

**Rudimentary Evolution Simulation using a Genetic Algorithm**

MComp Honours Computer Science (G405)

Supervisor: Dr Phillip Lord

May 2021

Word Count: 3040

Devon Foister (180282173)

## Abstract

## Declaration

“I declare that this dissertation represents my own work except, where otherwise stated.”

## Acknowledgements

Table of Contents

[Abstract 2](#_Toc69402703)

[Declaration 3](#_Toc69402704)

[Acknowledgements 4](#_Toc69402705)

[Section 1: Introduction 7](#_Toc69402706)

[1.1 Context 7](#_Toc69402707)

[1.2 Motivation 8](#_Toc69402708)

[1.2.1 The Problem 8](#_Toc69402709)

[1.2.2 My Approach 8](#_Toc69402710)

[1.3 Aims & Objectives 9](#_Toc69402711)

[1.3.1 Aim 9](#_Toc69402712)

[1.3.2 Objectives 9](#_Toc69402713)

[1.4 Project Schedule 10](#_Toc69402714)

[1.4.1 Schedule Explanation 11](#_Toc69402715)

[Section 2: Background 12](#_Toc69402716)

[2.1.Genetic Algorithms 12](#_Toc69402717)

[2.2 Physics Simulation 12](#_Toc69402718)

[2.3 Collision Detection and Resolution 12](#_Toc69402719)

[2.1.4 Graphical Interfaces in C++ 12](#_Toc69402720)

[2.1.5 C++ as a Programming Language 12](#_Toc69402721)

[2.4 Summary 13](#_Toc69402722)

[Section 3: Design & Implementation 14](#_Toc69402723)

[3.1 Introduction 14](#_Toc69402724)

[3.2 Planning 14](#_Toc69402725)

[Genetic Algorithm Planning 14](#_Toc69402726)

[3.3 Tools & Technologies 14](#_Toc69402727)

[3.3.1 C++17 14](#_Toc69402728)

[3.3.2 Visual Studio 2019 15](#_Toc69402729)

[3.3.3 Eigen 15](#_Toc69402730)

[3.3.4 SFML 15](#_Toc69402731)

[3.4 Application Development Methodology 15](#_Toc69402732)

[3.5 Development 15](#_Toc69402733)

[3.5.1 The Algorithm 15](#_Toc69402734)

[3.5.2 The Physics Engine 18](#_Toc69402735)

[3.5.3 Collision Detection & Resolution 18](#_Toc69402736)

[3.5.4 The Renderer 18](#_Toc69402737)

[3.5.5 Constants Header 18](#_Toc69402738)

[3.6 Validation 18](#_Toc69402739)

[3.7 Summary 19](#_Toc69402740)

[Section 4: Evaluation 20](#_Toc69402741)

[4.1 Overview 20](#_Toc69402742)

[4.2 Testing Solution 20](#_Toc69402743)

[4.3 Results 20](#_Toc69402744)

[4.4 Summary 20](#_Toc69402745)

[Section 5: Conclusion 21](#_Toc69402746)

[5.1 Satisfaction of Aims & Objectives 21](#_Toc69402747)

[5.2 What is established and what can be done further? 21](#_Toc69402748)

[5.3 What Went Well? 21](#_Toc69402749)

[5.4 Even Better If? 21](#_Toc69402750)

[References 22](#_Toc69402751)

# Section 1: Introduction

## 1.1 Context

To understand genetic algorithms, you need to take a step back and first look at evolutionary computing. Evolutionary computing is a research area within computer science, inspired by the process of natural evolution. Although you probably have a solid grasp on the concept of evolution, the Introduction to Evolutionary Computing describes it as follows:

A given environment is filled with a population of individuals that strive for survival and reproduction. The fitness of these individuals is determined by the environment, and relates to how well they succeed in achieving their goals. In other words, it represents their chances of survival and of multiplying. [1]

This style of problem solving is known as trial-and-error or generate-and-test, where you have a collection of possible solutions of varying qualities that establish how close they are to the ideal solution of a particular problem. Depending on their quality, they are then used as a basis for creating new solutions. This step is repeated until, ideally, a perfect solution is found. Most of the time it is more likely to be a very good solution, rather than perfect. There is a clear link between evolution and this method of problem solving.

Environment 🡨🡪 Problem Individual 🡨🡪 Possible Solution Fitness 🡨🡪 Quality

Genetic algorithms in particular follow a very specific structure. The algorithm starts by generating the initial population of individuals. These individuals each possess a chromosome, which in turn contains a selection of genes. As these terms will be used frequently throughout this paper, here is a diagram to aid in understanding. In this context, A1 to A5 are the individuals.

Figure 1.1 – Population, Chromosomes and Genes [2]

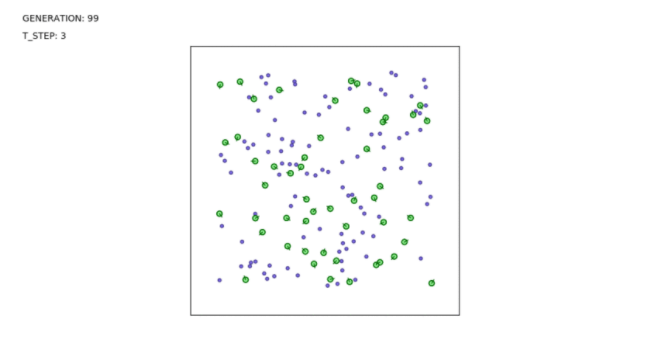
Next, each individual has their fitness calculated which ranks them depending on their ability to solve the given problem. The selection phase is next which selects the fittest individuals to have their genes passed on to the next generation. This process places the selected individuals into pairs, which then become parents.

The next two stages in the algorithm are the reproduction stages, the first being crossover. Crossover takes each pair of parents and creates offspring by combining their genes to create new chromosomes. It is called crossover as this switching of genes normally takes place at different places, dependent on one or many crossover points. This concept and its different variations will be discussed later. Mutation is the next step in the algorithm. For each new offspring, there is a low probability for each of its genes to be mutated. This could be as simple as flipping a bit or inverting entire subsets of chromosomes. Once again, this will be expanded on later.

This new population of individuals will then have their fitness values calculated and the evolution cycle continues. The algorithm will have a termination criterion that is activated when either a suitable solution is converged upon, the population has not improved for X generations or a specified generation limit has been reached.

## 1.2 Motivation

The motivation for this project began after I had taken a module at Newcastle University, titled Biologically Inspired Computing. After reading through the lecture notes for genetic algorithms [3] I wanted to see the algorithm used in the context of natural evolution. At this point, I had no idea of the extent and reach of evolutionary algorithms and just found the subject interesting. Around this same time, I was still struggling to come up with a topic for my third-year dissertation. This is when the idea of creating an evolution simulation was spawned. I then had the idea of making it interactive in order to showcase what a genetic algorithm does and make the program enjoyable to use for others who are interested in genetic algorithms. I know that the project is possible due to similar concepts being achieved by others such as Nathan Rooy [4] who managed to achieve a similar result, with the use of neural networks. Here is an screenshot from Rooy’s simulation.



### 1.2.1 The Problem

When choosing to study biologically inspired computing, I was expecting the topic to be more based around nature itself, and simulating nature within the scope of computing. Mainly due to my own poor research, I found this was not the case, so I decided to relate it more to nature myself, learning about evolution and genetic algorithms in the process. The genetic algorithms I have used so far involve simply changing parameters and watching numbers pop up in a console. I did not find this very engaging and thought it would be more interesting if I could actively see what the algorithm was doing over several generations. This could make the subject more interesting for those who want to learn about genetic algorithms, those with an interest in biologically inspired computing or people who find the idea of evolution and tiny virtual organisms interesting.

### 1.2.2 My Approach

Similar projects, such as one by Nathan Rooy [4], tend to ignore negative factors in an organism’s environment such as predators or poison. This project will, at the very least, include poison in the environment to add another factor for the organism to adapt to. My project will also be interactive, allowing the user to alter the genetic algorithm to produce vastly different results. This could be used in teaching to provide a more exciting approach to genetic algorithm optimisation. I will be using C++’s graphics library with the help of a manual from Stanford University WRONG [4] to display the entities on screen along with information pertaining to the algorithm. I will be using a collection of external resources [5] as well as University lecture material [6] in order to detect and handle collisions between organisms and food/poison, affecting the organism’s health accordingly. Each organism will have a chromosome, comprising of several genes. These genes will contain important information on the organisms, such as size, speed, food/poison perception radius. There will also be random variables used in movement calculations that gage the general intelligence of an organism. Each generation, the healthiest organisms will be carried over to the next generation, along with a selection of their offspring, generated from the genetic algorithm. The end result should be a self-sustaining colony of organisms.

## 1.3 Aims & Objectives

### 1.3.1 Aim

Explore how a genetic algorithm can be manufactured to simulate evolution amongst artificial organisms, providing an interactive interface where parameters can be altered to adjust the effectiveness of evolution.

### 1.3.2 Objectives

1. Theorize and create a genetic algorithm that increases the calculated fitness of an organism, effectively evolving it.

This is crucial to the entire project and will need to be completed early on to ensure the rest of the objectives are met on time.

1. Implement a GUI which displays organisms, other entities, and the current statistics of the genetic algorithm.

This is one of the more straightforward objectives but utterly new territory for me. Once implemented, it will not need to be modified much at all.

1. Create a movement system whereby organisms choose where to move based on their size, quality of their vision, and nearby food and poison positions.

The movement itself will be easy to achieve but calculating how to move the organisms will take a lot of research and be based on the organism's genes.

1. Create a collision detection system that allows organisms to interact with food/poison entities and have their health affected accordingly. This will be visible in the program.

This is important for the organisms to be able to react to the environment around them.

1. Program a death and reproduction cycle which will remove the previous organisms, reproduce them using the genetic algorithm and create a new set of organisms with new chromosome values.

This is key for the actual evolution to occur as there will need to be many generations for the organisms to evolve and become more suited to the environment. The genetic algorithm will manage the reproduction.

1. Introduce an interactive component to the GUI that would allow a user to modify the algorithm's specific parameters to change how fast or effective the evolution of organisms is.

This is another crucial objective as it makes up the entire interactive portion of the project. This will most likely consist of sliders for mutation/crossover chance and the option to change the selection method the algorithm uses.

## 1.4 Project Schedule

Figure 1.1 Project Plan Diagram

### 1.4.1 Schedule Explanation

The diagram above details the project plan that I have used throughout the course of the year, beginning in November 2020, and ending in May 2021. The diagram was created using tomsplanner.com in the early stages of the project. It is split into 5 sections:

* **Programming**

Any task that involves using Visual Studio and creating the actual simulation in which this project is aiming to produce. This will often have overlap with the genetic algorithm and design section, indicated by a flag of the respective colour of these sections.

* **Genetic Algorithm**

This section is for any task related to the genetic algorithm used on the project. The algorithm needed to be built from the ground up, tailored specifically to work hand in hand with the C++ program. This meant careful planning was needed, as well as achieved in a suitable time in order for the rest of the project to run according to plan.

* **Design**

These tasks relate to general designing of aspects of the project. This ended up being a very small part of the project in the end as I kept the visual front of the program very simple in order to spend more time on the complicated backend.

* **Testing**

Any task related to ensuring correct functionality of the program and algorithm. This was a long and important part of the project as every single algorithm method and aspect of the physics engine had to be tested. Although a lot of testing was done at the tail end of the project, testing was continuous throughout due to adopting an Agile development methodology.

* **Research/Documentation**

Constant research and documentation have taken place throughout the lifespan of the project. Despite there being start and end dates for each of the learning tasks, I have been continually learning new things since beginning this dissertation. Documentation has been frequent, with me writing down how I’m implementing features as I create them and keeping the aims and objectives noted down ready for the final writeup.

As you can see from the dependency arrows in Figure 1.1, nearly every piece of the project had to be completed in a set order, or else continuing would become impossible. Later I will discuss how this affected the rate in which the project was completed and whether I would change my approach in the future.

# Section 2: Background

This section of the

## 2.1.Genetic Algorithms

Newcastle University CSC3423 Genetic Algorithms Lecture [1]. This lecture gave me the inspiration to start this project and goes in depth into the history of genetic algorithms (GA) as well as in depth information on different selection methods, mutation/crossover methods and many different use cases. Along with the speaking aspect of the lecture, this gave me a solid groundwork to begin planning my own GA.

Genetic Algorithm: A Tutorial Review [6]. This short paper outlines the features of a genetic algorithm as well as the basic pseudocode for a functioning algorithm. Each component of a GA is described in detail, such as the selection methods and crossover/mutation methods for reproduction. It also outlines the different GA expressions and explains each one. I used this resource for my early research but still come back to it occasionally as a reminder of certain aspects.

Evolving Simple Organisms using Deep Learning from Scratch [2]. This blog post which I found quite recently tries to solve a similar problem in scratch using a neural network. This was helpful as it showed, at the very least, that the problem is solvable. It also led me to an interesting question. How am I going to make an organism better or worse at actually navigating to the food? The author used a neural network and used a normalized value representing the direction to the nearest food as an input. The direction the organism turns is then calculated within the neural network. Reading his blog, I ended learning about neural networks and their usefulness in a situation like this. Although I do not plan on using neural networks, I believe I can create a similar result for determining the angle an organism rotates. I can use a simple function to determine the angle from where the organism is facing to the food/poison, then calculate a turning angle based on the values stored in the organism’s chromosomes.

## 2.2 Physics Simulation

## 2.3 Collision Detection and Resolution

Jeffrey Thompson Collision Detection [4]. Another piece of coursework I have done this year saw me experimenting with collision detection in C++. This website aided me in calculating how to detect whether circles and squares were colliding. Using this I was able to create functions in C++ that achieved this task. The website extends to all polygons if this becomes necessary in the project. I will definitely be visiting this website throughout the development process of the project.

### 2.1.4 Graphical Interfaces in C++

Stanford graphics library manual [3]. Stanford University has a publicly accessible manual for the graphics library in C++. The website contains detailed descriptions on each function in the library.

### 2.1.5 C++ as a Programming Language

CSC3221 Programming for Games Lectures [5]. These lectures provided me with solid C++ knowledge and groundwork for me to build upon. I already had quite a good grasp on object-oriented programming fundamentals, but these lectures and notes helped to reinforce this.

## 2.4 Summary

# Section 3: Design & Implementation

## 3.1 Introduction

## 3.2 Planning

### Genetic Algorithm Planning

Although I wanted the algorithm to be customisable by the user, there were a couple of factors that either could not be changeable or I felt were not necessary for the project. One of these factors was the population initialisation. The two main methods for this were random initialisation and heuristic initialisation.

The problem with the heuristic approach was that there is not a known heuristic for the problem I am trying to solve. I also did not want a fully random approach as I soon found out while experimenting with the early stages of the simulation, extreme values can ruin the entire simulation. For example, if one organism has an incredibly high speed and size it will hit other organisms, in turn, pushing them towards or away from food and poison they otherwise would have found. Although realistically this could happen, this population is specifically designed to ignore each other. The approach I went with was to use a mixture of both. I experimented with a reasonable range of values that a gene could have to begin with and then the population was initialized with a range of these semi-random values.

The other factor I addressed before the project took place was the population model for the algorithm. The two most popular approaches are the steady state model whereby just a couple of organisms are replaced in each generation with new offspring, and the generation model where the entire population is replaced each generation. I found that the difference between the choices was too drastic, with steady state being too slow for a problem where generations can last over a minute long. Meanwhile, generational runs the risk of deviating away from the best solution by removing potentially perfect organisms. Once again, I compromised to a solution, I believed would be better. By default, each generation would keep 50% of its population through the selection process and new offspring will be created from this 50% through crossover and mutation.

I had to also decide on a fitness function suitable for this problem. Initially it would just take the populations current health values when the generation ends and use that, however I ran into an issue. What if the entire population died before the generation ended? Slightly less drastically, what if any 10% of the population died before the generation ended. It would not make sense that two organisms that died at different times would have the same fitness value. So, I decided to track each organism’s lifetime and current health, combining these to create the fitness value. This means that dead organisms will be ranked by the time they stayed alive and living organisms will be ranked by their current health.

## 3.3 Tools & Technologies

### 3.3.1 C++17

Before I began this project, I was not sure which language I was going to use. Java was my preferred and most used language, but I had recently started using C++, and there were a few things I liked compared to Java.

C++ is a compiled language, meaning it is much faster than other languages and offers much greater control due to being a strongly-typed unsafe language. As cplusplus.com forum user, Albatross states:

*"C++ is a language that expects the programmer to know what he or she is doing, but allows for incredible amounts of control as a result."* [x]

The language also supports dynamic type checking, which has helped create the collision classes.

Java's garbage collection system constantly checks whether memory is still in use, whereas C++ trusts the programmer only to allocate and use what they need to. In a simulation where there are many calculations taking place every frame, this is rather important.

SFML, which provides the GUI for the project, was also primarily created for the C and .NET languages, which helped in the decision-making process.

### 3.3.2 Visual Studio 2019

Visual Studio (VS) was my IDE of choice as I have used it many times for previous projects I have undertaken. The inbuilt IntelliSense feature has saved hours that would have been spent searching through class documentation by providing a robust code-completion aid. The debugger is also the best I have used and has been very useful to me throughout the year. Git integration out of the box has also made source control and backing up the project a swift process.

### 3.3.3 Eigen

The Eigen header library was used for Vector2 representation and calculations. This was needed for representing positions and the forces used on elements of the simulation. I was tempted to create my own vector class; however, it seemed more sensible to use a simple header library and fill in the blanks where needed as the time would be better spent elsewhere.

### 3.3.4 SFML

Simple and Fast Multimedia Library provides the ability to apply a simple interface to Windows applications to help with the development of games. There are five modules within the library, of which I used three. These three were the system, window, and graphic modules.

## 3.4 Application Development Methodology

As I am constantly learning and applying new knowledge to the project, I felt the waterfall approach would not be an effective methodology

## 3.5 Development

### 3.5.1 The Algorithm

The algorithm is called at the end of a generation within the main evolution simulation class. It is constructed with multiple arguments, including:

* A vector of all organisms in the population.
* An enumeration to determine the selection type.
* An enumeration to determine the crossover type.
* An enumeration to determine the mutation type.

The constructor of the algorithm class then sets the respective properties from these parameters. The constructor then calls the ‘computeAlgorithm()’ method which calls each algorithm function in the following order.

sortPopulationByFitness

This method simply sorts the population vector by their fitness values, which are calculated by adding together the lifetime of the organism and its current health when the generation ended.

#### selectionProcess -> crossoverProcess -> mutationProcess

These methods check the corresponding type properties and call the respective function. For example, if the mutation\_ parameter was set to swap mutation, mutationProcess would call the ‘mutationSwap()’ function.

if (mutation\_ == MutationType::SWAP) mutationSwap();

The crossover process method also randomises the selected population and pairs them, in preparation for mating via crossover.

#### createNewPopulation

#### selectionRouletteWheel

This function keeps a variable can contains the total of all fitness values from the population. A random value between zero and the total fitness is then created. Next, a loop goes through the population, adding their fitness value to a temporary variable and comparing this temporary variable against the random value. If the random value is less than or equal to the temporary variable, this organism is added to the list of organisms to be mated and placed in the next generation.

#### selectionTournament

This function first loops through the number of participants in a tournament (the tournament size) and adds that number of random organisms from the population to a temporary vector. This vector is then sorted by the fitness values of the organisms within. Finally, the N fittest organisms from that vector are added to the list of organisms to be mated and placed in the next generation, where N is the number of survivors. Both the tournament size and survivor amount can be modified.

#### selectionRandom

This is quite a simple method. It calculates 10 random, unique numbers between 0 and the population size. These are used as indexes for the population vector. N number of organisms from the population are then added to the new list of organisms, based on the random indexes. N in this case would be the pre-determined size of organisms to be selected for crossover and mutation.

#### crossoverUniform

The chromosomes from both organisms in a mating pair are looped through. For each gene, in both chromosomes, there is a 50% chance they are swapped in the offspring. A diagram detailing from tutorialspoint.com is below. [x] These offspring are then placed, with the parents, in a vector ready to be passed to the mutation function. This process is repeated for each mating pair.



#### crossoverSinglePoint

For single point crossover, a random point along the chromosomes of the parents is chosen using the program’s random engine. The tail end of each chromosome is then swapped in the offspring. Both new organisms and the parents are then moved to a vector to be passed to the mutation function. This process is repeated for each mating pair.



#### crossoverMultPoint

The multiple point crossover function is very similar to the one above, except this time there are two dividing points on a chromosome. To get these separate points, I created a vector containing all indexes in a chromosome. The vector is then shuffled, and all but the first two values are removed. This remaining vector would now contain the two dividing points.



#### mutationScramble & mutationInversion

The mutation methods start off by doing a probability check to see if the mutation chance condition is met. For scramble mutation, a vector containing all indexes in a chromosome is created. This vector is shuffled, and the first two values are sorted and placed into a separate vector. These two values act as a subset of the chromosome. The genes within the subset are also randomly shuffled and replace the values previously in the subset of the chromosome. This is then repeated for each organism in the population.

Inversion mutation is the same as scramble in the sense that a random subset of each chromosome is altered. The difference is, instead of scrambling the subset, it is instead inverted.

#### mutationRandomValue

#### mutationSwap

Swap mutation is quite simple. If the probability check is successful, then a gene in the chromosome is swapped with another gene. This is then repeated as if it is only done once, the other mutation methods are more likely and more capable of having a larger impact on the population.

### 3.5.2 The Physics Engine

### 3.5.3 Collision Detection & Resolution

### 3.5.4 The Renderer

The renderer class contains all the drawing functions needed for the program as well as holding the window object itself. The four drawing functions included are:

* DrawCircle: used for drawing food and poison elements.
* DrawBox: used for drawing the outer walls of the environment.
* DrawOrganism: used for drawing the population.
* DrawHealth: used for drawing the organism’s current health above it.

Main.cpp is where the program begins, creating a new renderer object. Using the SFML library, a window is then created at a resolution of 1280 by 720 and a framerate limit is set. I chose 720p as 16:9 is the most common aspect ratio for modern computer monitors and laptop. Also, even cheaper laptops tend to have a resolution no smaller than 1366 by 768. In the worst case where the resolution is smaller than this, the program can be resized, only altering the simulation for a few frames.

At this point a new EvolutionSimulation object is created, with the window object passed into its constructor, allowing elements to be drawn outside of main.cpp. Next the main game loop begins.

### 3.5.5 Constants Header

The constants file holds all global variables used throughout the program. Throughout the development of the project, it would store variables that I had not yet decided how to use or where to use them. Some of the variables it holds include:

* Population size.
* Number of food pieces.
* Number of poison pieces.
* The window resolution.
* The framerate limit.
* Default mutation chance.
* Default crossover chance.
* Default tournament size.
* Default survivor amount.
* Default generation time (in seconds).

## 3.6 Validation

## 3.7 Summary

# Section 4: Evaluation

## 4.1 Overview

## 4.2 Testing Solution

## 4.3 Results

## 4.4 Summary

# Section 5: Conclusion

## 5.1 Satisfaction of Aims & Objectives

## 5.2 What is established and what can be done further?

## 5.3 What Went Well?

## 5.4 Even Better If?

# References

[1] – Bacardit, J 2020, *L02 – Genetic Algorithms,* lecture notes, Biologically-Inspired Computing CSC3423, Newcastle University, delivered 18th October 2020.

[2] – Rooy, N 2017, *Evolving Simple Organisms using a Genetic Algorithm and Deep Learning from Scratch with Python,* Nathan Rooy, viewed 20th January 2021, <<https://nathanrooy.github.io/posts/2017-11-30/evolving-simple-organisms-using-a-genetic-algorithm-and-deep-learning/>>

[3] – Stanford University 2020, *graphics.h,* Stanford University, viewed 12th January 2021, <<https://web.stanford.edu/class/archive/cs/cs106b/cs106b.1126/materials/cppdoc/graphics.html>>

[4] – Thompson, J 2020, *Collision Detection,* Jeffrey Thompson, viewed 20th December 2020, <<http://www.jeffreythompson.org/collision-detection>>

[5] – Speirs, N 2020, *C++ Lecture Material*, lecture notes, Programming for Games CSC3221, Newcastle University, delivered 29th September 2020.

[6] – Mukhopadhyay, D, Balitanas, M, Farkhod, A, Jeon, S & Bhattacharyya, D 2009, ‘Genetic Algorithm: A Tutorial Review’, *International Journal of Grid and Distributed Computing,* vol. 2, no. 3, pp. 25-32.