identifying problem difficulty with fitness landscape structure, outlining the advantages of the evolvability metrics developed in this paper.

2.3 Fitness landscape structure and problem difficulty

The difficulty of searching in a given problem search space is clearly related to the structure of the fitness landscape. Early work proposed that the correlation structure of the landscape was a key feature corresponding to problem epistasis, or landscape ruggedness (Kauffman, 1993; Weinberger, 1990; Manderick et al., 1991; Lipsitch, 1991; Hordijk, 1996). Typically, the correlation structure of the landscape is derived in terms of parent-offspring fitness correlation, or the fitness-distance autocorrelation function along a random non-adaptive walk. Similarly, fitness-distance correlation uses the correlation of genotype fitnesses with the fitness and distance of known optima as a metric (Jones and Forrest, 1995).

Related work defines the landscape in terms of the information required to describe the landscape. Fourier analysis decomposes the landscape into its Fourier components; the ruggedness of the system is measured by the number of Fourier components, especially those of high-frequency, required for the decomposition (Stadler and Wagner, 1998). Epistasis variance calculates the level of variance in the system with respect to a set of first order approximations to the system (Davidor, 1991). Similarly, information content analysis relates the ruggedness of the system to the fitness entropy encountered along a random walk (Vassilev, 1997; Vassilev et al., 2000)

However, the majority of such techniques assume that the difficulty of searching in a given space is solely a function of the search space ruggedness, measured through such correlation or information analysis as described above. No account is taken of the neutrality in the space; as described in section 2.2 the correlation functions do not change when the degree of neutrality is varied. In the remainder of this paper, we introduce a set of measures based on the characteristics of the fitness landscape surrounding sample solutions, and show that such measures can predict the difficulty of searching in both tunably rugged and tunably neutral landscapes.

2.4 Sampling the fitness landscape

In many hard problems, the distribution of solution fitnesses may be markedly non-normal, with fitnesses additionally distributed heterogenously throughout the search space. In such problems, the set of solutions collected through random sampling methods will not accurately represent the set of solutions seen during optimisation. All measures of problem difficulty based on random sampling will thus not accurately reflect the difficulty of reaching different fitness levels in the space. One extreme example is investigated by Smith et al. (2001a), where analysis of an evolutionary robotics search space shows that fewer than 0.0001% of solutions in a random sample have fitness above 50% of the maximum in a neural network robot control problem, despite this fitness being relatively easy to reach using optimisation techniques. Two spaces that differ only in high fitness regions may show markedly different times to find good solutions through optimisation, but problem difficulty metrics based on random sampling will not show these differences (Smith et al., 2001a)

One potential approach is to bias the random sample procedure, keeping only some set percentage of solutions at each fitness. Even this method may fail to collect solutions above some fitness level in reasonable time, and it may be necessary to perform some kind of directed search process to collect the sample. Clearly there is some point at which the time taken to collect such a sample may well approach a significant fraction of the time taken to solve the problem. For instance, if the sample required to characterise the problem involves collecting solutions at or near the optimum, we will have effectively solved the problem merely in the act of description. A useful analogy could be drawn with Marr's type II systems; the system may not be reducible to a simpler level of description than the system itself (Marr, 1976). By contrast, type I systems can be reduced to a simpler description, e.g. a fitness landscape which can usefully be reduced to a single correlation length description.

In section 7 we collect samples though simple hill-climbing, and show that the problem difficulty metrics based on the biased sample set make the same predictions as those based on unbiased random samples. Although the landscapes used in this paper have normal fitness distributions, verifying that the metrics are robust to sample bias is important if we are to go on to use them on other problems with highly skewed fitness distributions (Smith et al., 2001b).

In the next section, we introduce the notion of evolvability as the capacity of a solution to evolve, closely tied to the fitness landscape neighbouring that solution. We then derive a set of solution and population evolvability metrics, and use them to predict the difficulty of searching in a given fitness landscape.

3 Evolvability and the transmission function

Evolvability is loosely defined as the capacity to evolve, alternatively the ability of an individual or population to generate fit variants (Altenberg, 1994; Marrow, 1999; Wagner and Altenberg, 1996). Thus evolvability is more closely allied with the *potential* for fitness than with fitness itself; two equal fitness individuals or populations can have very different evolvabilities (Turney, 1999). Typically, researchers use some definition of evolvability based on the offspring of current individuals or populations: in this paper we follow Cavalli-Sforza and Feldman (1976) and Altenberg (1994) in using the *transmission function* of all possible offspring from a parent to define a set of metrics of evolvability, see section 3.1 for further details.

It is often argued that there may be long-term trends for evolvability to increase during evolution (see e.g. Wilke, 2001; Turney, 1999). However, as evolvability is more directly related to fitness potential than fitness itself, long-term change cannot be due to straight fitness selection. Thus any trend towards change in evolvability can only be understood through some second order selection mechanism, by which evolution tends to select solutions that have a more evolvable genetic system (Dawkins, 1989; Kirschner and Gerhart, 1998).

Researchers in both biology and evolutionary computation typically link evolvability with the local structure of the search space. For example, Burch and Chao (2000) shows that RNA virus evolvability can be understood in terms of the mutational neighbourhood, while many evolutionary computation researchers (see e.g. Ebner et al., 2001; Marrow, 1999) argue that changing the properties of the search space (through such mechanisms as adding neutrality) can affect evolvability as evidenced by the speed of evolution. The interest in evolvability for evolutionary computation practitioners is thus tied closely to work on the ruggedness and modality of the search space, argued to primarily influence the ease of finding good solutions in the space (Weinberger, 1990; Hordijk, 1996; Jones and Forrest, 1995; Naudts and Kallel, 2000).

Recent work has emphasised that in addition to landscape ruggedness and modality, search space neutrality may have impact on the population dynamics of evolution (section 2.2). This factor may not be predicted by standard measures based on the landscape ruggedness and local modality, but may be measurable through change in evolvability. For example, recent artificial evolution research has shown that evolvability can change during neutral epochs; populations tend to move to “flatter” areas of the fitness landscape where fewer mutations are deleterious (Wilke et al., 2001; Wilke, 2001). This can clearly have an impact on the speed of search, but may not be picked up by the standard landscape ruggedness and modality measures.

Other biological research in evolvability is also of relevance to evolutionary computation, e.g. the work on adaptation to change in environment through such mechanisms as alleles providing increased mutation rates (Taddei et al., 1997; Sniegowski et al., 1997). However, in this paper we focus on evolvability in terms of the properties of the solutions’ local search space. The next section outlines the offspring transmission function, and defines a simple set of evolvability metrics.

3.1 The transmission function

In this paper, we follow the definition of evolvability as the ability of individuals and populations to produce fit variants, specifically the ability to both produce fitter variants, and to not produce less fit variants. This definition is intimately tied in with research on the *transmission function* T (Altenberg, 1994; Cavalli-Sforza and Feldman, 1976), and the population offspring probability distribution function ϕ from all possible applications of the genetic operators to the parent(s):

$$\phi(g, f) = \int\int\int\int \psi(h, k, h', k')\, T(g, f : h, k, h', k')\, dh\, dk\, dh'\, dk' \tag{1}$$

or the probability ϕ (with parental selection function ψ) of obtaining offspring genotype g and phenotype f, over all parents of genotypes h, h' and phenotypes k, k'. The transmission function T is the probability density function of obtaining g, f given h, k, h', k' (Cavalli-Sforza and Feldman, 1976).

In the absence of recombination, only a single parent h, k is required to produce offspring through mutation:

$$\phi(g, f) = \int_{-\infty}^{\infty} \psi(h, k)\, T(g, f : h, k)\, dh\, dk \tag{2}$$

or the probability of obtaining offspring g, f over all parents h, k with selection ψ. In this paper, we focus on the offspring of a set of single genotypes (saved during the course of evolutionary runs), so do not integrate over the set of all possible parents. Similarly, the selection function can be omitted as we pre-select the parent. Since we are interested only in the offspring phenotypes f, and not the offspring genotypes g, we can refer to the transmission function $T(f : h, k)$ as short-hand for the probability density function of offspring fitnesses from a single parent h, k.

The transmission function thus encompasses both the operators and the representation; instead of referring to good and bad genetic operators or good and bad representations, we can talk about the effectiveness of the transmission function. Thus the evolvability of an individual or population, i.e. their ability to generate fit variants, is simply a property of the individual or population transmission function. The next section derives measures for the evolvability of an individual solution in terms of this transmission function for continuous variables.

3.2 Evolvability metrics: Continuous variables

The evolvability of a solution genotype h and fitness k is directly tied to the probability of that solution not producing offspring of lower fitness. Thus we derive our first metric of evolvability E_a:

$$E_a = \frac{\int_k^{\infty} T(f : h, k)\, df}{\int_{-\infty}^{\infty} T(f : h, k)\, df} \tag{3}$$

or the probability that the offspring fitness f is greater or equal to the current fitness k, i.e. the mutation is non-deleterious. Since the transmission function $T(f : h, k)$ is a probability density function, the infinite integral sums to unity, so we have:

$$E_a = \int_k^{\infty} T(f : h, k)\, df \tag{4}$$

Low fitness solutions may have a larger E_a than high fitness solutions, simply due to the increased number of better mutations. The second evolvability metric E_b uses only the offspring fitnesses:

$$E_b = \int_{-\infty}^{\infty} f\, T(f : h, k)\, df \tag{5}$$

or the expected offspring fitness from genotype h. Note, this value is fitness dependent, so should not be compared across genotypes without reference to their original fitness. A further problem with both E_a and E_b is their dependence on the entire set of offspring fitnesses; the fraction of offspring that are significantly fitter than the parent may be extremely small. The third measure reflects this dimension of evolvability, looking only at the top C'th percentile of the offspring fitnesses:

$$\begin{aligned} E_c &= \frac{100}{C} \int_{F_c}^{\infty} f\, T(f : h, k)\, df && (6) \\ \text{where } \mathrm{F_c} \text{ defined by } \int_{F_c}^{\infty} T(f : h, k)\, df &= \frac{C}{100} && (7) \end{aligned}$$

or the expected fitness of only the top C'th percentile of fitnesses. A similar measure E_d (not shown) calculates the expected fitness of the bottom C'th percentile of offspring.

The next section extends the continuous analysis presented above to the discrete set.

3.3 Evolvability metrics: The discrete set

Consider the fitness landscape as a directed graph (V, E), with vertices V (genotypes) connected by edges E (defined by the genetic operators). The set G of offspring from a parent genotype h, k is thus defined by the vertices connected to the parent vertex:

$$G(h, k) = \{g \in V : E(h, k) = g\} \tag{8}$$

The fitness function F maps each vertex on to a single fitness, so similarly we define the set of offspring with fitness $F(g)$ greater than some fitness c:

$$G_c^+(h, k) = \{g \in V : E(h, k) = g, F(g) \geq c\} \tag{9}$$

The probability of the offspring fitness being higher or equal to the parent fitness, or E_a, is simply the fraction of the set with $F(g) \geq k$:

$$E_a = \frac{|G_k^+(h, k)|}{|G(h, k)|} \tag{10}$$

As in the previous section, the mean fitness of the offspring solutions, or E_b, is simply the mean fitness of all members of the set:

$$E_b = \frac{\sum_{g \in G(h,k)} F(g)}{|G(h, k)|} \tag{11}$$

The mean fitness of the set of offspring with fitness in the top C'th percentile is similarly defined:

$$E_c = \frac{\sum_{g \in G_{F_c}^+(h,k)} F(g)}{|G_{F_c}^+(h, k)|} \tag{12}$$

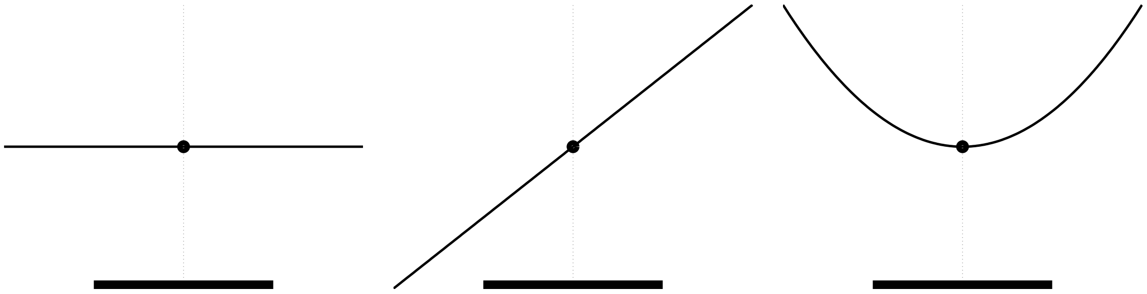
$$\text{where } \mathrm{F_c} \text{ defined by } |G^{+}_{F_c}(h,k)| \quad = \quad \frac{C\,|G(h,k)|}{100} \tag{13}$$

The mean fitness of the set of offspring with fitness in the bottom percentile can be defined through the set $G^{-}_{F_d}(h,k)$ of offspring with fitness below some fitness F_d.

The next section applies the metrics to a set of simple cases, where the parent genotypes lie at different points in a hypothetical landscape.

3.4 Simple evolvability examples

The metrics derived in the previous two sections are here applied to a set of simple cases, showing their ability to discriminate between solutions lying at different points in a hypothetical landscape, including flat plateaus, local optima, and hill-sides. Figure 3 shows three such sample cases.



(a) $F(x) = 0$ (b) $F(x) = kx$ (c) $F(x) = kx^2$

Figure 3: Three 1-dimensional landscapes $F(x)$, with the parent genotype x_0 shown by the solid circle lying at $x = 0$ (in all cases, $F(0) = 0$). The mutation operator $\mu(x, x_0)$ is a probability distribution function, producing offspring x_1 lying in a uniform distribution around x with range d, shown by the thick bar below each landscape, centred on x_0. See text for the derived evolvability in each landscape.

Through applying the continuous evolvability metrics (equations 4-7) to the continuous landscapes defined by the functions $F(x)$, with the mutation operator probability distribution around the parent solution $\mu(x, x_0)$, we obtain the following results for the evolvability of parent solutions x_0 (and offspring solutions x_1):

$$E_a \equiv P\left(F(x_1) \geq F(x_0)\right) = \int_{-\infty}^{\infty} \mu(x, x_0)\, G^{+}(x, x_0)\, dx \tag{14}$$

$$\text{where } G^{+}(a,b) = \begin{cases} 1 & F(a) \geq F(b) \\ 0 & \text{else} \end{cases} \tag{15}$$

$$E_b \equiv \, < F(x_1) > \, = \int_{-\infty}^{\infty} F(x)\, \mu(x, x_0)\, dx \tag{16}$$

$$E_c = \frac{100}{C} \int_{-\infty}^{\infty} F(x)\, \mu(x, x_0)\, G^{+}(x, F_c)\, dx \tag{17}$$

$$\text{where } \int_{-\infty}^{\infty} \mu(x, x_0)\, G^{+}(x, F_c)\, dx = \frac{C}{100} \tag{18}$$

Similarly, the mean fitness over the bottom percentile E_d, can be defined using the step function $G^{-}(a,b)$.

Table 1 gives the results for the four evolvability metrics derived on the three landscapes shown in figure 3, for the uniformly distributed mutation operator $\mu(x) = U(x_0 - d, x_0 + d)$, and a parent genotype $x_0 = 0$. No single metric gives enough information to correctly identify the nature of the fitness landscape surrounding the genotype x_0, but the combination of the four metrics allows us to do so. The evolvability data for the first landscape tells us that no mutations are deleterious, that the expected offspring fitness is equal to the current fitness, and that both the top and bottom quartile of offspring fitness are equal to the current fitness: we conclude the landscape neighbouring the current solution must be a flat plateau. We can similarly identify the hill-side and local optima landscapes from the relevant data. The approach can also be used on problems with higher dimensional landscapes, although the offspring distributions may need to be approximated through sampled applications of the mutation operator(s).

Metric	Plateau $F(x) = 0$	Hill-side $F(x) = kx$	Optima $F(x) = kx^2$
E_a	1	0.5	1
E_b	0	0	$kd^2/3$
$E_{c=25}$	0	$3kd/4$	$37kd^2/48$
$E_{d=25}$	0	$-3kd/4$	$kd^2/48$

Table 1: Derived values for the evolvability of the three sample landscapes shown in figure 3. See text for details.

In the next section, we show how the evolvability can be calculated over a population of solutions with the same fitness, to give plots predicting the difficulty of reaching certain fitness levels in a given landscape.

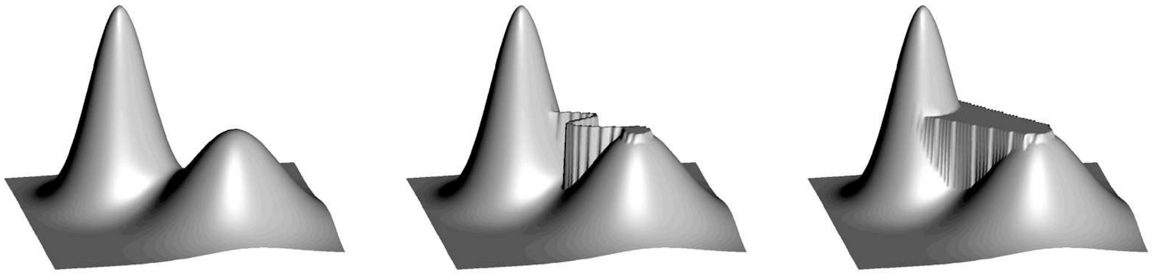
3.5 Population evolvability and fitness

The previous section described how the evolvability metrics could be calculated over the fitness neighbourhood for a single solution genotype. We define the same evolvability metrics over a sampled population of solutions, by simply defining the metrics as calculated over the sum of population transmission functions, i.e. we take the distribution of offspring fitnesses from all members of the sample and calculate the evolvability metrics. For the discrete case this translates to taking the population set of offspring defined over the combined sets of offspring from all members of the population.

Two important ideas emerge from this definition of population evolvability. First, we can compare entire populations simply by comparing their metrics of evolvability. This is not explored further in this paper, but has been used by Smith et al. (2001d,c) to investigate the behaviour of populations during neutral epochs, in particular whether the populations are moving to more evolvable areas of space during such neutral epochs. Second, we can take samples of equal fitness (in practice, we take samples of nearly equal fitness, lying in some range) to build up a portrait of the landscape at different fitnesses. For each equal fitness sample of solutions we can calculate evolvability, and plot the evolvability metric against solution fitness.

Figure 5 shows the probability of a neighbouring solution being of equal or higher fitness for all three landscapes shown in figure 4. The key feature is the behaviour at a fitness of just under 0.5, the height of the local *deceptive* peak. When the two peaks are unconnected, the probability of finding a fitter neighbour at this fitness drops sharply to nearly 40%, the lowest value seen until we reach fitnesses of nearly 100%. However, when the peaks are connected by a single neutral pathway, this value rises sharply to a value of over 60%, and when the peaks are connected by a broad plateau, the corresponding value is over 90%. Thus we see the strength of the population evolvability approach; a single problem difficulty value defined over the entire space, e.g. correlation length, would be unlikely to highlight such features. However, plotting some value over different fitnesses shows up the differences clearly.

It should be noted that the idea of plotting some difficulty metric over fitness was used by Rosé et al. (1996) in their density of states approach. However in this paper we are focusing on the potential for reaching higher fitnesses, i.e. their accessibility, rather than simply the number of solutions of given fitnesses. In many spaces it may well be that there are very small numbers of solutions of a certain fitness, but that



(a) Unconnected peaks (b) Single neutral pathway (c) Broad neutral plateau

Figure 4: Three two-dimensional model fitness landscapes showing the possible advantage of neutrality in a simple landscape with one globally-optimal and one (nearly) locally-optimal deceptive peak.

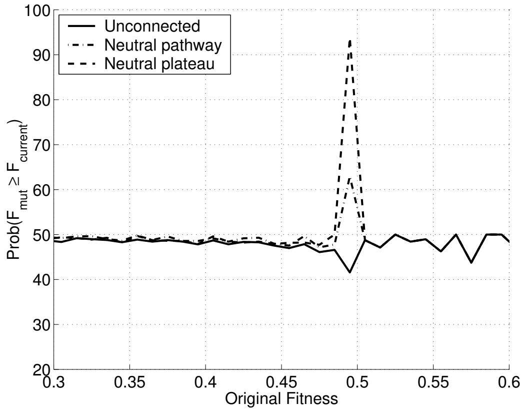


Figure 5: The probability of finding a non-deleterious mutation, E_a, over the three landscapes shown in figure 4. Over most of the fitness range, the graphs are identical. However, the increase at a fitness of 0.5, the height of the locally optimal peak, are highlighted for the neutral pathway (figure 4(b)) and neutral plateau (figure 4(b)) landscapes when compared to the unconnected peaks landscape (figure 4(a)).

those solutions are easy to reach through optimisation. Or vice-versa, there may be large numbers of solutions of a certain fitness, all of which are extremely difficult to reach through optimisation. Thus it may be more useful to look at the accessibility of solutions at certain fitness levels, rather than the number of such fitness solutions.

The next section goes on to outline the NK and terraced NK landscapes.

4 The NK and terraced NK landscapes

4.1 Kauffman's tunably rugged NK landscapes

Consider a genotype consisting of $N > 0$ loci, with each locus having some state drawn from an alphabet of size A. In addition, each locus epistatically interacts with $0 \leq K < N$ linked loci. The genotype fitness is defined as the mean fitness over the N loci, with each locus fitness uniquely determined by both its own state, and the state of the K loci it is epistatically linked to. Thus each locus has a total number of A^{K+1} assigned fitnesses (one fitness for each of the possible states for the locus and K linked loci), typically drawn from a uniform distribution over $[0, 1]$, and the entire landscape is defined by the table of NA^{K+1} fitnesses. In the remainder of this paper, we take a binary alphabet $A = 2$.

This NK landscape is a generalised version of the spin-glass model found in statistical physics (see, e.g. Binder and Young, 1986). As K increases from the unimodal $K = 0$ landscape, the *ruggedness* of the landscape increases, up to the maximally rugged random $K = N - 1$ landscape, corresponding to the random energy spin-glass model (Derrida, 1981). Early work by Weinberger (1990, 1991) has shown that increasing ruggedness of the system is well predicted by the decrease in correlation length for the system, and that the number of locally optimal peaks increases dramatically with K; this has long been the benchmark result arguing that landscape ruggedness is the key feature for problem difficulty. Figure 6 shows the correlation lengths derived from random walks in the $N = 25$, $K = [0, 1, 2, 6, 12, 18, 24]$ landscapes; increasing K clearly produces more rugged landscapes with shorter correlation lengths.

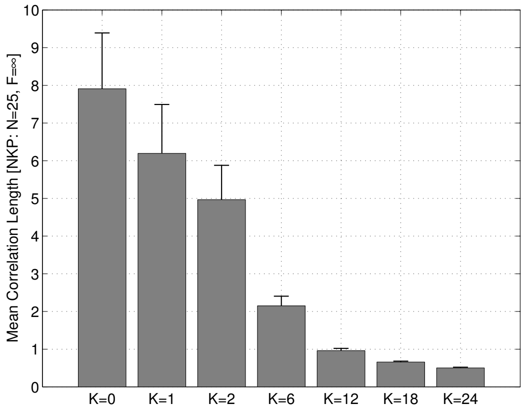


Figure 6: Correlation lengths calculated over random walks on Kauffman's NK landscapes (mean and standard deviation over 1000 walks shown). Shown for $N = 25$, $K = [0, 1, 2, 6, 12, 18, 24]$. The correlation lengths decrease with increasing K, showing the increasing ruggedness in the system.

The next section goes on to describe tunably neutral extensions to Kauffman's NK systems, developed by Barnett (1998) and Newman and Engelhardt (1998).

4.2 Tunably neutral NK landscapes

It has been argued by many researchers that other features of problem difficulty are not adequately represented by measures of landscape ruggedness. One such feature is landscape neutrality (section 2.2), and two different methods have been proposed to incorporate a tunable level of neutrality into the NK landscapes.

Barnett (1998) has argued that in real systems, many loci will not contribute to fitness, thus some fraction $0 \leq p \leq 1$ of the NA^{K+1} locus fitnesses are taken to equal zero. The neutrality of the system is thus represented by p, which can be tuned from zero (the standard NK landscape) to 1.0 (all genotypes have fitness zero). The resulting NKp landscapes are thus both tunably rugged and tunably neutral. Similarly, Reidys and Stadler (2001) derive the more general additive fitness landscape class, of which the NKp landscapes are a special case. Barnett goes on to show that changing the level of neutrality on the system has large impact on the population dynamics of evolutionary search, but produces no change in the correlation length (Barnett, 1997, 1998).

A second approach taken is to discretise the possible locus fitnesses, through only allowing each locus to take one of F fitnesses, or *terraces* (Newman and Engelhardt, 1998). Varying this terrace parameter F from an infinite value (corresponding to the standard NK landscape), down to the minimal value of 2 (where each locus has a fitness of 0 or 1) thus decreases the number of possible fitnesses, so increasing the fraction of equal fitness neutral solutions (the actual number of distinct solution fitnesses will be equal to $1 + (N - 1)(F - 1)$). Newman and Engelhardt (1998) go on to show that increasing the neutrality in the system reduces the evolutionary time required to find higher fitness solutions. In figure 8, we show that this speed difference is not predicted by the change in correlation length; the correlation length for given K is constant for all levels of neutrality.

In the remainder of this paper, we investigate the evolvability of the terraced NK landscapes, with varying degrees of ruggedness and neutrality.

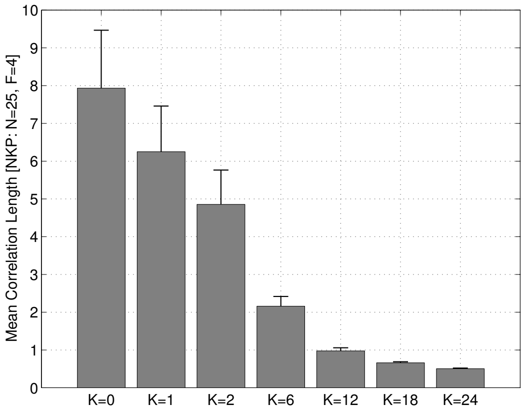
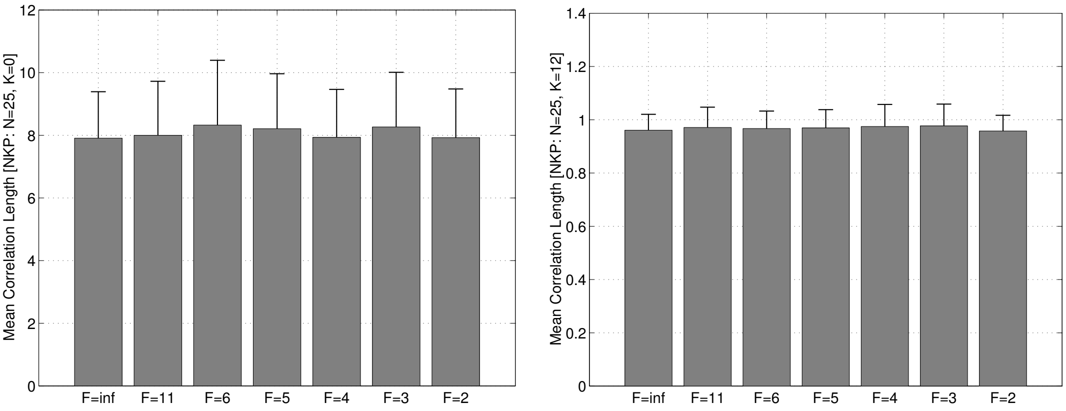


Figure 7: Correlation lengths calculated over random walks on the neutrally terraced NK landscapes (mean and standard deviation over 1000 walks shown). Shown for $N = 25$, $K = [0, 1, 2, 6, 12, 18, 24]$, $F = 4$. The correlation lengths are identical to the corresponding lengths shown in figure 6 (calculated on the non-neutral NK landscape), and decrease with increasing K, showing the increasing ruggedness in the system. Thus the correlation length measures still predict increasing search difficulty, even in the presence of significant levels of neutrality in the system.



(a) Terraced NK landscape, $K = 0$

(b) Terraced NK landscape, $K = 12$

Figure 8: Correlation lengths calculated over random walks on the neutrally terraced NK landscapes (mean and standard deviation over 1000 walks shown). Shown for $N = 25$, $K = [0, 12]$, $F = [\infty, 11, 6, 5, 4, 3, 2]$. The correlation lengths are equal for a given K, despite the varying amount of neutrality in the system (shown by the varying number of neutral terraces F). Thus correlation lengths do not predict the difference in time required to find high fitness solutions.

5 Evolvability of a tunably rugged landscape

For a given genotype length N, the overall distribution of *genotype* fitnesses will be identical for all K, equal to the normal distribution $N(M_N, \sigma_N^2)$ obtained from the mean of N *loci* fitnesses sampled from a uniform $[0, 1]$ distribution (M_N and σ_N^2 are the mean and variance respectively). As N increases, the mean of this genotype fitness distribution M_N will stay constant at $M_N = 0.5$, but the standard deviation will decrease with the number of samples: $\sigma_N = \sigma/\sqrt{N}$. Thus, as our genotypes increase in length, the fraction of extremely fit and unfit genotypes will decrease, as the variance of the population decreases (note that some NK models address this issue through scaling the genotype fitnesses by $\sqrt{N}$ (see e.g. Sibani and Pederson, 1999)).

Although the overall distribution of genotype fitnesses is equal for all K over a given genotype length N, the distribution of these fitnesses over the search space is decidedly not equal for all K. For the unimodal $K = 0$ landscape, all high fitness genotypes are grouped in a small volume of the search space, while for the maximally multi-modal $K = N - 1$ landscape, the fitness distribution over the search space is random. In general, the distribution of fitnesses neighbouring a solution of given fitness is normal with mean and deviation dependent on N, K and the current solution fitness (see e.g. Weinberger, 1990; Stadler and Schnabl, 1992). From this it is possible to derive the expected fitnesses (and the time taken on both adaptive and random walks) at which local optima are reached for various N and K (again, see e.g. Weinberger, 1990; Stadler and Schnabl, 1992). In the next section we derive analytic and empirical results for the evolvability measures when applied to the NK landscapes.

5.1 Analytically derived evolvability for NK landscapes

In this section, we focus on the probability that an offspring derived from a single bit mutation of the parent has a higher (or equal) fitness than the parent, i.e. the first evolvability metric E_a (section 3.1), as a function of the parental fitness. The other evolvability metrics derived in section 3 can be similarly treated. Consider a parent genotype of fitness F_0, the mean of the N locus fitnesses f_i drawn from a uniform distribution over $[0, 1]$:

$$F_0 = \frac{1}{N}\sum_{i=1}^{N} f_i \text{ where } f_i \in U[0,1] \tag{19}$$

Now, the probability that the offspring fitness F_1 is not lower than the parent fitness, is simply the probability that the $K + 1$ loci affected by a single bit mutation do not, on average, decrease in fitness:

$$E_a \equiv P(F_1 \geq F_0) = P\left(\left(\frac{1}{K+1}\sum_{i=1}^{K+1} f_i\right) \geq F_0\right) \text{ where } f_i \in U[0,1] \tag{20}$$

the probability that the mean of $K + 1$ uniformly distributed samples is not smaller than the current fitness. For the unimodal $K = 0$ we can solve trivially:

$$P(F_1 \geq F_0) = P\left(f \geq F_0\right) = 1 - F_0 \text{ where } f_i \in U[0,1] \tag{21}$$

For $K \gg 0$, the mean of affected loci fitnesses tends to a normal distribution with mean $M_{K+1} = 0.5$, and deviation $\sigma_{K+1} = \sigma/\sqrt{K+1}$ (where σ is the deviation of loci fitnesses as $K \rightarrow \infty$, assumed to be non-zero and finite). For a normal distribution $N(M, \sigma^2)$, the probability density function n is given by:

$$\begin{aligned} n &= \frac{1}{\sigma}\phi\left(\frac{x-M}{\sigma}\right) && (22) \\ \text{where } \phi(z) &= \frac{1}{\sqrt{2\pi}}\exp\left(-0.5z^2\right) && (23) \end{aligned}$$

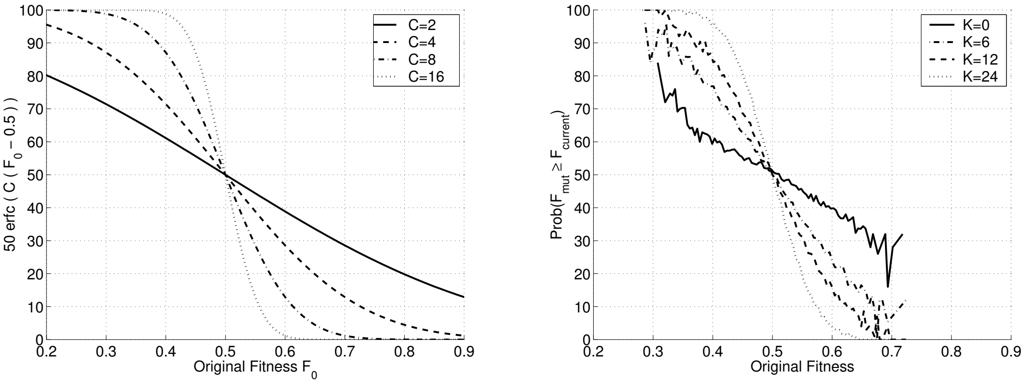
The probability that a random variable from this distribution is greater than some value F_0 is given by the integral of the probability density function over the relevant limits (with mean $M = 0.5$ and deviation $\sigma_{K+1} = \sigma/\sqrt{K+1}$):

$$\begin{aligned} P(F_1 \geq F_0) &= \frac{\sqrt{K+1}}{\sigma}\int_{F_0}^{\infty}\phi\left(\frac{\sqrt{K+1}(x-0.5)}{\sigma}\right)dx && (24) \\ &= \sqrt{\frac{K+1}{2\pi\sigma^2}}\int_{F_0}^{\infty}\exp\left(\frac{-(K+1)}{2\sigma^2}(x-0.5)^2\right)dx && (25) \end{aligned}$$

Which is simply the complementary error function $\text{erfc}(x)$:

$$\mathrm{erfc}(x) \equiv \frac{2}{\sqrt{\pi}}\int_x^{\infty} \exp\left(-z^2\right) dz \tag{26}$$

$$\text{and } P(F_1 \geq F_0) = 0.5\ \mathrm{erfc}\left(\sqrt{\frac{K+1}{2\sigma^2}}\,(F_0 - 0.5)\right) \tag{27}$$



(a) Analytically derived probability of a non-deleterious mutation, E_a

(b) Empirically derived probability of a non-deleterious mutation, E_a

Figure 9: Analytically and empirically derived probabilities of a non-deleterious mutation, E_a. Calculated on the NK landscape with $N = 25$, $K = [0, 6, 12, 24]$, for a random sample set of solutions. All random sample sets used in this paper consist of 1000 individuals sampled from each of 100 generated landscapes; a total of 100,000 sampled solutions. Note that for each set of N, K, there is an infinite number of fitness lookup tables that can be generated, thus an infinite number of possible landscapes. For this reason, we sample both a set of individuals and a set of landscapes for each value of N, K.

Note that an equivalent result to equation 27 is derived by Stadler and Schnabl (1992) in order to calculate the probability of solutions of given fitness being local optima.

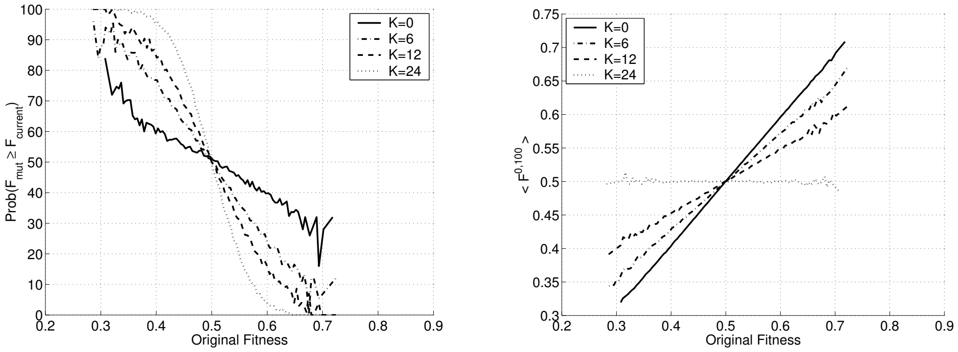
5.2 Empirically derived evolvability for NK landscapes

Figure 9 shows data generated from equation 27 compared against data derived from empirical sampling of simulated NK landscapes ($N = 25$, $K = [0, 1, 2, 6, 12, 24]$), showing good agreement between the analytically and empirically derived data. Both sets of data predict that as K increases, the probability of finding a fitter mutant *increases* for parent fitnesses below the population mean of 0.5. Only for parent fitnesses above this mean value of 0.5 does the probability of reaching a fitter mutant favour the lower K landscapes. This can be understood by considering that a single bit flip mutation can affect the fitness F_0 by a fraction of order $O(\frac{K+1}{N})$. Low K landscapes are thus highly correlated[2], and offspring fitnesses are close to parent fitnesses. For high K landscapes, the offspring-parent fitnesses are less correlated, thus offspring fitnesses *on average* are close to the population mean of 0.5, and the distribution of genotype fitnesses is essentially random in space.

The other evolvability measures can be derived in similar fashion, and give good agreement with the data derived from empirical simulation of the NK landscapes. Figure 10 shows empirical data for the evolvability metrics E_a, E_b, E_c, E_d. We see that the expected fitness is higher for the $K = 0$ landscape only with parent solution fitnesses above 0.5, strengthening the probability metric results derived above.

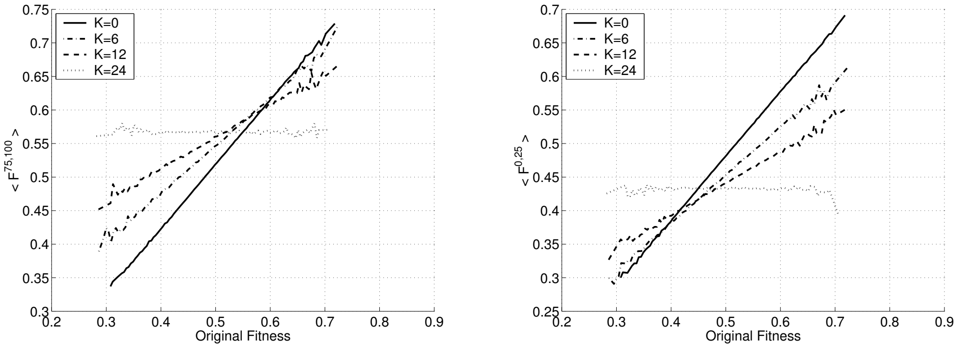
---

[2] The offspring-parent correlation is simply $\rho = 1 - \frac{K+1}{N}$ (Weinberger, 1990), with the correlation length $\tau = -1/\ln(\rho)$.



(a) Probability of a non-deleterious mutation, E_a

(b) Expected fitness over all mutations, E_b



(c) Expected fitness over top quartile of mutations, E_c

(d) Expected fitness over bottom quartile of mutations, E_d

Figure 10: The metrics of evolvability derived in section 3.1 applied to the NK landscapes with $N = 25$, $K = [0, 6, 12, 24]$. The evolvability metrics were calculated from a random sample set of solutions.

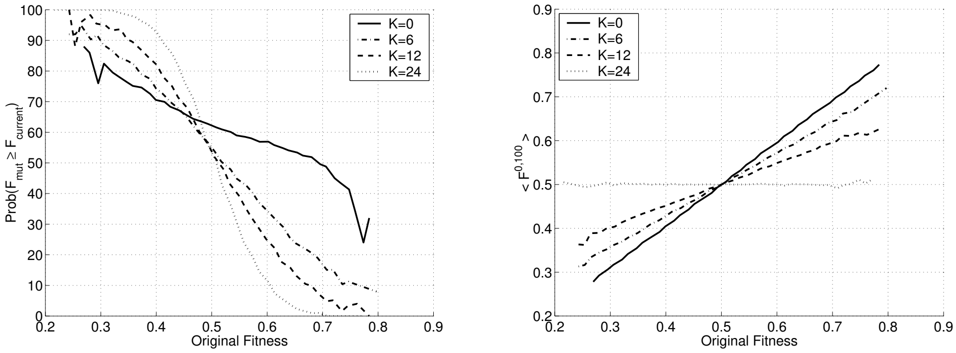
Similarly, the top and bottom quartile expected fitnesses are higher for $K = 0$ only for high fitness parent solutions. Thus the evolved metrics derived in section 3.1 do indeed predict the difficulty of searching in the NK landscapes, and in fact give more information than the correlation lengths alone. From the metrics shown in figure 10, we predict that fitnesses below 0.5 will be found more easily through search in highly-rugged NK landscapes (although of course these fitnesses will also be easily found through random search). We further predict that reaching solutions of fitness above 0.5 will be significantly easier in the $K = 0$ landscape.

The next section applies the evolvability analysis to the tunably neutral terraced NK landscapes.

6 Evolvability of a tunably neutral landscape

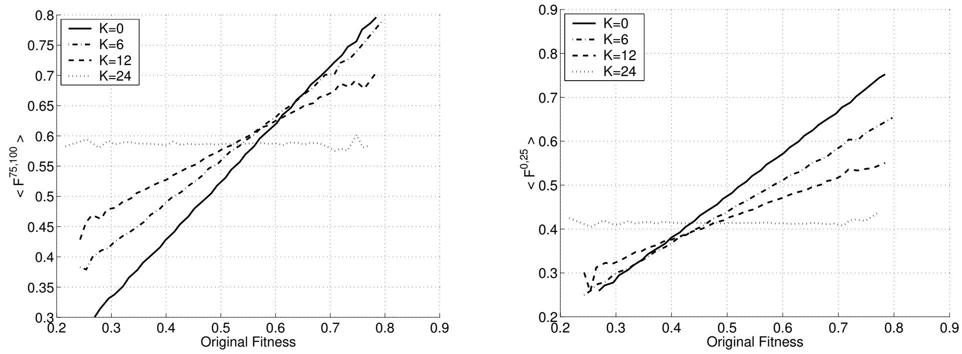
In the previous section, we have seen how we can discriminate between landscapes of varying ruggedness using the metrics of evolvability derived in section 3.1. In this section, we apply the metrics to the tunably neutral terraced NK landscape (section 4), in order to discriminate between landscapes of varying ruggedness in the presence of neutrality, and landscapes of varying neutrality.

Figure 11 shows the four metrics applied to the terraced NK landscape, with the number of possible loci



(a) Probability of a non-deleterious mutation, E_a

(b) Expected fitness over all mutations, E_b



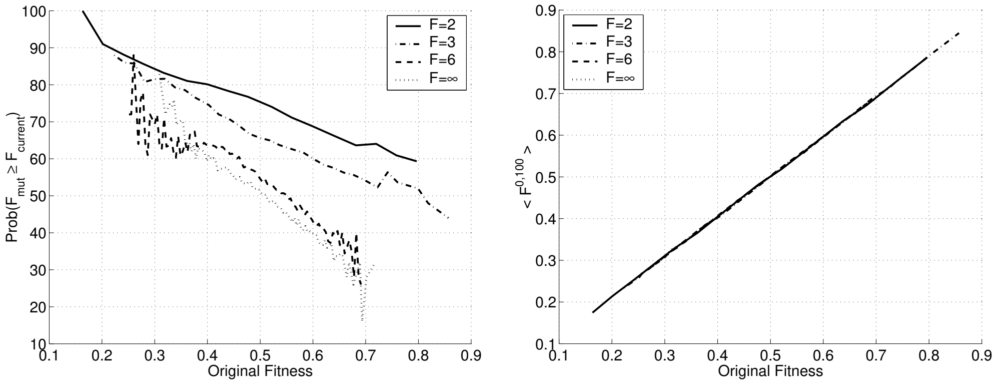
(c) Expected fitness over top quartile of mutations, E_c

(d) Expected fitness over bottom quartile of mutations, E_d

Figure 11: The metrics of evolvability derived in section 3.1 applied to the terraced NK landscapes with $N = 25$, $K = [0, 6, 12, 24]$, $F = 4$. The evolvability metrics were calculated from a random sample set of solutions.

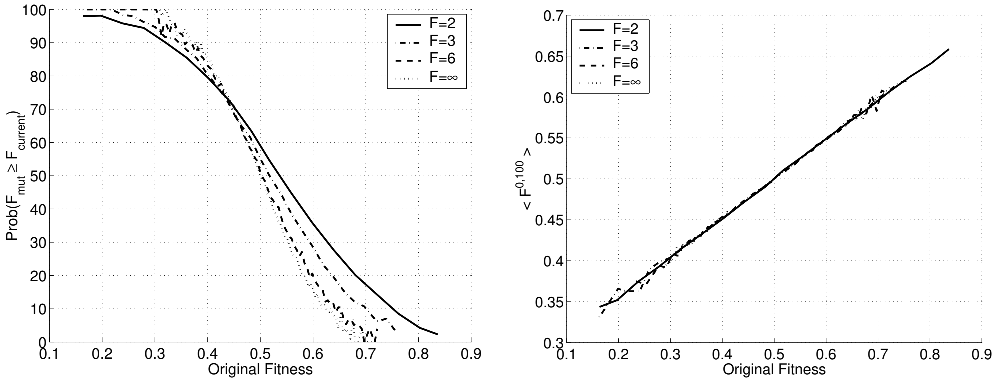
fitness terraces fixed at $F = 4$ ($N = 25$ and $K = [0, 6, 12, 24]$). We see that the presence of a large degree of neutrality in the system does not affect the ability of the evolvability metrics to distinguish between differing levels of ruggedness in the system, as seen with correlation lengths (figure 7).

Figure 12 shows the metrics applied to the terraced NK landscapes with varying neutrality and fixed ruggedness (the probability of obtaining a non-deleterious mutation, and the expected fitness of all mutations, for $N = 25$, $K = [0, 12]$, $F = [2, 3, 6, \infty]$). We see that the expected mutation fitness (figures 12(b) and 12(d)) does not change with differing levels of neutrality in the system, but that the probability of obtaining a non-deleterious mutation (figures 12(a) and 12(c)) does show such change. As $F \to 2$, neutrality increases and the probability of obtaining a non-deleterious mutation increases. For $K = 0$, even at high fitnesses there are still on average $1/F$ neutral mutations. For high K, this probability tends to zero at high fitnesses as all $K+1$ loci fitness affected by the mutation need to show a neutral mutation. However, this decrease in the probability of finding non-deleterious mutations is slower for landscapes with more neutrality. The difference is significant: at a fitness of 0.6, roughly 37% of mutations in the $F = 2$ landscape are non-deleterious compared with roughly 13% such mutations for the non-neutral $F = \infty$ landscape. At a fitness of 0.7, the corresponding percentages are roughly 18% and 0%. Thus, in the highly neutral $F = 2$ landscape, the probability of the search process reaching a local optimum is significantly smaller than the probability of reaching a local optimum in the non-neutral $F = \infty$. Rather than sticking in local optima, the search process can explore more of the space along neutral networks, eventually reaching higher fitness solutions.



(a) Probability of a non-deleterious mutation, E_a, for $K = 0$

(b) Expected fitness over all mutations, E_b, for $K = 0$



(c) Probability of a non-deleterious mutation, E_a, for $K = 12$

(d) Expected fitness over all mutations, E_b, for $K = 12$

Figure 12: The metrics of evolvability derived in section 3.1 applied to the terraced NK landscapes with $N = 25$, $K = [0, 12]$, $F = [2, 3, 6, \infty]$. The evolvability metrics were calculated from a random sample set of solutions.

Thus the metrics of evolvability do indeed differentiate between landscapes of both varying ruggedness (with constant neutrality) and varying neutrality (with constant ruggedness). In particular, the metrics make three main predictions backed up by empirical evidence (Weinberger, 1990, 1991; Newman and Engelhardt, 1998). First, they predict that for NK landscapes of different ruggedness, it will be easier to reach fitnesses below 0.5 in highly rugged landscape. Second, fitnesses above 0.5 in such landscapes will be easier to reach in a non-rugged landscape. Third, increasing levels of neutrality in the terraced NK system are not seen with measures of correlation length, but are seen in the evolvability metrics, in particular the numbers of non-deleterious mutations are significantly higher for more neutral landscapes.

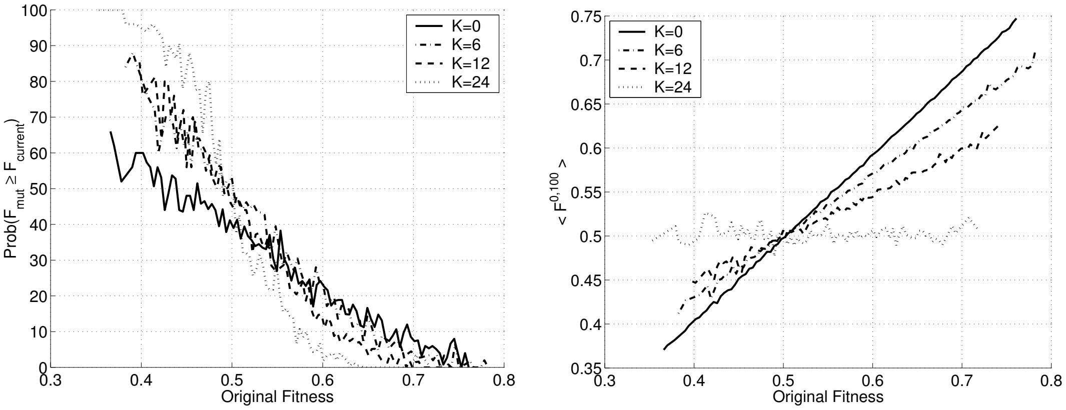
In the next section, we show that the evolvability of samples collected during simple hill-climbing optimisation show the same features as for the random samples used in the previous two sections. This is crucial for problems with extremely skewed solution fitness distributions, for which random sampling is inappropriate and biased sampling techniques must be used. If difficulty measures are to be applied to such problems, they must be robust when calculated over such biased samples.

7 Online sampling evolvability

In the previous sections we have investigated empirically derived evolvabilities for the tunably rugged and tunably neutral terraced NK landscapes through random sampling of the space of all solutions. This random sampling technique works well with the NK landscapes where solution fitnesses are defined as the linear sum of all loci fitnesses; due to the central limit theorem, the solution fitnesses will be normally distributed. However, in many problems, such normally distributed solution fitnesses will not be encountered, and metrics based on random sampling of the space will in general be less successful at predicting problem difficulty (see e.g. Smith et al., 2001a).

With such skewed solution fitness distributions, it may be necessary to bias the collected sample through only keeping a percentage of solutions found at each fitness, and define the problem difficulty metric over this biased sample. With even more extremely skewed distributions, it may be necessary to collect a biased sample through some direct search optimisation procedure such as a simple hill-climber. For example, Smith et al. (2001a) find only 0.0001% of random solutions have fitness above 50% of the maximum in a neural network robot control problem, despite this fitness being relatively easy to reach using optimisation techniques.

In this section, we show that the evolvability metrics presented in the previous sections still predict the difficulty of search in the terraced NK landscapes, when based on a biased sample collected using a (1+1) evolutionary strategy hill-climber (Rechenberg, 1973). 100 runs of the hill-climber were performed for each parameter setting (generating a new landscape for each run); from an initial random solution, random mutations were applied (using both single bit mutation, and mutation probability per bit gives similar results), with non-deleterious mutations accepted and deleterious mutations rejected. All new encountered genotypes were saved for analysis, and the hill-climber stopped after $1,000$ mutations had been tried. The following analysis uses the saved samples over each parameter setting.



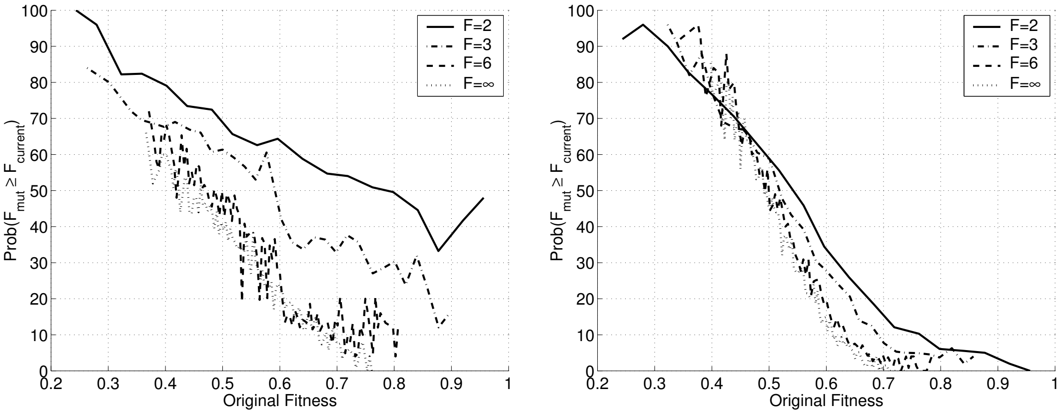
(a) Probability of a non-deleterious mutation, E_a

(b) Expected fitness over all mutations, E_b

Figure 13: The metrics of evolvability derived in section 3.1 applied to the NK landscapes with $N = 25$, $K = [0, 6, 12, 24]$. The evolvability metrics were calculated from a sample set of solutions collected during hill-climbing.

Figure 13 shows the probability of a non-deleterious mutation, and the expected mutation fitness over the NK landscape with $N = 25$, $K = [0, 6, 12, 24]$, for the biased hill-climber sample. As seen with the random sample analysis (section 5), the $K = 0$ landscape shows the highest probability of obtaining a fitter mutation, and the highest expected fitness, only for fitnesses above 0.5. However, the differences are not as significant as for the unbiased sample; the $K = 6$ landscape shows similar values to the $K = 0$ landscape. We hypothesise that this is due to the hill-climber moving towards areas of the $K = 6$

landscape which are more evolvable than on average across the space.



(a) Probability of a non-deleterious mutation, E_a, for $K = 0$

(b) Probability of a non-deleterious mutation, E_a, for $K = 12$

Figure 14: The metrics of evolvability derived in section 3.1 applied to the terraced NK landscapes with $N = 25$, $K = [0, 12]$, $F = [2, 3, 6, \infty]$. The evolvability metrics were calculated from a sample set of solutions collected during hill-climbing.

Figure 14 shows the probability of a non-deleterious mutation for the terraced NK landscapes with $N = 25$, $F = [2, 3, 6, \infty]$, $K = [0, 12]$, for the biased hill-climber sample. Again we see the same results as for the unbiased random sample (section 6); as expected the more neutral landscapes with lower numbers of terraces F show greater probability of non-deleterious mutations. This holds for landscapes of both low and high ruggedness.

The results for the biased hill-climber sample over both the tunably rugged (figure 13) and the tunably neutral landscapes (figure 13) are equivalent to the results seen for the unbiased random sample used in sections 5 and 6. The evolvability metrics are indeed robust over the use of biased and unbiased samples. In the final section, we discuss the results and relevance of the research presented in this paper.

8 Discussion

In this paper we have developed a set of measures based on evolvability statistics of the fitness landscape surrounding sampled solutions. Averaging the measures over a sample of equal fitness solutions allows us to build up a fitness portrait of the fitness landscape, which we have shown can be used to predict the difficulty of finding good solutions in a set of tunably rugged and tunably neutral landscapes.

The primary aim of this work is to develop a set of metrics that can be used to predict the difficulty of finding good solutions in hard problem spaces. In previous work we have shown that standard metrics based on random sampling of the space, such as the autocorrelation function, fail when applied to spaces with highly skewed fitness distributions (Smith et al., 2001a). Although such techniques can be used with biased sample sets, it is unclear to what extent their underlying assumptions are violated by the heterogenous anisotropic nature of certain problem spaces, e.g. evolutionary robotics landscapes. It is also unclear how such metrics might be applied to samples of unconnected solutions, i.e. not collected through successive applications of mutation operators, when there is no clear way of defining distance between two solutions (in the evolutionary robotics problems we are investigating, solution representations are not of fixed length, and the same solution may be represented by several permuted genotypes). To this end, this paper presents the development of a set of metrics that can be easily averaged over the search space, without having to define a distance metric.

A second aim of the work is to develop a set of metrics that can be used to identify the areas of space which are difficult for optimisation to reach. In the sort of problem space exemplified by evolutionary robotics and evolutionary hardware experiments, it may be very simple to reach certain fitness levels (even if such levels of fitness are not represented by random search), but significantly harder to reach higher levels through optimisation. Although simply looking at the distribution of fitnesses may give us some idea of where these fitness levels are, it will say nothing about the *accessibility* of such fitness levels. As shown in this paper, the use of biased samples collected through simple hill-climbing allows us to focus directly on the accessibility of solutions, rather than simply their density over the search space. Identification of inaccessible "bottleneck" fitness levels will certainly be of use when designing more evolvable representations with which it proves easier to find good solutions.

References

Altenberg, L. (1994). The evolution of evolvability in genetic programming. In Kinnear Jr, K., editor, *Advances in Genetic Programming*, chapter 3, pages 47–74. MIT Press.

Barnett, L. (1997). Tangled webs - evolutionary dynamics on fitness landscapes with neutrality. Master's thesis, School of Cognitive and Computing Sciences, University of Sussex.

Barnett, L. (1998). Ruggedness and neutrality: The NKp family of fitness landscapes. In Adami, C., Belew, R., Kitano, H., and Taylor, C., editors, *Artificial Life VI: Proceedings of the Sixth International Conference on Artificial Life*, pages 18–27. MIT Press / Bradford Books.

Binder, K. and Young, A. (1986). Spin glasses: Experimental facts, theoretical concepts and open questions. *Reviews of Modern physics*, 58(4):801–976.

Burch, C. and Chao, L. (2000). Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature*, 406:625–628.

Cavalli-Sforza, L. and Feldman, M. (1976). Evolution of continuous variation: Direct approach through joint distribution of genotypes and phenotypes. *Proceedings of the National Academy of Sciences, USA*, 73:1689–1692.

Coyne, J., Barton, N., and Turelli, M. (1997). Perspective: A critique of Sewall Wright's shifting balance theory of evolution. *Evolution*, 51(3):643–671.

Davidor, Y. (1991). Epistasis variance: A viewpoint on GA-hardness. In Rawlins, G., editor, *Foundations of Genetic Algorithms (FOGA'1)*, pages 22–35. Morgan Kaufmann.

Dawkins, R. (1989). The evolution of evolvability. In Langton, C., editor, *Artificial Life: Proceedings of the Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*, volume VI of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 201–220. Addison-Wesley, Redwood, CA.

Derrida, B. (1981). Random energy model - an exactly solvable model of disordered-systems. *Physical Review B*, 24(5):2613–2626.

Ebner, M., Langguth, P., Albert, J., Shackleton, M., and Shipman, R. (2001). On neutral networks and evolvability. In *Proceedings of the 2001 Congress on Evolutionary Computation: CEC2001*, pages 1–8. IEEE, Seoul, Korea.

Eldredge, N. and Gould, S. (1972). Punctuated equilibria: An alternative to phyletic gradualism. In Schopf, T., editor, *Models in Paleobiology*, pages 82–115. Freeman, San Francisco.

Elena, S., Cooper, V., and Lenski, R. (1996). Punctuated evolution caused by selection of rare beneficial mutations. *Science*, 272:1802–1804.

Fogel, L., Owens, A., and Walsh, M. (1966). *Artificial Intelligence through Simulated Evolution.* John Wiley, New York.

Forst, C., Reidys, C., and Weber, J. (1995). Neutral networks as model-landscapes for RNA secondary-structure folding landscapes. In Moran, F., Moreno, A., Merelo, J., and Cachon, P., editors, *Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life, ECAL95*, pages 128–147. Springer.

Gould, S. and Eldredge, N. (1977). Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology*, 3:115–151.

Grüner, W., Giegerich, R., Strothmann, D., Reidys, C., Weber, J., Hofacker, I., Stadler, P., and Schuster, P. (1996). Analysis of RNA sequence structure maps by exhaustive enumeration: I Neutral networks, II Structures of neutral networks and shape space covering. *Monathefte Chem.*, 127:355–374, 375–389.

Harvey, I. and Thompson, A. (1996). Through the labyrinth evolution finds a way: A silicon ridge. In Higuchi, T., editor, *Proceedings of the First International Conference on Evolvable Systems: From Biology to Hardware (ICES'96)*, pages 406–422. Springer-Verlag.

Holland, J. (1992). *Adaptation in Natural and Artificial Systems.* MIT Press / Bradford Books, 2nd edition.

Hordijk, W. (1996). A measure of landscapes. *Evolutionary Computation*, 4(4):335–360.

Huynen, M. (1996). Exploring phenotype space through neutral evolution. *Journal of Molecular Evolution*, 43:165–169.

Huynen, M., Stadler, P., and Fontana, W. (1996). Smoothness within ruggedness: The role of neutrality in adaptation. *Proceedings of the National Academy of Sciences, USA*, 93:394–401.

Jones, T. and Forrest, S. (1995). Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In Eshelmann, L., editor, *Proceedings of the Sixth International Conference on Genetic Algorithms (ICGA95)*, pages 184–192. Morgan Kaufmann, CA.

Kauffman, S. (1993). *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford University Press.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution.* Cambridge University Press.

Kirschner, M. and Gerhart, J. (1998). Evolvability. *Proceedings of the National Academy of Sciences, USA*, 95:8420–8427.

Koza, J. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* MIT Press / Bradford Books.

Lawler, E., Lenstra, J., Rinnooy Kan, A., and Schmoys, D. (1985). *The Travelling Salesman Problem.* John Wiley, New York.

Lipsitch, M. (1991). Adaptation on rugged landscapes generated by iterated local neighboring genes. In Belew, R. and Booker, L., editors, *Proceedings of the Fourth International Conference on Genetic Algorithms (ICGA91)*, pages 128–135. Morgan Kaufmann.

Manderick, B., Weger, M., and Spiessens, P. (1991). The genetic algorithm and the structure of the fitness landscape. In Belew, R. and Booker, L., editors, *Proceedings of the Fourth International Conference on Genetic Algorithms (ICGA91)*, pages 143–150. Morgan Kaufmann.

Marr, D. (1976). Artificial intelligence - a personal view. Technical Report AIM 355, MIT AI Lab.

Marrow, P. (1999). Evolvability: Evolution, computation, biology. In Wu, A., editor, *Proceedings of the 1999 Genetic and Evolutionary Computation Conference Workshop Program (GECCO-99 Workshop on Evolvability)*, pages 30–33.

Naudts, B. and Kallel, L. (2000). A comparison of predictive measures of problem difficulty in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 4(1):1–15.

Newman, M. and Engelhardt, R. (1998). Effects of selective neutrality on the evolution of molecular species. *Proceedings of the Royal Society of London, B*, 265:1333–1338.

Provine, W. (1986). *Sewall Wright and Evolutionary Biology.* Chicago Press.

Rechenberg, I. (1973). *Evolutionsstrategie.* Friedrich Frommann Verlag, Stuttgart.

Reidys, C. and Stadler, P. (2001). Neutrality in fitness landscapes. *Applied Mathematics and Computation*, 117(2-3):321–350.

Rosé, H., Ebeling, W., and Asselmeyer, T. (1996). The density of states - a measure of the difficulty of optimization problems. In Voigt, H.-M., Ebeling, W., Rechenberg, I., and Schwefel, H.-P., editors, *Proceedings of the Fourth Conference on Parallel Problem Solving from Nature: PPSN IV*, volume 1141 of *Lecture Notes in Computer Science*, pages 208–217. Springer-Verlag.

Shackleton, M., Shipman, R., and Ebner, M. (2000). An investigation of redundant genotype-phenotype mappings and their role in evolutionary search. In *Proceedings of the 2000 Congress on Evolutionary Computation: CEC2000*, pages 493–500. IEEE, San Diego, USA.

Sibani, P. and Pederson, A. (1999). Evolution dynamics in terraced NK landscapes. *Europhysics Letters*, 48:34–52.

Smith, T., Husbands, P., and O'Shea, M. (2001a). *Not* measuring evolvability: Initial exploration of an evolutionary robotics search space. In *Proceedings of the 2001 Congress on Evolutionary Computation: CEC2001*, pages 9–16. IEEE, Seoul, Korea.

Smith, T., Husbands, P., and O'Shea, M. (2001b). Evolvability, neutrality, and search difficulty. Submitted to *BioSystems* special issue on evolvability.

Smith, T., Husbands, P., and O'Shea, M. (2001c). Neutral networks and evolvability. In *5th European Conference on Artificial Life, ECAL2001.* Springer. To appear.

Smith, T., Husbands, P., and O'Shea, M. (2001d). Neutral networks in an evolutionary robotics search space. In *Proceedings of the 2001 Congress on Evolutionary Computation: CEC2001*, pages 136–145. IEEE, Seoul, Korea.

Sniegowski, P., Gerrish, P., and Lenski, R. (1997). Evolution of high mutation rates in experimental populations of *E. coli*. *Nature*, 387:703–705.

Stadler, P. (1996). Landscapes and their correlation functions. *Journal of Mathematical Chemistry*, 20:1–45.

Stadler, P. and Schnabl, W. (1992). The landscape of the traveling salesman problem. *Physics Letter A*, 161:337–344.

Stadler, P. and Wagner, G. (1998). Algebraic theory of recombination spaces. *Evolutionary Computation*, 5:241–275.

Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P., and Godelle, B. (1997). Role of mutator alleles in adaptive evolution. *Nature*, 387:700–702.

Thompson, A. (2001). Neutrality in evolutionary hardware experiments. In preparation.

Turney, P. (1999). Increasing evolvability considered as a large-scale trend in evolution. In Wu, A., editor, *Proceedings of the 1999 Genetic and Evolutionary Computation Conference Workshop Program (GECCO-99 Workshop on Evolvability)*, pages 43–46.

van Nimwegen, E., Crutchfield, J., and Huynen, M. (1999). Neutral evolution of mutational robustness. *Proceedings of the National Academy of Sciences, USA*, 96:9716–9720.

Vassilev, V. (1997). Information analysis of fitness landscapes. In Husbands, P. and Harvey, I., editors, *Proceedings of the Fourth European Conference on Artificial Life: ECAL97*, pages 116–124. MIT Press / Bradford Books.

Vassilev, V., Fogarty, T., and Miller, J. (2000). Information characteristics and the structure of landscapes. *Evolutionary Computation*, 8(1):31–60.

Vassilev, V. and Miller, J. (2000). The advantages of landscape neutrality in digital circuit evolution. In Miller, J., Thompson, A., Thomson, P., and T., F., editors, *Proceedings of the Third International Conference on Evolvable Systems: From Biology to Hardware (ICES'2000)*, volume 1801 of *Lecture Notes in Computer Science*, pages 252–263. Springer-Verlag.

Wagner, G. and Altenberg, L. (1996). Complex adaptations and the evolution of evolvability. *Evolution*, 50(3):967–976.

Weinberger, E. (1990). Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological Cybernetics*, 63:325–336.

Weinberger, E. (1991). Local properties of Kauffman's N-K model: A tunably rugged energy landscape. *Physical Review A*, 44(10):6399–6413.

Wilke, C. (2001). Adaptive evolution on neutral networks. *Bulletin of Mathematical Biology*. Submitted.

Wilke, C., Wang, J., Ofria, C., Lenski, R., and Adami, C. (2001). Evolution of digital organisms lead to survival of the flattest. *Nature*. Submitted.

Wright, S. (1932). The role of mutation, inbreeding, crossbreeding and selection in evolution. In Jones, D., editor, *Proceedings of the Sixth International Congress on Genetics*, volume 1, pages 356–366. Reprinted in Ridley, M. (1997). *Evolution*. pp32-40. Oxford University Press.