

Weasel World: a simple artificial environment for investigating open-ended evolution

David Cornforth¹, David G Green² and Jenny Awburn¹

¹*School of Environmental and Information Sciences, Charles Sturt University, PO Box 789, Albury, NSW, Australia.*

Email: dcornforth@csu.edu.au

²*Faculty of Information Technology, Monash University, Clayton 3800 Australia*

Email: david.green@infotech.monash.edu.au

Abstract

Weasel World is an artificial life environment for simulating open ended evolution. It consists of a genetic algorithm embedded in a pseudo spatial context, composed of the text of Shakespeare's play *Hamlet*. The novelty of Weasel World is that it combines features from many evolutionary algorithms with an operation that increases genotype length, placement of individuals within a spatial context, the notion of territory, and interaction with global environmental parameters. Individuals interact with others in their territory, competing for the right to reproduce in a modified tournament selection. Weasel World demonstrates phase changes, clustering and self-organization, as well as showing the importance of considering the interaction between local phenomena and global constraints. The evolutionary process undergoes different phases arising from independent growth, a competitive phase and a lack of nutrient.

1. Introduction

In his book, *The Blind Watchmaker*, Richard Dawkins (1988) illustrates the power of natural selection to produce complex results through the example of a child typing Shakespeare. Taking the line "methinks it is like a weasel" out of the play Hamlet, he points out that the chances of a child typing this are virtually zero. However, if each time the child types a correct letter, that letter becomes fixed, then the child's typing will quickly converge on the targeted line.

As a simple analogy for what evolution does, the above example is fine. However, it falls short of being a good model because the selection is directed towards a fixed target. In reality, evolution does not proceed in such a simple linear way, but behaves more like a growing bush (Gould, 1989), with branching and pruning producing a host of different forms.

The initial inspiration for this study was to devise an alternative to Dawkins's example that would behave more like real evolution. We call the result *Weasel World* after the line Dawkins used. Instead of a single line of text, the environment for Weasel World (WW) is the entire play *Hamlet*. Populations of organisms ("weasels") within this environment can breed and evolve to converge on any part of the text.

The WW model provides a platform for investigating many questions about evolutionary processes. For instance, one of the key questions in the study of Artificial Life is how complex behaviour arises in natural systems, and how it may be simulated (Bedau et al 2001). Complex behaviour arises from interactions between simple elements of a system (Green, 1993), and includes clustering, modularity, and phase changes. Many models have successfully demonstrated emergent complex behaviour from simple systems, for example, cellular automata (Wolfram, 1984), but have generally not examined the interaction between open-ended evolution, local phenomena and global constraints. In addition, other work (Green et al., 2000) has shown that phase changes have an important role in explaining how more complex structures arise in complex systems.

In this study we investigate some of the features exhibited by the Weasel World model. In particular, we use it to show (1) the existence of phase changes in a simple model of open-ended evolution, (2) the role of phase changes in building complexity, and (3) how such events are influenced by global constraints, external to the individuals in the population.

2. Background

Genetic Algorithms (Goldberg, 1989; Holland, 1975) are simplified models of evolution. A population of possible solutions is used, where each candidate solution is coded as a genotype. Evolution takes place over a number of generations, in which solutions are replaced by superior ones. Each generation typically proceeds by applying operators such as mutation, crossover, and selection based on relative fitness.

Typically, genetic algorithms are utilised in complex problem domains to search for a defined target or fixed goal (He & Yao, 2001), in directed evolution such as the example from Dawkins mentioned above, and in optimization tasks. The ability of evolution to produce "solutions" to complex problems of survival has inspired many computational analogies (Fogel, 1998).

In contrast, our motivation of building this model was to provide a test bed where we can investigate processes common in biological evolution, rather than to solve a specific goal. Our model is "open-ended" in the sense that it does not have a specific goal or function to be optimised. In this way it is similar to models such as Tierra (Ray, 1991), which was able to demonstrate emergence effects such as the spontaneous appearance of parasitic organisms.

3. Weasel World

Weasel World is an implementation of a simple open ended evolutionary algorithm, which features evolving string matching entities within a one-dimensional spatial environment. It consists of the following components:

- a genotype representation
- a genotype to phenotype mapping
- a fitness function

- selection mechanism
- reproduction and genetic operators

3.1 *Genotype representation and genotype to phenotype mapping*

The obvious representation to use for solving what is essentially a string matching problem, is a string of symbols of length n $\langle x_1, x_2, \dots, x_n \rangle: x_i \in \{a, b, \dots, z\}$, so that genotype and phenotype are identical. For Weasel World, this set is extended to use all characters included in the *Hamlet* text, for example punctuation, spaces, brackets etc. We have selected a fixed genotype representation, where the genotype is made up of a number of genes, each of which is a character that exists in the artificial environment (refer section 4).

Most implementations of evolutionary algorithms use a fixed length encoding, where the length of the genotype is set at initialization and does not change during the evolution of the model. While this is sufficient for a fixed target, it is seldom so for open ended evolution (Harvey, 1992), where individuals in the population are free to grow in complexity.

In Weasel World, individuals (or ‘weasels’) begin with a set genotype length but are then able to grow in length (within certain constraints described below). The evolutionary process thus leads to weasels of differing lengths, with longer strings being a means to achieve greater competitiveness. Selecting a variable length encoding has obvious implications for the reproduction operators - see the following section.

3.2 *Fitness function*

We did not implement a traditional external fitness function. Rather, we implemented a means of comparison or competition between weasels that takes into account several factors:

- phenotype length;
- number of matches to the text;
- ‘clumpiness’, measuring how contiguous the matching characters are; and
- ratio, the number of matches divided by the phenotype length.

Competition proceeds as follows:

1. For agents with equal phenotype lengths, the weasel with the highest ratio wins. If ratios are identical, the weasel with the highest number of matches wins. Otherwise, the weasel with the higher clumpiness measure wins.
2. For agents with different phenotype lengths, either weasel with a ratio of 1.0 (number of matches equal to phenotype length) wins. If both have a ratio of 1.0, the longer one wins. If neither is the case, the weasel with the higher ratio wins. Otherwise the weasel with higher clumpiness wins.

3.3 *Selection mechanism*

This is achieved using a kind of tournament selection. Each weasel is given the opportunity to reproduce by examining its territory to look for others, its rivals. If there are no rivals, the weasel is classed a winner and reproduces as described below. Otherwise, it selects a rival at random from within its territory, and they compete as described above, with the winner of the tournament winning the right to reproduce.

3.4 *Reproduction and genetic operators*

Standard evolutionary operators are crossover and mutation. However, the use of a variable length genotype requires other operators. Harvey (1992) has discussed the use of a length extension operator, which increases the length of the genotype by adding a random gene to the end. Harvey suggests this operator should be controlled so that the length of individuals in the population should increase very gradually, with the population being almost completely converged before the next increase in length occurs.

The traditional crossover operator was not included in the model used in this study, as this was viewed as an unnecessary complication for the purposes of investigating phase changes. A consequence of trying to find exact matches for a text is that once individuals have achieved a relatively modest number of matches, combining individuals from separate locations is likely to produce children with a much lower fitness and hence, is fairly unproductive.

Instead, only asexual reproduction was implemented. Weasels that have no rivals (i.e. no other weasels in their territory) and weasels that are the winners of a competition as described in section 3.2 reproduce by creating an exact copy of themselves in the new generation. Thus, some weasels will be copied several times and some will not be copied at all. Population size is fixed throughout each evolution. Once the new generation is created, they are subject to the mutation and lengthening.

Mutation occurs by altering characters in the genes randomly. Characters are assigned in accordance with the same probabilities as for initial creation (see Section 4). Rather than a fixed mutation rate, the rate is specified relative to genotype length, which Harvey (1992) suggests is more appropriate for variable length genotypes. The given mutation parameter is divided by the genotype length and for each character in the genotype, if the result is less than a random double between 0 and 1 inclusive, that character will be mutated. For example with a genotype length 5 and mutation parameter 1, each character will have a 1 in 5 chance of being mutated. In this way the given parameter is effectively the number of mutations *per* genotype, rather than a fixed rate.

The increase-length operator, suggested by Harvey (1992), increases genotype length by exactly one character. A random gene is created in accordance with original character assignment probabilities. The gene is appended to the end of the genotype so as not to disrupt the phenotype. For example, if the agent phenotype is “methinks it is like a weasel” and any character is inserted after the word ‘is’, the agent will immediately lose half its matches and will likely die off. If the given length rate parameter is less than a random double between 0 and 1 inclusive, the genotype will have its length increased by one.

Global parameters

Several global variables affect the evolutionary process. These are the nutrient and the territory ratio.

The rate at which weasels may grow, and the maximum size they can become, may be controlled by the available nutrient level. A global source represents nutrients available via natural processes, but limited in quantity. This is modelled using a simple variable, and adding a fixed amount each iteration. At each iteration of the model, weasels consume nutrient according to their size. When the nutrient levels fall below zero, the lengthening operator is disabled so that weasels can no longer grow, thus limiting the size of the weasels. Because of this, they will not be able to expand their territory, and must remain in their clusters.

As the weasels grow in length of DNA due to the lengthening operator, their territory also grows. Territory is controlled by a parameter called ‘territory ratio’. This is multiplied by the phenotype length (which is equal to genotype length in this implementation) to determine the size of the territory. The weasel exists in the centre of the territory, i.e. if the weasel length is 5 and territory ratio 100, the weasel’s territory will consist of 250 characters to its left, the 5 characters it currently occupies, and 250 characters to its right. The text is considered to wrap so the territory may begin at the end of the text and wrap around to the beginning. The maximum territory size is the length of the text.

Given the variable length genotype, territory size will be different for each weasel. When the territories of two creatures grow large enough for them to meet, they come into competition for reproduction. As the population grows in size, this causes a phase change - local competition gives way to global competition. Depending upon the ratio of the size of weasels to the size of their territory, the system demonstrates convergence to a single population, identical and centred at one location, or several clusters that are geographically isolated.

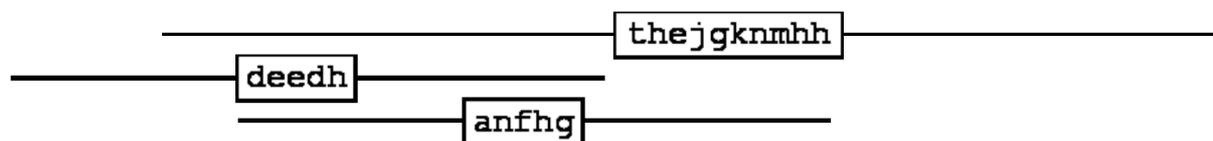


Figure 1. Illustrating the territory of individuals of varying size. The string “deedh” is able to find one rival within its neighbourhood, “anfhg”, so reproduces only if it is fitter than this one. The string “anfhg”, although having the same size neighbourhood as “deedh”, has two rivals, due to the juxtaposition of strings in the environment. The string “thejgknmhh” has more rivals due to its larger territory size.

4. Experiments

Initially, we wished to confirm the operation of the algorithm. We would expect the number of matches to rise over time with no apparent limit. Having achieved this, the aims of the experiments are:

1. To demonstrate phase changes as the territory increases in size and individuals become more connected.
2. To check that this phase change is more sharply defined if the individuals are evenly distributed in the environment initially.
3. To investigate the effects of an environmental limit on the convergence and behaviour of the population during the phase change.

For all experiments, individuals were initialized with characters chosen at random from the set {abcdefghijklmnopqrstuvwxyz ,'-[];?!():"&1} - 42 characters including punctuation and the space character. This is a complete list of characters appearing in the electronic text version of *Hamlet* (Project Gutenberg, 1999), minus additions involved in conversion to e-text, for example copyright information. The probability of any character being selected is directly proportional to its existence in the text. For example, the space character and the letter ‘e’ are the most common characters and are more likely to be selected randomly than say, the character ‘z’. Individual experiments are listed in Table 1 below. A short explanation follows.

Table 1. Parameters used in the experiments.
The abbreviation *k* is used for a thousand.

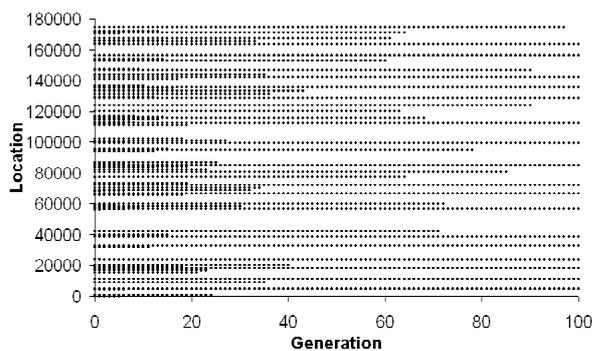
Parameter	Experiment		
	1	2	3
Population size	100	100	100
Initial distribution	random	even	random
Nutrient	disabled	disabled	10,000
Initial no. of characters	5	5	5
Lengthening rate	1	1	1
Mutation rate (per genotype)	1	1	1
Territory ratio	100	100	100
Iterations	100	100	250

In the first experiment, we initialized 100 individuals with 5 characters each, randomly spaced in the environment, with a territory ratio size 100. We disabled the nutrient requirement. Weasels were permitted to grow for 100 generations. The second experiment is a repeat of the first, except that we initialized the environment so that the population was evenly spread. In the third experiment, we enabled the nutrition requirement. All experiments were repeated 20 times.

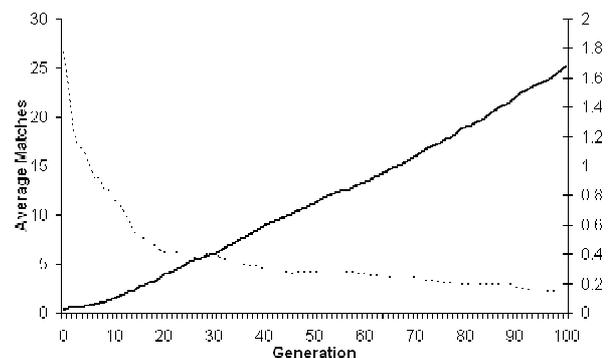
5. Results

As expected, preliminary experiments confirmed that the model was able to discover large contiguous sections of the text. For example, a randomly selected run produced the text segment shown below. Notice the lack of capitalization, and the broken word at the end of the segment:

*of other nations; they clip us drunkards and with swinish phrase soil our addition;
and indeed it takes from our achievements, though perform'd at height, the pith
and marrow of our attribute. so oft it chances in parti*



(a)



(b)

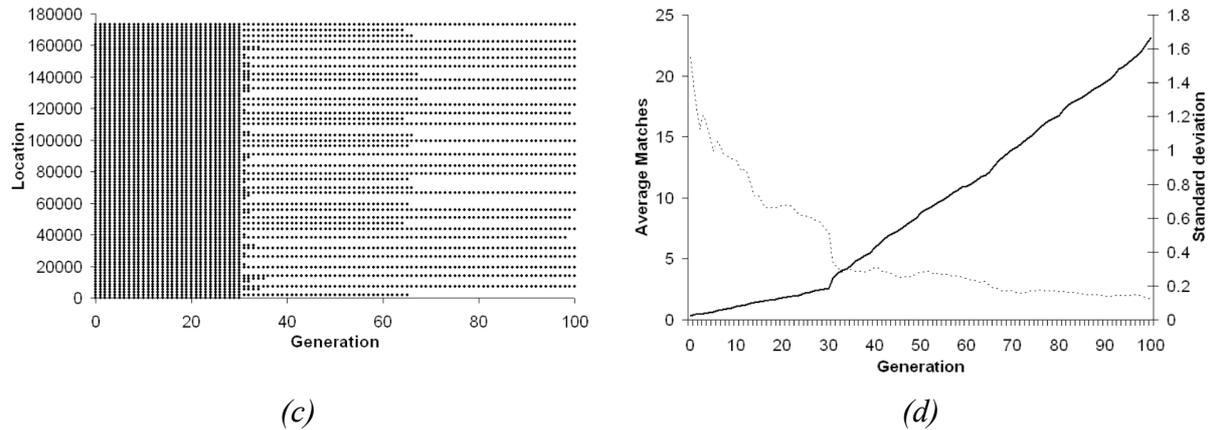


Figure 2. Results from experiments 1 and 2. Strings located at random throughout the text are removed after generation 10 for the remainder of the run (a); the change of gradient in the average matches (b) marks a phase change as the model shifts from independent growth to competition. In contrast, strings located evenly display a much sharper phase transition, shown by the simultaneous removal of strings at many locations (c), and by the sudden change of gradient in the graph of the average number of matches (d).

Figure 2 compares results from experiments 1 and 2. The results shown are each from a single run, but are typical of the results obtained from the 20 runs. Figure 2a shows the location of individuals over time for a random distribution of weasels. Horizontal lines indicate sites where individuals are located in the text of Hamlet. The vertical axis represents the position within the text, while the horizontal axis represents time. Horizontal lines stretching over the entire width of the graph represent locations in the text where there is a continuous presence of at least one string. Horizontal lines that fail to stretch to the right hand side of the graph illustrate locations in the text where individuals have died due to competition with others. After about 10 generations, some of the lines terminate, indicating a loss of weasels at some locations. After 100 iterations of the model, the number of locations where weasels reside has markedly decreased, corresponding to a reduction in the diversity of strings. The locations that remain are populated by a number of similar strings.

Figure 2b shows the number of matches for the same run as figure 2a. The average number of matches (solid line) increases slowly at first, as this is driven by correct characters added by the lengthening operator. After about 10 generations, there is an increase in rate as territory size increases sufficiently for individuals to detect the presence of rivals. This represents a phase change, as the population gradually shifts from isolated individuals to a connected population. Following this point, selection drives the increase in matches, as strings with a lower number of matches are eliminated. The loss of diversity is apparent from the normalised standard deviation (dotted line).

In spite of the fact that a change of phase occurs, the random distribution tend to mask this phenomenon. In experiment 2, we distributed weasels evenly at initialisation, in order to make the phase transition more sharply defined. As the strings grow longer, territory expands until the transition from individual to competitive behaviour happens simultaneously for all strings. This is shown by the sudden loss of multiple locations in figure 2c, and by the sudden change of gradient of the graph of average matches in figure 2d. These results are shown for a single run, but are typical of the results found for the 20 runs. This happens at around generation 30, and is accompanied by a large drop in the standard deviation. Weasels increase their territory in line with their size (phenotype length). When weasels encounter others in their territory, they begin contests for the right to reproduce. The number of groups decreases as some

weasels die out, to be replaced by others. This confirms the Competitive Exclusion Principle (Hardin, 1960).

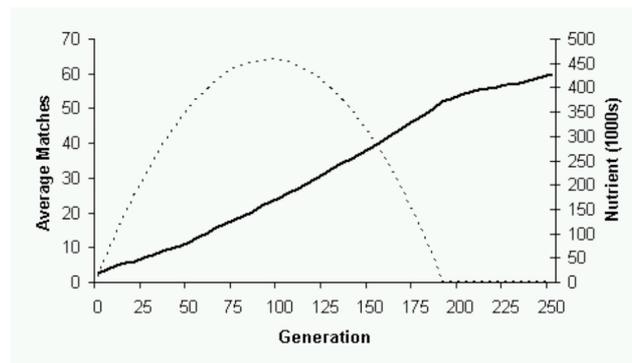


Figure 3. Evenly distributed weasels with mutation rate 1, lengthening rate 1 and nutrient 10,000.

The results of the third experiment are shown in Figure 3. Here we introduced the nutrient to investigate the effect of an environmental limit. The average number of matches (solid line) shows the characteristic change of gradient as the population passes the first phase change, at around 15 generations. The amount of nutrient available (dotted line) increases for a time, then decreases as the growing strings increase their rate of consumption. A second phase change occurs at around generation 200, when the nutrient runs out. At this point the increase length operator is disabled, reducing the rate of the population finding matches. The rise in matches continues, but can no longer be attributed to the addition of characters through the lengthening operator. Now the strings are fixed in length, but an increase in matches is still possible due to mutation and selection.

6. Discussion

In this study, we have presented a simple, novel model for investigating open ended evolution. In Weasel World, creatures increase their genotype, and hence their complexity, continuously. Our main findings are:

- Individuals in the population self organize into groups or clusters.
- This is because a group can evaluate more candidate solutions in a given time than an individual.
- Phase changes are apparent as the population re-organizes in response to changes in the size of individual territory.
- Convergence of the population can be controlled by an external constraint.

Open ended evolution was made possible by the use of a variable length genotype. Variable-length genotypes are not often studied but how they are essential to open-ended evolution and hence artificial life. We used a simple string matching problem to demonstrate a viable model using mutation and increase length operators only. The notion of territory allowed us to use a novel version of tournament selection that is more realistic. Results show that creatures are dependant upon environmental constraints for controlling their development.

In future work we plan to investigate potential applications of this work. It has not escaped our notice, for instance, that the Weasel World model provides a potential research tool for investigating patterns within data of various kinds, including text and protein sequences.

Acknowledgements

This work was supported in part by a grant from Charles Sturt University under the Early Career Researcher scheme. The simulations were carried out on the Thurgoona Cluster Computing Centre.

References

- Bedau M. A., McCaskill J. S., Packard N. H., Rasmussen S., Adami C., Green D. G., Ikegami T., Kaneko K., & Ray T. S. (2000). Open Problems in Artificial Life, *Artificial Life* 6: 363–376.
- Dawkins R. (1988). *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design*, W. W. Norton & Company, U.S.A.
- Fogel, D. B. (1988). An evolutionary approach to the travelling salesman problem, *Biological Cybernetics* 60(2): 139-144.
- Gould, S. J. (1989). *Wonderful Life*. Penguin, London.
- Green, D. G. (2000). Self-Organization in complex systems, in *Complex Systems*, Bossomaier T. R. J. & Green D. G. (Eds), Cambridge University Press: 7–41.
- Green D.G., Newth D. & Kirley M. (2000). Connectivity and catastrophe - towards a general theory of evolution, in M.A. Bedau et al. (eds.), *Artificial Life VII: Proceedings of the Seventh International Conference*: 153-161, Cambridge Mass. MIT Press.
- Green D.G., Newth D., Cornforth D. & Kirley M. (2001). On Evolutionary Processes in Natural and Artificial Systems, in *Proceeding of the Fifth Australia-Japan Joint Workshop on Intelligent and Evolutionary Systems*: 1-10, Dunedin, New Zealand: The University of Otago.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc.
- Hardin G. (1960). The Competitive Exclusion Principle, *Science* 131: 1292-1297.
- Harvey I. (1992). Species Adaptation Genetic Algorithms: A Basis for a Continuing SAGA, in *Proceedings of First European Conference on Artificial Life*, F. J. Varela et al. (eds), MIT Press/Bradford Books.
- He J. & Yao X. (2001). Drift Analysis and Average Time Complexity of Evolutionary Algorithms, *Artificial Intelligence* 127(1): 57-85.
- Holland J. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- Project Gutenberg (1999). *The Tragedy of Hamlet, Prince of Denmark*, W. Shakespeare, URL: <http://www.gutenberg.net/etext/1787> (etext).
- Ray T. S. (1991). Evolution and optimization of digital organisms, in Billingsley K.R. et al (eds), *Scientific Excellence in Supercomputing: The IBM 1990 Contest Prize Papers*, Athens, GA, 30602: The Baldwin Press, The University of Georgia: 489-531.
- Langton C. G. (1986). Studying artificial life with cellular automata. *Physica D* 22: 120–149.
- Wolfram S. (1984). Cellular Automata as Models of Complexity. *Nature* 311: 419-424.