A Comparison of Neutral Landscapes – NK, NKp and NKq

Nicholas Geard[1] (nic@itee.uq.edu.au)
Janet Wiles[1, 2] (j.wiles@itee.uq.edu.au)
Jennifer Hallinan[3] (j.hallinan@imb.uq.edu.au)
Bradley Tonkes[1] (btonkes@itee.uq.edu.au)
Ben Skellett[1] (ben@itee.uq.edu.au)

[1]School of Information Technology and Electrical Engineering
[2]School of Psychology
[3]Institute for Molecular Biosciences
The University of Queensland, Brisbane Q 4072, Australia

Abstract – Recent research in molecular evolution has raised awareness of the importance of selective neutrality. Several different models of neutrality have been proposed based on Kauffman's well-known NK landscape model. Two of these models, NKp and NKq, are investigated and found to display significantly different structural properties. The fitness distributions of these neutral landscapes reveal that their levels of correlation with non-neutral landscapes are significantly different, as are the distributions of neutral mutations.

In this paper we describe a series of simulations of a hill climbing search algorithm on NK, NKp and NKq landscapes with varying levels of epistatic interaction. These simulations demonstrate differences in the way that epistatic interaction affects the 'searchability' of neutral landscapes.

We conclude that the method used to implement neutrality has an impact on both the structure of the resulting landscapes and on the performance of evolutionary search algorithms on these landscapes. These model-dependent effects must be taken into consideration when modelling biological phenomena.

I. INTRODUCTION

In recent years, a new view of evolutionary dynamics based on neutral mutations and subsequent genetic drift has arisen to challenge the traditional 'hill climbing' paradigm. In Darwin's view of evolution, any mutation to an organism's genotype resulted in a change in fitness which would be selected for or against. The new view has been prompted by research into molecular evolution that suggests that a large proportion of mutations at a molecular level are selectively neutral [9]. The evolutionary computation community is re-evaluating the idea that the end result of evolution is a population converged upon a local optimum in its fitness landscape.

The new view proposes that population clusters drift genetically across layers of uniform fitness until such time as an individual chances upon a rare beneficial mutation. This advantageous mutation lifts it to a higher fitness level and gives its offspring a selective advantage that will eventually spread throughout the population resulting in a rise in average population fitness. Much of the inspiration for the investigation of neutral landscapes has come from research into the structure of RNA folding landscapes in molecular biology [4, 7]. The discovery of neutrality in the search spaces of several real world problems, such as the evolution of neural nets and hardware evolution has further established the validity of this approach [6]. While it is clear that there is a potential role for neutrality in search and optimisation applications, the emphasis in this study was on neutrality in relation to biological modelling.

Several models of neutrality have been proposed that generalise Kauffman's NK landscape model by adding a tuneable level of neutrality. These include the NKp ('probabilistic' NK) model developed by Barnett [1] and another model developed by Newman & Engelhardt [10], which we term, analogously, NKq ('quantised' NK). The NKp and NKq models, while both based on the NK landscape model, employ distinctly different approaches to modelling neutrality. Due to these differences, the fitness landscapes generated by the two neutral models should be expected to display significantly different structural properties. These different landscape models have not previously been compared.

In this paper we examine several structural properties of landscapes both with and without neutrality. Initially, the distribution of fitness values and distribution of neutral mutations are investigated. Next, the results of a series of simulations of a simple search algorithm on a variety of NK, NKp and NKq landscapes are reported. Finally, we discuss these results and their implications for our understanding of evolutionary search spaces.

II. SELECTION AND NEUTRALITY

Traditionally, natural selection combined with variability in individual fitness has been held to be the driving force behind evolution [2]. It was thought that differences between individuals at the genotypic level were reflected at the phenotypic level and hence in an individual's reproductive fitness. A useful metaphor for evolutionary search was the concept of a *fitness landscape* – a high dimensional search space based on the set of all possible genotypes in which the 'height' of a particular point is determined by the genotype's fitness [14].

The view that each mutation results in a change to fitness leads to fitness landscapes that are visualised as rugged and hilly [8] – hence the hill climbing paradigm came to dominate thinking about evolutionary dynamics. It was thought that selection would pull a population to the nearest fitness optimum, where it would remain trapped until either

the fitness landscape was altered (e.g., via the 'shifting balance' theory [14]) or until some individuals succeeded in making a long jump away from their current position (e.g., via a 'macro mutation' operator). This interpretation presents a problem for biological modelling. In nature, populations generally don't become trapped on local fitness optima – even when certain suboptimal traits become fixed in a population, a species is not necessarily prevented from undergoing further evolution.

The traditional view of evolution has been challenged by recent investigations into molecular evolution, which suggest that the majority of mutations have no selective effect [4, 7]. The mapping from genotype to phenotype (and hence to fitness) is therefore a many to one mapping and there arises the possibility of *neutral mutations* between genotypes of equal fitness occurring in the absence of selective pressure. The resulting fitness landscape is significantly different from the rugged and hilly model. If the frequency of neutral mutations is high enough, a *neutral layer* of genotypes may exist across which a population may drift until an individual discovers a relatively rare beneficial mutation. Rather than becoming trapped at local optima, populations may be able to escape via a sequence of neutral mutations leading to a more rewarding region of the fitness landscape.

Computational studies reveal that the evolutionary dynamics of populations on neutral landscapes follow a characteristic pattern. Long periods of stasis occur in which the population explores the current neutral layer alternating with rapid fitness increases when an individual discovers a transition point to a fitter neutral layer [11]. This pattern of evolution reflects the phenomenon of punctuated equilibria observed in several biological populations [3].

The NK landscape model was initially developed to model the fitness landscapes resulting from systems with various levels of interaction between the components, for example a genotype with epistatic linkages between genes [8]. By altering the level of epistatic interaction, it is possible to generate a range of landscapes from a smooth peak with a single optimum to a rugged terrain with many local optima. The basic NK model was further developed into models incorporating neutrality by Newman & Engelhardt [10] and independently by Barnett [1]. These models have been used to investigate the biological phenomenon of neutrality and to explore the potential of neutrality to improve the efficiency of evolutionary algorithms as a search and optimisation technique [13].

III. LANDSCAPE MODELS

Landscapes in the NK family have two primary parameters; N, the length of the genotype, and K, the number of epistatic linkages between genes. Each gene contributes to the total fitness of the genotype in a manner dependant on its allele and on the alleles of the other K genes to which it is linked. It is assigned a *fitness table* mapping each of the 2^{K+1} possible allele combinations to a random fitness value. In the original NK model, this fitness value is a real number in the range [0, 1] (Figure 1). The fitness of the entire genotype is given by the average of the N fitness contributions and also falls in the range [0, 1]. The epistatic linkages may be either to the K nearest genes (neighbourhood interaction) or to K genes at random locations (random interaction). There are unlikely to be any neutral mutations in an NK landscape, since any single mutation will result in a different fitness contribution being made, and the probability of two of these fitness contributions being equal is exceedingly small.

The NKp landscape model is identical to the NK model except that there is a probability, p, that a certain allele combination makes no contribution to a genotype's fitness. The NKp model is implemented by setting a proportion, p, of a genotype's fitness tables to zero (Figure 1). If p is sufficiently high, the majority of entries in a fitness table will be zero. Therefore the majority of allele combinations will make no contribution to fitness. When an allele combination that makes no contribution to the fitness is mutated, there is a good chance that the mutation will be neutral.

The NKq landscape model is also similar to the NK model except that the fitness contributions are integers drawn from the range [0, q). That is, there are q discrete levels (0, 1, ... q) (Figure 1). The total fitness of the genotype is scaled by a factor of $1/(q-1)$ to bring the fitness within the range [0, 1]. A neutral mutation occurs when the new allele combinations containing the mutated gene make the same fitness contributions as the old allele combinations.

	NK	NKp	NKq
0 0 0	0.23	0.0	0
0 0 1	**0.74**	**0.0**	**1**
0 1 0	0.56	0.56	1
0 1 1	0.15	0.0	0
1 0 0	0.27	0.0	0
1 0 1	0.42	0.0	0
1 1 0	0.89	0.89	1
1 1 1	0.65	0.0	1

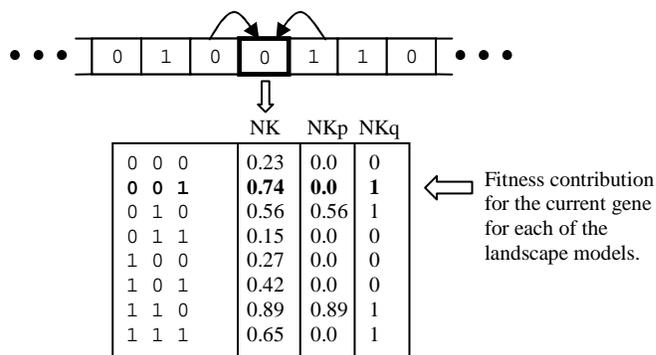⇐ Fitness contribution for the current gene for each of the landscape models.

Figure 1: A portion of the genotype and a single fitness table for a neighbourhood NK model. In this example, $K = 2$, therefore each gene is epistatically linked to the genes on either side. Each gene has its own fitness table with one entry for each possible combination of alleles. Fitness values are shown for NK, NKp and NKq landscapes. For the NK landscape, the entries are real values drawn from the range [0, 1]. For the NKp landscape, a proportion p (here $p = 0.75$) of the entries are set to zero. For the NKq landscape, the entries a quantised to q (here $q = 2$) discrete levels.

IV. STRUCTURE OF NEUTRAL LANDSCAPES

A. *Fitness Distribution*

We initially investigated the way in which the fitness distribution of a landscape changes depending on the implementation of neutrality. Values for the expected fitness distribution of an NK, NKp and NKq landscape were determined theoretically (see [12] for full details).

The theoretical results were illustrated by plotting the fitness of a sample of genotypes from a relatively small NK landscape ($N = 8$, $K = 7$) as well as the corresponding NKp (with $p = (N – 1)/N$, in this case, $p = 0.875$) and NKq (with $q = 2$) landscapes (Figure 2). The genotypes were ordered according to their fitness values. In the conversion from NK to NKp and NKq, the original fitness tables were retained and modified. In the NKp case a proportion, p, of the entries were set to zero. In the NKq case, the entries were rounded to the nearest level of quantisation.

(a) NK landscape ($N = 8$, $K = 7$)

(b) NKp landscape ($N = 8$, $K = 7$, $p = 0.875$)

(c) NKq landscape ($N = 8$, $K = 7$, $q = 2$)
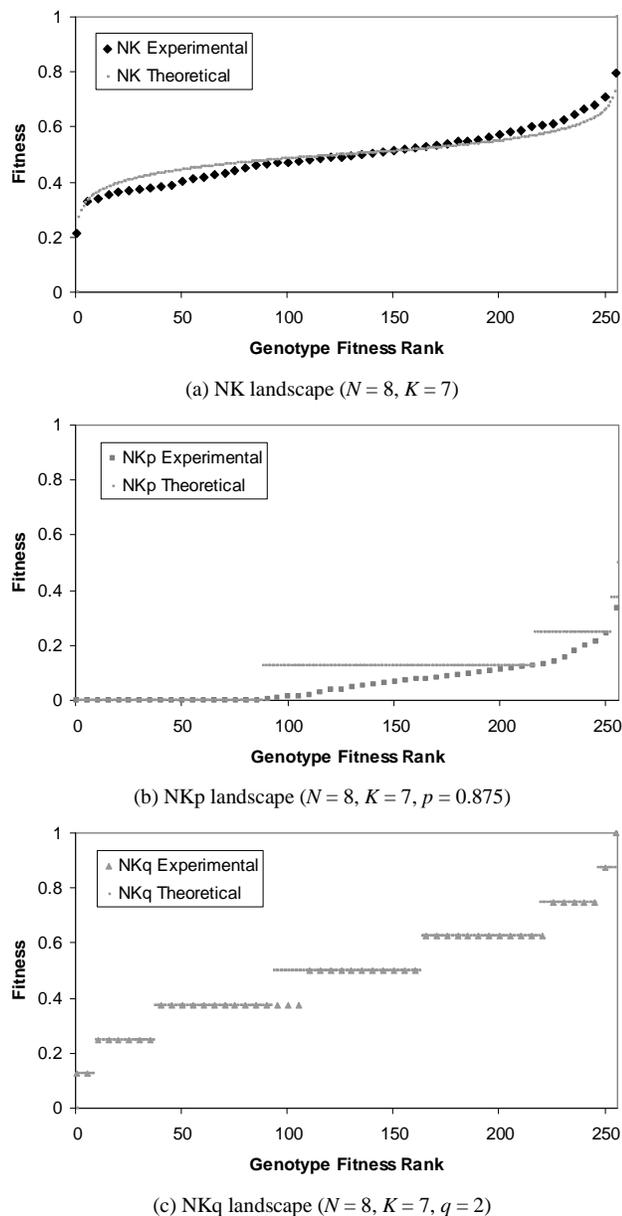
Figure 2: Theoretical and experimental fitness distributions for maximally rugged NK, NKp and NKq landscapes. The experimental values were generated by taking 50 samples from a single landscape. The theoretical values in graph (b) represent the upper limit of a fitness range in the series $x = 0$; $0 < x < 0.125$; $0.125 < x$ 0.25; etc.

For each landscape the maximally rugged case ($K = N – 1$) is illustrated. This extreme value was chosen because the expected distribution of fitness values is independent of K. Furthermore, high values of K result in a larger pool of random numbers from which the fitness contributions are drawn, and hence provide a fitness distribution with a lower variance.

From these graphs, it can be seen that the average fitness of the NKp landscape is significantly decreased with respect to the NK model significantly and the lower half of the landscape is flattened into a zero-fitness floor. The NKq landscape model results in a quantisation of the fitness into discrete layers.

Given an understanding of the individual fitness distributions, the next step was to investigate the level of correlation between each of the landscape models. To achieve this, the three fitness distributions (using every genotype, rather than a sample) were plotted on the same graph, using the initial NK landscape genotype ordering (Figure 3).
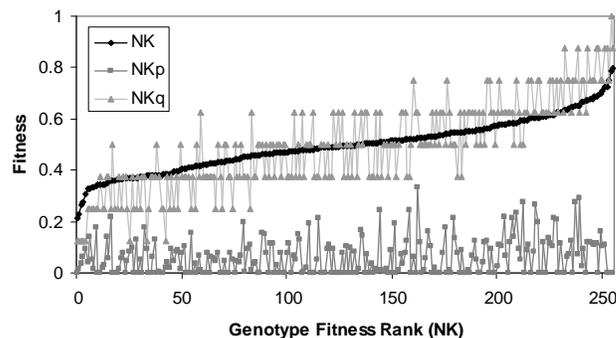

Figure 3: Fitness distribution of NK, NKp and NKq landscapes. The genotypes of the NK landscape are ordered by fitness. The NKp and NKq fitness distributions were created by retaining the same genotype ordering and modifying the entries in the fitness table to reflect the changes to the landscape.

It can be seen that the NKq landscape fitness distribution correlates reasonably closely with that of the NK landscape. It is also apparent, however, that quantisation of the fitness contributions does not necessarily result in an even quantisation of final genotype fitness. The fitness distribution of the NKp landscape is completely uncorrelated with that of the NK landscape.

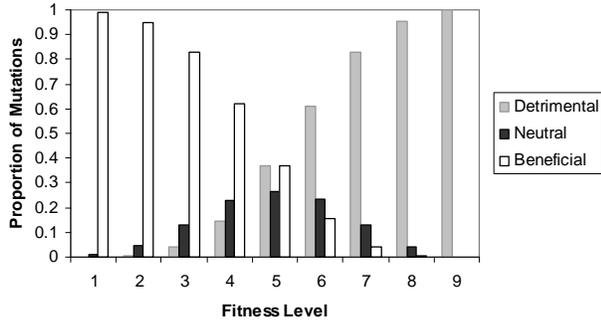B. *Distribution of Neutral Mutations*

In order to investigate how neutrality is distributed across the landscape in each of the models, several neutral landscapes ($N = 8$), both smooth ($K = 2$) and maximally rugged ($K = 7$), were used and all genotypes were enumerated. For each genotype, all possible one-bit mutations were tested to see whether they were selectively beneficial, detrimental or neutral. Twenty independent landscapes were investigated and the average proportion of each type of mutation was calculated (Figure 4).
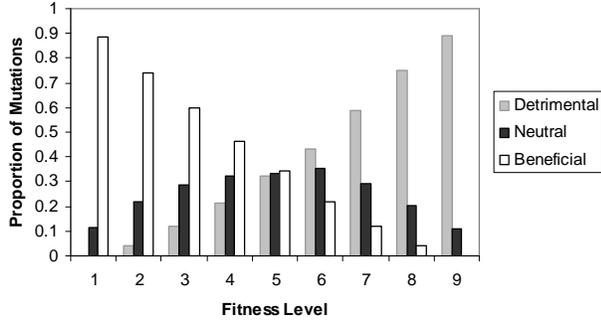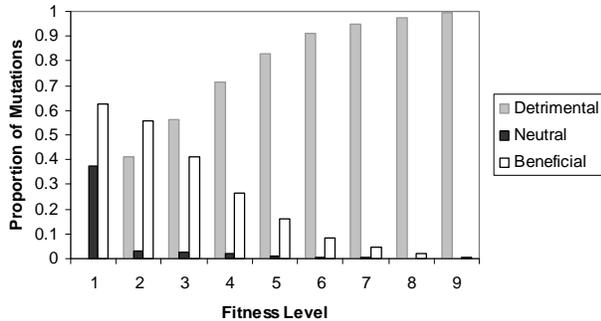
(a) Smooth NKq landscape ($N = 8$, $K = 2$, $q = 2$)
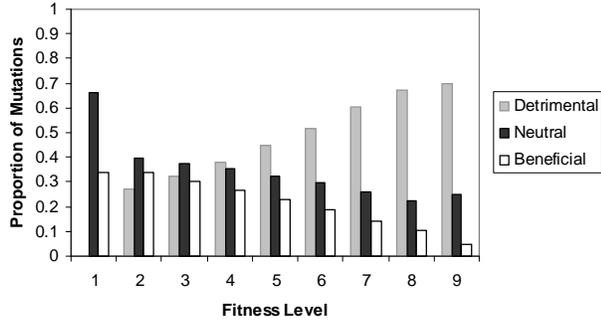
(b) Smooth NKp landscape ($N = 8$, $K = 2$, $p = 0.875$)

(c) Rugged NKq landscape ($N = 8$, $K = 7$, $q = 2$)

(d) Rugged NKp landscape ($N = 8$, $K = 7$, $p = 0.875$)

Figure 4: The distribution of neutral mutations on a range of smooth (a and b) and rugged (c and d) NKq (a and c) and NKp (b and d) landscapes. As expected, the proportion of detrimental mutations increases and the proportion of beneficial mutations decreases as fitness increases. The two models show striking differences in the distribution of neutral mutations however.

For the NKq landscape, the bins used were based upon the discrete fitness levels of the landscape (when $N = 8$ and $q = 2$, there are 9 possible fitness levels). To enable comparison, the genotypes on the NKp landscapes were binned into an equivalent number of equally sized fitness ranges. The range of fitness values on an NKp landscape varies stochastically, therefore the size of the bins also varies depending on the total fitness range of the landscape.

C. *Comparative Simulations*

Finally, a series of comparative simulations were run in which a population of independent hill climbers searched a variety of NK, NKp and NKq landscapes. The landscape parameters used were $N = 100$ and K varying between 0 and 16. The landscape parameters were chosen to provide a representative sample from completely smooth to moderately rugged. At higher levels of epistatic interaction, the increasing randomness of the landscape means that search behaviour is less interesting. On the NKp landscapes, p was set at $(N - 1)/N$ (in this case, $p = 0.99$), and on the NKq landscapes, q was set at 2. The hill climbing algorithm operates as follows: At each generation, a new individual is created by mutating each gene (i.e. flipping each bit) with probability $\mu = 1/N$ (in this case, $N = 100$, therefore $\mu = 0.01$). Hence the new individual will differ from the current individual at a single gene on average. This new individual replaces the current individual if its fitness is greater than *or equal to* that of the current individual. This last condition is necessary in order that a hill climber be able to make progress across neutral layers as well as up fitness slopes.

A population of 200 independent hill climbers was used, each of which was initialised at a random location in the landscape and allowed to search for 1500 generations. The population size was chosen to ensure that a reasonable sample of search paths over the landscape could be explored. The maximum number of the generations was chosen after a series of trials showing that the average population fitness stabilised well before this point. For each run the best fitness was the maximum fitness in the population in the final generation and the average fitness was the mean population fitness in the final generation. The final results were averaged over 20 simulation runs (Figure 5).

V. DISCUSSION

It is widely accepted that the addition of neutrality to an NK landscape has an effect on its structure [1, 10]. The simulations in this study show that the method used to implement neutrality is also significant. In terms of landscapes, the addition of neutrality is often seen as converting the continuous slopes of NK landscapes into plateaux. Such landscapes would be produced by quantising the final fitness values of an NK landscape. Both NKp and NKq landscapes are the result of neutral mutations at the level of individual genes, and the resulting landscapes are not simple plateaux superimposed on an original NK landscape, as can be seen from Figures 2 and 3.
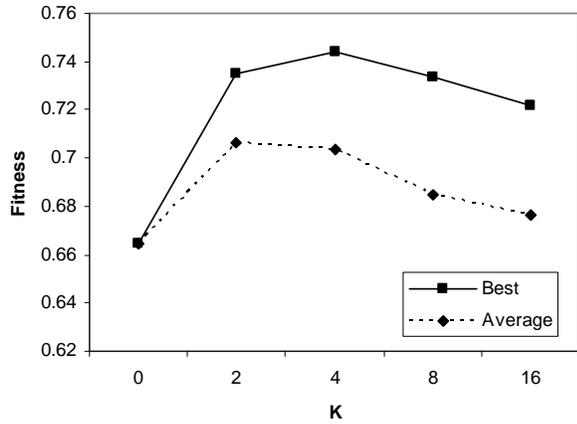
(a) NK
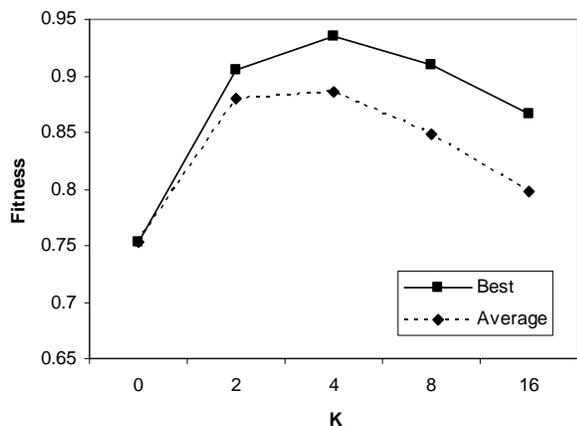
(b) NKq
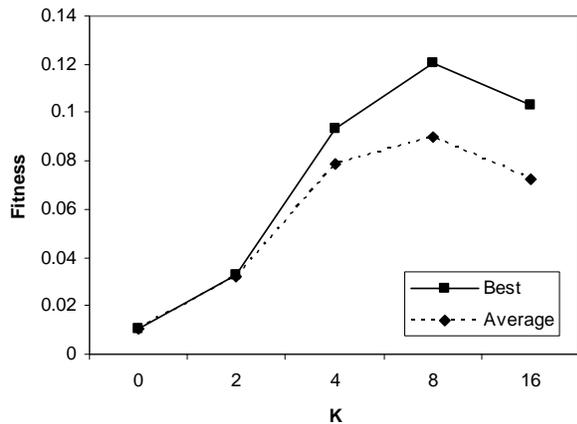
(c) NKp

Figure 5: Best and average fitness found by a population of hill climbers on a variety of (a) NK, (b) NKq (q = 2) and (c) NKp (p = 0.99) landscapes (N = 100). See text for simulation parameters. Whereas hill climbers on the NK and NKq landscapes discover the highest fitness values when $K = 4$, hill climbers on the NKp landscape discover the highest fitness values when $K = 8$. The differences in the y-axis scales are due to the different fitness ranges that occur in each of the landscapes.

The fitness values found on NKq landscapes are quite closely correlated with those on NK landscapes and hence intuitions of neutral terraces may be approximately correct. However, the fitness values found on NKp landscapes are not

organised into terraces and are uncorrelated with the underlying NK landscape (Figure 3). A notable data point in the example in Figure 3 is the genotype resulting in the global optimum on the NK landscape, which has a fitness of zero on the NKp landscape.

NKp and NKq landscapes also differ considerably in the distribution of neutral mutations. Neutral mutations on smooth NKq landscapes are concentrated around the landscape average fitness value and become slightly less common at higher and lower points on the landscape (Figure 4a). Neutral mutations on smooth NKp landscapes are most common at the lowest fitness levels and become less common as fitness increases (Figure 4b). This finding is in agreement with analytical results reported by Barnett [1]. Neutral mutations on rugged NKq landscapes are also concentrated around the landscape average but become significantly less common as fitness increases or decreases (Figure 4c). Neutral mutations on maximally rugged NKp landscapes are found exclusively on a flat, zero-fitness 'floor', with no neutrality at higher levels (Figure 4d)[1].

The simulation results provide evidence of a difference in the way that the level of epistatic interaction affects the 'searchability' of a neutral landscape. The absolute fitness values of each landscape cannot be compared directly since differences in the landscape models result in variations in the range of possible fitness values. For example, on an NKq landscape with $q = 2$, more extreme fitness values will tend to occur, as intermediate fitness contributions will be replaced by 0 or 1. On an NKp landscape, however, a smaller number of positive fitness contributions are being averaged over the same number of genes, and therefore the fitness values will be lower.

A general trend is apparent, however, as the level of epistatic interaction increases and the landscapes change from smooth to rugged. Kauffman recognised that low levels of epistatic interaction have the effect of 'buckling' the landscape and resulting in higher fitness values than are possible in the $K = 0$ case [8]. When $K = 0$, only two possible fitness contributions are generated for each gene and, while each of these genes can be optimised independently, there is a chance that for some genes, the choice will be between two low fitness alleles. As K increases, the pool of possible fitness contributions for each gene becomes larger, increasing the chance of a gene being able to contribute a high fitness value. The down side of increasing the level of epistatic interaction is that it may not always be possible to make the optimal choice ('frustration' in the system increases). Optimising one gene may result in others with which it interacts being set to suboptimal values. As K increases further, the number of compromises that have to be made also increases (the 'complexity catastrophe' described by Kauffman [8]).

---

[1] In figure 3(d), some neutral mutations do appear at higher levels in the landscape. It has been determined from theoretical analysis that no strictly neutral mutations occur above the zero-fitness floor. These 'neutral' mutations are due to a loss of precision in the averaging process and correspond to 'nearly neutral' mutations.

At a certain point, the benefit of having a greater number of fitness contributions from which to choose can be balanced against the need to compromise on these choices, and the resulting landscape will contain higher fitness values. For NK and NKq landscapes with $N = 100$, the optimal fitness values occur for $K = 4$. By contrast, for the NKp landscape the optimal fitness values occur at a much higher level of epistatic interaction ($K = 8$). Other population based algorithms using a variety of mutation rates and recombination operators (single point, two point and uniform) were also tested and found to return comparable results (full details can be found in [5]).

VI. CONCLUSION

It is apparent that the way in which neutrality is implemented in a model of neutral evolution has a critical impact on both the structure of the resulting landscapes and on the performance of evolutionary search algorithms on these landscapes. While NKq is qualitatively similar to NK in several respects, NKp differs considerably from both NK and NKq. NKq landscapes relate more closely to the intuitive idea of a 'terraced' NK landscape in which rugged hillsides have been flattened into locally smooth ledges. NKp landscapes, on the other hand, capture some of the features of 'lethal' mutations in biology. It is plausible that selective neutrality operates at several different levels in biological organisms, from nucleotide sequences up to entire organisms, and therefore different computational models may be appropriate in different situations.

ACKNOWLEDGEMENTS

REFERENCES

[1] Barnett, L., Ruggedness and neutrality - the NKp family of fitness landscapes. In C. Adami, R. K. Belew, H. Kitano and C. Taylor (Eds.), *ALIFE VI, Proceedings of the Sixth International Conference on Artificial Life* (pp. 18-27). Cambridge, MA: The MIT Press, 1998.

[2] Darwin, C., *The Origin of Species by means of Natural Selection.* Oxford, NY: Oxford University Press, 1996.

[3] Eldredge, N. & Gould, S. J., Punctuated equilibria: an alternative to phyletic gradualism. In T. J. M. Schopf (Ed.), *Models in Paleobiology*. pp. 82-115. San Francisco: Freeman, 1972.

[4] Forst, C. V., Reidys, C. & Weber, J., Evolutionary Dynamics and Optimization: Neutral Networks as Model-Landscapes for RNA Secondary-Structure Folding-Landscapes. In F. Moran, A. Moreno, J. J. Merelo and P. Chacon (Eds.), *Third European Conference on Artificial Life*, number 929 in Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag, 1995.

[5] Geard, N., *An exploration of NK landscapes with neutrality*. Unpublished Technical Report, School of Information Technology and Electrical Engineering, The University of Queensland, 2001.

[6] Harvey, I. & Thompson, A., Through the labyrinth evolution finds a way: A silicon ridge. In T. Higuchi (Ed.), *Proceedings of The First International Conference on Evolvable Systems: From Biology to Hardware (ICES96)*. Berlin: Springer-Verlag, 1996.

[7] Huynen, M. A., Stadler, P. F., & Fontana, W., Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA*, 93, 397-401, 1996.

[8] Kauffman, S. A., *The Origins of Order - Organization and Selection in Evolution*. New York: Oxford University Press, 1993.

[9] Kimura, M., *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge University Press, 1983.

[10] Newman, M. & Engelhardt, R., Effect of neutral selection on the evolution of molecular species. *Proc. R. Soc. London B.*, 256, 1333-1338, 1998.

[11] van Nimwegen, E., Crutchfield, J. P., & Mitchell, M., Statistical dynamics of the Royal Road genetic algorithm. In A. Eiben and G. Rudolph, (Eds.), *Theoretical Computer Science, Special issue on Evolutionary Computation*, 1998.

[12] Skellett, B., Geard, N., Wiles, J., The Distribution of Neutral Layers in NKp and NKq Landscapes. In preparation.

[13] Stewart, T., Extrema Selection: Accelerated Evolution on Neutral Networks. In *Proceedings of the 2001 Congress on Evolutionary Computation*, pp.256-260, IEEE Press, Piscataway, NJ, USA, May 2001.

[14] Wright, S., The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution. In D. F. Jones (Ed.), *Intl. Proceedings of the Sixth International Congress on Genetics* (vol. 1, pp. 356-366). Brooklyn, NY: Brooklyn Botanic Garden, 1932.