

# A Forward Spatial Simulator on Evolutionary Game Theory

*Aggelos Koropoulos*

Thesis submitted in partial fulfillment of the requirements for the  
*Masters' of Science degree in Computer Science and Engineering*

University of Crete  
School of Sciences and Engineering  
Computer Science Department  
Voutes University Campus, 700 13 Heraklion, Crete, Greece

Thesis Advisor: *Dr. Pavlos Pavlidis*

---

This work has been performed at the University of Crete, School of Sciences and Engineering, Computer Science Department.

The work has been supported by the Foundation for Research and Technology - Hellas (FORTH), Institute of Computer Science (ICS).



UNIVERSITY OF CRETE  
COMPUTER SCIENCE DEPARTMENT

**Title**

Thesis submitted by  
**Aggelos Koropoulos**  
in partial fulfillment of the requirements for the  
Masters' of Science degree in Computer Science

THESIS APPROVAL

Author: \_\_\_\_\_  
Aggelos Koropoulos

Committee approvals: \_\_\_\_\_  
Pavlos Pavlidis  
Researcher C, Thesis Supervisor

\_\_\_\_\_  
Ioannis Tsamardinos  
Professor, Committee Member

\_\_\_\_\_  
Ioannis Tollis  
Professor, Committee Member

Departmental approval: \_\_\_\_\_  
Antonios Argyros  
Professor, Director of Graduate Studies

Heraklion, March 2017



# A Forward Spatial Simulator on Evolutionary Game Theory

## Abstract

Individuals best adapted to their biotic and abiotic environments are more likely to survive and reproduce. Until today, studies that link behavior with the genome evolution are sparse. A framework to link behavior of individuals to the evolution of populations is the evolutionary game theory (EGT). The main idea behind this theory is that individuals with different phenotypes represent different strategies regarding interactions between individuals participating in an evolutionary game. These interactions affect access to food supplies and mates. This received payoff is translated to fitness in population genetics, and determines the reproduction rate of an individual. EGT games are commonly presented as matrix games, which have guided the research on social dilemmas. Usually two strategies are studied: cooperation (synergy) and competition. Cooperation enables higher payoff than individuals could achieve on their own. However, the tendency to antagonize the other players and benefit from their behavior is often appealing to secure resources, especially in the cases when resources are limited.

Simulators are important for studying EGT and its effect on the structure of populations and genome evolution as they make associations between parameters and outcomes possible. To date, most of the algorithms for population genetics studies make a multitude of simplifying assumptions. For example, the widely-used software *ms* allows for population substructure, homologous recombination, but it does not assume any spatial heterogeneity. Thus, the population is assumed to live in an ideal space where all members of a deme are allowed to mate with the same probability with any member of the same deme and without otherwise interacting with each other. Furthermore, the behavior of an individual has not been linked to its fitness value. To tackle this issue, we implemented *FEG* a forward-in-time spatial simulator featuring a predator interaction model under the concept of EGT.

*FEG* is available as an open source software (GPL V3.0), which can be obtained from <https://github.com/aggelosk/game>.



# Χωρικός Προσομοιωτής σε Εξελικτική Θεωρία Παιγνίων

## Περίληψη

Τα άτομα που προσαρμόζονται καλύτερα στο βιοτικό και αβιοτικό τους περιβάλλον είναι πιο πιθανό να επιβιώσουν και να αναπαραχθούν. Μέχρι σήμερα, μελέτες που συνδέουν τη συμπεριφορά με την εξέλιξη του γονιδιώματος σπανίζουν. Η εξελικτική θεωρία παιγνίων (*EGT*) προσφέρει ένα πλαίσιο για τη σύνδεση της συμπεριφοράς των ατόμων με την εξέλιξη των πληθυσμών. Η κύρια ιδέα πίσω από αυτή τη θεωρία είναι ότι τα άτομα με διαφορετικούς φαινότυπους αντιπροσωπεύουν διαφορετικές στρατηγικές όσον αφορά τις αλληλεπιδράσεις μεταξύ ατόμων που συμμετέχουν σε ένα εξελικτικό παιχνίδι. Αυτές οι αλληλεπιδράσεις επηρεάζουν την πρόσβαση σε τρόφιμα και συντρόφους. Αυτή η απολαβή μεταφράζεται σε δαρβινική αρμοστικότητα στην πληθυσμιακή γενετική και καθορίζει τον ρυθμό αναπαραγωγής ενός ατόμου. Συνήθως, στα Εξελικτικά Παιχνίδια (*EG*) μελετώνται δύο στρατηγικές: η συνεργασία και ο ανταγωνισμός. Η συνεργασία επιτρέπει υψηλότερη απολαβή από ό, τι τα άτομα μπορούν να επιτύχουν μόνα τους. Ωστόσο, η τάση να ανταγωνίζονται τους άλλους παίκτες και να επωφελούνται από τη συμπεριφορά τους είναι συχνά ελκυστική για την εξασφάλιση πόρων, ειδικά στις περιπτώσεις όπου οι πόροι είναι περιορισμένοι.

Οι προσομοιωτές είναι σημαντικοί για τη μελέτη της *EGT* και της επίδρασής της στη δομή των πληθυσμών και την εξέλιξη του γονιδιώματος καθώς καθιστούν δυνατή τη σύνδεση των παραμέτρων και των αποτελεσμάτων. Μέχρι σήμερα, οι περισσότεροι από τους αλγόριθμους για μελέτες πληθυσμιακής γενετικής κάνουν ένα πλήθος απλουστευμένων υποθέσεων. Για παράδειγμα, το ευρέως χρησιμοποιούμενο λογισμικό *ms* επιτρέπει την υποδομή του πληθυσμού, τον ομόλογο ανασυνδυασμό, αλλά δεν υποθέτει καμία χωρική ετερογένεια. Έτσι, ο πληθυσμός θεωρείται ότι ζει σε ένα ιδανικό χώρο όπου όλα τα μέλη ενός πληθυσμού επιτρέπεται να μοιράζονται με την ίδια πιθανότητα με οποιοδήποτε μέλος του ίδιου πληθυσμού, χωρίς αλληλεπίδραση μεταξύ τους. Επιπλέον, η συμπεριφορά ενός ατόμου δεν συνδέεται με την αρμοστικότητά του. Για να αντιμετωπιστεί αυτό το ζήτημα, υλοποιήσαμε το *FEG* έναν προσομοιωτή χωρικών δεδομένων που περιλαμβάνει ένα μοντέλο αλληλεπίδρασης θηρευτών υπό την σκοπιά της Εξελικτικής Θεωρίας Παιγνίων. Το *FEG* είναι διαθέσιμο ως λογισμικό ανοικτού κώδικα (*GPL V3.0*) στο <https://github.com/aggelosk/game>.





## Ευχαριστίες

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω όσους συνέβαλλαν στην εργασία αυτή με τον ένα ή με τον άλλο τρόπο. Καταρχάς θέλω να ευχαριστήσω τον δρ. Παυλίδη που με πήρε στο εργαστήριο του, μου έδειξε έναν τομέα με τον οποίο πιθανώς θ' ασχοληθώ στο υπόλοιπο της καριέρας μου και μου έδωσε τη δυνατότητα να ασχοληθώ με μία δική μου ιδέα στηρίζοντάς με όποτε χρειαζόταν όλο αυτό τον καιρό. Θα ήθελα ακόμα να ευχαριστήσω τους γονείς μου για τη συνολική στήριξη που μου προσέφεραν και που μου προσφέρουν ακόμα. Έστερα, όλα τα μέλη του εργαστηρίου, τώρα αλλά και συνολικά στα τελευταία τέσσερα χρόνια για το πολύ όμορφο κλίμα που υπήρχε καθόλη την διάρκεια της παρουσίας μου. Πιο συγκεκριμένα, θα ήθελα να ευχαριστήσω τον Αντώνη τον Κιούκη για την βοήθεια, αλλά πρωτίστως για την παρέα από το πρώτο έτος του προπτυχιακού μου, το Στέφανο Παπαδαντωνάκη για την ανταλλαγή ιδεών συν τη συνύπαρξη στο γραφείο και το Σπύρο Χαβλή ο οποίος παρότι είναι σε διαφορετικό εργαστήριο και δεν είχε καμία απολύτως υποχρέωση να βοηθήσει ήταν πάντα εκεί όταν χρειάστηκε κάτι. Τέλος, θα ήθελα να ευχαριστήσω τη Δήμητρα Τσακίρη για την τεράστια υποστήριξη τα τελευταία 2 χρόνια και για την υπέρμετρη υπομονή και ανοχή στην γκρίνια μου σε περιπτώσεις που μ είχε πάρει από κάτω. Δεν είμαι ιδιαίτερα υπέρ των ευχαριστιών μέσω κειμένου, ούτε καλός σε αυτό, οπότε σίγουρα αδικώ αυτούς που ανέφερα γιατί δεν μπορώ να περιγράψω το μέγεθος της συμβολής τους, όχι μόνο στην εργασία καθεαυτό αλλά γενικότερα. Σίγουρα επίσης, παραβλέπω αυτή τη στιγμή κάποια άτομα που θα ήθελα και όφειλα να αναφέρω σ αυτό το σημείο. Όπως και να χει, προς τους αναφερόμενους, αλλά και σε όσους θα έπρεπε να αναφερθούν, απλά ευχαριστώ.



*στους γονείς μου*



# Contents

<b>Table of Contents</b>	<b>i</b>
<b>List of Tables</b>	<b>iii</b>
<b>List of Figures</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Population Genetics Background . . . . .	1
1.1.1 Selection Detection . . . . .	1
1.1.2 Phenotype . . . . .	3
1.2 Simulations & Simulated Data . . . . .	3
1.2.1 Other Simulators . . . . .	4
1.3 Evolutionary Game Theory . . . . .	5
1.3.1 Strategies . . . . .	6
<b>2 Methods</b>	<b>9</b>
2.1 Evolutionary Game Description . . . . .	9
2.2 Simulator . . . . .	11
2.2.1 Setting the Game . . . . .	11
2.2.2 Life cycle . . . . .	11
2.2.3 Genome representation . . . . .	11
2.2.4 Burn-in . . . . .	12
2.2.5 Strategy Selection . . . . .	13
2.2.6 Predator fitness calculation . . . . .	14
2.2.7 Predator reproduction . . . . .	14
2.2.7.1 Parental selection . . . . .	14
2.2.7.2 Genome Inheritance . . . . .	15
2.2.7.3 Recombination . . . . .	15
2.2.7.4 Mutations . . . . .	16
2.2.8 Events . . . . .	17
2.2.8.1 Population Bottlenecks . . . . .	17
2.2.8.2 Adding and removing preys . . . . .	17
2.2.9 Sampling and Output . . . . .	19

<b>3</b>	<b>Results</b>	<b>21</b>
3.1	Population genetics summary statistics . . . . .	21
3.2	Two Strategy Games . . . . .	22
3.2.1	Synergy - Ignore . . . . .	22
3.2.2	Synergy - Competition . . . . .	24
3.2.3	Ignore - Competition . . . . .	26
3.3	Strategy - Percentage of Aggression level . . . . .	27
3.4	Three Strategy Games . . . . .	28
3.5	Detecting Selection . . . . .	31
<b>4</b>	<b>Discussion - Future Work</b>	<b>37</b>
	<b>Bibliography</b>	<b>39</b>

# List of Tables

1.1	Typical two strategy game using the Synergy and Defect Strategies. $p$ is the reward each player would receive on his own. $s$ is the benefit received from a co-operator. $d$ is the cost a Synergy player pays to attempt to co-operate. . . . .	7
1.2	Typical two strategy game using the Synergy and Competition Strategies. $p$ is the reward each player would receive on his own. $s$ is the gain from mutual synergy. $c$ is the gain from exploiting a player that attempts to co-operate. $\delta$ is a small payoff from mutual competition.	7
2.1	Proposed game. $p$ is the reward each player would receive on his own. $s$ is the gain from mutual synergy. $\epsilon$ is the small gain an individual gets from a synergizing player. $c$ is the gain from exploiting a player that tries to synergize. $\zeta$ is the gain a competitive player gets from a solo player. $\delta$ is a small payoff from mutual competition.	10
3.1	Description of a subset of the summary statistics generated from <i>CoMuStats</i> . . . . .	22
3.2	Two strategy game using the Ignore and Competition Strategies. $p$ is the reward each player would receive on his own. $s$ is the gain from mutual synergy. $c$ is the small gain from exploiting an ignore player. $\delta$ is a small payoff from mutual competition. . . . .	26
3.3	Parameters used for simulating the various scenarios. The Initial column described the initial number of predators selecting each strategy. $R$ symbolizes a random initialization. The three numbers are for Synergy, Ignore and Competition players accordingly. $THR_S$ is the Synergy threshold and $THR_C$ the Competition threshold. Finally, $mut\_rate$ describes the mutation rate and $rec\_rate$ the recombination rate. The first 3 blocks on the table represent the two strategy runs to test for ESS. The runs labeled as Select and Neutral are for selection detection. In the neutral scenario the game was played for 0 generations. For the rest, it was played for 3000 generations. . . . .	36





# List of Figures

2.1	Life Cycle of a Predator . . . . .	12
2.2	Bottleneck demographic scenarios (top panel) may result in similar genealogies to a selective sweep (bottom panel). Both models may produce very short coalescent trees. As we move from the selection site, selective sweeps produce genealogies with long internal branches. Similarly, bottlenecks may produce genealogies with very long internal branches if the ancestral population size is large. . . .	18
2.3	ms output example taken from the Hudson paper . . . . .	19
3.1	Comparison of common summary statistics values between <i>FEG</i> and Hudson's ms using the same parameter values. The distribution of summary statistics values is nearly identical for the two programs.	23
3.2	Evolution of a Synergy - Ignore population. Each row represents a different initialization. From top to bottom, the initial population consists of ( i) co-operators only, (ii) ignore players only, (ii) half of each and (iv) randomly selected. Left figure shows a mutation rate of 0.005 whereas the right figure is for a mutation rate of 0.03. Synergy threshold was set to 0.5 for the half and random initialization. For the Synergy and Ignore ones, left figure showcases a synergy threshold of 0.2 and the right of 0.8. . . . .	25
3.3	Effect of the competition threshold in an Ignore - Competition population with random initialization. Threshold ranges from 0.2 to 0.8. . . . .	27
3.4	Evolution of an Ignore - Competition population. Each row represents a different initialization. From top to bottom, the initial population consists of (i) ignore players only, (ii) competitors only, (ii) half of each and (iv) randomly selected. . . . .	28
3.5	How increasing the percentage of aggression correlated to each strategy affects the, on average, appearance of that strategy in the population. For initial populations that unanimously select each of the strategies. Top row shows Synergy initialization, middle one shows Ignore initialization and the bottom row shows Competition. Left figures represent a mutation rate of 0.005 while right figures represent a mutation rate of 0.03. . . . .	29

3.6	How increasing the percentage of aggression correlated to each strategy affects the appearance of that strategy. initial populations of equal representation of each strategy can be seen in the top figures and for random initialization at the bottom figures. Left figures represent a mutation rate of 0.005 while right figures represent a mutation rate of 0.03. . . . .	30
3.7	Three strategy game with synergy as the initial strategy. The first figure is for balanced ranges of aggression (0.35 synergy threshold and 0.65 competition threshold). The second figure shows the results of having one strategy, synergy in this case, correspond to a higher percentage of aggression. . . . .	31
3.8	Three strategy game with balanced aggression levels (0.35 synergy threshold and 0.65 competition threshold). Each plot showcases a different initialization. In the top plots, the initial population unanimously selects a single strategy. The left plot starts with a population solely comprised of ignore players, whereas the right plot starts with a population of competitors. The bottom two plots are with random (left) and equal (right) number of predators choosing each of the three strategies. . . . .	32
3.9	Comparison of the likelihood score distributions produced by SweeD for a model under neutrality vs other models created using <i>FEG</i> . The plot on the left showcases a model featuring reproduction that is only influenced by the distance between the predators. The plot on the right showcases a model featuring an evolutionary game but no spatial characteristics. For both, the Site Frequency Spectrum of the neutral model was given as an input to SweeD. . . . .	33
3.10	Comparison of common summary statistics values amongst various scenarios simulated using <i>FEG</i> The first column represents a model of neutrality. The second a model where the population is solely influenced by spatial characteristics. The third column represents a scenario where competitors are clearly dominant against co-operators. The final column represents a scenario where a balance between the two strategies is the Evolutionary Stable Strategy. . . . .	35

# Chapter 1

## Introduction

### 1.1 Population Genetics Background

Population genetics deals with genetic differences within and between populations. Studies in this branch of evolutionary biology examine phenomena such as adaptation, speciation, and population structure. There are two main forces which drive the evolutionary process of populations, selection and genetic drift [39, 21, 12].

In the population level, the transmission of genetic material to the next generation is governed by both stochastic and deterministic processes. In finite populations, the change of allelic frequencies due to stochastic sampling over generations can be modeled by the binomial distribution and is called random genetic drift [39, 21, 22]. Due to genetic drift, identical initial populations, in exactly the same environment, may become dramatically different over time.

On the other hand, directional selection, or positive selection is a mode of natural selection in which an extreme phenotype is favored over other phenotypes, causing the allele frequency to shift over time in the direction of that phenotype, deterministically. Under directional selection, the advantageous allele increases as a consequence of differences in survival and reproduction among different phenotypes. In natural populations, both random genetic drift and positive selection operate on allelic frequencies simultaneously. Genetic drift becomes more prominent in small populations, whereas positive selection becomes more important in large populations.

#### 1.1.1 Selection Detection

Some genetic traits make it more likely for an organism to survive and reproduce. This effect is called directional selection [39, 21] in population genetics and is described by defining fitness as an individual's probability of survival and reproduction in a particular environment. Positive selection, also known as Darwinian selection, occurs when an allele is favored by natural selection. The frequency of the favored allele increases in the population and, due to genetic hitchhiking [49], neighboring linked variation diminishes, creating so-called selective sweeps. Such

a process leaves traces in genomes that can be detected in a future time point. Detecting traces of positive selection in genomes is achieved by searching for signatures introduced by selective sweeps, such as regions of reduced variation, a specific shift of the site frequency spectrum, and particular Linkage Disequilibrium (LD) patterns in the region. A variety of approaches can be used for detecting selective sweeps, ranging from simple implementations that compute summary statistics, to more advanced statistical approaches, e.g, Bayesian approaches [69], maximum likelihood-based methods [56], and machine learning methods [55, 62, 70, 66]. When a strongly beneficial mutation occurs and spreads in a population, the frequency of linked neutral (or weakly negatively selected) variants will increase.

The selective sweep model predicts that in recombining chromosomal regions diversity vanishes at the site of selection immediately after the fixation of the beneficial allele. Due to recombination, genetic diversity is predicted to increase as a function of the distance to the selected site (scaled by the selection coefficient and the recombination rate). As a result, the genetic diversity is maintained due to recombination in genomic regions that are in the proximity of a selective sweep: Single Nucleotide Polymorphisms (SNPs) are not generated by novel mutations, but are old mutations that escaped selection because of recombination. This result is also roughly correct in finite populations [33]. Further signatures of the hitchhiking effect include (i) shifts in the Site Frequency Spectrum (SFS) of polymorphisms such as an excess of low- and high-frequency derived alleles [9, 16], and (ii) an elevated level of LD in the early phase of the fixation process of a beneficial mutation [72]. It is important to note that the aforementioned signatures of a selective sweep are predicted when (i) fixation of the beneficial mutation has just been completed; (ii) recombination rate is positive, *i.e.*, the chromosome is recombining; (iii) the population size is approximately constant over time; (iv) the population is isolated; (v) no gene conversion has occurred in the proximity of the beneficial mutation. Despite the relatively strict assumptions of the selective sweep model, several tests have been developed that exploit the properties of the hitchhiking effect to map recent, strong, positive directional selection along recombining chromosomes of several species.

Searching for strong positive selection in the genomes of individuals of a natural population has been the focus of a multitude of studies over the past years [2, 52, 67, 4, 77, 80, 81]. The goals of these studies have been (i) to provide evidence of positive selection, (ii) estimate the strength of selection, and (iii) localize the targets of selection. Thus, these studies aim to provide insights into the genetical mechanisms of adaptation either in wild populations or during domestication. A long-term goal is that the genes that experienced recent and strong positive selection could be identified and the associated functions and phenotypes characterized.

### 1.1.2 Phenotype

Selection acts on the phenotype, the composite of an organism's observable characteristics [40]. Hence, population genetic models assume relatively simple relationships to derive the phenotype from an individual's genotype. When two or more clearly different phenotypes exist in the same population of a species, the trait is called polymorphic.

Phenotypic traits which are determined by many genes of small effect unlike Mendelian traits [41], where the phenotypic state is controlled by just one gene, are called quantitative traits [41, 18, 11]. Quantitative traits can be of essentially continuous variability within a population. They can also be discrete but countable with some traits characterized by only a small possible set of states. Many of the traits affecting the probability that a predator will capture a prey, are quantitative traits. The social behaviour of a predator in regard to other predators of the same population can also be considered, from a certain point of observation, as a quantitative trait. In this work, we choose to study the effect of the aggression of a predator as a means of interaction with other predators under the concept of evolutionary game theory.

## 1.2 Simulations & Simulated Data

A main goal in population genetics is to understand how, the previously described, adaptive and non-adaptive processes, such as selection and random genetic drift, respectively, drive the evolutionary process. Population geneticists attempt to reach this goal by building mathematical models and developing statistical methods to test hypotheses based on the analysis of real data. With the availability of real data, it becomes evident that simple mathematical models are inadequate to model real data. This often leads to erroneous results and conclusions due to the fact that the models are too simplistic. A plausible solution to this issue is the production of simulated data. Simulating data plays a crucial role in modern population genetics studies. Simulations (i) allow researchers to study scenarios that are not tractable mathematically, (ii) to test inference algorithms and (iii) facilitate and inspire researchers to develop analytical models for complex genomic data [79]. Two of the main categories of genetic data simulations are forward-time approaches and backward-time approaches [79]. In forward-time simulators an initial population is constructed at some time point in the past and it evolves forward in time until the present-day. The main advantage of forward-time simulators is flexibility, as it allows the generation of complex models. The main disadvantage is that the whole population needs to be tracked, thus they are expensive in terms of computational resources.

With the introduction of coalescent theory [37, 28, 74], backward in time approaches were implemented [61, 43, 27, 63, 23]. The major advantage of backward simulators over the forward simulators is that only a sample needs to be tracked (backwards in time), which is typically much smaller than the population size, thus

both time and space complexity of the algorithm is greatly reduced. However, the flexibility of backward simulators is limited, thus only relatively simple scenarios can be implemented. Backward (coalescent) simulators have been widely used for the analysis of neutral models [12], including scenarios with population size changes or population subdivision. Due to their computational efficiency, coalescent simulators can be used for obtaining estimates of parameters of interest [30, 12, 14].

### 1.2.1 Other Simulators

Several simulating tools precede the one proposed in this work. Here, an example of several notable ones in each category (Forward-in-time, Coalescent, Spatial) is given.

The widely-used coalescent simulation software *ms* [29] allows for population substructure, homologous recombination, gene conversion, and population size changes. However, like most to date algorithms for population genetics studies, it makes a multitude of simplifying assumptions, as it does not assume any spatial heterogeneity. Thus, the population is assumed to live in an ideal space where all members of a deme are allowed to mate with the same probability with any other member of the deme. Another widely used coalescent simulator is CoalHMM [27], where the coalescent process is modeled as a Hidden Markov Model (HMM) along the alignment. This allows the development of efficient inference algorithms compared to sampling based approaches. It models the ancestry of neighboring nucleotides back in time using a continuous time approach.

A program which simulates a forward in time model is SFS CODE [24]. Through SFS CODE the simulation of several populations where each of them can experience its own demographic history, while following a variety of mutation models. Male and Female sexes are maintained to allow biased sex - ratios to be simulated. An additional forward in time simulator which, by using its own programming language, allows for spatial characteristics is SLIM3 [20]. It is a scriptable simulator [20], that gives a user the ability to construct various models by writing code in its innate programming language, causing the tool, according to the authors, to be less restrictive than tools based on command line parameter input and allows for changes without requiring the user to write code on his own. Despite this flexibility which SLIM3 offers, for the tool to properly function, the user needs to learn a programming language that has no use beyond SLIM3 and end up writing his own code. A software that features both a forward and a backward in time approach, and still allows for spatial heterogeneity is SPLATCHE2 [65]. SPLATCHE2 splits the plane into cells forming a 2-dimensional grid. The first phase of the software, comprises a population which enters the grid and evolves until a user-defined time point. Then, the second phase uses the demographic history simulated in the first phase to generate genetic diversity for a sample of the population. Importantly, the compiled version of SPLATCHE2 is freely available for academic purposes but the source code is not available.

### 1.3 Evolutionary Game Theory

Individuals best adapted to their environments are more likely to survive and reproduce. As long as there is some variation between them which is heritable, there will be an inevitable selection of individuals with the most advantageous characteristics. If the variations are indeed heritable, differential reproductive success leads to a progressive evolution of particular populations of a species, and populations that evolve to be sufficiently different eventually become different species. One of the mathematical tools to study the evolution of species is the evolutionary game theory (EGT), proposed by Maynard Smith and Price [71], which linked Darwinian fitness and species evolution. The main idea is that individuals with different phenotypes represent different strategies regarding interactions between each other. The result of these interactions is a change of the degree of evolutionary adjustment by achieving access to food supplies and mates. This received payoff is what is called fitness in population genetics, and determines the reproduction rate of that individual. Most games are played between two players [71, 50, 31, 8] and only a few researchers have focused on multiplayer games instead [57, 19].

One of the main goals in EGT is to find an equilibrium strategy. An evolutionary equilibrium or, as it is most often called, an evolutionary stable strategy (ESS) [71, 6, 19], can be perceived in two different ways. The first is to find the probability distribution of all pure strategies in a monomorphic population. The second is to find the distribution of individuals adopting pure strategies in a polymorphic population. In this work, the latter is followed. Once the population at question reaches an equilibrium, if it does so, there should be either a dominant strategy that all the individuals have adopted or a non-changing percentage of each strategy. In multiple strategy games, the equilibrium does not need to feature every possible strategy available to the players. One or more strategies may prove to be dominant to the rest. ESS theory can be extended to sexual populations, populations in which genetic or geographic relatedness is important, finite populations, multi-species populations, and populations in which strategies transmitted from one generation to the next are subject to change because of mutational or learning effects.

The effects of encounters between individuals in the population are assumed summarized by a payoff matrix. For  $n = 2$ , for example, the payoff matrix  $A$  is a  $2 \times 2$  table recording the outcomes to the user of any of tactics 1 or 2 resulting from contests with opponents using the same set of tactics. Rows correspond to the individual's choice of tactic and columns to his opponent's choice. By convention, outcomes of contests contribute additively to eventual reproductive success, with all reproductive successes being scaled (equally) as to keep the total population size fixed [26]. If all individuals use a common strategy, the population will be monomorphic, while populations in which strategy diversity exists will be polymorphic.

The players in the majority of the Evolutionary Game Theory field are either cells [6, 31] or molecules [8], as research, especially in the earlier days, revolved

around simpler organisms. As far as more complex organisms are considered, several research has been published on predator - prey models [7, 59]. Previous models of predator-prey co-evolution, involving continuous traits, have been of two types. ESS models [10, 25, 8] have not included the dynamics of evolving traits. They have nonetheless assumed that traits and trait-value distributions will integrate to frequency-dependent fitness surfaces. Despite studying phenotypic quantitative genetic models [39, 68], only a few very simple and specific models were investigated. Investigations have been made in predator - prey models that assume evolution occurs exclusively by the introduction of new mutants with small effects [48, 47]. Population dynamic equilibrium is assumed to be reached between the invasions of each new mutant form.

The games in Game Theory are usually played once [64]. Each player selects a strategy and receives a payoff according to that strategy in respect to the other players choices. In EGT, games are mostly repeated, *i.e.*, a base game is played multiple times by the same players. This, (i) gives the players the opportunity to have a second chance at perhaps attempting a different strategy that could be more profitable after receiving feedback from a first instance of the game and (ii) enables for more complex behaviour as a predator may plan ahead [15].

Instead of having the population live in an idea space, the plays of an EG may reside in an environment that assumes spatial heterogeneity [51]. This, adds spatial characteristics to the model [38, 12]. Thus, not all players will take part in a single global game. There will be several "sub-games" taking place on the landscape, since players can only interact in proximity to each other. In the spatial game, the analogue of the frequencies with which different strategies are adopted at an ESS is the proportion of different strategies present at the spatial equilibrium distribution [35]. If the spatial distribution is continually changing, the average proportion of the different strategies.

### 1.3.1 Strategies

Both in game theory and in EGT, two strategies have been the prime focus of researchers [64, 17, 50, 46, 44]. Co-operation and competition. There are multiple ways to perceive how cooperation and competition works. For example, according to Nowak [50],

*a cooperator is someone who pays a cost,  $c$ , for another individual to receive a benefit,  $b$ . A defector has no cost and does not deal out benefits.*

One can see that the term competitor is nowhere to be found in the above statement. *EGT* occasionally considers as a competitor someone who simply opts to not simply not cooperate, often described by the term defector.

A different, yet equally plausible, way to view co-operation between individuals is that co-operation makes better outcomes possible for all than any could obtain on their own [64, 44]. In predator species for instance, such as wolf packs, one can



Player1/Player2	<i>S</i>	<i>C</i>
<i>S</i>	$p - d + s, p - d + s$	$p - d, p + s$
<i>C</i>	$p + s, p - d$	$p, p$

Table 1.1: Typical two strategy game using the Synergy and Defect Strategies.  $p$  is the reward each player would receive on his own.  $s$  is the benefit received from a co-operator.  $d$  is the cost a Synergy player pays to attempt to co-operate.

easily imagine a scenario where multiple predators co-operating with each other would result in capturing a larger number of preys. This point of view is the one examined in this study. However, the tendency to antagonize the other players and benefit from their behavior is often appealing as a means to securing resources, especially in the cases when they are limited. The competitor in this scenario can either be a defector, or a more aggressive version. One that actively tries to take resources away from other players. The defector by acting individually, does not grant the other player(s) the benefit of the co-operation. The aggressive competitor actively tries to take advantage of other players. This could be by stealing resources already gathered by others or attacking an individual of the same species and population.

Player1/Player2	<i>S</i>	<i>C</i>
<i>S</i>	$p + s, p + s$	$p - c, p + c$
<i>C</i>	$p + c, p - c$	$\delta, \delta$

Table 1.2: Typical two strategy game using the Synergy and Competition Strategies.  $p$  is the reward each player would receive on his own.  $s$  is the gain from mutual synergy.  $c$  is the gain from exploiting a player that attempts to co-operate.  $\delta$  is a small payoff from mutual competition.



# Chapter 2

## Methods

In this thesis, we will study genome evolution as a result of stochastic processes (random genetic drift) and selection acting on the behavior of individuals in an Evolutionary Game (EG) context. Since a mathematical theory that incorporates both stochasticity and selection in an EG is still not developed, we have implemented a simulator to study the problem.

### 2.1 Evolutionary Game Description

In the game that we propose, the players will be predators hunting prey. In spite of having both co-existing in the same environment, the work described here, only focuses on how predators interact with each other and not on their interaction with the prey species. We study how the inter-species interaction between the predators affects the structure of the population. In this simulator, the predators' game will not be a standard repeated one. At each time step, the game is repeated in next generations by the descendants of the previous generations. The genotype influencing the strategy of the parents is transferred to the offsprings. Mutation and recombination may happen, resulting in offsprings with different and/or mixed strategies, respectively. After the game is played for a given number of generations, a sample of individuals will be obtained to study its ancestry and analyze their genome to determine the footprints that such a process leaves on the genome. The preys and the predators will reside in a 2-dimensional continuous space. This, adds spatial characteristics to the model [38, 12] and thus, not all players will take part in a single global game. There will be several "sub-games" taking place on the landscape, since players can only interact in proximity to each other.

Thus far, the majority of the research in the field of Evolutionary Game Theory has been focused on games with only two strategies where individuals either cooperate (S) or compete (C) against each other in the attempt to maximize their fitness [15, 73, 46, 78]. The 2-strategy games have numerous advantages such as simplicity and fast computation times. However, limiting complex (predator) species to a binary decision making system should also limit the results which we

can deduce from simulating complex scenarios. Therefore, in an attempt to tackle this issue, what is proposed in this work is the addition of a third strategy to try make make the game more realistic. Besides cooperating with other individual or competing against them, an individual will have the option to Ignore (I) others and try to survive by himself. In this game, via Synergy (S) players attempt to achieve a greater payoff that individually possible. As stated before, competitors can either be defectors that opt to simply not synergize with others or aggressive competitors that try and take advantage of the other players. Instead of selecting either of the two, in this game, they each become a separate strategy. The Ignore player simply uses the Ignore strategy to avoid interaction with other predators. The Competitor will actively antagonize other predators to try and maximize his payoff.

Predators using the ignore strategy will receive a certain payoff  $p$ . Predators that might adopt the Synergy strategy will receive an increased payoff  $p + s$  from interacting with each other. Then, a slightly declined payoff  $p - \epsilon$  will be awarded to co-operative individuals due to their interaction with Ignore players. The Ignore players will receive a payoff of  $p + \epsilon$  due to the help received. Finally, the minimum payoff a predator with the Synergy strategy will receive a payoff of  $p - c$  from attempting to co-operate with a player that is actively trying to get advantage of the situation, a Competitor. The Competitor will then receive the maximum possible payoff  $p + c$ . When a Competitor attempts to take advantage of an ignore player, it is way less successful that in the previous scenario, as the Ignore player does not bother interacting. A payoff of  $p + \zeta$  is then awarded to the competitor and one of  $p - \zeta$  to the predator using the ignore strategy. Finally, for a Competitor the worst possible outcome is when facing another player with the same strategy. As both players attempt to antagonize each other rather that capture a prey they both receive the minimum payoff possible in this game,  $\delta$ . The above is only true in two player games, but serves as an example of how the different strategy players interact. Nevertheless, the principle remains the same for multiple players interacting. The payoff matrix for a three strategy game can be seen in Table 2.1.

Player1/Player2	<i>S</i>	<i>I</i>	<i>C</i>
<i>S</i>	$p + s, p + s$	$p - \epsilon, p + \epsilon$	$p - c, p + c$
<i>I</i>	$p + \epsilon, p - \epsilon$	$p, p$	$p - \zeta, p + \zeta$
<i>C</i>	$p + c, p - c$	$p + \zeta, p - \zeta$	$\delta, \delta$

Table 2.1: Proposed game.  $p$  is the reward each player would receive on his own.  $s$  is the gain from mutual synergy.  $\epsilon$  is the small gain an individual gets from a synergizing player.  $c$  is the gain from exploiting a player that tries to synergize.  $\zeta$  is the gain a competitive player gets from a solo player.  $\delta$  is a small payoff from mutual competition.

## 2.2 Simulator

### 2.2.1 Setting the Game

A prey and a predator population are initialized on a map. The map is a continuous-space square of certain user-defined dimensions. After creating the map, preys are set in certain positions on the map. The number and the position of preys can be constant during simulation or preys may be erased/added after a certain number of generations. After the map is generated and populated with preys, predators are set on various random positions on the map. The user has the option to choose a starting behavioral strategy for some or all initial predators. The genomic area affecting the strategy is defined, and the genotype of each individual is initialized accordingly. The positions on the genome influencing strategies can be in any place (thus, recombination may be important) rather than necessarily being grouped together.

As a data structure, each predator is attributed by (i) its latitude and longitude on the map, (ii) its genotype, (iii) its phenotype, (iv) its strategy and (v) its fitness. The game in which a predator will participate in, will determine its fitness based on its strategy. Depending on predator's position on the map, it will be grouped with other predators around a prey. Not all players will take part in a single global game. There will be several "sub-games" taking place on the landscape, since players can only interact in proximity to each other while grouping around a prey.

### 2.2.2 Life cycle

The life cycle of each predator in the model can be described as follows: (i) birth, (ii) participation in a game, (iii) fitness evaluation, (iv) reproduction and (v) death. Once born, offspring's position is a function of parents' position. The genotype of the offspring is based on parents' genotype following the basic rules of genetics (mutation and recombination). The genotype defines the strategy/phenotype and the position of the predator determines the game it will participate in. The result of the game will result in a payoff for the participants and the fitness of each individual will depend on this payoff. The parents of the next generation will be chosen based on their fitness value and their position. Generations are discrete and after reproduction parents are replaced by their offsprings.

### 2.2.3 Genome representation

The genotype of each individual comprises  $l$  nucleotides from the  $A, C, G, T$  alphabet. Thus, in a population of  $N$  individuals it is possible that at maximum four nucleotide states are present at each genomic location. In a population genetics context, however, in which a single species is studied, we assume that at maximum one mutation may occur at each genomic location. This assumption is called 'the infinite site model' and is valid when the genome under study is long

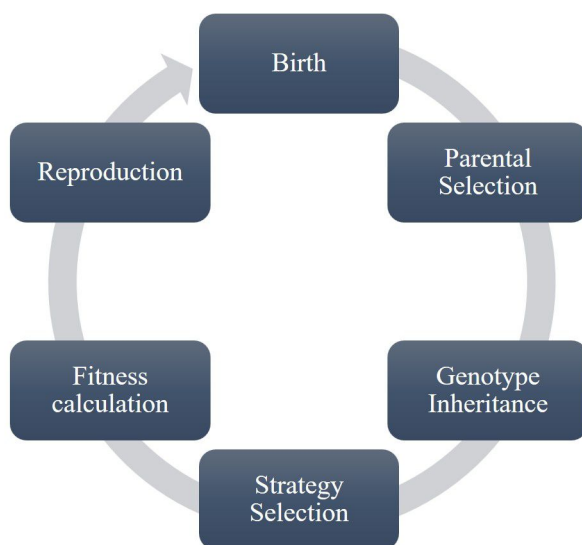


Figure 2.1: Life Cycle of a Predator

and the mutation rate small. In most biological species, the mutation rate is about  $10^{-8} - 10^{-7}$  mutations per base pair, per generation and the genome is hundreds of thousands to several millions of base pairs (3,000 millions base pairs for humans).

In this system, we do not encode all four states, but we use '0' for the ancestral state and '1' for the derived state. Indeed, most population genetics tools [29, 34, 56, 3] instead of using the nucleotide bases A, C, G, T, use a binary system. Furthermore, in most population genetics studies, only the polymorphic sites need to be included in analyses. Neglecting the monomorphic sites and encoding the DNA base pairs as binary ancestral/derived states reduces the information that needs to be stored and features a way to compress the genotype as a way to vastly improve the memory footprint of the tool. Instead of storing each base as a character with a '0' or '1' value, a whole area of the genotype can be encoded as an unsigned integer of a certain size (by using the *stdint* library [13]). By using a vector of such unsigned numbers, whole genome sequences or large areas of a chromosome can be efficiently stored. Modifications on the genotype, (See Section 2.2.7.2), can be applied by using binary operations which considerably speed up the process.

#### 2.2.4 Burn-in

Results of forward-in-time simulators depend on the initial state of the population. To alleviate this issue, a burn-in phase precedes the actual process. During burn-in, predators reproduce neutrally: instead of determining their fitness by participating in a game, they randomly reproduce for a certain number of generations in which all predators are equally fit. During burn-in, mutation and recombination events may take place and the evolutionary process is solely influenced by genetic drift.

After the burn-in phase is complete, the dependency of the simulation results on the initialization of the population has been diminished.

### 2.2.5 Strategy Selection

Each of the predators in the first generation (generation 0) opts to act based on its initial strategy. A strategy can be either randomly selected or specifically appointed to any predator. This initialization process helps users evaluate various scenarios as it impacts the evolutionary process of that population of predators. There is a 6-step process to determine the strategy of each predator at each generation after the first. As mentioned before, every new predator is the offspring of two predators from the previous generation.

Thus, (i) the first step is to determine the genotype of the offspring based on the genotypes of the parents. The child can inherit an exact copy of the genotype of each of the parents or, in cases of recombination, the child inherits from both parents, with each of them giving it a segment of its genotype. In terms of how the code works, an array of unsigned integers representing the genotype is passed from one generation to the next. Each unsigned integer can be translated into a binary number of length equal to the number of bits describing the unsigned number. As stated before, the 0's in the array are positions in the genome where no mutation has taken place whereas the 1's indicate a position which has had a mutation. After the genotype of the newborn child has been set, a mutation may modify a site's state (see Section 2.2.7.4).

Past the completion of steps (i) and (ii), the genotype of the child is known. Now, based on the known genotype (iii) we find the positions of the genotype that affect the strategy. From these positions (iv) the phenotype can be determined. This is done by either estimating the aggression of a predator as a means of interaction with other predators or by a probabilistic model based on different areas of the genome. For the first method, we determine the phenotype by simply counting the elements in the genotype array with a value of 1 and dividing by the length of the array. The percentage number  $P$  is then (v) fed to a Normal distribution with it as a mean value and a user defined variance  $var$ . An integer  $P'$  drawn for this distribution determines the percentage of aggression of the predator at question, as displayed in equation 2.1. Finally, (vi) we categorize each predator to a certain strategy depending on its percentage of aggression  $P'$ . If  $P'$  is below a certain threshold  $THR_S$  we consider the predator to be of the lowest possible aggression and thus the strategy it goes with is Synergy. If  $P'$  is greater than  $THR_S$  but below the threshold  $THR_C$ , the strategy of the predator is Ignore and if  $P'$  exceeds threshold  $THR_C$  then we deem the predator as very aggressive and therefore competitive to other predators.

$$P' = Gaussian(P, var) \quad (2.1)$$

The probabilistic model states that (v) the areas of the genome affecting the

strategy are divided into 3, user-defined sized, parts. Each part of a strategy-affecting area is assigned to a different strategy. The left part to  $S$ , the middle to  $I$  and the right to  $C$ . We then proceed to count the number of mutations, aka 1's in each part. These 3 numbers, each divided by their sum determine the probability of each strategy being played. (vi) One of the strategies is then randomly selected based on the above values.

Both models of strategy selection are based on user-defined parameters. Hence, aggression thresholds of the first method, and the areas of influence on the probabilistic model can be set in a way to make it impossible for one strategy to be selected making it possible to simulate a two strategy game.

### 2.2.6 Predator fitness calculation

After the strategy of each predator becomes known, around each prey a game begins between predators in close proximity. Meaning that for every prey in the game the predators that exist inside a certain range are found. These predators are the ones that compete for access to this specific prey. Thus, they are the participants in the evolutionary game around this prey. The fitness of each predator is determined based on the success they have in the game (payoff). Depending on the strategies of all the players that participated in the game, a payoff ( $P_s$ ,  $P_i$  and  $P_c$ ) is granted for each of the three strategies. Then each player receives the payoff of the selected strategy divided by the number of players that selected that strategy since they equally split it. As mentioned before, in *FE*G the predators' game will not be a standard repeated one because at each time step the game is repeated in next generations by the descendants of the previous generations.

### 2.2.7 Predator reproduction

Since the fitness of each predator in the current generation is set, the process of creating the next generation can start. To improve the time efficiency and the memory footprint of the code, the new generation (next) will overwrite the previous one, aka the one before the current. As the population size remains constant following the Wright Fisher model [29], with the exception perhaps of bottleneck events, described in Section 2.2.8.1, the number of predators in the new generation is already known.

#### 2.2.7.1 Parental selection

For each of the newly born predators, the first step is to choose two predators from the current generation as parents. The common approach is to separately select two individual predators from the population. Individuals with a higher fitness have a higher chance to be chosen. Still, with this approach, any predator with a non-zero fitness has at least a slight opportunity to be paired with any other predator from the same generation regardless of any other criteria. The approach followed here, differs from the one described above in two important ways. (i)



Since there is a spatial parameter in this simulation, predators closer to each other should have a higher chance to reproduce with each other and, on the contrary, if the distance between two predators is greater than a certain threshold the two predators should be unable to reproduce with each other. The threshold above which the reproduction of the pair is prohibited is equal to the range that is used for predators to find prey. It was considered logical and fair that the distance a predator is able to travel to track prey should be the same as when attempting to reproduce. Thus, since a few pairs already have been eliminated as possible parents, instead of selecting two parents individually each pair of predators has its own probability of being selected. The total fitness of the pair is the product of the both predators fitness divided by their distance. After the fitness of each pair is calculated, one of the pairs with non-zero fitness is chosen.

$$PairFitness = (Fit_1 * Fit_2) / (distance + 1) \quad (2.2)$$

### 2.2.7.2 Genome Inheritance

For each child predator created in the new generation, after choosing its two parents, what these parents pass on to the child needs to be determined. For terms of simplicity, in addition to time and memory efficiency, the predator genes are considered to be haploids instead of diploids. A haplotype is a group of genes within an organism that was inherited together from a single parent. A haplotype can describe a pair of genes inherited together from one parent on one chromosome, or it can describe all of the genes on a chromosome that were inherited together from a single parent. Thus, in the majority of cases, the child predator inherits his genotype from one of the two parents.

### 2.2.7.3 Recombination

*FEG* also accounts for recombination events during the birth of a predator. If recombination occurs, the child has inherited both parents. During this biological process, a child gets different regions of its chromosome from each parent [17, 21]. A cut happens somewhere along the genotype, at a randomly selected position. We call that position cutting point. The part of the genotype left of the cutting point comes from one parent and the right part from the other. Without loss of generality, let's assume that the first parent passes on the left part of the genotype and the second parent the other. In *FEG*, since each area of the genotype, consisting of numerous nucleotide bases, is described in the form of an unsigned integer, the general area of the cutting point is first determined. For the area at question, we choose one of the bits describing the nucleotide bases, and name it cutting base. The recombination happens at that exact place. The number is then divided in two pieces. The most significant bits (*MS*), which are the ones left of the cutting base and the rest, the least significant bits (*LS*), including the cutting base. The *MS* bits are inherited from the first parent. This is achieved via using the *AND*

binary operation with a number functioning as a mask that has set bits only in the *MS* positions. A similar process is followed for the *LS* bits, inherited from the second parent.

#### 2.2.7.4 Mutations

Besides selection, genetic drift and recombination an additional phenomenon which can be observed in populations are the existence of mutations [21]. A mutation is ultimately essential for adaptive evolution in all populations as it causes an altering of the construction of a gene produced by the modification of single base units in DNA. In *FEG* mutations may take place right after a predators genome has been determined. The user-defined parameter *mut\_rate* determines the probability that a mutation occurs on the predator at focus. Different methods [42] might opt to have a parameter determining the probability that a mutation occurs at each site of the genotype of an individual separately following Kimura's method [36]. On the contrary, in this work the probability of a mutation describes an individual's genotype as a whole. When a mutation occurs, one of the unsigned integers describing the genome is selected at random. Once selected, an exclusive OR operation (*XOR*) is performed with a mask with only one bit set, the one where the mutation occurs. The result of the *XOR* operation is the genome of the child.

---

#### Algorithm 1 Forward in Time Process for a Single Generation

---

**Data:** predators in current generation, preys, areas of the genome affecting the strategy, etc

**Result:** predators in the next generation

```

while  $i < prey\_num$  do
  find_predators_in_range(preys[i]);
  if found then
    | grant_payoff();
  end
   $i++$ ; # move to the next prey
end
create_fitness_map();
while  $i < new\_predator\_num$  do
  choose_parents(); # selects the parents based on the fitness of the pair
  set_position(); # based on the parents positions
  set_genome(); # either copies the genome from 1 parent or recombination
  happens. Also applies any mutations that occur.
end

```

---

### 2.2.8 Events

During the forward in time process there are several events that might occur besides the ones described in the previous sections of this chapter. Each event may happen at specific, user-defined, point in time.

#### 2.2.8.1 Population Bottlenecks

Demography poses severe challenges on the selection detection process due to the fact that it may generate SNP patterns that resemble the signatures of genetic hitchhiking. In recombining chromosomes, selective sweep detection becomes feasible mainly due to two factors: (i) the fixation of the beneficial mutation, and (ii) the fact that coalescent events occur at a higher rate in the presence of a sweep than they do in its absence. Besides recombination and the existence of positive selection, there are other means to produce similar demographic signatures, such as population bottlenecks [60]. Assume a bottleneck event that is characterized by three phases: (i) a recent phase of large effective population size, (ii) a second phase, prior to the first one, of small population size (the bottleneck phase), and (iii) an ancestral period of large population size. During the bottleneck phase the effective population, which is the size of an ideal population that would have the same level of genetic variance as the real population size [32], is reduced.

It is due to the decrease of the effective population size in the bottleneck phase that a high rate of coalescent events occur in a relatively short period of time. Furthermore, lineages can escape the bottleneck, passing to the ancestral phase of large effective population size, and therefore requiring more time to coalesce. In a recombining chromosome, genomic regions that are characterized by short coalescent trees due to massive coalescent events may alternate with genomic regions with lineages that have escaped the bottleneck phase. Such alternations can generate SNP patterns that are highly similar to those generated by a selective sweep, yielding the detection process very challenging, if not infeasible [54].

Simpler bottleneck events, featuring a single change in population size also exist. *FEG* implements bottleneck events in this simpler form. At a certain generation, an expansion or a reduction in population size may occur. This, means that a generation with a population of  $X$  predators, will give birth to the next generation  $X'$  predators, where  $X'$  is not equal to  $X$ . Every future generation from that point will have a population size of  $X'$  unless an additional bottleneck event takes place. By allowing for multiple bottleneck events, complex bottleneck scenarios, such as the one described at the beginning of this section can be modeled.

#### 2.2.8.2 Adding and removing preys

An additional way to impact the structure of the population of predators is to change the population of preys. In every generation, the number of games played is equal to the number of preys. To participate in one of these games, each predator needs to be around at least one prey in close proximity. A change in the preys,

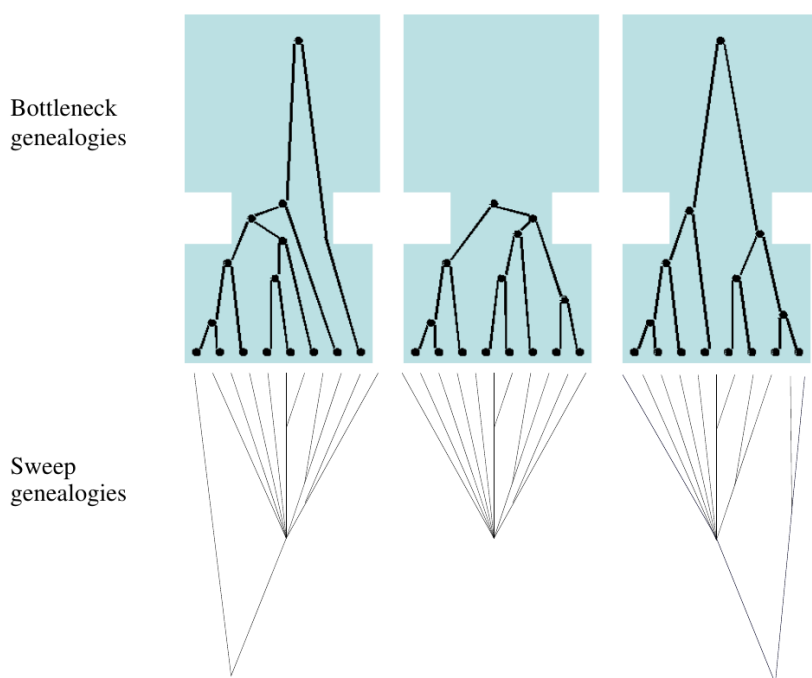


Figure 2.2: Bottleneck demographic scenarios (top panel) may result in similar genealogies to a selective sweep (bottom panel). Both models may produce very short coalescent trees. As we move from the selection site, selective sweeps produce genealogies with long internal branches. Similarly, bottlenecks may produce genealogies with very long internal branches if the ancestral population size is large.

would have an immediate impact on the predators close by. Therefore, to study scenarios under this category, *FEG* supports prey related events. These events can be either the addition of a new prey or the removal of an existing one. The position of the prey to be added/removed can be either random or specified. If specified, the prey is added to a specific position or a specific prey is removed. If not, either a prey appears at a random location in the map, or one of the existing is randomly selected to disappear. Adding a prey to a position causes a major shift to the predators around it. Now, there is one more game to be played and predators participating in it. This will have four effects. (i) The predators in the new game will receive a certain payoff with the strategy each one uses having a different outcome, more or less favourable, that if they participated in another game. (ii) It is possible that predators that would be unable to interact with each other due to their positions may both coalesce in this new prey creating new population dynamics. (iii) The predators participating in this game would, most likely, have participated in a different one. Thus, the result of at least one other game is influenced, changing the fitness of predators in this game. (iv) If we consider predators clustering

around each prey, by the passage of each generation several sub-populations may be formed. Theoretically, in each of these sub-population a different strategy could be beneficial to its members. The addition of a new prey might therefore, create a new sub-population. Removing an existing prey would cause similar ripples in the structure of the population as the addition of one. Furthermore, the removal of an existing prey might, in an extreme but plausible scenario, cause the predators of this specific sub-population to suddenly become stranded from their primary resource leading them to extinction impacting the total population the same way a single generation bottleneck would. The effective population of the first generation post the removal event, is reduced by the number of predators unable to participate in any game. Hence, each of the remaining predators has a higher chance of leaving an offspring since the gene pool is reduced.

### 2.2.9 Sampling and Output

During the passage of generations, both during the burn-in phase and the regular one, the structure of the population changes. Bottleneck and prey events further advocate this change. To make it possible to track these changes, *FEG* allows for sampling events to take place. Such an event takes a random sample out of the current generation of the population and prints the genotype of the selected predators. Also, it gives as output the percentile strategies from that instance of the population. The output file follows the format used by *ms* [29], to be consistent and comparable with other simulation tools. An example of the output format can be seen in Figure 2.3.

```
ms 4 3 -t 5.0
1779988551

//
segsites: 5
positions: 0.0227 0.5520 0.6190 0.9200 0.9459
10001
00010
00000
01100

//
segsites: 4
positions: 0.6760 0.7866 0.9056 0.9606
0101
1000
0101
0110
```

Figure 2.3: *ms* output example taken from the Hudson paper



## Chapter 3

# Results

*FEG* is a spatial simulator featuring an evolutionary game theory model of reproduction for a predator species. *FEG* aims at a complex model of evolution that includes both the ability to simulate quantitative phenotypic traits and the effect of population structure on genomic diversity, while simultaneously featuring a method of selection based on the concept of evolutionary game theory. Various models of evolution were simulated using *FEG*. One of the main goals in EGT is to find an equilibrium strategy, *i.e.*, an evolutionary stable strategy on a population level, translated as the percentage of individuals in a population adopting each strategy. Since the strategy is determined by random processes affecting the genotype, such as mutation and recombination not all game repetitions result in the same outcome.

### 3.1 Population genetics summary statistics

The raw data, generated by *FEG* in *ms* output format are summarized in a vector of summary statistics,  $S$ , calculated from each polymorphic dataset.  $S$  can be used for demography inference and for selection detection since summary statistics in  $S$  are affected by the demographic changes the population has experienced and by the action of positive selection. We used the software *CoMuStats* [53] to calculate a multitude of summary statistics from the *FEG* simulations, such as Tajima's  $D$  [75], Wall's  $B$  and  $Q$  statistics [76], the site frequency spectrum [21], and others. For example, Tajima's  $D$  takes negative values in an expansion scenario as well as in regions where positive selection has acted. In contrast, it takes both positive and negative values, spread throughout the genome when a demographic bottleneck has occurred. Summary statistics are calculated by the CoMuStats software, which is a part of *CoMuS* (Coalescent of Multiple Species) [53].

Currently, the most widely-used software for generating simulated polymorphic data is Hudson's *ms* [29]. Models of neutrality, as well as models of selection, can be simulated via *ms*. *FEG* also features both models. The interaction between predators models a special kind of selection, where the most fit individuals are

Summary Statistic	Definition
$\theta_W$	Watterson's estimator of $\theta$ using the number of segregating sites and the sample size.
Tajima's D	computed as the difference between two measures of genetic diversity: the mean number of pairwise differences and the number of segregating sites, each scaled so that they are expected to be the same in a neutrally evolving population of constant size.
B and Q	the number of pairs of adjacent segregating sites that are congruent
SFS	the number of segregating sites where the derived allele occurs $i$ times out of $n$ samples

Table 3.1: Description of a subset of the summary statistics generated from *Co-MuStats*

those who receive the greatest payoff from the game. Since, the behavior of an individual is based on its genome, the behavior is transferred to the next generation, affected by the recombination and mutation processes. A burn-in phase, influenced by random genetic drift, precedes the game to guarantee independence of the initialization conditions. By using the same parameter values we run both software and compared the summary statistics describing the results. In Figure 3.1 it can be seen that indeed *FEG* is capable of producing a neutral model with results almost identical to that of Hudson's ms.

## 3.2 Two Strategy Games

The majority of evolutionary games involve a two strategy model. Hence, before including a third strategy to the model, scenarios with only two strategies were simulated. All three pairwise combinations (Synergy vs Ignore, Ignore vs Competition and Synergy vs Competition) were examined. We examined the role of the following parameters for each simulated scenario: (i) the mutation rate per individual genome, since the change of strategies emerges from mutations occurring in the population, (ii) the initial strategies of predators and (iii) the threshold for which a predator shifts from one strategy to another. For each set of parameters, we run a total of 100 simulations. One of the questions those simulations aim to answer is what is the ESS in each scenario.

### 3.2.1 Synergy - Ignore

Via this set of simulations, the effect of the co-operator vs Ignore player model on the predator population was tested. Thus, the competitor threshold  $THR_C$  was set to  $> 1$  for all simulations making it impossible for a predator to play the game



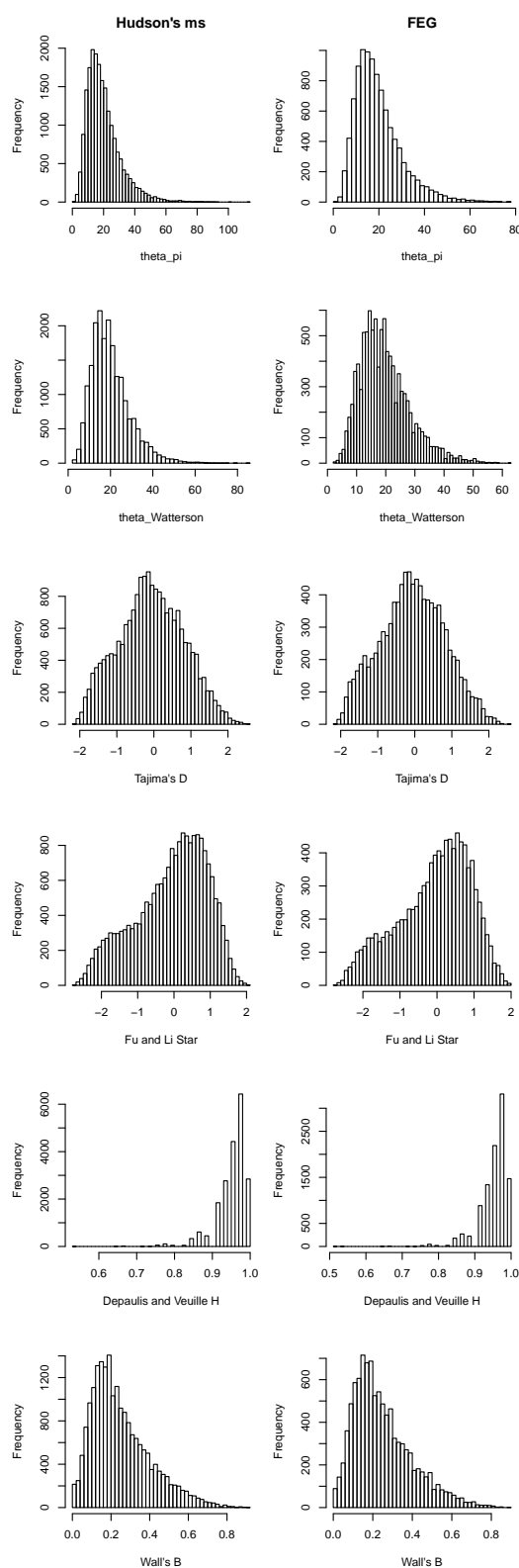


Figure 3.1: Comparison of common summary statistics values between *FEG* and Hudson's *ms* using the same parameter values. The distribution of summary statistics values is nearly identical for the two programs.

using this strategy. It would require more than a 100% of '1's in the genotype. The synergy threshold  $THR_S$  ranged from 0.2 to 0.8 to make it increasingly tougher for a predator to deviate from the Synergy strategy. Four different ways to initialize the population were examined. (i) First, all the predators were initialized as Synergy players, to test whether aggression would emerge in the population. (ii) Second, the opposite scenario was examined, initializing all predators as Ignore players, to test whether a less aggressive strategy would emerge. The third scenario featured (iii) simulations that started with exactly half the population as Synergy players and the other half as Ignore. The goal is to study whether the initial Synergy/Ignore (S/I) state will continue exist in the population in an equilibrium or, if not, towards which direction the population will shift. Finally, (iv) a scenario with completely random initialization was implemented. In Figure 3.2, the scenario that is more suitable for co-operators, with a mutation rate of 0.05 and a  $THR_S$  of 0.8 is shown when the population is initialized with a single strategy. The opposite scenario, where ignore players are most likely to persist, with a mutation rate of 3 and a  $THR_S$  of 0.2 is also examined. For random and balanced initialization both  $THR_C$  and  $THR_S$  are 0.5.

When the initial population comprises only the Synergy strategy, on average, the 81.84% of the population ends up selecting the Synergy strategy. Predators with the Ignore strategy do emerge in the population (18.16% on average) (Figure 3.2). In the opposite scenario, where the initial population purely consists of Ignore players, the Ignore strategy is only represented by more than 50% of the population, with a 69.37% representation, when the Synergy threshold is up to 0.35. For all other scenarios, given adequate time, synergy again appears to be the dominant strategy with 65.67% of the populations selecting it. When a random initialization is applied, synergy players are the ones that appear to be more fit and comprise 65.62% of the population. The same outcome can be seen for a balanced initialization with synergy players being the 83.86%.

### 3.2.2 Synergy - Competition

In the Synergy - Competition model, the Synergy threshold  $THR_S$  and the competition threshold  $THR_C$  are always equal (thus, Ignore strategy is not possible). Competition is favored in an environment with all Synergy players, and Synergy players cannot easily emerge in an all competitive environment. For this scenario, we started from an initial population of Synergy players with 0% aggression and tested whether competition would emerge. Since too many mutations would be required to shift from the all-0 genotype to a genotype associated with competitive behavior, the  $THR_C$  in this scenario was set to 0.05 (i.e., for a region of 32 bits, 2 mutations are adequate to obtain competitive behavior). Two different parameters were tested. First, the mutual competition parameter which is the penalty for competitors interacting with other competitors was set to either (i) 0.02 to create a competitor-friendly environment or to (ii) 0.1 to penalize the existence of multiple competitors in the same environment. For the first scenario

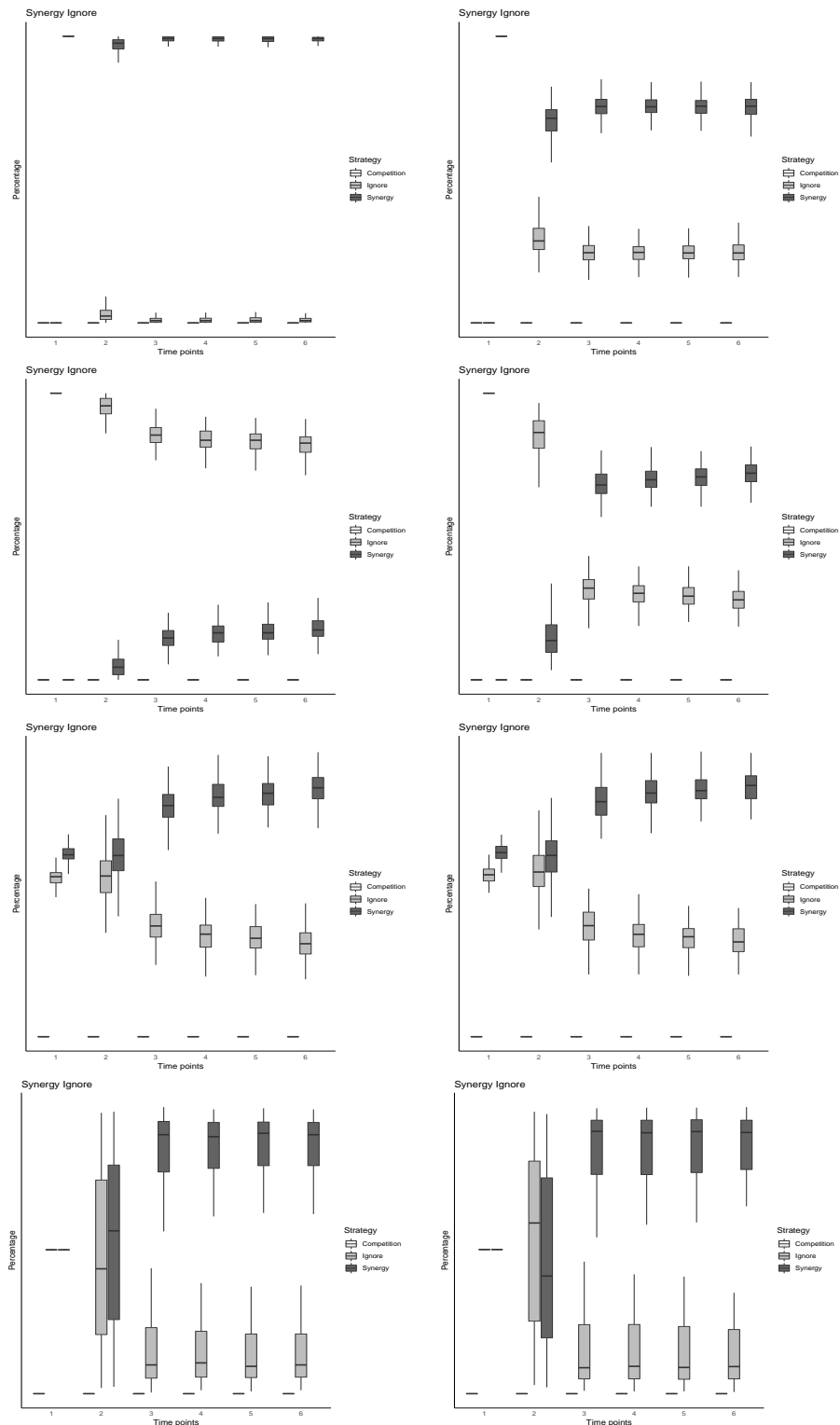


Figure 3.2: Evolution of a Synergy - Ignore population. Each row represents a different initialization. From top to bottom, the initial population consists of (i) co-operators only, (ii) ignore players only, (ii) half of each and (iv) randomly selected. Left figure shows a mutation rate of 0.005 whereas the right figure is for a mutation rate of 0.03. Synergy threshold was set to 0.5 for the half and random initialization. For the Synergy and Ignore ones, left figure showcases a synergy threshold of 0.2 and the right of 0.8.

competitors were, as expected completely dominant, ending up comprising the whole population in 99.8% of the cases. For the 0.02% of the cases not enough mutations were accumulated for a competitor to emerge, thus resulting in a 100% Synergy player ratio. For the second scenario, again with the same exception of an all-synergy population due to lack of mutations, competitors did emerge in the population. When competitor exceed a certain percentage of, on average, 52.19%, the penalty for mutual competition exceeds the gain from exploiting synergy players. Therefore, an equilibrium state is reached where Synergy and Competition players co-exist, suggesting a scenario of balancing selection driven by frequency dependent selection, as seen in Figure 3.10.

### 3.2.3 Ignore - Competition

Competitors are predators that attempt to take advantage of other players. In contrast, Ignore players, simply, refuse to co-operate. Either of them is the counterpart of Synergy in *EGT* studies. To our knowledge, a game studying these two strategies has not been studied yet. Both of the Ignore and Competition strategies can utilize the Synergy players to increase their own fitness. However, in a game involving only Competition and Ignore players, the result is unknown.  $THR_S$  in this scenario is set as  $< 0$ , to prevent a predator for adopting the Synergy strategy. Again, the competition threshold  $THR_C$  ranged from 0.2 to 0.8 to make it increasingly tougher for a predator to adopt the Competition strategy by random mutations. Starting with a population of all Ignore predators, the question is whether a more aggressive strategy will emerge in the population. An initial population of only competitors tests the opposite direction: whether predators with Ignore traits can emerge and survive on a fiercely competitive environment. Balanced and random initialization scenarios are also examined.

Player1/Player2	$S$	$C$
$S$	$p, p$	$p - c, p + c$
$C$	$p + c, p - c$	$\delta, \delta$

Table 3.2: Two strategy game using the Ignore and Competition Strategies.  $p$  is the reward each player would receive on his own.  $s$  is the gain from mutual synergy.  $c$  is the small gain from exploiting an ignore player.  $\delta$  is a small payoff from mutual competition.

In the scenario where the initial population is competitive, ignore players emerge but only as, on average, the 12.24% of the population, regardless of the mutation rate and the  $THR_C$ , as seen in Figure 3.4. For the opposite scenario, where the initial population solely consists of players with the ignore strategy, they remain the 90.40% of the population, with only a few competitors emerging, suggesting that the result depends on the initial conditions of the population.

In a scenario with random initialization, we can see that both strategies exist in population, but competitors do perform worse and decrease to comprise 40.99%

percentage of the population with the passing of generations. By increasing the  $THR_C$ , the percentage of ignore players also increases, as seen in Figure 3.3. For a balanced initialization, again Ignore players perform better reaching 58.80% of the population.

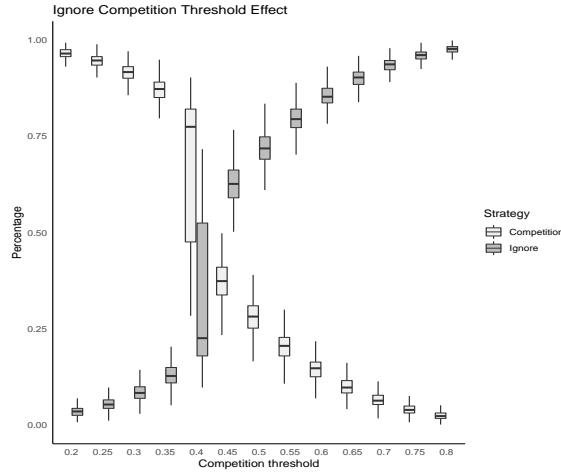


Figure 3.3: Effect of the competition threshold in an Ignore - Competition population with random initialization. Threshold ranges from 0.2 to 0.8.

### 3.3 Strategy - Percentage of Aggression level

The strategy/phenotype of a predator is derived from its genotype. The percentage of mutations ('1s') in the genotype areas that affect the phenotype determines the selected strategy. Thus, two factors play a major role on the population evolution: (i) the mutation rate; the higher the mutation rate, the more rapid strategy shifts and (ii) the thresholds  $THR_S$  and  $THR_C$  which determine the 'genotypic space' available for each strategy. These two factors can be considered as neutral factors because they are not related to the benefit (payoff) associated with each strategy.

In Figure 3.5 we can see that, for the scenario in which all predators are initialized as Synergy or Ignore players, mutation rate does have an impact. Similarly, in a scenario of either random initialization or equal predator distribution of the three strategies across the population, the mutation rate again does not seem to impact the process as seen Figure 3.6.

In those scenarios where the initial strategy was Competition, the mutation rate does affect the results. For a low mutation ratio, the 63.51% of the population remain as competitors. For a large mutation rate (3% chance per individual for a mutation to take place), gradually setting a higher  $THR_S$  causes the population to shift from competitors to synergy players.

With Synergy as an initial strategy for the whole population, the greater the genotypic percentage the synergy strategy occupies, the more difficult it becomes to

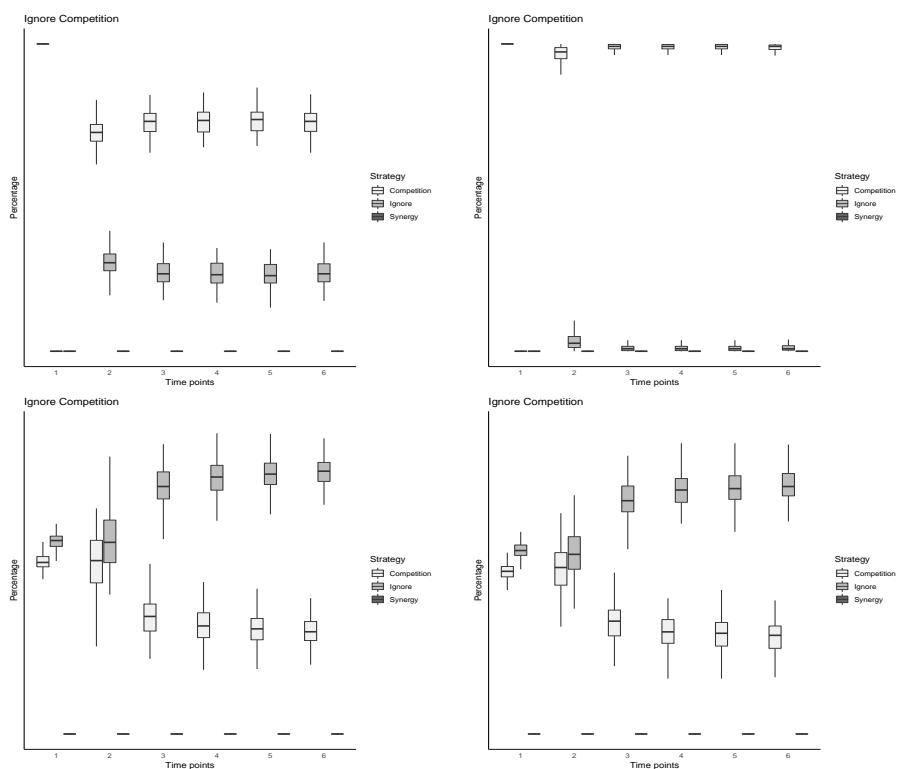


Figure 3.4: Evolution of an Ignore - Competition population. Each row represents a different initialization. From top to bottom, the initial population consists of (i) ignore players only, (ii) competitors only, (ii) half of each and (iv) randomly selected.

shift from the Synergy into more aggressive strategies independent of the mutation rate. The Ignore strategy is the one with the most peculiar behaviour. Above a certain range it appears to consist, on average, a smaller part of the population.

### 3.4 Three Strategy Games

After studying all the possible pairs of strategies, the effect of all three strategies co-existing in the population is examined. Again, the effect of (i) the mutation rate `mut_rate`, (ii) the initial strategies of predators and (iii) the shift thresholds from one strategy to another were examined. For each set of parameters, we ran 100 simulations.

The percentage of aggression usually determines the strategy of the majority of the population. Given a significantly higher percentage than the other two, each of the three strategies may become the one selected by most of the predators. In the scenarios, however, where the range of each strategy is balanced, or close to balanced, the strategy that seems to gain an advantage is the Ignore

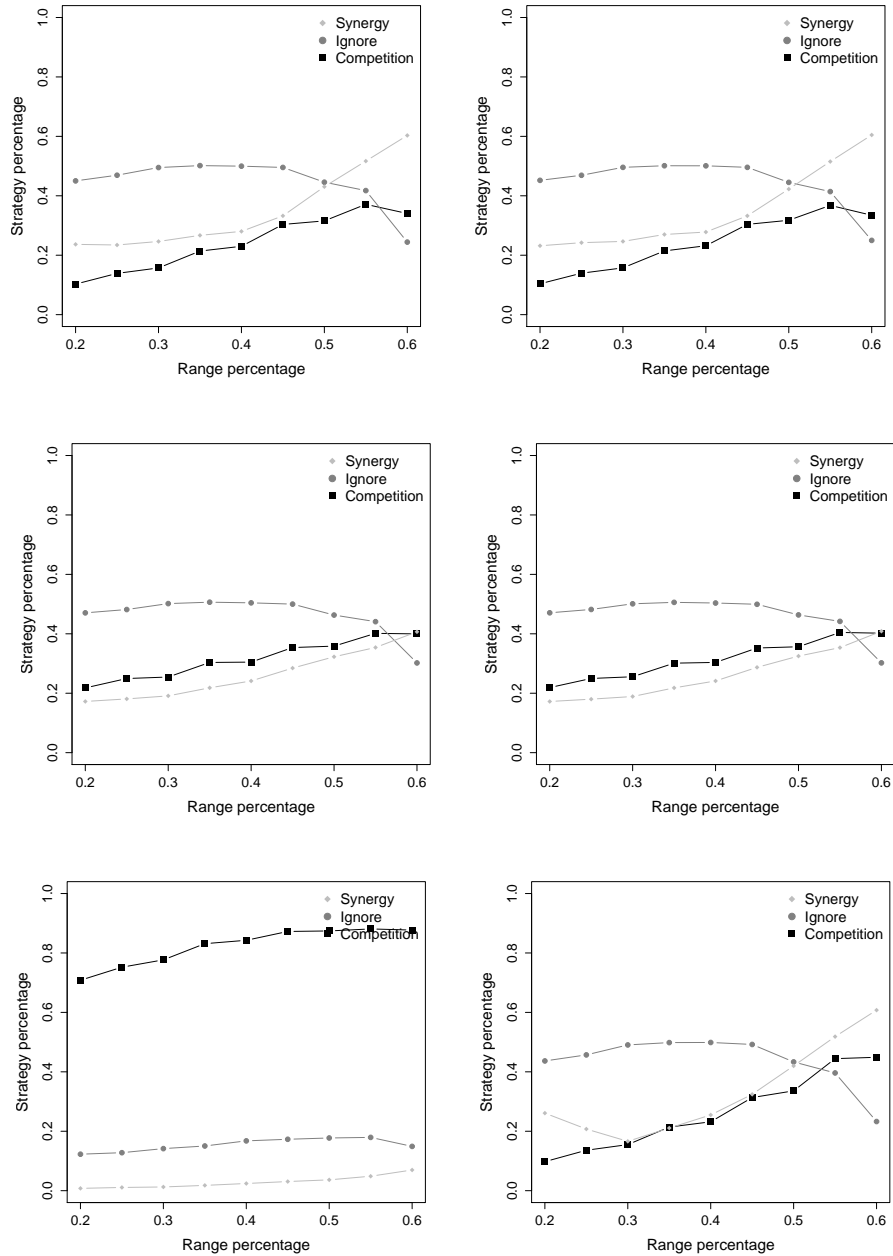


Figure 3.5: How increasing the percentage of aggression correlated to each strategy affects the, on average, appearance of that strategy in the population. For initial populations that unanimously select each of the strategies. Top row shows Synergy initialization, middle one shows Ignore initialization and the bottom row shows Competition. Left figures represent a mutation rate of 0.005 while right figures represent a mutation rate of 0.03.

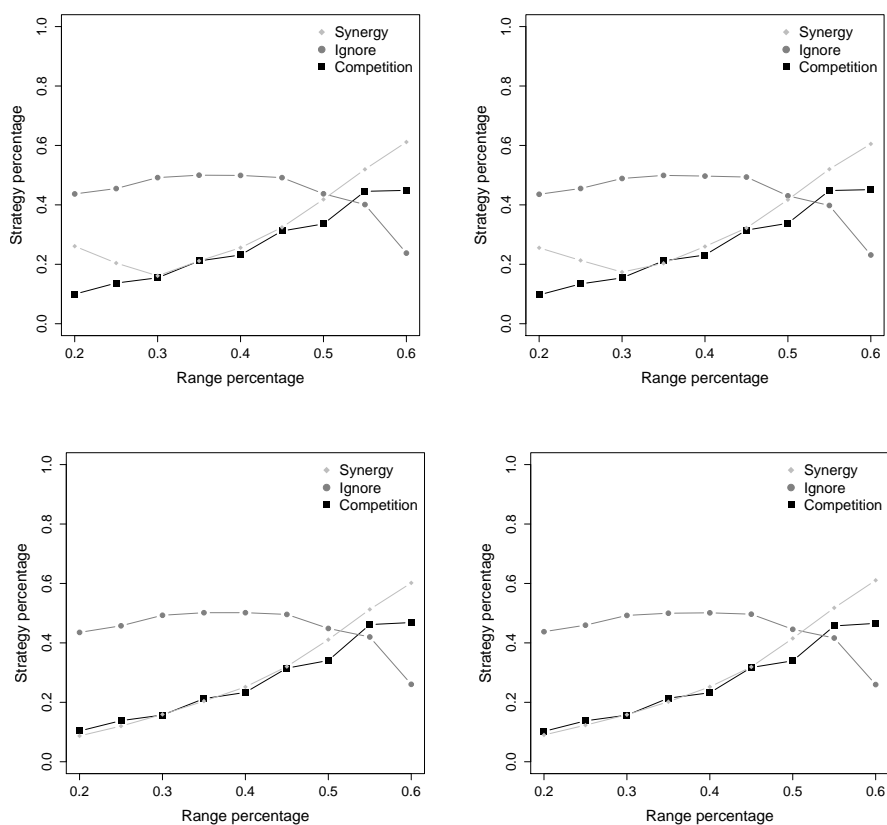


Figure 3.6: How increasing the percentage of aggression correlated to each strategy affects the appearance of that strategy. Initial populations of equal representation of each strategy can be seen in the top figures and for random initialization at the bottom figures. Left figures represent a mutation rate of 0.005 while right figures represent a mutation rate of 0.03.



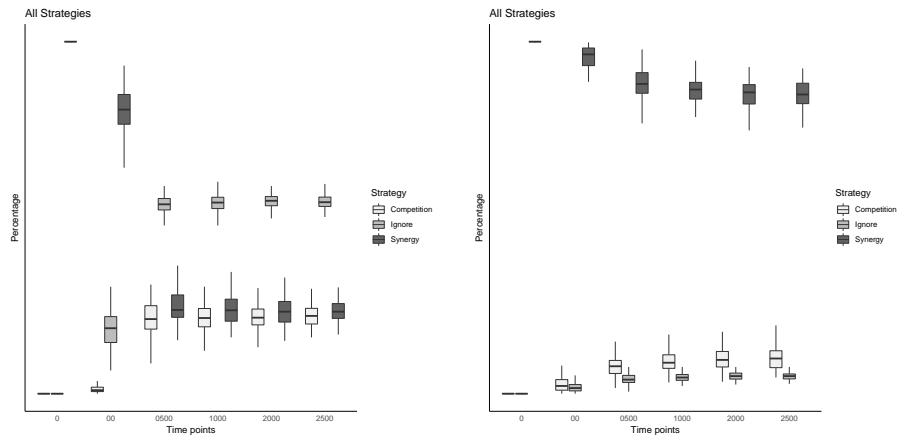


Figure 3.7: Three strategy game with synergy as the initial strategy. The first figure is for balanced ranges of aggression (0.35 synergy threshold and 0.65 competition threshold). The second figure shows the results of having one strategy, synergy in this case, correspond to a higher percentage of aggression.

strategy 3.7. Ignore players seem to thrive in these scenarios, possibly due to being the middle ground. The randomness when deriving the phenotype from the genotype of a predator, caused by the Gaussian equation 2.1 that is used during this process, might give an advantage to the Ignore strategy. Players whose genotype corresponds to a Synergy or a Competition phenotype may, in rare occasions, end up choosing the Ignore strategy instead. Naturally, a predator whose genotype corresponds to an ignore player may also end up choosing either of the three strategies depending on the aggression percentage and the randomness factor.

### 3.5 Detecting Selection

A tool that can be used for Selective Sweep detection is SweeD [56]. SweeD can parse various input file formats including the output format of *FEG*. In this work, we use SweeD to detect traces of selection in the genotype of predators sampled after the completion of a simulation, i.e., when the game has been finished. Using genotypic data, SweeD calculates a likelihood score for the possible existence of selection in regions of sampled genotypes. Running *FEG* under a neutral model 3.1, provides likelihood scores under the null hypothesis, thus it provides values of the statistic under the null hypothesis of no selection that can be used to perform hypothesis testing and calculate threshold values ???. We consider simulation of the second model that have a likelihood score that exceeds 95% of the likelihood scores derived from the neutral model to be under selection. Therefore, Using neutral selection results to calculate the threshold value at a 0.05 level of significance, the estimated 41.56% of the runs are considered to be under selection in the *FEG* model (i.e, the true positive rate). Such a low true positive rate may be

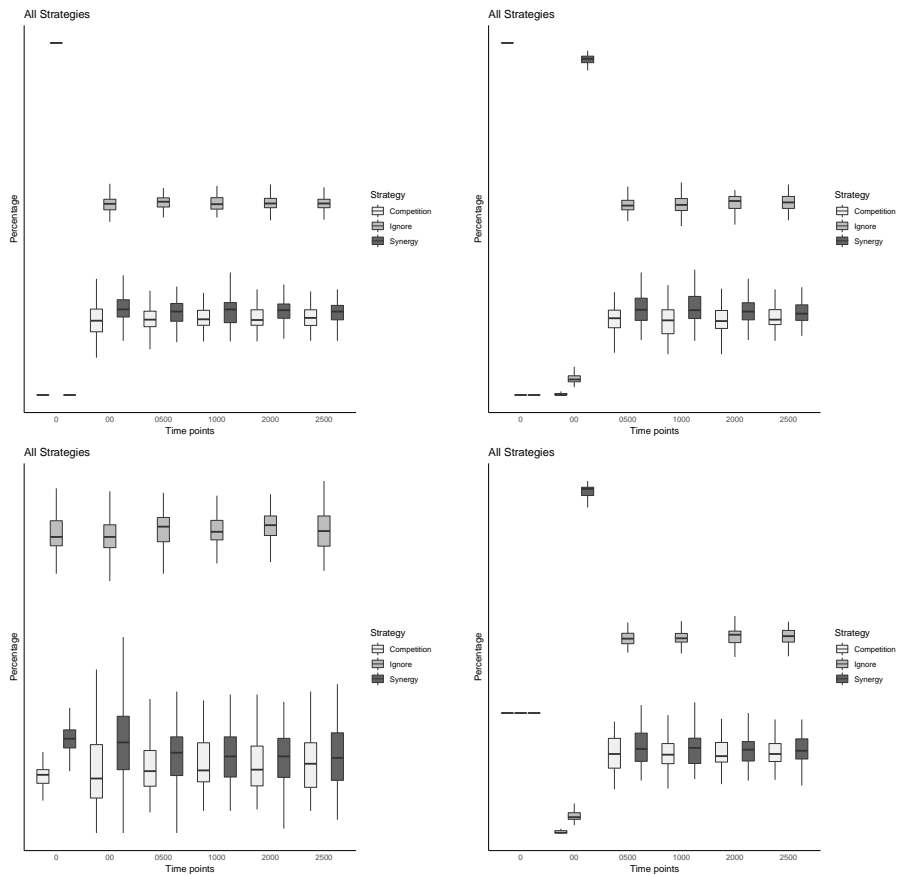


Figure 3.8: Three strategy game with balanced aggression levels (0.35 synergy threshold and 0.65 competition threshold). Each plot showcases a different initialization. In the top plots, the initial population unanimously selects a single strategy. The left plot starts with a population solely comprised of ignore players, whereas the right plot starts with a population of competitors. The bottom two plots are with random (left) and equal (right) number of predators choosing each of the three strategies.

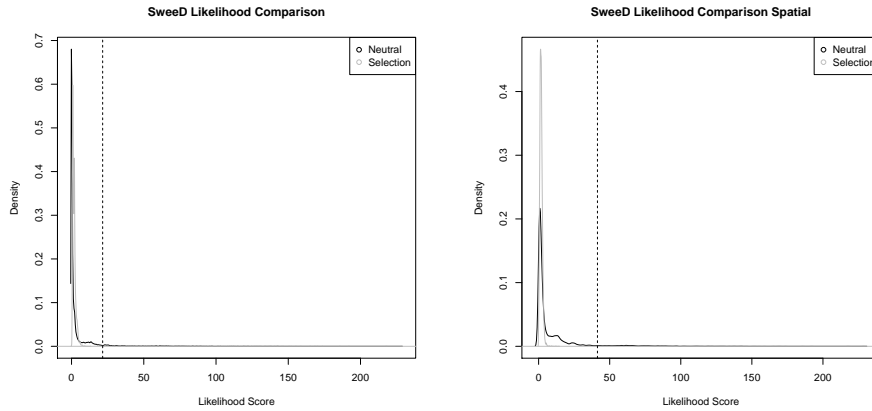


Figure 3.9: Comparison of the likelihood score distributions produced by SweeD for a model under neutrality vs other models created using *FEG*. The plot on the left showcases a model featuring reproduction that is only influenced by the distance between the predators. The plot on the right showcases a model featuring an evolutionary game but no spatial characteristics. For both, the Site Frequency Spectrum of the neutral model was given as an input to SweeD.

associated with random genetic drift due to small effective population size used in the simulations (i.e., increased levels of stochasticity).

A related question is which evolutionary factor generates the selective sweep signal captured by SweeD. Since *FEG* is characterized by both a spatial and an EGT component, either or both of these features could be the cause for the selection signal. The spatial factor is, of course, a non-adaptive factor that is known to increase the false positive rate of selection inference. To answer this question, we performed two additional experiments. One that only features the spatial aspects without the predators participating in an EG and a second that only has the EG part and lacks the spatial characteristics. For both models compared, the Site Frequency Spectrum of the neutral model was given as an input to SweeD. For the spatial only model, we used a single prey, that could be reached from anywhere in the map. Thus, the only parameter influencing the reproduction is the distance between the predators, that will however be reduced as the generations progress. Due to the increased fitness a pair of predators closer to each other receive, the population will end up in a cluster where all predators are very close to each other, thus leaving behind a similar footprint to the neutral model. For the EG only model, selection should exist. In the scenario simulated, featuring co-operators vs competitors, competitors are shown to be dominant. Starting from a population solely comprised of co-operators, competition emerges in almost every simulation. However, using SweeD to detect those traces of selection, did not have the expected results, with the comparison between the neutral model and the one featuring the EG, showing hardly any evidence of selection.

In a Synergy - Competition comparison where all predators are initialized with a zero-level of aggression, the appearance of competitive predators, and the ESS that will persist in the population depend on certain parameters. These parameters are (i) the threshold that determines whether the switch in strategies happens, (ii) the mutation rate that allows for a more or less rapid shift in the population structure and (iii) the payoff that each synergy grants to the player selecting it. The latter is what we opt to examine here, thus keeping the other two fixed to specify on the payoff shift impact on the population. Since the model revolves around co-operators and competitors the payoff of each predator is influenced by (i) the gain synergy players receive from co-operating with each other, (ii) the exploit of the synergy players by the competitors and (iii) the antagonism loss that competitors receive from co-existing.

Setting different values on these parameters should result in different ESS. Competitors are by definition [71] favorable vs co-operators. However, a majority of competitors might change this due to the impact of mutual competition. If the antagonism loss ends up being greater than the exploit from the synergy players, the number of co-operators should increase ending up in a scenario of balancing selection, where both strategies co-exist in a population. On the contrary, if the above condition is never met, a population of purely competitive predators should be the result. Both of these scenarios can be simulated by *FEG* and their effect on the population can be seen in Figure 3.10.

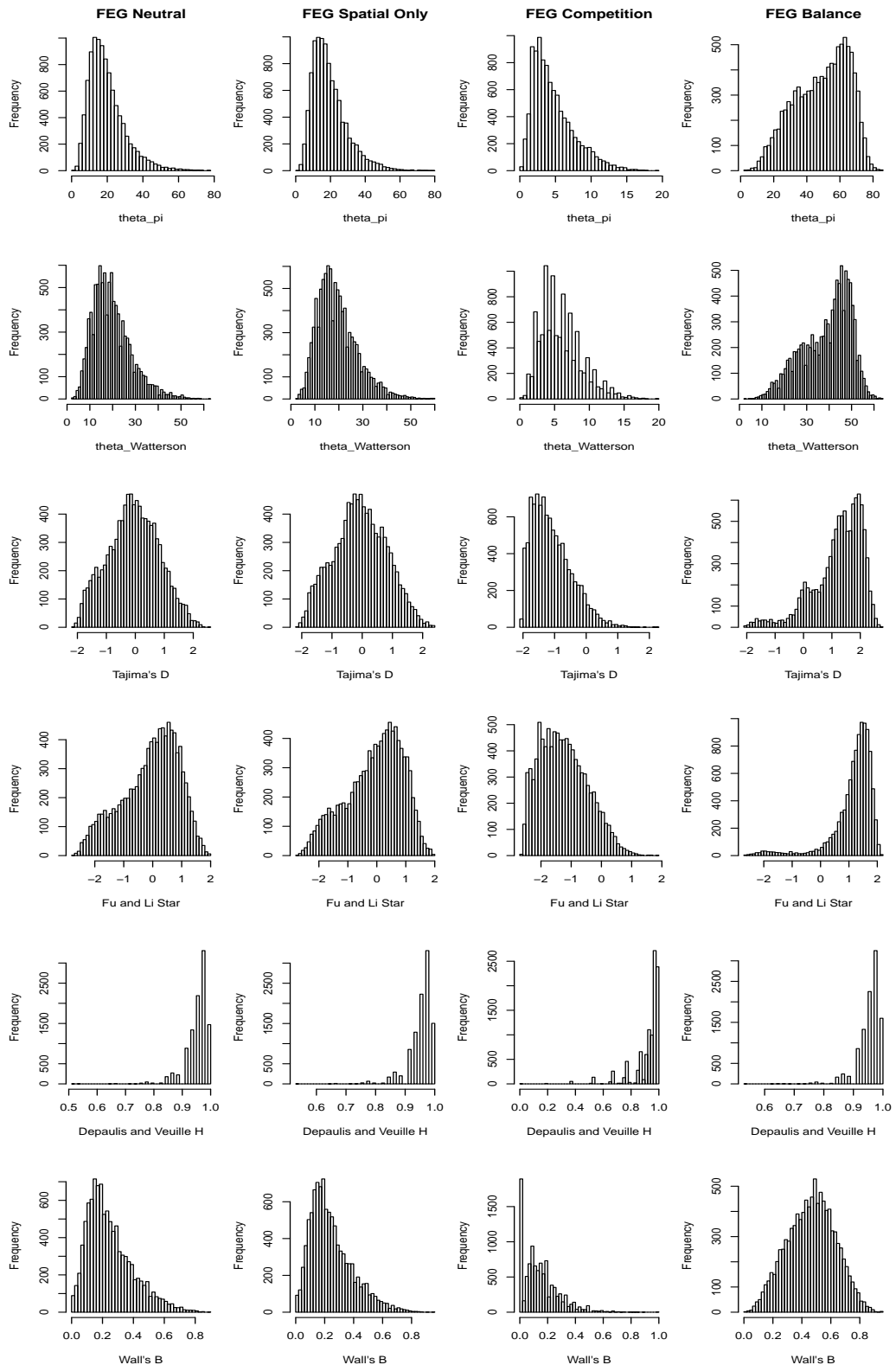


Figure 3.10: Comparison of common summary statistics values amongst various scenarios simulated using *FEG*. The first column represents a model of neutrality. The second a model where the population is solely influenced by spatial characteristics. The third column represents a scenario where competitors are clearly dominant against co-operators. The final column represents a scenario where a balance between the two strategies is the Evolutionary Stable Strategy.

Name / Parameters	Initial	$THR_S$	$THR_C$	mut_rate	rec_rate	burn
Syn - Ign SynA	500/0/0	0.2	1.1	0.005	0.001	500
Syn - Ign SynB	500 0 0	0.8	1.1	0.03	0.001	500
Syn - Ign IgnA	0 500 0	0.2	1.1	0.005	0.001	500
Syn - Ign IgnB	0 500 0	0.8	1.1	0.03	0.001	500
Syn - Ign HalfA	250 250 0	0.5	1.1	0.005	0.001	500
Syn - Ign HalfA	250 250 0	0.5	1.1	0.03	0.001	500
Syn - Ign RandA	R R 0	0.5	1.1	0.005	0.001	500
Syn - Ign RandB	R R 0	0.5	1.1	0.03	0.001	500
Syn - Com SynA	500 0 0	0.2	0.2	0.005	0.001	500
Syn - Com SynB	500 0 0	0.8	0.8	0.03	0.001	500
Syn - Com ComA	0 0 500	0.2	0.2	0.005	0.001	500
Syn - Com ComB	0 0 500	0.8	0.8	0.03	0.001	500
Syn - Com HalfA	250 0 250	0.5	0.5	0.005	0.001	500
Syn - Com HalfA	250 0 250	0.5	0.5	0.03	0.001	500
Syn - Com RandA	R 0 R	0.5	0.5	0.005	0.001	500
Syn - Com RandB	R 0 R	0.5	0.5	0.03	0.001	500
Ign - Com SynA	0 500 0	0.0	0.2	0.05	0.001	500
Ign - Com SynB	0 500 0	0.0	0.8	0.03	0.001	500
Ign - Com ComA	0 0 500	0.0	0.2	0.005	0.001	500
Ign - Com ComB	0 0 500	0.0	0.8	0.03	0.001	500
Ign - Com HalfA	0 250 250	0.0	0.5	0.005	0.001	500
Ign - Com HalfA	0 250 250	0.0	0.5	0.03	0.001	500
Ign - Com RandA	0 R R	0.0	0.5	0.005	0.001	500
Ign - Com RandB	0 R R	0.0	0.5	0.03	0.001	500
Syn - Com Select	500 0 0	0.05	0.05	0.02	0.1	500
Syn - Com Neutral	500 0 0	0.05	0.05	0.02	0.1	3500

Table 3.3: Parameters used for simulating the various scenarios. The Initial column described the initial number of predators selecting each strategy. R symbolizes a random initialization. The three numbers are for Synergy, Ignore and Competition players accordingly.  $THR_S$  is the Synergy threshold and  $THR_C$  the Competition threshold. Finally, mut\_rate describes the mutation rate and rec\_rate the recombination rate. The first 3 blocks on the table represent the two strategy runs to test for ESS. The runs labeled as Select and Neutral are for selection detection. In the neutral scenario the game was played for 0 generations. For the rest, it was played for 3000 generations.

## Chapter 4

# Discussion - Future Work

The interplay between theoretical and empirical models has been the focus of population genetics over the past decades. Complex stochastic models describe how population parameters such as recombination or mutation rates might lead to certain features of genetic polymorphisms [45]. The goal, broadly stated, is to estimate the model parameters that, under a series of a priori assumptions, would produce patterns of variation similar to the ones we observe in nature. Individuals that co-exist in a population interact with each other, in an attempt to secure resources and reproduce. One way to study these interactions is from the perspective of *EGT*, where each individual acts accordingly to a set of possible strategies. These strategies can be determined by its phenotype. During this work, we created *FEG* a forward in time spatial simulator to model the effect of *EGT* on a population of predators moving around a set of preys in the map. The pool of possible strategies consists of the Synergy, Ignore and Competition strategies.

The main goal, as already stated, was the implementation of a tool that can simulate a non-binary strategy EG with spatial characteristics. Along the development of the tool, there are several other accomplishments of this work. The various scenarios described in section 3 above, demonstrate the impact of *EGT* in population genetics. Furthermore, the simultaneous existence of both ignore players and competitors in an EGT game has, to our knowledge, never been studied in a spatial environment. The results described here, highlight the effect of quantitative traits on the structure of populations. We considered the aggression of a predator to be a quantitative trait. By shifting the synergy and competition thresholds accordingly, each of the three strategies is given a certain ‘genotypic space’. Increasing that range makes more probable for the corresponding strategy to emerge in the population. The initial population composition also affects, as expected, the resulting state of the population. From the scenarios described we concluded that a co-operator thrives along a majority of other co-operators. The emergence of competitors leads to either a population of purely Competitors or to a state of balance between Synergy and co-operation depending on the decrease in payoff due to mutual competition. Ignore players perform significantly better

versus competitors but are in a disadvantage vs co-operators. Interestingly, in a three strategy game, the ignore players seem to be advantageous against the other two strategies.

Preys in this model, are considered of constant value and position and do not reproduce or deviate over time, since the model is oriented around predators rather than preys or on the co-existence of the two species. There has been a lot of research on predator-prey co-evolution [1, 10, 68, 7]. Previous models of predator-prey co-evolution involving continuous traits have been classified in two types. ESS models [10, 25, 8] have not included the dynamics of evolving traits, but have assumed that traits and trait-value distributions will conform to frequency-dependent fitness surfaces. Despite studying phenotypic quantitative genetic models [39, 68], only a few very simple and specific models were investigated. Investigations have been made in predator - prey models that assume evolution occurs exclusively by the introduction of new mutants with small effects [48, 47]. Population dynamic equilibrium is assumed to be reached between the invasions of each new mutant form. Thus, a future path of this work will take into account the prey populations and the co-evolution of the two species. Studying quantitative Host - Parasite models [5, 58] is especially important, as they may provide insight to fields such as epidemiology. As one can see, a large part of the future work proposed here, involves further additions or modifications to the current simulation algorithm. At a certain point, a more general solution should be attempted, one not necessarily limited to *EGT*. This simulator already features a forward-in-time approach, a continuous spatial environment and a reproduction model based on quantitative phenotypic traits. A population genetics library, i.e, a low level library which would allow for the easy construction of various functional simulator could be based on the work described in this M.Sc. thesis.



# Bibliography

- [1] Peter A Abrams and Hiroyuki Matsuda. Fitness minimization and dynamic instability as a consequence of predator–prey coevolution. *Evolutionary Ecology*, 11(1):1–20, 1997.
- [2] Joshua M Akey, Michael A Eberle, Mark J Rieder, Christopher S Carlson, Mark D Shriver, Deborah A Nickerson, and Leonid Kruglyak. Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS biology*, 2(10):e286, 2004.
- [3] Nikolaos Alachiotis and Pavlos Pavlidis. Raisd detects positive selection based on multiple signatures of a selective sweep and snp vectors. *Communications Biology*, 1(1):79, 2018.
- [4] Erik C Andersen, Justin P Gerke, Joshua A Shapiro, Jonathan R Crissman, Rajarshi Ghosh, Joshua S Bloom, Marie-Anne Félix, and Leonid Kruglyak. Chromosome-scale selective sweeps shape *caenorhabditis elegans* genomic diversity. *Nature genetics*, 44(3):285, 2012.
- [5] Roy M Anderson and Robert M May. Regulation and stability of host-parasite population interactions. i. regulatory processes. *Journal of animal ecology*, 47(1):219–247, 1978.
- [6] Lars Arve Bach\*, David JT Sumpter, Jan Alsner, and Volker Loeschcke. Spatial evolutionary games of interaction among generic cancer cells. *Journal of Theoretical Medicine*, 5(1):47–58, 2003.
- [7] H Clark Barrett. Adaptations to predators and prey. *The handbook of evolutionary psychology*, pages 1–18, 2015.
- [8] Katrin Bohl, Sabine Hummert, Sarah Werner, David Basanta, Andreas Deutsch, Stefan Schuster, Günter Theißen, and Anja Schroeter. Evolutionary game theory: molecules as players. *Molecular BioSystems*, 10(12):3066–3074, 2014.
- [9] John M Braverman, Richard R Hudson, Norman L Kaplan, Charles H Langley, and Wolfgang Stephan. The hitchhiking effect on the site frequency spectrum of dna polymorphisms. *Genetics*, 140(2):783–796, 1995.

- [10] Joel S Brown and Thomas L Vincent. Coevolution as an evolutionary game. *Evolution*, 41(1):66–79, 1987.
- [11] Steven X Cadrin, Lisa A Karr, and Stefano Mariani. Stock identification methods: an overview. In *Stock identification methods*, pages 1–5. Elsevier, 2014.
- [12] B Charlesworth and D Charlesworth. Population genetics from 1966 to 2016. *Heredity*, 118(1):2, 2017.
- [13] Mike Cowlshaw. The decnumber c library. Available at [www2.hursley.ibm.com/decimal/decnumber/pdf](http://www2.hursley.ibm.com/decimal/decnumber/pdf), 2006.
- [14] Katalin Csilléry, Olivier François, and Michael GB Blum. abc: an r package for approximate bayesian computation (abc). *Methods in ecology and evolution*, 3(3):475–479, 2012.
- [15] Patrick De Leenheer, Anushaya Mohapatra, Haley A Ohms, David A Lytle, and JM Cushing. The puzzle of partial migration: Adaptive dynamics and evolutionary game theory perspectives. *Journal of theoretical biology*, 412:172–185, 2017.
- [16] Justin C Fay and Chung-I Wu. Hitchhiking under positive darwinian selection. *Genetics*, 155(3):1405–1413, 2000.
- [17] Joseph Felsenstein. The evolutionary advantage of recombination. *Genetics*, 78(2):737–756, 1974.
- [18] Anne M Glazier, Joseph H Nadeau, and Timothy J Aitman. Finding genes that underlie complex traits. *science*, 298(5602):2345–2349, 2002.
- [19] Chaitanya S Gokhale and Arne Traulsen. Evolutionary multiplayer games. *Dynamic Games and Applications*, 4(4):468–488, 2014.
- [20] Benjamin C Haller and Philipp W Messer. Slim 2: Flexible, interactive forward genetic simulations. *Molecular biology and evolution*, 34(1):230–240, 2016.
- [21] Daniel L Hartl, Andrew G Clark, and Andrew G Clark. *Principles of population genetics*, volume 116. Sinauer associates Sunderland, 1997.
- [22] Mikko Heino. Quantitative traits. In *Stock Identification Methods*, pages 59–76. Elsevier, 2014.
- [23] Garrett Hellenthal and Matthew Stephens. mshot: modifying hudson’s ms simulator to incorporate crossover and gene conversion hotspots. *Bioinformatics*, 23(4), 2007.

- [24] Ryan D Hernandez. A flexible forward simulator for populations subject to selection and demography. *Bioinformatics*, 24(23):2786–2787, 2008.
- [25] W Gord S Hines and J Maynard Smith. Games between relatives. *Journal of Theoretical Biology*, 79(1):19–30, 1979.
- [26] WGS Hines. Evolutionary stable strategies: a review of basic theory. *Theoretical Population Biology*, 31(2):195–272, 1987.
- [27] Asger Hobolth, Ole F Christensen, Thomas Mailund, and Mikkel H Schierup. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden markov model. *PLoS genetics*, 3(2):e7, 2007.
- [28] Richard R Hudson. Properties of a neutral allele model with intragenic recombination. *Theoretical population biology*, 23(2):183–201, 1983.
- [29] Richard R Hudson. Generating samples under a wright–fisher neutral model of genetic variation. *Bioinformatics*, 18(2):337–338, 2002.
- [30] Richard R Hudson et al. Gene genealogies and the coalescent process. *Oxford surveys in evolutionary biology*, 7(1):44, 1990.
- [31] Sabine Hummert, Katrin Bohl, David Basanta, Andreas Deutsch, Sarah Werner, Günter Theißen, Anja Schroeter, and Stefan Schuster. Evolutionary game theory: cells as players. *Molecular BioSystems*, 10(12):3044–3065, 2014.
- [32] M Husemann, FE Zachos, RJ Paxton, and JC Habel. Effective population size in ecology and evolution, 2016.
- [33] Norman L Kaplan, RR Hudson, and CH Langley. The” hitchhiking effect” revisited. *Genetics*, 123(4):887–899, 1989.
- [34] Jerome Kelleher, Alison M Etheridge, and Gilean McVean. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS computational biology*, 12(5):e1004842, 2016.
- [35] Timothy Killingback and Michael Doebeli. Spatial evolutionary game theory: Hawks and doves revisited. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263(1374):1135–1144, 1996.
- [36] Motoo Kimura. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*, 16(2):111–120, 1980.
- [37] John FC Kingman. On the genealogy of large populations. *Journal of applied probability*, 19(A):27–43, 1982.

- [38] Michał Krześlak and Andrzej Świerniak. Multidimensional extended spatial evolutionary games. *Computers in biology and medicine*, 69:315–327, 2016.
- [39] Russell Lande. Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 30(2):314–334, 1976.
- [40] Russell Lande and Stevan J Arnold. The measurement of selection on correlated characters. *Evolution*, 37(6):1210–1226, 1983.
- [41] Eric S Lander and David Botstein. Mapping mendelian factors underlying quantitative traits using rflp linkage maps. *Genetics*, 121(1):185–199, 1989.
- [42] Gregory I Lang and Andrew W Murray. Estimating the per-base-pair mutation rate in the yeast *saccharomyces cerevisiae*. *Genetics*, 178(1):67–82, 2008.
- [43] Guillaume Laval and Laurent Excoffier. Simcoal 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, 20(15):2485–2487, 2004.
- [44] Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. Multi-agent reinforcement learning in sequential social dilemmas. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 464–473. International Foundation for Autonomous Agents and Multiagent Systems, 2017.
- [45] Junrui Li, Haipeng Li, Mattias Jakobsson, SEN Li, PER SjöDin, and Martin Lascoux. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Molecular ecology*, 21(1):28–44, 2012.
- [46] David Liao and Thea D Tlsty. Evolutionary game theory for physical and biological scientists. i. training and validating population dynamics equations. *Interface Focus*, 4(4):20140037, 2014.
- [47] Paul Marrow and Chris Cannings. Evolutionary instability in predator-prey systems. *Journal of theoretical biology*, 160(2):135–150, 1993.
- [48] Paul Marrow, Richard Law, and C Cannings. The coevolution of predator—prey interactions: Esss and red queen dynamics. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 250(1328):133–141, 1992.
- [49] John Maynard Smith and John Haigh. The hitch-hiking effect of a favourable gene. *Genetical research*, 23(1):23–35, February 1974.
- [50] Martin A Nowak. Five rules for the evolution of cooperation. *science*, 314(5805):1560–1563, 2006.

- [51] Martin A Nowak and Robert M May. Evolutionary games and spatial chaos. *Nature*, 359(6398):826, 1992.
- [52] Dorcas J Orengo and Montserrat Aguadé. Detecting the footprint of positive selection in a european population of drosophila melanogaster: multilocus pattern of variation and distance to coding regions. *Genetics*, 167(4):1759–1766, 2004.
- [53] S Papadantonakis, P Poirazi, and P Pavlidis. Comus: simulating coalescent histories and polymorphic data from multiple species. *Molecular ecology resources*, 16(6):1435–1448, 2016.
- [54] Pavlos Pavlidis, Stephan Hutter, and Wolfgang Stephan. A population genomic approach to map recent positive selection in model species. *Molecular ecology*, 17(16):3585–3598, 2008.
- [55] Pavlos Pavlidis, Jeffrey D Jensen, and Wolfgang Stephan. Searching for footprints of positive selection in whole-genome snp data from non-equilibrium populations. *Genetics*, 2010.
- [56] Pavlos Pavlidis, Daniel Živković, Alexandros Stamatakis, and Nikolaos Alachiotis. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular biology and evolution*, page mst112, 2013.
- [57] Jorge Pena. Group-size diversity in public goods games. *Evolution: International Journal of Organic Evolution*, 66(3):623–636, 2012.
- [58] Rachel M Penczykowski, Anna-Liisa Laine, and Britt Koskella. Understanding the ecology and evolution of host–parasite interactions across scales. *Evolutionary applications*, 9(1):37–52, 2016.
- [59] Jérôme Pinti and André W Visser. Predator-prey games in multiple habitats reveal mixed strategies in diel vertical migration. *The American Naturalist*, 193(3):E65–E77, 2019.
- [60] S Piry, G Luikart, and JM Cornuet. Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of heredity*, 90:502–503, 1999.
- [61] David Posada and Carsten Wiuf. Simulating haplotype blocks in the human genome. *Bioinformatics*, 19(2):289–290, 2003.
- [62] Marc Pybus, Pierre Luisi, Giovanni Marco Dall’Olio, Manu Uzkudun, Hafid Laayouni, Jaume Bertranpetit, and Johannes Engelken. Hierarchical boosting: a machine-learning framework to detect and classify hard selective sweeps in human populations. *Bioinformatics*, 31(24):3946–3952, 2015.

- [63] Sebastian E Ramos-Onsins and Thomas Mitchell-Olds. Mlcoalsim: multilocus coalescent simulations. *Evolutionary Bioinformatics*, 3:117693430700300020, 2007.
- [64] Anatol Rapoport. Prisoner’s dilemma—recollections and observations. In *Game Theory as a Theory of a Conflict Resolution*, pages 17–34. Springer, 1974.
- [65] Nicolas Ray, Mathias Currat, Matthieu Foll, and Laurent Excoffier. Splat2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, 26(23):2993–2994, 2010.
- [66] Roy Ronen, Nitin Udpa, Eran Halperin, and Vineet Bafna. Learning natural selection from the site frequency spectrum. *Genetics*, 195(1):181–193, 2013.
- [67] Carl-Johan Rubin, Michael C Zody, Jonas Eriksson, Jennifer RS Meadows, Ellen Sherwood, Matthew T Webster, Lin Jiang, Max Ingman, Ted Sharpe, Sojeong Ka, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, 464(7288):587, 2010.
- [68] Irma Saloniemi. A coevolutionary predator-prey model with quantitative characters. *The American Naturalist*, 141(6):880–896, 1993.
- [69] Joshua G Schraiber, Steven N Evans, and Montgomery Slatkin. Bayesian inference of natural selection from allele frequency time series. *Genetics*, pages genetics–116, 2016.
- [70] Daniel R Schrider and Andrew D Kern. S/hic: Robust identification of soft and hard sweeps using machine learning. *PLoS genetics*, 12(3):e1005928, 2016.
- [71] J Maynard Smith and George R Price. The logic of animal conflict. *Nature*, 246(5427):15–18, 1973.
- [72] Wolfgang Stephan, Yun S Song, and Charles H Langley. The hitchhiking effect on linkage disequilibrium between linked neutral sites. *Genetics*, 2006.
- [73] Andrzej Świerniak, Michał Krześlak, Sebastian Student, and Joanna Rzeszowska-Wolny. Development of a population of cancer cells: observation and modeling by a mixed spatial evolutionary games approach. *Journal of theoretical biology*, 405:94–103, 2016.
- [74] Fumio Tajima. Evolutionary relationship of dna sequences in finite populations. *Genetics*, 105(2):437–460, 1983.
- [75] Fumio Tajima. Statistical method for testing the neutral mutation hypothesis by dna polymorphism. *Genetics*, 123(3):585–595, 1989.
- [76] Jeffrey D Wall. Recombination and the power of statistical tests of neutrality. *Genetics Research*, 74(1):65–79, 1999.

- [77] Ming-Shan Wang, Rong-wei Zhang, Ling-Yan Su, Yan Li, Min-Sheng Peng, He-Qun Liu, Lin Zeng, David M Irwin, Jiu-Lin Du, Yong-Gang Yao, et al. Positive selection rather than relaxation of functional constraint drives the evolution of vision during chicken domestication. *Cell research*, 26(5):556, 2016.
- [78] Claus O Wilke, Jia Lan Wang, Charles Ofria, Richard E Lenski, and Christoph Adami. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 412(6844):331–333, 2001.
- [79] Xiguo Yuan, David J Miller, Junying Zhang, David Herrington, and Yue Wang. An overview of population genetic data simulation. *Journal of Computational Biology*, 19(1):42–54, 2012.
- [80] Yang Yuan, Qijun Zhang, Shuiyun Zeng, Longjiang Gu, Weina Si, Xiaohui Zhang, Dacheng Tian, Sihai Yang, and Long Wang. Selective sweep with significant positive selection serves as the driving force for the differentiation of japonica and indica rice cultivars. *BMC genomics*, 18(1):307, 2017.
- [81] Zebin Zhang, Yaxiong Jia, Pedro Almeida, Judith E Mank, Marcel van Tuinen, Qiong Wang, Zhihua Jiang, Yu Chen, Kai Zhan, Shuisheng Hou, et al. Whole-genome resequencing reveals signatures of selection and timing of duck domestication. *GigaScience*, 7(4):giy027, 2018.