Imperial College
London

Evolution of Artificial Life

Investigating Optimisation in Gene Space

Botond Branyicskai-Nagy
CID: 01871330

Supervisor: Dr David Clements

Assessor: Dr Tim Evans

5958 words

Department of Physics
Imperial College London

July 2024

Declaration of work undertaken

I hereby declare that I have completed the following work in collaboration with my project partner. Except where specific reference is made to the work of others, the contents of this report are original. Throughout its duration, my partner and I contributed equally to the success of the project by splitting the coding work and analysis between us. We also regularly discussed our progress and results with our supervisor on a weekly basis.

Botond Branyicskai-Nagy
July 2024

Abstract

This project investigates the optimisation problem of evolution by using a numerical agent-based simulation. The results provide valuable insights into the dynamics of this process in nature. After validation using the Lotka-Volterra model, we examined reproductive strategies and vision in a single-species system. Our findings confirmed hypotheses about the effect of evolutionary pressures on reproduction rates and optimal gene values for food distribution. We also observed the positive impact of mutations on survival rates, as well as coevolutionary dynamics in predation, namely through speed and colour. These results aim to demonstrate the adequacy of the evolutionary algorithm to meet the criteria enforced by natural selection and the role of mutations and pressures in shaping the genetics of populations over time. This project provides a possible starting point for further investigation into computational modelling of evolution.

Table of contents

Chapter 1

Introduction

1.1 Evolution's optimisation problem

Optimisation problems are at the heart of our physical world: be it the shortest path for light, the least energy for a molecule, or the lowest elevation for a waterfall. In each of these, we know there is a well-defined functional that is being minimised, the action, which can be solved exactly – at least by nature if not us. But we like to optimise things too, in fact it is not so far-fetched to observe that most things we do as people can be viewed as a minimisation problem. Businesses maximise profits while minimising spending, the police tries to minimise crime, students maximise their marks. These are examples of very hard problems, where an optimal solution is impossible to find, yet, stocks mostly tend upwards, crime drops, and most students pass their exams. With no clear analytic function to minimise, in some cases the best method is trial and error, rejecting the worst choices and moving towards the best.

When Darwin published *On the Origin of Species*, he implicitly recognised yet another optimisation problem: evolution. The novelty lies in the subtitle, *by Means of Natural Selection*, where he goes against the status quo of a top-down selective algorithm by intelligent design (a Creator), in favour of a bottom-up, self-organised process, run by the survival and reproduction of the fittest. Mathematically, this implies there is an algorithm (natural selection) that assigns a *fitness* value to each possible organism – giving the probability of its survival in its environment [1]. In the space of all traits, given independent (orthogonal) axes – body height, hair colour, etc. – each individual organism is a point. For our purposes, we can assume there is a one-to-one correspondence between genes and traits, so that a genotype encodes a specific, well-defined phenotype – such that the gene *height:* 2 m will mean the organism will be 2 metres tall – which is not the case in general in nature. The axes are then the gene-values, defining a multi-dimensional *gene space*. Adding a dimension for the fitness, and its value for each of these points defines the function to be extremised, called the *fitness landscape* [2]– if there were only two possible traits/genes, this would be the surface of a 2D histogram (see Fig. 1.1). In a selective environment, this landscape will have clear local maxima – indicating highly successful organisms (in terms of reproduction) – and minima – impossible, infeasible or impractical combinations of genes for an individual. Realised organisms *test* their point on the landscape, and if they are higher up on the fitness scale, the more likely they are to reproduce by definition, resulting in more individuals at higher altitudes and movement of the most common gene point towards the peak of the landscape over time - hence slowly maximising the fitness function. In some specific applications, such as recent breakthroughs in chemistry [3], this approach to optimisation has been proven to be adequate.

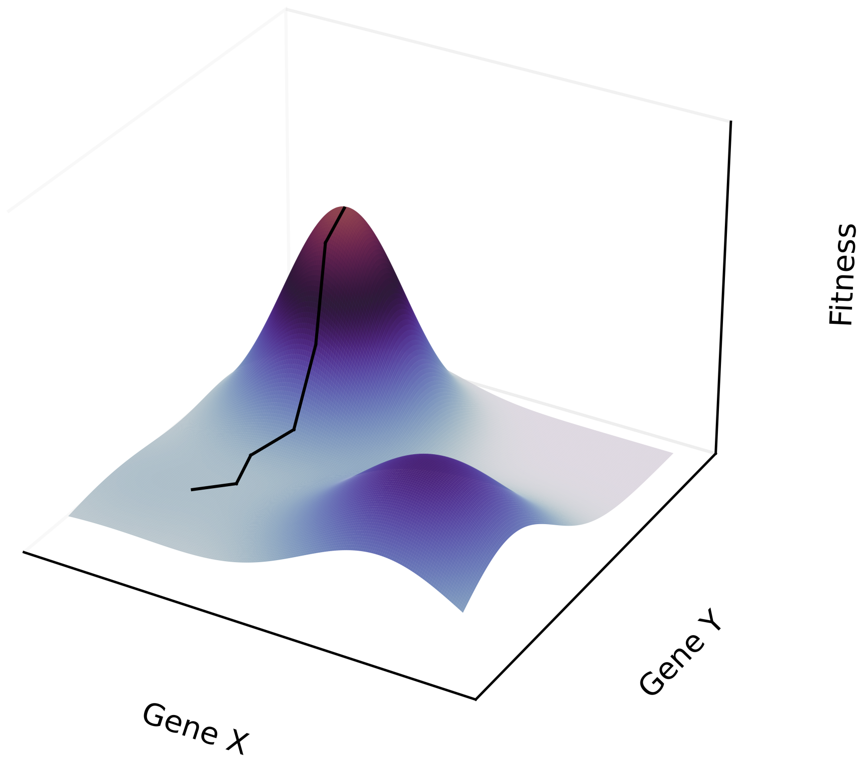


Fig. 1.1 Illustration of a theoretical fitness landscape in gene space, with two local maxima and an optimisation algorithm finding one of them (in black).

In practice, we never have access to the true underlying fitness landscape, which in general changes over time - and in fact, is a function of the genetics of the organisms trying to find its maximum. This makes it difficult to determine whether the global optimum has been reached, or if the system is trapped at a local maximum - we postulate stability over time is a good indicator, but it is not a definitive one either. Our aim is to investigate key evolutionary processes by means of a numerical simulation, exploring adaptation to different environments, examining how mutations affect survival rate and drawing conclusions as to the adequacy of nature's optimisation algorithm for navigating the impossible problem of developing and sustaining life on Earth.

Chapter 2

The simulation

2.1 Initial setup

To set up an environment where the genetic processes of a population can be tracked, we first need to consider what the essential elements are in a real ecosystem, with enough detail so that natural selection and its emergent properties can be observed, but taking care not to introduce so much complexity that would make it hard to isolate cause and effect relationships.

The simulation takes place on a discrete square grid of positions, that can be mapped onto the surface of a torus (see Fig. 2.1) - a choice made to ensure all locations are indistinguishable. Our organisms are modelled as simple agents operating in an environment using primitive logic. Agents can move one cell up, down, left or right to find food scattered across the toroidal planet. Initially, their movement is randomised, constantly moving but changing direction with a probability of 0.2 every step. This is an arbitrary value that was found to be optimal for this setup and is the same for all agents, however there are examples of other work where the rotation probabilities (separately for each direction) are chosen as variable genes, resulting in spinners or line-runners depending on the spatial food distribution [4, 5].

As their energy level decreases linearly with each timestep, for the agents to stay alive they have to find food often enough that they do not starve between sources. Once they reach a certain energy level, as set by their reproduction threshold, they divide into offspring, splitting the food available equally. This is done asexually, meaning it requires only one individual to make copies of itself - as is the case with bacteria. In nature, this always means splitting into two halves, analogous to cell division, but to capture the dynamics of as many types of organisms as possible, in our case they have the ability to produce an arbitrary number of offspring at reproduction. The argument can be made that if we chose to stick with modelling bacteria, this is equivalent to multiple subsequent cell divisions in the same timestep - with the caveat of odd numbers not resulting in equal food splitting.

The first genes we introduce are the integer values for the energy (food) threshold for reproduction and the number of offspring to produce (illustrated in Fig. 2.2). Each agent therefore has a specific value assigned for these that remains constant during its lifetime. At reproduction, these values are passed on to the offspring, with a small

(~ 0.1) probability of mutating (changing) by a small amount either higher or lower - we can choose to set this probability to zero to observe the dynamics with no mutations. The uniform restriction of $\{g \in \mathbb{N} \mid g > 1\}$ was enforced for all genes g to adhere to realistic considerations.

The reasoning behind this choice of genes is straightforward: there is no immediately apparent reason to pick any specific value over another by default and it serves as a good starting point to allow these to vary by evolution. While they are independently specified, there are effective constraints on the ratio of the two genes, as we will see below.

To investigate the genetic behaviour of other traits, we decided to look at environments where we could set clear expectations on certain properties that would likely be of advantage. One example of this is a planet with a so-called Garden of Eden, where most food objects are spawned: here, if an agent can see where the nearest source is, its chances of survival are greatly increased. Hence the addition of a sight gene (see Fig. 2.2), along with movement towards that source, allows us to test this hypothesis. Similarly, moving by a cell every timestep is expected to provide a competitive advantage against, say, once every third timestep - giving us a gene for the agent speed, albeit given as an inverse, and with a maximum value of 1 cell/timestep. Both of these must have an associated energy cost, as is the case in nature, to avoid runaway values that are unrealistic, as well as introducing a trade-off that means there is - in theory - an optimal cost to value balance for the given environment. In the case of the former two genes, this was implicitly included as the more offspring are produced, the less food they will start out with, and similarly with the threshold level.

2.2 Predators

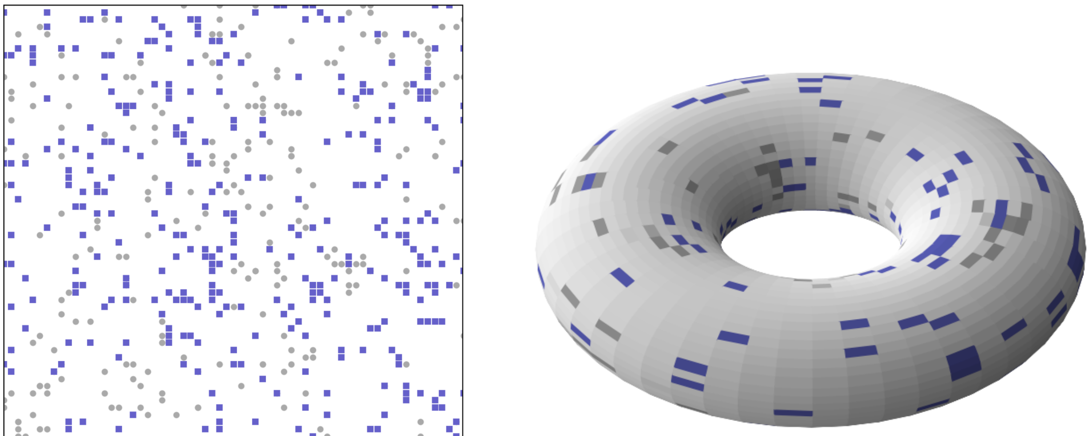
In nature, most ecosystems include complex predator-prey relationships. To better understand how these might have evolved, our simplified model can be extended to include a new type of agent, a carnivore, that hunts for the herbivorous kind described up to this point. Their movement, logic, reproduction and set of genes differ only in that they see the herbivores - their prey - the same way the herbivores see food - and therefore move towards the nearest such organism seen. To level the field, the prey are capable of fleeing if the predator is within their field of view, which takes precedence over food-finding (unless they are very close to starvation).

With inspiration from nature - toxic frogs, snakes and certain fish - *colour* was added as a gene, recognising the fact that it has no relevance outside of the predator-prey relationship. For prey, it can be interpreted as visual appearance in the form of brightness or vibrancy, and for predators it indicates a preference for hunting prey with a colour similar to its own value. The preference peaks at the matching value and drops off as a Gaussian in the probability for choosing to chase the given prey or not. As opposed to the other four genes, here there is an upper limit as colours move between 1-16 in integer steps - again, to avoid runaway values, but without an associated energy cost in this case.

2.3 Parameters and variables

There are a few general parameters to be tuned for the simulation to represent a realistic, controlled environment. The first of these is the size of the grid, which needs to be large enough to accommodate the desired number of agents without a significant amount of spatial overlap - which was chosen to be allowed based on considerations of alternatives introducing unnecessary complexity to the model. A related pair of parameters are the unit amount of food, which must give an agent enough steps to traverse to the next source on average, and the starting energy that each agent has.

Just as in any real ecosystem, the relative spatial distribution of resources plays an important role in the evolution of the gene space: the difference between lifeforms developed in the savanna and the jungle is striking. In practice, here this means specifying the initial spatial configuration - location and density of food and agents - as well as the first agents' genetic background, and these will have an effect on the evolution of the system. These are the initial conditions for the fitness landscape, and the subsequent food input procedure determines its shape throughout the simulation. As we wanted to investigate how different environments drive the genetics in different ways, and the effects of varying the amount of food in the system as a function of time, the simulation had to allow for this as well.



(a) (b)

Fig. 2.1 (a) Flattened and (b) 3-dimensional visualisations of a typical map, agents shown in purple and food in grey. Toroidal topology was chosen to eliminate any special positions such as corners and edges, that would introduce further unwanted complexity. This is equivalent to saying that an agent reaching the edge of the flattened square map can step onto the opposite edge (ie. the top edge is connected to the bottom, and the left edge to the right).

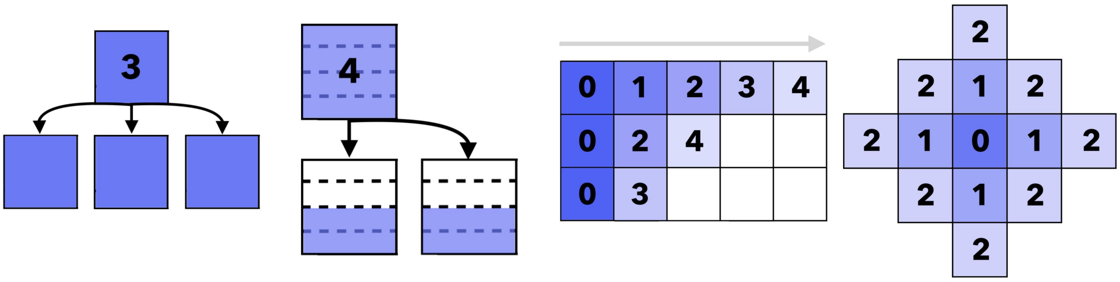


Fig. 2.2 Gene definitions from left to right: (a) *number of offspring*: one agent splits into $N > 1$ (3 in example above) agents by asexual reproduction (b) *reproduction threshold*: in food units; when threshold reached, reproduction occurs and food is split equally among offspring (c) *(inverse) speed*: time steps taken to move one square (d) *sight*: largest von Neumann distance of food objects that are noticed by agent (at position 0), determines movement

Chapter 3

Results

3.1 Population Dynamics

3.1.1 Lotka-Volterra Validation

Before we can conduct any meaningful experiments, we need to verify that we have indeed built a representative population model. The Lotka-Volterra (LV) equations [6] are used for modelling predator-prey population dynamics in real ecosystems. They are an example of the more general Kolmogorov models from probability theory. The change of the two populations in time are given by a coupled pair of differential equations, which we have to modify slightly to fit our setup:

$$\begin{aligned} \frac{dx}{dt} &= \alpha - \beta xy, \\ \frac{dy}{dt} &= \delta xy - \gamma y \end{aligned} \tag{3.1}$$

where x and y are the prey and predator populations respectively. α is the (constant) rate of food input into the system, γ is the starvation rate (sometimes called basal metabolic rate) of predators and finally β and δ are factors describing the interactions between the two species: the decrease in prey and the increase in predator populations respectively. In the first instance, we will use this to model the food-agent system as a prey-predator relationship. The food is not mobile, but more importantly it cannot starve, which is the reason the original equations have to be modified to Eq. (3.1). This, of course assumes a continuous domain, and so it has to be further adapted to the discrete nature of our simulation which works in steps of Δt. It is also an implicit assumption of the LV model that the organisms remain identical throughout, ie. it does not account for mutations - therefore we have to turn them off for this comparison.

To validate the experimental method, we ran a series of simulations with different initial conditions, tracking the food and agent populations, and compared these against the Lotka-Volterra prediction (using x to predict Δy, to obtain y and vice versa), with remarkable agreement in most cases (see Fig. 3.1). An evident difference is the relative smoothness of the LV graph, but this is expected as the model assumes an infinitely large system and is derived from an analytic formula, whereas the simulation is inherently finite and discrete. As we increase the number of agents, we expect - and find - that the agreement improves as fluctuations become negligible. Taking $\Delta t \to 0$ would also be necessary but the increased computation time sets an effective limit, and the current setup gives clear and reliable results regardless.

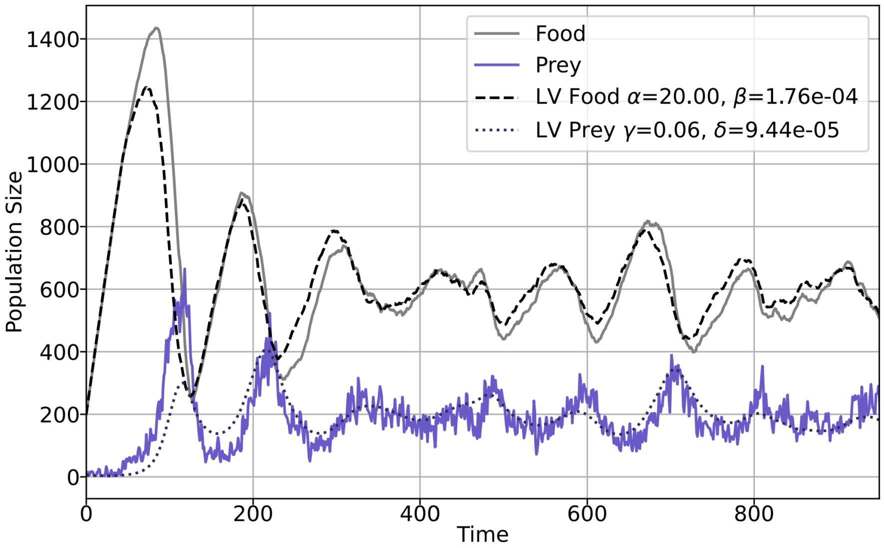
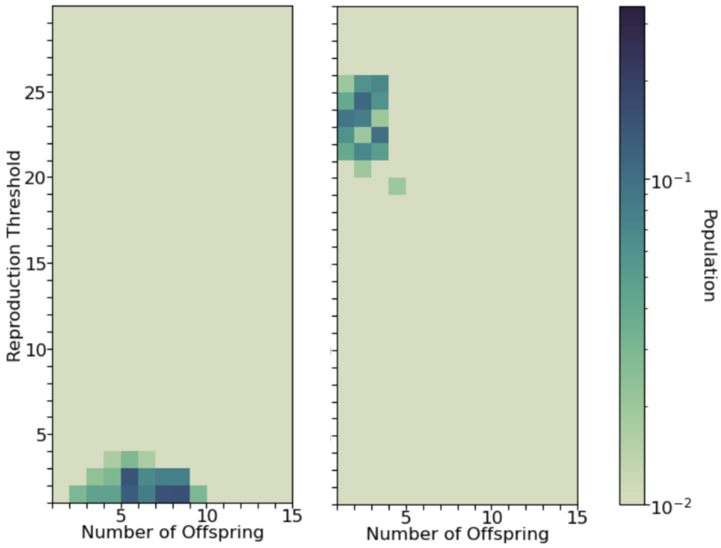


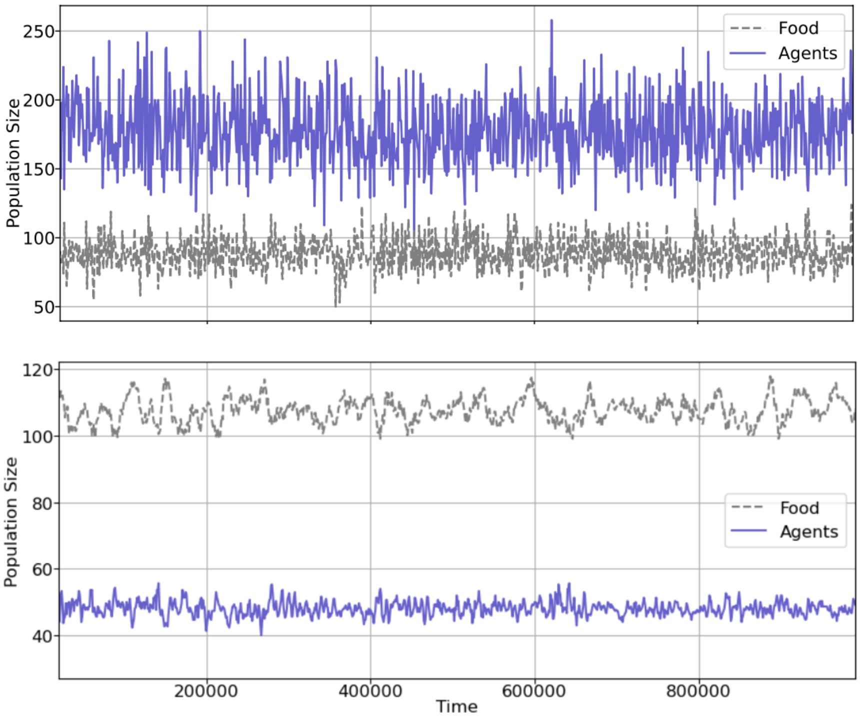
Fig. 3.1 Amount of food and agent population over time, compared with Lotka-Volterra predictions (dotted and dashed lines). General agreement is confirmed, minor fluctuations are due to finite system effects and discrete time. Interaction parameters β, γ, δ were computationally fitted by minimising variance, best values are quoted on the plot, associated fitting uncertainties were relatively low: $\sigma_\beta = 0.86 \cdot 10^{-4}$, $\sigma_\gamma = 4 \cdot 10^{-3}$, $\sigma_\delta = 0.89 \cdot 10^{-5}$ as approximated from the inverse of the Hessian. Parameter α is fixed as it is given by constant food input.

3.1.2 Reproductive strategies

With the simulation's agreement with the LV model confirmed, we can move on to investigating evolutionary processes in the context of our first two genes: number of offspring and reproduction threshold. We know from biology - life history theory, or r/K selection specifically - that there are two main *strategies* observed in nature when it comes to these genes - rather their biological analogues - depending on the environment. Species that produce a great number of offspring during their life cycle tend to give little help and resources to them, are physically weaker and on average live shorter than those where a parent only has a few of its young to take care of. Tuna, practising an example of the former, r-strategy, produce millions of eggs of which only a few



(a)



(b)

Fig. 3.2 r/K selection: (a) Two separate simulated populations' (final) gene space distributions for number of offspring and reproduction threshold. The two occupy different regions in this space, with the r-strategists (left) low on the vertical axis but with more offspring, and the opposite for K-strategists (heat map on right). (b) Corresponding population sizes for the r- and K-populations (top and bottom, respectively) with the characteristic high-amplitude fluctuations in the former case, lower for the latter. Both average to their environment's carrying capacity, which do not match here as food density is lower in the K-strategist's environment.

hatch and reach adulthood [7], in contrast to a whale, a K-strategist, that typically only gives birth to one calf at a time [8]. There is a greater uncertainty of survival for the tuna, which gives rise to high amplitude oscillations in their population size. In comparison, the whales' dynamics follow a more stable curve, but both of them must average out to the given environment's carrying capacity K (from the German Kapazitätsgrenze) as this is the point of equilibrium.

Depending on the environmental variables - namely food density (or food input) - our simulation reproduces both strategists, as the more suitable one slowly outcompetes the other, and the gene space evolves towards the associated region (see Fig. 3.2), resulting in the characteristic population-oscillations for each. It is also possible in some cases to observe stable populations of both types cohabiting, giving a bimodal distribution in gene space. In other words, two species are present in the same environment - using a definition purely based on having similar genetics.

3.2 Adaptation

3.2.1 Evolutionary pressures and reproduction rate

The notion of r/K selection theory, more frequently referred to as fast/slow life history, can be extended beyond a binary classification into a continuous spectrum. For clarity, it is useful to analyse the adapted populations in terms of the ratio of the two genes in question, reproduction threshold and number of offspring. This quantity is a measure of how much energy is given to each offspring at birth (or division) and is therefore directly related to the minimum life expectancy of that agent (using an energy usage of 1 unit per timestep, they are numerically equivalent). Assuming a final, stable distribution in gene space has been reached, the ratio is an indicator of the most suitable amount of energy to store in that specific environment to have the best reproductive success - a fitness landscape maximum.

Given the evident correspondence between energy stored and the average travel distance needed to find food, we expect the food density (per unit area) to have a direct effect on the optimal value of the ratio of genes mentioned above. As such, we can look at this dependence over multiple runs of the simulation, varying the food input for each. We also repeat each run to obtain a statistical uncertainty, the error of the mean gene-value ratio. The result is indeed a monotonically decreasing graph, initially indicating a power law relationship but stagnating at 1 above a given density. Finite size effects are likely at play, such as overcrowding and resulting numerical artefacts

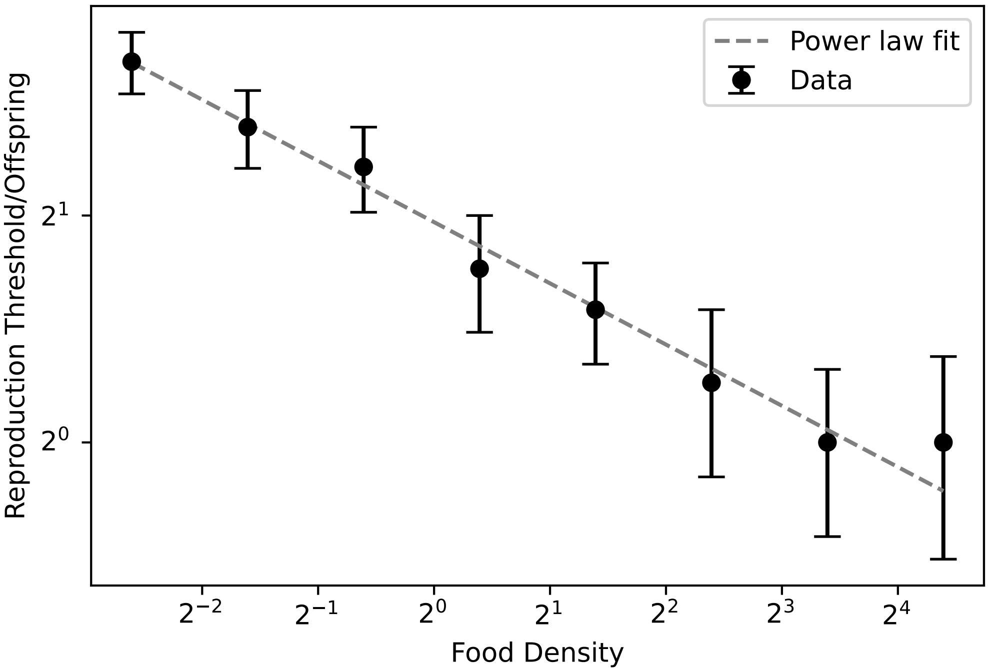


Fig. 3.3 Ratio of reproduction threshold to number of offspring gene-values in environments of various (constant) food densities. Population mean and uncertainty computed in the latter stage of the systems' evolution, where stability was numerically evaluated (by maximum fluctuations). As the linear log-log plot suggests, there seems to be a power law relationship, which is confirmed by performing a numerical fit of the form ax^b: $a = 1.96 \pm 0.04, b = -0.27 \pm 0.01$, with a high p-value (from χ^2 test) of 0.57.

being amplified for these very high densities. This shows us the spectrum between K-strategists' success in low density environments - where the need to store energy creates evolutionary pressure - and r-strategists prevailing as density decreases. This is analogous to the scale of large and stronger animals (eg. rhinoceros) to small, weak, short-lived ones (such as fruit flies). The exact power law (with exponent $\sim -1/4$) is not readily understood, as the relationships between food storage, average path between sources, overall agent population are all factors that need to be taken into account, therefore it is not a trivial dependence.

It is important to note that due to the higher number of offspring, r-strategists possess a wider genetic spread and therefore are more capable of adaptation. The faster life cycle also means they - as a population - can respond more quickly to sudden environmental changes, where K-strategists are virtually left waiting for the next generation. This serves as a plausible explanation for the magnitudes higher fraction of species following the former strategy, but their counterparts typically being more resilient individually.

3.2.2 Sight and spatial food distribution

The locations of food sources are also in direct relation with the sight gene - which determines movement, with the agent making a beeline to the nearest source. As with energy storage, we expect the distance to which it is advantageous to see - without incurring to high an energy cost spent on vision - to decrease with higher food density in the environment. A low number of sources per unit area means the average shortest path between these points is larger, therefore if the agent is incapable of seeing any food sources within the distance it can still traverse, it will be at a disadvantage. Finally, the cost for sight introduces an upper limit to gene, as there will be an optimal balance where the ratio of advantage gained to energy cost is at a maximum - we expect this to be the mean value in the genetically stable population.

Indeed, the results in Fig. 3.4 confirm our hypothesis, with a decreasing graph that flattens to 1, the minimum allowed value. The boundary is between 1/2 and 1/3 food units/cell, where even within a von Neumann distance of 1, there is likely to be 1 to 2 food sources, meaning there is no real advantage to seeing any further. As to the exact relationship, we see it is close to an inverse-square law, but this has no intuitive explanation. If we instead try to find the distribution based on the assumption that the required (optimal) probability of finding food in the agent's field of view is a constant of the simulation, irrespective of its environment, using a binomial distribution of the form

$$\mathbb{P}(n > n_{min}) = \sum_{n=n_{min}}^{A} \frac{A!}{(A-n)!\; n!} \rho^n (1-\rho)^{A-n} \tag{3.2}$$

for $A, n, n_{min} \in \mathbb{N}$ and $A > n_{min}$, or in the continuous limit, where $A \gg 1$ and up to a normalisation constant N

$$N \int_{n=n_{min}}^{A} \frac{\Gamma(A+1)}{\Gamma(A-n+1)\Gamma(n+1)} \rho^n (1-\rho)^{A-n} dn \tag{3.3}$$

for area in sight A, number of food units n in the sight area, with n_{min} being the minimum needed for survival (on average), and ρ the probability of finding food in a randomly chosen cell. Upon solving the integral, a numerical fit can be attempted, as before, however due to the high uncertainties in the data, and the lack of knowledge about either the minimum probability or n_{min} giving two degrees of freedom, the range of possible values was found to be too large to draw any meaningful conclusions from. Further analysis of the behaviour of these parameters is needed to establish the complete dependence of A on ρ.

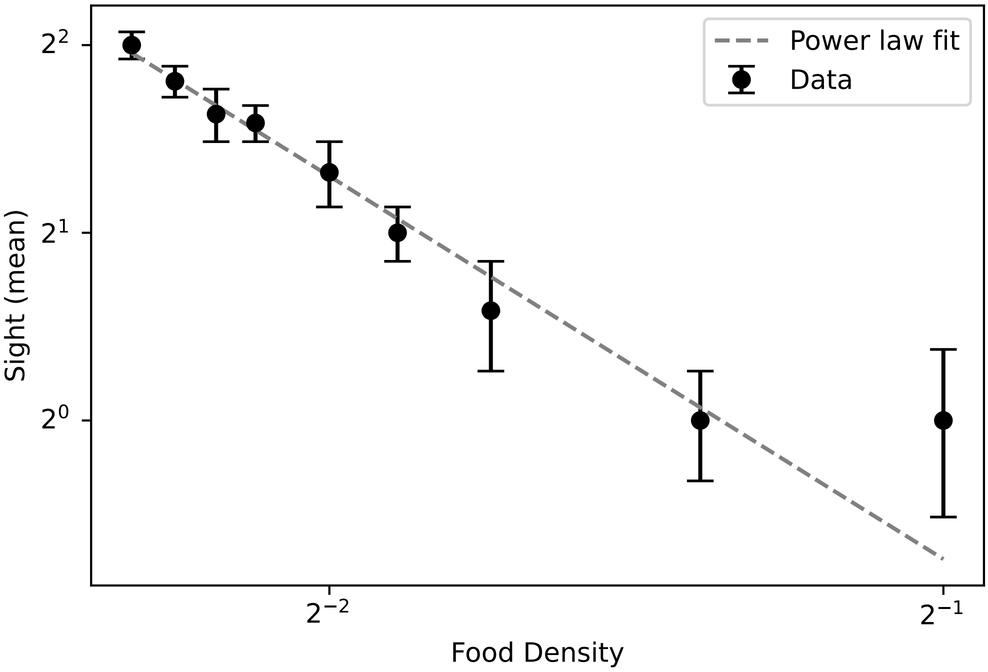


Fig. 3.4 Sight gene values in environments of various (constant) food densities. Population mean and uncertainty computed in stable regime. The linear log-log plot suggests a power law relationship, confirmed by the best numerical fit to the data, - excluding the last point which is a clear outlier to this region - with ax^b: $a = 0.15 \pm 0.04, b = -2.04 \pm 0.17$, again giving a high p-value (from χ^2 test) of 0.43.

3.2.3 Effect of mutations

As we have seen so far, competition drives selection when two or more gene-points coexist. It is important to note that if we start with a spread of values in gene space, we can observe selection even if the mutation probability is zero. The mutations provide a mechanism by which the spread is created *naturally*, but they are not strictly necessary in the context of simulating selection. There are cases in nature where this can occur, if a new species were to be introduced to an otherwise stable ecosystem (by means of migration, or otherwise) the more suited organism could out-compete the weaker far faster than the evolutionary timescale. In this case, any mutations would be negligible as there is simply not enough time to develop traits that provide any

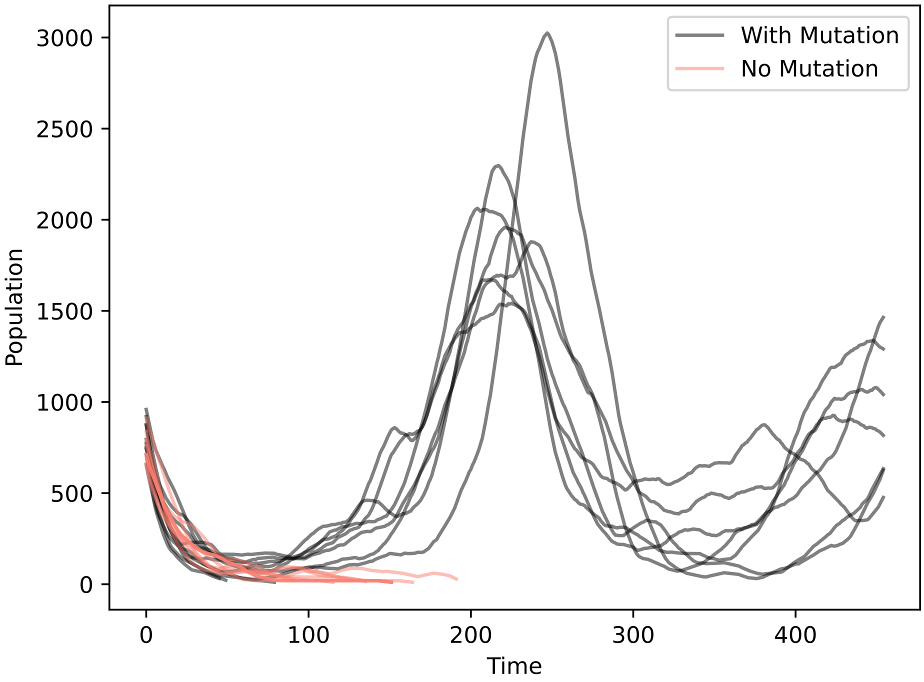


Fig. 3.5 Comparison of the survival of non-mutating and mutating populations in the same, highly selective environment for sight. Out of 10 runs of the simulation for each case, 7 survive by adaptation when the mutation probability is at 20%, while none do if mutations are not allowed.

significant advantage. From a computational standpoint, evolution works in discrete steps, generations, which over millions of years effectively approaches a continuous process; over a single lifetime however, it has no effect.

Whether a mutation is useful - ie. provides an advantage - depends on the environment. To isolate the benefit of having mutations in the first place, we can construct a space where the selection pressure is so high, that the chance of survival is zero, unless a species can adapt. This experiment requires the initial gene space to contain sub-optimal values only, to enforce adaptation. We chose an environment where the food input is very low, but the total area is large, resulting in low food density. The scarcity of food sources means randomly moving agents will rarely find enough to survive and reproduce, making the environment highly selective for seeing further (the sight gene). As we expect, repeating the experiment with and without allowing mutations shows there is a clear tendency for populations that mutate to adapt and survive, while ones with a fixed set of genes never do (see Fig. 3.5).

Even though in reality mutations are an inherent part of DNA replication, we again can make the argument that a sudden change in the environment can rule out any genetic adaptation, as that would require evolutionary time-scales. Therefore, the benefit of mutations is observed in nature in the same way as discussed in our experiment, only if the environment does not suddenly change to a highly selective one, but does so over a long period of time - allowing many generations to pass.

3.3 Coevolution

3.3.1 Predation

The introduction of predators to the ecosystem brings with it countless possible experiments, however we will focus on just a couple that demonstrate the process of coevolution most clearly. The concept itself is rather simple: taking two cohabiting species (living in the same environment), changes in the genes of one induces the other to adapt too. As one of the species can be considered part of the other's environment, this is just a special case of adaptation. Examples are readily observed in nature, a class of these being mutually beneficial processes, such as insect- and bird-pollinated flowers that have evolved to match the animals' spectral (colour) sensitivity, its preferred physical shape and nectar sugar concentration, among others. However, here we will be concerned with the predator-prey instance, which is observed in every such relationship and is the source of one of the strongest evolutionary pressures to exist.

In order to validate this extension of our model, we again turn to the Lotka-Volterra equations, augmenting them to be suitable for the dynamics of a system with 3 species. The hierarchy consists of the population of predators z hunting for prey y, who in turn are foraging for food x. We again have a system of coupled differential equations

$$\begin{aligned} \frac{dx}{dt} &= \alpha - \beta xy \\ \frac{dy}{dt} &= -\gamma y + \delta xy - \epsilon yz \\ \frac{dz}{dt} &= -\zeta z + \eta yz \end{aligned} \tag{3.4}$$

where constant parameters α through δ are the same as those defined in Eq. (3.1), ϵ describes how hunting reduces the prey population, ζ is the starvation factor for predators and η is related to how the predators benefit from hunting. As before, the LV model assumes no mutations, but an added difficulty is finding stable solutions where neither species goes extinct. Setting $y, z > 0$ for all t yields the following survival condition on the parameters [6]:

$$\alpha\eta \geq \zeta\beta. \tag{3.5}$$

Since β has been estimated above (see Fig. 3.1), and is not significantly changed by the addition of predators, if we approximate ζ and η, it is possible to tune the food input α to obtain a stable system. This was done successfully (see Fig. 3.6) in a large number of simulations, confirming the model's agreement with analytic considerations, as far as population dynamics are concerned.

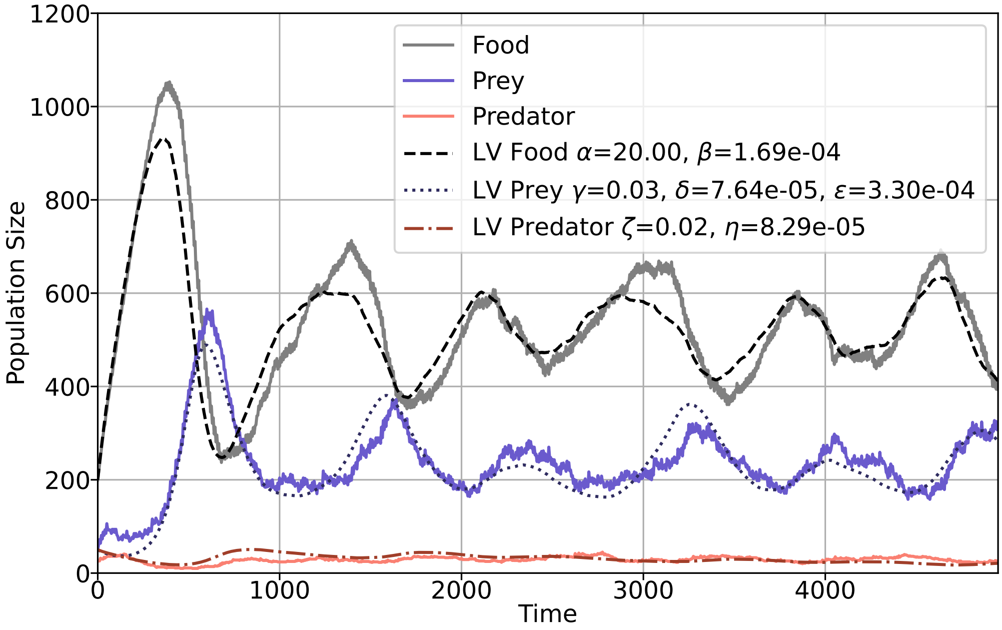


Fig. 3.6 Three-species Lotka-Volterra model prediction as compared with recorded population dynamics. We see generally good agreement, with minor fluctuations likely due to finite system effects and discrete time. Interaction parameters β through η were computationally fitted by minimising variance, best values are quoted on the plot. The associated fitting uncertainties were determined from the inverse Hessian to be relatively low: $\sigma_\beta = 0.78 \cdot 10^{-4}$, $\sigma_\gamma = 9 \cdot 10^{-3}$, $\sigma_\delta = 0.12 \cdot 10^{-5}$, $\sigma_\epsilon = 1.4 \cdot 10^{-4}$, $\sigma_\zeta = 7 \cdot 10^{-3}$, $\sigma_\eta = 0.92 \cdot 10^{-5}$. Parameter α is fixed as it is given by the constant food input.

3.3.2 Speed

The gene space distributions for the two species, predator and prey, are coupled, as can be argued by making the observation that the predators' traits are part of the

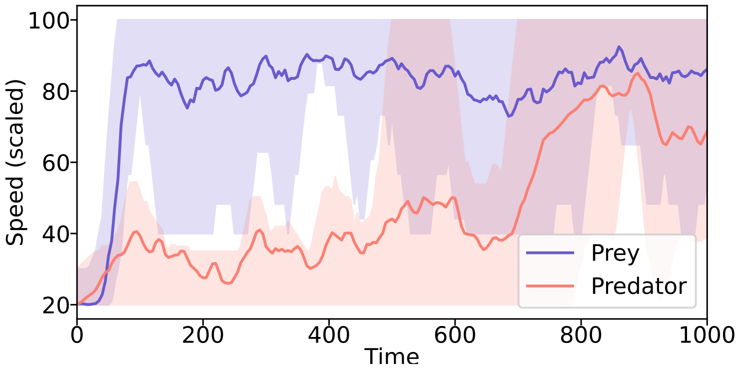


Fig. 3.7 Mean speed gene values (scaled to percentage of maximum) in a coexisting population of predator and prey agents. Shaded area indicates 75th to 25th percentiles as recorded from the data. Moving average (mean) over time was taken to make the plot clearer. Prey evolve quickly to the maximum value, followed by a slower adaptation from predators.

prey's environmental parameters and vice versa. Thus, if one of them changes (by adaptation), the fitness landscape for the other species will morph, its maxima shifted.

If we think of cheetahs hunting gazelles in the savanna, we might wonder why they ended up as two of the fastest animals on the planet. Posing a similar question in the context of our simulation, we can reproduce the same dynamics and conclude that the gazelle must have driven the increase in speed (see Fig. 3.7), with the cheetah forced to follow suit. This can intuitively be understood as there is normally many more prey than predators and therefore they reproduce and mutate more frequently. Analogous to the biological and physical constraints on these animals, there is also an upper limit to the agents' speed (at one cell per timestep) that forms a barrier in the evolutionary arms race. As there is no further possibility of further development along this axis, and there are no other options in this model for the prey to increase its chances of survival (as there would be in nature), a stable gene distribution sets in. Unless the system is disturbed in some manner - eg. a sudden reduce in food input - the location and height of the (global) peak in the fitness landscape (along the speed dimension at least) has been found, the gradient in the direction of lower speeds is negative and therefore a disadvantageous mutation.

In reality this of course never happens so clearly as environments always contain more than two actors and countless numbers of parameters changing in time (eg. temperature, evolving vegetation). Without a stable maximum, the optimisation problem becomes insurmountable by means of evolution. It is only within certain time frames that the gene space distribution of a population can be considered relatively stable.

3.3.3 Colour

As a demonstration of the point made about no stable maximum, we look at a gene with relative meaning only: colour, as defined in Section 2.2. The results (see Fig. 3.8) differ from those for speed, in that they exhibit an oscillation, with the prey leading by about a quarter period (although the oscillation is irregular). Higher population resulting in more mutations taking place helps explain why the prey are faster to adapt.

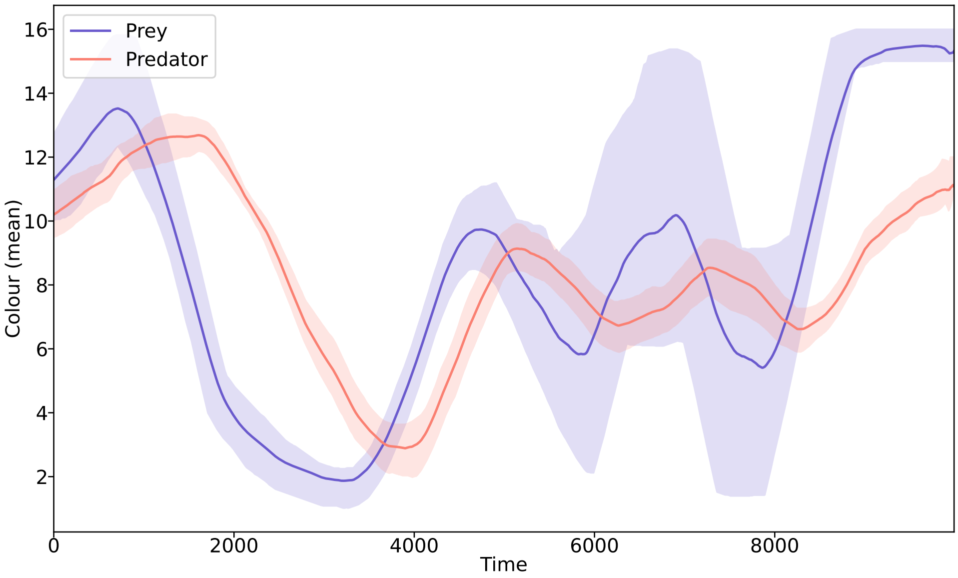


Fig. 3.8 Mean colour gene values in a coexisting predator-prey agent population. Shaded area indicates 75th to 25th percentiles as recorded from the data. Moving average (mean) over time was taken to make the plot clearer. Due to lack of an absolute best-value, the relative distance of the gene's value for the two species drives an oscillation. Prey adapt faster as a consequence of their larger population.

From the point of view of the prey, the fitness landscape (along the colour gene axis) is completely determined by the colour preference genes of the predators. Thus, there is pressure to move away from their population's mean value, which - if there happens to be overlap - can happen in either direction (toward 1 or 16). The opposite is true from the predators' perspective: those with a preference covering most of the prey colour distribution will be better at hunting, providing them with a reproductive advantage. The result is that the chase continues in gene space.

As before, there are examples of similar behaviour observed in nature, particularly in ecosystems of amphibians and reptiles [9]. Certain species of poisonous frogs have developed bright, vibrant colours to discourage predators to avoid them - who in turn have learned not to bother with hunting these species. The same gene space profile can also arise for fish such as guppy [10] that, as a result of runaway sexual selection, are now too vibrant to blend into their environment, making them easy targets that is likely to put less conspicuous individuals at a comparative advantage.

Chapter 4

Conclusion

4.1 Discussion of uncertainties

The main source of uncertainty in this study relates to the simulation itself. While the simulation provides a useful tool for studying population dynamics and adaptation, it is also subject to a number of limitations and assumptions that could impact the accuracy and reliability of the results. For example, the simulation may not fully capture the complexity and variability of real-world ecological systems, and it relies on a simplified model and assumptions that, inevitably, does not fully reflect the complexity of the organisms being studied. As such, it is important to consider the potential impact of these limitations on the conclusions drawn from this study.

Another potential source of uncertainty relates to the numerical measurements and recorded data. While measurement errors are generally not an issue in computer simulation, the inherently discrete nature of the model can introduce unwanted effects and therefore influence the results. We have made every effort to highlight cases where this is a significant factor, as well as minimising other sources of variability that could impact the accuracy of the results eg. by running multiple simulations with the same parameters and averaging the results over runs, as well as quoting relevant uncertainty measures when performing numerical fitting.

Finally, it is worth noting that this project focuses on a specific set of parameters and variables, and as such, the conclusions drawn may not necessarily be applicable in more complex environments. For example, the results may be influenced by the specific combinations of traits that are being selected for.

4.2 Conclusion

In this project, we explored evolution's optimisation problem through our numerical agent-based simulation, extracting population level dynamics and investigating adaptation and coevolution. We also observed and examined reproductive strategies, confirmed hypotheses about environmental pressures in a single-species system and their effect on the genetics of the population over time - including relationships between optimal gene values and food distribution. Finally, we looked at the impact of predation in the context of speed and colour, as examples of coevolution.

The results of our simulation provide valuable insights into the complex mechanism of evolutionary optimisation in nature. Through isolating individual aspects of the process, our findings demonstrate the feasibility of adaptation to meet the criteria enforced by natural selection in the case of simple organisms, and the critical role of mutations and evolutionary pressures. There are countless possible extensions, such as the investigation of speciation using machine learning, expanding on the concept of colour (eg. through camouflage effect), or extensions to the model with other traits (strength, size etc.) that result in new behaviour. A natural route to increasing the accuracy of the model would be to enable more complex organisms, allowing their logic to develop through eg. a simple neural network approach to their brain. Therefore, this project provides a possible starting point for further investigation into computational modelling of evolution.

Acknowledgements

I would like to express my gratitude to my project supervisor for providing us with valuable guidance and support throughout this project. They have introduced me to concepts that have not only influenced my approach to this project but also impacted my way of thinking in general.

I would also like to thank my project partner for their collaboration and contribution to this project.

Finally, I am grateful to my partner for her unwavering support and stimulating discussions about evolution and biology, as her insights have greatly enriched my understanding and thinking throughout this project.

References

[1] Stefan Boettcher and Allon Percus. Nature's way of optimizing. *Artificial Intelligence*, 119(1):275–286, 2000. ID: 271585.

[2] David M. McCandlish. Long-term evolution on complex fitness landscapes when mutation is weak. *Heredity*, 121(5):449–465, -11 2018.

[3] Philip A. Romero and Frances H. Arnold. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology*, 10(12):866–876, -12 2009.

[4] Ingo Berg. Simulating evolution with a computer. [Online]. Available from: https://beltoforion.de/en/simulated_evolution/index.php, 2020. [Accessed 15 March 2023].

[5] A. K. Dewdney. Computer recreations. *Scientific American*, 251(6):14–26, 1984. http://www.jstor.org/stable/24969495.

[6] Erica Chauvet, Joseph E. Paullet, Joseph P. Previte, and Zac Walls. A Lotka-Volterra three-species food chain. *Mathematics Magazine*, 75(4):243–255, -10-01 2002. https://doi.org/10.1080/0025570X.2002.11953139.

[7] NOAA Fisheries. Western atlantic bluefin tuna. [Online]. Available from: https://www.fisheries.noaa.gov/species/western-atlantic-bluefin-tuna, Fri, 03/03/ - 15:57 2023. [Accessed 6 May 2023].

[8] NOAA Fisheries. North atlantic right whale calving season 2023. [Online]. Available from: https://www.fisheries.noaa.gov/national/endangered-species-conservation/north-atlantic-right-whale-calving-season-2023, Wed, 02/08/ - 00:06 2023. [Accessed 6 May 2023].

[9] Kyle Summers and Mark E. Clough. The evolution of coloration and toxicity in the poison frog family (Dendrobatidae). *Proceedings of the National Academy of Sciences*, 98(11):6227–6232, 2001. doi: 10.1073/pnas.101134898; 06.

[10] Jean-Guy J. Godin and Heather E. McDonough. Predator preference for brightly colored males in the guppy: a viability cost for a sexually selected trait. *Behavioral Ecology*, 14(2):194–200, 2003. https://doi.org/10.1093/beheco/14.2.194.