# Neutral search spaces for artificial evolution: a lesson from life

**Rob Shipman[1], Mark Shackleton[1], Marc Ebner[2], Richard Watson[3]**

[1] BT Labs at Adastral Park, Admin 2-5, Martlesham Heath, Ipswich, IP5 3RE, UK. rob.shipman@bt.com, mark.shackleton@bt.com

[2] Universität Würzburg, Lehrstuhl für Informatik II, Programmiersprachen und Programmiermethodik, Am Hubland, 97074 Würzburg, Germany. ebner@informatik.uni-wuerzburg.de

[3] Brandeis University, Volen Center for Complex Systems, Mail Stop 18, Waltham MA 02454-9110, USA. richardw@cs.brandeis.edu

**Abstract**

Natural evolutionary systems exhibit a complex mapping from genotype to phenotype. One property of these mappings is neutrality, where many mutations do not have an appreciable effect on the phenotype. In this case the mapping from genotype to phenotype contains redundancy such that a phenotype is represented by many genotypes. Studies of RNA and protein molecules, the fundamental building blocks of life, reveal that this can result in neutral networks - sets of genotypes connected by single point mutations that map into the same phenotype. This allows genetic changes to be made while maintaining the current phenotype and thus may reduce the chance of becoming trapped in sub-optimal regions of genotype space. In this paper we present a redundant mapping and explore its properties by performing random walks on the neutral networks in its genotype space. We investigate whether the properties found in nature's search space can be engineered into our artificial evolutionary systems. The mapping, based on a random Boolean network (RBN), was found to give promising results.

## 1. Introduction

Natural evolution differs in many respects from the evolutionary algorithms typically employed today. One such difference is highlighted by the neutral theory of evolution. According to this theory a considerable fraction of all mutations are neutral and only a minute fraction of the remainder are actually beneficial (Kimura 1994). This results in a redundant genotype-phenotype mapping with typically many genotypes representing any given phenotype. The redundancy manifests itself in a number of different ways and at a number of different levels - from the genetic code, consisting of 64 codons mapping into only 20 amino acids, to the complex interplay of molecules forming an organism. A particularly important and thorough study of the effects of such redundancy was performed in the context of the folding of RNA, and to a lesser extent, protein molecules (Huynen 1996, Huynen, Stadler, and Fontana 1996). These studies revealed a number of interesting properties in the nature of the secondary structures that the primary structures folded into. There were a number of common secondary structures each represented by a very large set of primary sequences i.e. there was large-scale redundancy in the genotype-phenotype mapping. The density was such that these sets were often connected by single-point mutations forming so-called *neutral networks*. Thus, it was possible to traverse the set of genotypes through the simplest of mutations without changing the represented phenotype.

This brought about the possibility of neutral drift allowing larger areas of genotype space to be explored in search of more adaptive secondary structures. However, during such a process there is no pressure influencing movement to areas of genotype space in which more adaptive phenotypes can be found and there is thus a danger of prolonged periods of random drift. It is important, therefore, that the neutral networks representing each of the phenotypes are intertwined with many access points

between them. This encourages beneficial transitions from one network to another and minimizes the amount of time spent drifting randomly. Studies of RNA folding suggested that this was the case. As the neutral networks were traversed a relatively high, and roughly constant, number of new structures were discovered at each step (Huynen 1996).

These properties may have a significant impact on the evolvability of a system. Instead of becoming trapped in sub-optimal regions of genotype space, adaptation is able to continue through genetic changes that do not alter the phenotype but enable movement in genotype space to areas that are closer to genotypes representing potentially more adaptive phenotypes. Following on from previous studies (Ebner 1999, Shipman 1999) this work explores a redundant genotype-phenotype mapping with a view to ascertaining whether these fundamental properties of living systems can be encouraged in our artificial systems. This mapping was compared to several other redundant mappings in an extended version of this paper (Shipman et al 2000).

The next section details the mapping that was studied, section 3 details the methods that were used to ascertain its properties, section 4 presents our findings, section 5 gives a discussion and section 6 concludes.

## 2. The random Boolean network mapping

A random Boolean network consists of a fixed-length binary string together with a set of rule tables that define how each bit of the string is updated. Each bit has a number of inputs that yield an index into the rule table for that bit. This is illustrated in figure 1. In this work, the string length was fixed at 8 bits, giving $2^8 = 256$ possible phenotypes. The genotype encoded the initial state of the RBN, the rule table for each bit and the inputs for each bit. More information on random Boolean networks can be found in (Kauffman 1993).

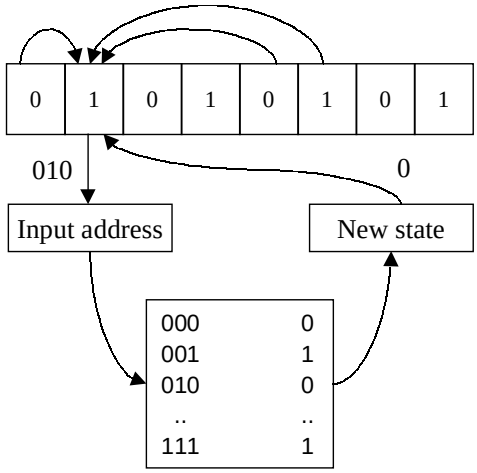



Figure 1. Illustration of the RBN mapping; the genotype specifies the rule table for each cell, the initial state of each cell and the inputs to each cell. For the example bit, the inputs specify that the third rule is used to give its next state.

In this work the number of inputs per cell was fixed at 3 and thus 9 bits were required to specify the inputs for each cell. This resulted in a total of 72 bits to specify the wiring of the 8-bit boolean network, which together with the 72 bits to specify the rule tables and initial state gave a genotype of length 144. To perform the mapping from genotype to phenotype the network was constructed and initialized using the information encoded in the genotype. It was then run for 20 time steps and the resulting binary string interpreted as one of the 256 phenotypes.

## 3. Evaluation of the mappings

The main tool used to evaluate the properties of the mapping described above was a random neutral

walk (Huynen 1996). This procedure allowed a measure to be made as to the benefit of neutral drift through assessing the number of new phenotypes encountered. The walk began by choosing a genotype mapping into a given phenotype at random. All one-point mutants of this genotype were assessed and the number of new phenotypes encountered was logged. These are termed innovations. In addition a list of neutral neighbors, i.e. one-point mutants mapping into the same phenotype, was formed. One of these neutral neighbors was chosen at random and the procedure was repeated for a given number of steps, 100 steps were used in this work. If no neutral neighbors are found, the walk remains in the same position and no further innovation is possible. A number of statistics were calculated using this procedure (see Ebner et al. 2000), the following two are reported on in this paper:

### 3.1 Total number of innovations

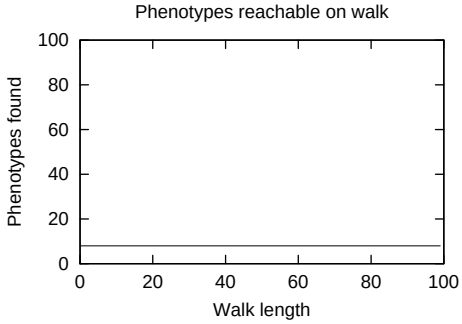
For each of the 256 phenotypes, a genotype mapping into that phenotype was chosen at random. A random neutral walk consisting of 100 steps was performed for each of these genotypes and the number of new phenotypes encountered at each step was logged. This procedure was repeated for four independent walks for each of the phenotypes. This resulted in a total of 1024 walks and the averaged cumulative number of innovations was plotted at each step of the walk for each mapping.

### 3.2 Phenotypic accessibility

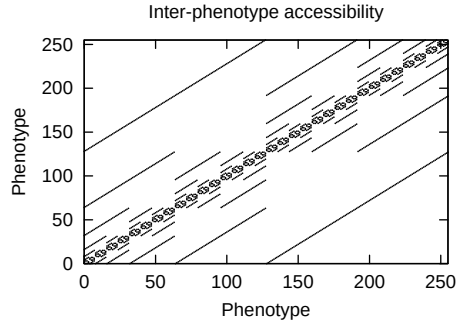
For the same set of 1024 walks described in the previous section an accessibility plot was formed that showed which phenotypes were encountered on the neutral walks for all 256 phenotypes. This resulted in a 256 by 256 plot with one axis showing the phenotype that neutral walks were being performed for and the other axis the phenotypes encountered on those walks. This plot gave some impression of the connectedness of phenotype space via the neutral pathways in genotype space.

## 4. Results

This section shows the statistics that were calculated for the RBN mapping together with the statistics for a direct encoding with no redundancy in order to allow comparison.




(a)



(b)

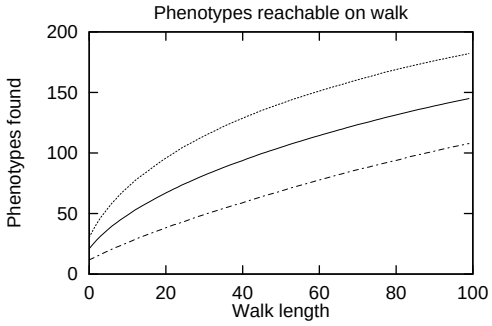
Figure 2. Results for a direct binary encoding. (a) The number of phenotypes found remains constant at 8, reflecting all possible one-point mutations. There are no neutral neighbors and thus there can be no innovation through neutral drift. (b) The accessibility is very sparse reflecting the fact that only 8 new phenotypes are accessible from each phenotype.

### 4.1 Direct binary encoding

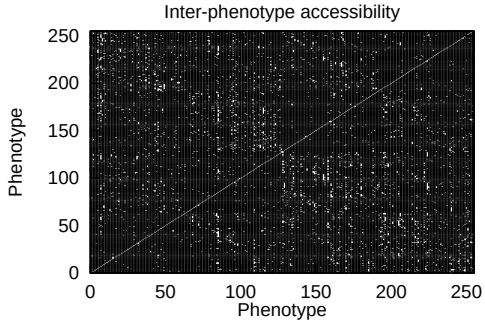
Figure 2 shows results for a direct encoding in which an 8-bit genotype directly specified the phenotype. In effect there was no genotype-phenotype mapping as is commonly the case in artificial evolution.

### 4.2 Random Boolean network mapping

Figure 3 shows that the RBN mapping increases the number of innovations on a neutral walk. The number of phenotypes found after 100 steps is very much greater than the direct encoding and the steep curve indicates the potential for further innovations. The density of the accessibility plot is also greater indicating even greater accessibility between the phenotypes.




(a)




(b)

Figure 3. Results for the RBN mapping. (a) The number of phenotypes found is higher than all the other mappings. A total of 145 different phenotypes were found after 100 steps along the neutral walk. The curve continues to rise indicating continuing innovation. (b) The accessibility plot is the densest of those presented in this paper with most transitions between phenotypes found.

## 5. Discussion

It is common practice in artificial evolution to use a direct one-to-one mapping between genotype and phenotype. An example of such a mapping is the direct binary encoding, the results for which were presented in section 4.1. In such a scenario the number of differing phenotypes accessible from any given genotype is restricted to the length of the genotype i.e. all one-point mutants. In many situations it may be common for none of these phenotypes to be better adapted than the current one and thus adaptation will effectively halt at a local optimum. This is a very different scenario to the seemingly

open-ended innovation found in natural systems. The mapping explored in this work shows that continuing innovation can be achieved with a suitable mapping between genotype and phenotype. Introducing the same kind of redundancy found in natural evolutionary systems into our artificial systems may increase the efficacy of artificial evolution through reducing the possibility of entrapment at local optima. However, the type of redundancy is crucial. The redundancy is only beneficial if it increases the accessibility between phenotypes, i.e. if it allows more new phenotypes to be discovered than would be the case for a direct encoding. In order for this to be the case neutral mutations must be possible that, although not immediately beneficial, "set-the-scene" for further mutations that are beneficial.

The success of a neutral mapping is dependent on the balance between structure and randomness. In order to increase the possibility of discovering a given phenotype, it would be desirable for many genotypes mapping into that phenotype to be randomly scattered throughout genotype space. Thus, from any point in that space it is likely that a required genotype will be in relatively close proximity. However this scattering cannot be entirely random, as it is important to maintain a relatively high number of neutral neighbors in order to encourage the formation of connected neutral networks and allow substantial neutral drift. The good results of the RBN mapping reflect a good balance between these two aspects. A large number of different phenotypes were discovered on neutral walks whilst an average of approximately 50% neutral neighbors was maintained.

## 6. Conclusion

This work has explored the properties of a redundant genotype-phenotype mapping that were constructed in an attempt to mimic the desirable properties found in nature's own redundant search space, evidenced by the work on RNA folding (Huynen, Stadler, and Fontana 1996) for example. The redundancy was found to be beneficial, in that movement on the resulting neutral networks allowed for the discovery of a larger number of phenotypes than would be the case for a direct encoding. Thus, the probability of entrapment at local optima when using this mapping would be reduced. It is possible that mappings such as this may be of real benefit in an artificial evolutionary system.

The results presented in this work used only a small number of phenotypes and a relatively small sample of the genotype space. With sizeable genotypes exhaustive enumeration of the space is impossible and some form of sampling is required. An intention of future work is to explore other statistics that can help to further reveal the properties of the spaces created by this and other mappings. The performance of the mapping on larger phenotype spaces and in the context of an adaptive fitness walk is also being explored (Shackleton et al. 2000), as is comparison to other redundant mappings (Shipman et al. 2000).

## References


Ebner, M. 1999. On the search space of genetic programming and its relation to nature's search space. In Proceedings of 1999 Congress on Evolutionary Computation, Volume 2, 1357-1361. IEEE Press.

Ebner, M., Shackleton, M., Shipman, R., and Watson, R. 2000. How redundant mappings influence the ability to adapt to a changing environment. Under submission.

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

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

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

Kimura, M. 1994. Population Genetics, Molecular Evolution, and the Neutral Theory: Selected Papers. The University of Chicago Press, Chicago.

Shackleton, M., Shipman, R., Ebner, M. 2000. An investigation of redundant genotype-phenotype mappings and their role in evolutionary search. To appear in Proceedings of the 2000 Congress on Evolutionary Computation.

Shipman, R. 1999. Genetic Redundancy: Desirable or Problematic for Evolutionary Adaptation? In Proceedings of the 4th International Conference on Artificial Neural Networks and Genetic Algorithms, 337-344, New York. Springer-Verlag.

Shipman, R., Shackleton, M., Ebner, M., Watson, R. 2000. Neutral search spaces for artificial evolution: a lesson from life. To appear in Proceedings of Artificial Life VII.