Evolutionary Theory and Computerised Genetic Algorithms

Computerised Genetic Algorithms have many applications. A Wikipedia article lists more than 80. But missing from that list is any mention of investigations into Evolutionary Theory. Perhaps biologists don't see the point of investigating the Theory of Evolution because they are convinced that Neo-Darwinism already fully explains the phenomenon of evolution. Unfortunately most of the world would disagree and not all the objections are on religious grounds.

A simple computerised genetic algorithm can be constructed with a few weeks effort and then simulations can be carried out as a way of exploring various facets of evolution. I have named my own model Geneworld and it is composed of 16 species (Pop1, Pop2, Pop3 etc.), each with a starting population of 16 organisms and each organism is composed of 16 genes. A gene is represented by a number and each gene in Geneworld is initially given a number between 590 and 660 chosen randomly. Another number, which is initially set at 625, represents the environment and the idea is that each gene attempts to mutate towards that target number. We will justify the existence of species later.

The measure of fitness of each gene is the difference between the value of that gene and the target. A smaller difference equates to a higher fitness. A fitness value of zero is therefore defined as perfect fitness.

In every generation of Geneworld there is a five percent chance that any particular gene will mutate and its value will then change by plus or minus ten. The fitness of each organism is then measured and two are chosen for reproduction. The chances of being chosen is weighted towards the organisms with a higher fitness. A random position on this pair of genomes is then chosen for crossover and the two resulting organisms are copied into a new temporary species. This process of reproduction is repeated until the new temporary species contains the same number of organisms as the old species and it then replaces the old species. The total fitness of each organism in Geneworld is then compared to two targets. A fitness score below the lower target (remember low values indicate higher fitness) will duplicate that organism and add that copy to the population of that species. A fitness score above the higher target will cause that particular organism to be deleted from the population. The number of organisms in each species can therefore vary. The maximum size of each species population is set at 32 but no lower limit is set and therefore a species can become permanently extinct.

So, the probability of a gene being mutated is 5% and the degree of mutation (the amount by which the value of that gene is changed) is initially set at ± 10. Let's try 500 generations of Geneworld and check the average value of each gene at the end of this trial.

Trial 1 shows that after 500 generations the average fitness value of each gene is around 5. 5 is about as near as the average can get to perfect fitness (a value of 0) because of the size of the degree of mutation (± 10).

Trial 1. Environmental target fixed at 625

generation no. is 600

average gene fitness for Pop1 = 4	No. of organisms in Pop1 is 32
average gene fitness for Pop2 = 5	No. of organisms in Pop2 is 32
average gene fitness for Pop3 = 4	No. of organisms in Pop3 is 32
average gene fitness for Pop4 = 4	No. of organisms in Pop4 is 32
average gene fitness for Pop5 = 5	No. of organisms in Pop5 is 32

average gene fitness for Pop6 = 6	No. of organisms in Pop6 is 32
average gene fitness for Pop7 = 6	No. of organisms in Pop7 is 32
average gene fitness for Pop8 = 6	No. of organisms in Pop8 is 32
average gene fitness for Pop9 = 5	No. of organisms in Pop9 is 32
average gene fitness for Pop10 = 5	No. of organisms in Pop10 is 32
average gene fitness for Pop11 = 5	No. of organisms in Pop11 is 32
average gene fitness for Pop12 = 7	No. of organisms in Pop12 is 32
average gene fitness for Pop13 = 6	No. of organisms in Pop13 is 32
average gene fitness for Pop14 = 5	No. of organisms in Pop14 is 32
average gene fitness for Pop15 = 6	No. of organisms in Pop15 is 32
average gene fitness for Pop16 = 4	No. of organisms in Pop16 is 32

The concept of environment, at least in this investigation, is referring to all the other organisms (or species) that could impact on the survival on each particular organism (or species). Evolution works because of the interaction between reproduction, competition and variety. These elements all relate to organisms (or their genes).

Let's now see what happens in a situation when the target (the environment) is variable. The target is now set to be either 600 or 650 and it fluctuates between these two figures on a random basis (five percent chance of it switching in each generation).

Trial 2 shows that the organisms of Geneworld have difficulty coping with this varying environment and in this trial all species are extinct after 512 generations.

Trial 2. Environmental target varies between 600 and 650

Generation no. is 512

average gene fitness for Pop1 = 0	No. of organisms in Pop1 is 0 (extinct)
average gene fitness for Pop2 = 0	No. of organisms in Pop2 is 0 (extinct)
average gene fitness for Pop3 = 0	No. of organisms in Pop3 is 0 (extinct)
average gene fitness for Pop4 = 0	No. of organisms in Pop4 is 0 (extinct)
average gene fitness for Pop5 = 0	No. of organisms in Pop5 is 0 (extinct)
average gene fitness for Pop6 = 0	No. of organisms in Pop6 is 0 (extinct)
average gene fitness for Pop7 = 0	No. of organisms in Pop7 is 0 (extinct)
average gene fitness for Pop8 = 0	No. of organisms in Pop8 is 0 (extinct)
average gene fitness for Pop9 = 0	No. of organisms in Pop9 is 0 (extinct)
average gene fitness for Pop10 = 0	No. of organisms in Pop10 is 0 (extinct)
average gene fitness for Pop11 = 0	No. of organisms in Pop11 is 0 (extinct)
average gene fitness for Pop12 = 0	No. of organisms in Pop12 is 0 (extinct)
average gene fitness for Pop13 = 0	No. of organisms in Pop13 is 0 (extinct)
average gene fitness for Pop14 = 0	No. of organisms in Pop14 is 0 (extinct)
average gene fitness for Pop15 = 0	No. of organisms in Pop15 is 0 (extinct)
average gene fitness for Pop16 = 0	No. of organisms in Pop16 is 0 (extinct)

Despite there being many ways in which a biological genome can re-arrange itself, no re-arrangement or mutation will be of any use unless it improves or maintains the fitness of that organism. The chances of an undirected or random change producing anything useful in the human genome of 3 billion base pairs is vanishing small. Even with comparatively miniscule genomes such as those in Geneworld the only sensible way in which a species can survive a varying environment is that its population contains at least some organisms that are adapted to each of the different

environments. Way back before the pre-Cambrian explosion when single-celled organisms found themselves in a desperate struggle to survive when being attacked by innumerable different enemies, evolution would have developed systems for maintaining useful variety within each identifiable species. What gave these organisms and their species a chance to 'learn' something about the opposition was that their enemies were not strikingly different from themselves. They were built by the same code and there was a strict limit on the types of chemicals that the code could build. It was this similarity and restricted range of the different types of enemy that allowed early life to 'learn', in an algorithmic sense, something about the likely opposition and therefore to produce offspring that were most likely to survive a varying environment. This contingent survival strategy was based on the experiences gleaned by each species from their evolutionary history. Organisms suited to survive rare types of enemy would only be needed in small quantities in the population. It is the limited range of possibility in the competitive environment of these simple organisms that allowed them to be prepared. Is this possible? Of course it is. The algorithm of life has the potential to find any survival mechanism, against any threat, provided that the survival mechanism is physically, chemically and logically possible. Survival mechanisms must also not violate the important principle of the selfish gene. We must never underestimate the power of Darwin's simple idea. Darwin didn't! All that is needed is for the genetic algorithm of life to simply evolve a hierarchy of selfish genes, some being able to dictate the outcome of reproduction or to control the types of possible mutation. All genes are selfish but some are more selfish than others! The system of sexual reproduction is a prime example of how genes have evolved to control the output of reproduction. As will be argued later, group selection and altruism are other examples of the evolution of genes which can control other genes.

Let us now give the species in Geneworld the chance to evolve 'mutator genes' that control the type of mutation which each 'body-building gene' is allowed. Each body-building gene has allocated to it a mutator gene. In Geneworld, as it is currently set, each new mutator gene ensures that its body-building gene can only mutate to a value of either 600 or 650. Let us now run Geneworld again but restrict the evolution of these mutator genes to the even-numbered species only. (Pop2, Pop4, etc.)

Trial 3 demonstrates that the even-numbered species, with their mutator genes, now have no problem in surviving the varying environment. It's all very artificial but we're trying to establish a principle here. Natural selection only directly operates on body-building genes and no matter what happens to body-building genes mutator genes always survive because they exist in the genomes of both winners and losers.

Trial 3. Varying environmental target with mutator genes existing in even-numbered species.

generation no. is 600

average gene fitness for Pop1 =0	No. of organisms in Pop1 is 0 (extinct)
average gene fitness for Pop2 = 24	No. of organisms in Pop2 is 32
average gene fitness for Pop3 = 0	No. of organisms in Pop3 is 0 (extinct)
average gene fitness for Pop4 = 24	No. of organisms in Pop4 is 32
average gene fitness for Pop5 = 0	No. of organisms in Pop5 is 0 (extinct)
average gene fitness for Pop6 = 24	No. of organisms in Pop6 is 32
average gene fitness for Pop7 = 0	No. of organisms in Pop7 is 0 (extinct)
average gene fitness for Pop8 = 19	No. of organisms in Pop8 is 32
average gene fitness for Pop9 = 0	No. of organisms in Pop9 is 0 (extinct)
average gene fitness for Pop10 = 20	No. of organisms in Pop10 is 31
average gene fitness for Pop11 = 0	No. of organisms in Pop11 is 0 (extinct)
average gene fitness for Pop12 = 24	No. of organisms in Pop12 is 32
average gene fitness for Pop13 = 0	No. of organisms in Pop13 is 0 (extinct)

average gene fitness for Pop14 = 22	No. of organisms in Pop14 is 32
average gene fitness for Pop15 = 0	No. of organisms in Pop15 is 0 (extinct)
average gene fitness for Pop16 = 23	No. of organisms in Pop16 is 32

SEX

It is difficult to demonstrate the usefulness of crossover within Geneworld as it is currently set. If we remove crossover (but continue to allow the more fit organisms to be given priority for reproduction) then, in the experiments so far completed, there is no deterioration in the survival rates when compared to experiments with crossover. So what use is crossover? Firstly, we have to state that the most important step in the creation of life from inanimate chemicals is the ability to self-replicate. And then, just as important, is the ability to evolve improved copying fidelity; there is little use in evolving a survival advantage and then not being able to pass on that advantage to the future. As a way of simulating less-than-perfect copying fidelity, let us now increase the degree of mutation from ± 10 to ± 55 and run Geneworld twice, first without crossover and second with crossover. We will dispense with the variable environment and fix the target back as 625.

Trials 4 and 5 clearly demonstrate the advantage of crossover. Crossover is evolution's method of eliminating copying errors. It was evolved to improve copying-fidelity, without which organisms would merely accumulate copying errors and suffer mutational meltdown and extinction.

Trial 4. Without crossover.

generation no. is 709

average gene fitness for Pop1 =0	No. of organisms in Pop1 is 0 (extinct)
average gene fitness for Pop2 = 0	No. of organisms in Pop2 is 0 (extinct)
average gene fitness for Pop3 = 0	No. of organisms in Pop3 is 0 (extinct)
average gene fitness for Pop4 = 0	No. of organisms in Pop4 is 0 (extinct)
average gene fitness for Pop5 = 0	No. of organisms in Pop5 is 0 (extinct)
average gene fitness for Pop6 = 0	No. of organisms in Pop6 is 0 (extinct)
average gene fitness for Pop7 = 0	No. of organisms in Pop7 is 0 (extinct)
average gene fitness for Pop8 = 0	No. of organisms in Pop8 is 0 (extinct)
average gene fitness for Pop9 = 0	No. of organisms in Pop9 is 0 (extinct)
average gene fitness for Pop10 = 0	No. of organisms in Pop10 is 0 (extinct)
average gene fitness for Pop11 = 0	No. of organisms in Pop11 is 0 (extinct)
average gene fitness for Pop12 = 0	No. of organisms in Pop12 is 0 (extinct)
average gene fitness for Pop13 = 0	No. of organisms in Pop13 is 0 (extinct)
average gene fitness for Pop14 = 0	No. of organisms in Pop14 is 0 (extinct)
average gene fitness for Pop15 = 0	No. of organisms in Pop15 is 0 (extinct)
average gene fitness for Pop16 = 0	No. of organisms in Pop16 is 0 (extinct)

Trial 5. With crossover.

generation no. is 800

average gene fitness for Pop1 = 22	No. of organisms in Pop1 is 30
average gene fitness for Pop2 = 23	No. of organisms in Pop2 is 32
average gene fitness for Pop3 = 20	No. of organisms in Pop3 is 32
average gene fitness for Pop4 = 17	No. of organisms in Pop4 is 14
average gene fitness for Pop5 = 20	No. of organisms in Pop5 is 12

average gene fitness for Pop6 = 22	No. of organisms in Pop6 is 32
average gene fitness for Pop7 = 24	No. of organisms in Pop7 is 28
average gene fitness for Pop8 = 21	No. of organisms in Pop8 is 22
average gene fitness for Pop9 = 25	No. of organisms in Pop9 is 30
average gene fitness for Pop10 = 22	No. of organisms in Pop10 is 32
average gene fitness for Pop11 = 24	No. of organisms in Pop11 is 18
average gene fitness for Pop12 = 23	No. of organisms in Pop12 is 26
average gene fitness for Pop13 = 20	No. of organisms in Pop13 is 32
average gene fitness for Pop14 = 21	No. of organisms in Pop14 is 32
average gene fitness for Pop15 = 27	No. of organisms in Pop15 is 20
average gene fitness for Pop16 = 20	No. of organisms in Pop16 is 32

ALTRUISM and GROUP SELECTION

Altruism and group selection were mentioned above as another possible use for mutator genes. Let's now set up a very contrived experiment to demonstrate the principle. Competition for survival so far was between organisms in one's own species. Let's now also set up competitions between organisms in different species. Two organisms from different species are chosen at random for a fight to the death, the winner being the one with the highest fitness. But let's only set up a system of altruism in the odd-numbered species (Pop1, Pop3, etc.). Organisms in these species who have not been selected for the deadly competition can now give up some of their fitness units to the competitor from their own species who is now fighting for survival. Several organisms can each sacrifice some their own fitness units to help their colleague and these altruistic contributions to the competitor's fitness will continue until she can overcome the opposition and claim victory. For this trial we will reduce the degree of mutation from the original ± 10 to ± 1. We will revert to the varying environment and we will run Geneworld for 200 generations. Trial 6 shows that all the species without this type of altruism become extinct. Altruistic organisms are contributing to the survival of their species by reducing their own individual survival chances. Altruism works! OK it only works within Geneworld but again we are trying to demonstrate a principle here.

Trial 6. Altruism in odd-numbered species with inter-species competition.

generation no. is 200

average gene fitness for Pop1 =26	No. of organisms in Pop1 is 32
average gene fitness for Pop2 = 0	No. of organisms in Pop2 is 0 (extinct)
average gene fitness for Pop3 = 28	No. of organisms in Pop3 is 32
average gene fitness for Pop4 = 0	No. of organisms in Pop4 is 0 (extinct)
average gene fitness for Pop5 = 26	No. of organisms in Pop5 is 32
average gene fitness for Pop6 = 0	No. of organisms in Pop6 is 0 (extinct)
average gene fitness for Pop7 = 20	No. of organisms in Pop7 is 32
average gene fitness for Pop8 = 0	No. of organisms in Pop8 is 0 (extinct)
average gene fitness for Pop9 = 24	No. of organisms in Pop9 is 32
average gene fitness for Pop10 = 0	No. of organisms in Pop10 is 0 (extinct)
average gene fitness for Pop11 = 26	No. of organisms in Pop11 is 32
average gene fitness for Pop12 = 0	No. of organisms in Pop12 is 0 (extinct)
average gene fitness for Pop13 = 29	No. of organisms in Pop13 is 32
average gene fitness for Pop14 = 0	No. of organisms in Pop14 is 0 (extinct)
average gene fitness for Pop15 = 37	No. of organisms in Pop15 is 32
average gene fitness for Pop16 = 0	No. of organisms in Pop16 is 0 (extinct)

Incidentally, if we remove the type of altruism involving inter-species competition and instead opt for a type of 'intra-species altruism' where the better fit organisms can take fitness units away from lesser fit comrades then the results are shown in Trial 7. Again the altruistic genes only exist in the odd-numbered species and here we get a very different result. This type of altruism is a form of primogeniture (where the eldest son gets everything and the others get nothing) and it seems that it might not be a good idea for the species as a whole. This presumably is because the strategy of 'putting all your eggs in one basket' can prove disastrous when the environment changes.

Trial 7. Altruism in odd-numbered species with intra-species competition.

generation no. is 200

average gene fitness for Pop1 =0	No. of organisms in Pop1 is 0 (extinct)
average gene fitness for Pop2 = 26	No. of organisms in Pop2 is 32
average gene fitness for Pop3 = 0	No. of organisms in Pop3 is 0 (extinct)
average gene fitness for Pop4 = 18	No. of organisms in Pop4 is 32
average gene fitness for Pop5 = 0	No. of organisms in Pop5 is 0 (extinct)
average gene fitness for Pop6 = 19	No. of organisms in Pop6 is 32
average gene fitness for Pop7 = 0	No. of organisms in Pop7 is 0 (extinct)
average gene fitness for Pop8 = 25	No. of organisms in Pop8 is 32
average gene fitness for Pop9 = 0	No. of organisms in Pop9 is 0 (extinct)
average gene fitness for Pop10 = 27	No. of organisms in Pop10 is 32
average gene fitness for Pop11 = 0	No. of organisms in Pop11 is 0 (extinct)
average gene fitness for Pop12 = 21	No. of organisms in Pop12 is 32
average gene fitness for Pop13 = 0	No. of organisms in Pop13 is 0 (extinct)
average gene fitness for Pop14 = 25	No. of organisms in Pop14 is 32
average gene fitness for Pop15 = 0	No. of organisms in Pop15 is 0 (extinct)
average gene fitness for Pop16 = 16	No. of organisms in Pop16 is 32

Very few people appreciate the power of Darwin's idea; very few people recognise the amazing potential of the DNA algorithm. There is no need to invoke the idea of copying errors: evolution would have come up with systems for creating variety way before the pre-Cambrian explosion. The argument that emerges from this article is that evolution is primarily about surviving a varying environment and therefore the necessity of creating a system of variety maintenance and generation. This system is what we mistakenly refer to as evolution. And in that sense there have been almost no major steps in evolution since the pre-Cambrian explosion.

The next question is how does evolution come up with completely new untested ideas? Darwin knew that variability was always available when needed. If we return to Geneworld and the situation, where individual genes are alternating between the values of 600 and 650, then occasionally new sequences of genes will occur on a purely probability basis. For example, the first four genes in any individual organism will occasionally be 600,600,600,600 and if this configuration, by chance, provides a better fit in the environment in which the organism finds itself it will be selected for transmission into the future as a four-gene group. A well-known phenomenon is the propensity for biological genomes is make extra copies of genes or other stretches of DNA. These extra copies could provide raw material for new genetic configurations. Individual genes have already learnt to co-operate with each other in order to build integrated bodies and these novel combinations of co-operating genes could create new characteristics which have not yet been tested by the environment. This idea of new configurations moves the key driver of organismic change from adaptation to pre-adaptation. Human migration works in a similar way – humans (and in fact all

animals) migrate to any empty niche in which they are, to some degree, pre-adapted in order to avoid competition and over-crowding.

THE ORIGIN OF SPECIES

Steve Jones, the British geneticist, said, “The one thing that The Origin of Species didn’t explain was the origin of species”. He couldn’t be more wrong. Darwin’s book explained natural selection and he argued that a major consequence of natural selection is the existence of species. To help understand how the possible endless variety of early life ‘condensed’ into species I have used a very pared-down version of Geneworld. We start with same 16 groups of 256 randomly chosen numbers between 590 and 650. We will dispense with the idea of genes for this trial and instead we can assume that each number represents an organism. And furthermore we will not initially regard Pop1, Pop2, etc. as species but merely groups of organisms who are close enough together to be in competition with each other. Organisms on the other side of the world for example are never likely to meet in face-to-face competition. In each generation of Geneworld two organisms from each of the different geographical areas are chosen at random for a fight to the death. The winner is chosen at random and the winner makes a copy of herself and the loser is eliminated from the population. There is no target environment, no mutation, in fact nothing except this simple knockout competition. The original 16 geographical areas, each composed of a similar set of randomly chosen numbers between 590 and 650, now become, after 500,000 generations:

Pop1
627,627,627,627,627 etc.
Pop2
630,630,630,630,630 etc.
Pop3
621,621,621,621,621 etc.
Pop4
628,628,628,628,628 etc.
Pop5
648,648,648,648,648 etc.
Pop6
639,639,639,639,639 etc.
Pop7
636,636,636,636,636 etc.
Pop8
615,615,615,615,615 etc.
Pop9
610,610,610,610,610 etc.
Pop10
651,651,651,651,651 etc.
Pop11
597,597,597,597,597 etc.
Pop12
617,617,617,617,617 etc.
Pop13
644,644,644,644,644 etc.
Pop14
595,595,595,595,595 etc.
Pop15
595,595,595,595,595 etc.

Pop16
622,622,622,622,622 etc.

So there we have it; species, without any form of adaptation. This result would have been achieved much quicker with differential environments but even without adaptation species are a natural consequence of the simple arithmetic of life.

In summary, copying errors are not the providers of the raw material on which natural selection depends. What has evolved, and this happened way before the pre-Cambrian explosion, is a system which we currently describe as Evolution. What we really now have is a system for creating variety and in this sense the real evolution, the evolution of this variety-generating system ended long ago. The avoidance of extinction, or survival, is the very raison d'être of life, and the only rational explanation of extinction-avoidance, is that genes have discovered a way to understand something of the variability of the possible environments in which they could find themselves. Remember, when we are talking about environments we are discussing environments composed of countless other organisms all fighting for survival. Darwin himself would never have accepted Neo-Darwinism. He always thought that there would eventually be revealed 'the laws of variation' and the variation which was needed as the raw material for natural selection 'would not be random'. I would invoke a form of 'Occam's Razor' to justify the above explanation of evolution: given a genetic algorithm such as Geneworld how else could extinction be avoided? The twentieth century definition of the theory of evolution states that evolution proceeds by the action of natural selection on the variety created by copying errors. The copying errors part of this theory is just not good enough and the only alternative explanation for the abundance and variety of different species is the explanation as given above.