2014

genomic and behavioural evolution in the artificial ecosystem simulation EcoSim

Marwa Fouad Khater
University of Windsor

Recommended Citation

Khater, Marwa Fouad, "genomic and behavioural evolution in the artificial ecosystem simulation EcoSim" (2014). *Electronic Theses and Dissertations*. 5204.
https://scholar.uwindsor.ca/etd/5204

Genomic and behavioural evolution in the artificial ecosystem simulation EcoSim

by

Marwa Fouad Khater

A Dissertation
Submitted to the Faculty of Graduate Studies
through School of Computer Science
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada,
2014

Genomic and behavioural evolution in the artificial ecosystem simulation EcoSim

by

Marwa Khater

APPROVED BY

Dr. G. Arhonditsis ,External Examiner
University of Toronto Scarborough

Dr. J. Ciborowski, External Reader
Department of Biological Sciences

Dr. Z. Kobti, Internal Reader
School of Computer Science

Dr. A. Ngom, Internal Reader
School of Computer Science

Dr. R. Gras, Advisor
School of Computer Science

September 11, 2014

Declaration of Co-Authorship

I hereby declare that this dissertation incorporates material that is result of joint research, as follows:

This dissertation also incorporates the outcome of a joint research undertaken in collaboration with Marwa Khater and Dorian Murariu under the supervision of professor Robin Gras. The collaboration is covered in Chapter 7 of the dissertation whenever a biological discussion is required. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contribution of co-authors was primarily through the provision of required background biological information.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my dissertation, and have obtained written permission from the co-author to include the above materials in my dissertation. I certify that, with the above qualification, this dissertation, and the research to which it refers, is the product of my own work.

Declaration of Previous Publication This dissertation includes five original papers that have been published/submitted for publication in peer reviewed conferences /Journals as follows:

Thesis Chapter	Publication title/ full citation	Publication status
Chapter 5	M. Khater, E. Salehi, and R. Gras, "The emergence of new genes in EcoSim and its effect on fitness," in Simulated Evolution and Learning. Springer, 2012, pp. 52-61.	Published
Chapter 5 Chapter 6	M. Khater and R. Gras, "Adaptation and genomic evolution in EcoSim," From Animals to Animats 12, pp. 219-229, 2012.	Published
Chapter 6	M. Khater, E. Salehi, and R. Gras, "Correlation between genetic diversity and fitness in a predator-prey ecosystem simulation," in AI 2011: Advances in Artificial Intelligence. Springer, 2011, pp. 422-431.	Published
Chapter 7	M. Khater, D. Murariu, and R. Gras. "Contemporary evolution and genetic change of prey as a response to predator removal." Ecological Informatics, 22:13-22, 2014.	Published
Chapter 7	M. Khater, D. Murariu, and R. Gras. "Predation risk tradeoffs in prey: effects on energy and behaviour". Ecological Complexity	Submitted

Abstract

Artificial life evolutionary systems facilitate addressing lots of fundamental questions in evolutionary genetics. Behavioral adaptation requires long term evolution with continuous emergence of new traits, governed by natural selection. We model organism's genomes coding for their behavioral model and represented by fuzzy cognitive maps (FCM), in an individual-based evolutionary ecosystem simulation (EcoSim). The emergent of new traits (genes) in EcoSim is examined by studying their effect on individual's fitness and well being. We examine how the new traits are used to predict the value of fitness using machine learning techniques. A comparison between the genomic evolution of EcoSim and a neutral model (a randomized version of EcoSim) is examined focusing on their respective genomic diversity. In order to further emphasize the importance of genetic diversity to adaptation and thus the well being of individuals, we were encouraged to study the effect that genetic diversity has on fitness. EcoSim gives us the chance to study the relation between species genetic diversity and average species fitness without the limits in environmental conditions and time scales found in biological studies, but in highly variable environments and across evolutionary time.

The ecological effects of predator removal and its consequence on prey behavior have been investigated widely. We investigated the effects of predation risk on prey energy allocation and fitness. Here the role of predator removal on the contemporary evolution of prey traits such as movement, reproduction and foraging was evaluated. Our study clearly shows that predation risk alone induces behavioural changes in prey which drastically affect population and community dynamics, A classification algorithm was used to demonstrate the difference between genomes belonging to prey co-evolving with predators and prey evolving in the absence of predation pressure. We argue that predator introductions to naive prey might be destabilizing if prey have evolved and adapted to the absence of predators. Our results suggest that both predator introduction and predator removal from an ecosystem have widespread effects on the survival and evolution of prey by altering their genomes and behaviour, even after relatively short time intervals.

Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Robin Gras for the continuous support of my PhD study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. He has always allocated time for me and was constantly giving me advice. I could not have imagined having a better supervisor and mentor for my PhD study. I would like to thank my committee members Dr. George Arhonditsis, Dr. Jan Ciborowski, Dr. Ziad Kobti and Dr. Alioune Ngom for accepting to allocate part of their valuable time to evaluate my research.

I would like to extend my gratitude towards Dorian Murariu, my research partner and biology cohort, who meticulously helped me in collecting all the required biological information needed for this research. His valuable insight and comments added depth to this thesis. I also like to thank all my team mates and other lab members specially Abbas Golestani, Morteza Mashayekhi and Yasaman Farahani who throughout this journey became dear friends. They gave me their comments and technical help when needed.

Special thanks must go to my husband and life-partner Rami Salem. He provided unconditional love, support and encouragement through both the highs and lows of my time in graduate school.

Finally, I like to thank and dedicate this thesis to my mother Dr. Sanaa Alassar who has always believed in me and constantly pushed me forward. I derive my strength through her.

This work was made possible by the facilities of Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca) and Compute/Calcul Canada.

Contents

Declaration of Co-Authorship iii

Abstract v

Acknowledgments vi

List of Tables xi

List of Figures xiii

1 Introduction 1

1.1 Motivation 1

1.2 Objective 3

1.3 Contributions of the thesis 4

1.4 Outline of thesis 6

2 Background and Literature Review 8

2.1 Artificial life 8

2.2 IBM 11

2.2.1 Tierra 12

2.2.2 Avida 13

2.2.3 Echo 15

2.2.4 Polyworld 15

2.2.5 Framsticks 16

2.3 Other predator-prey simulations 17

3 Individual based ecosystem simulation EcoSim 19

3.1 Purpose 19

3.2 Entities, state variables, and scales 20

3.3 Process overview and scheduling 21

3.4 Design concepts 22

3.4.1 Basic principles 22

3.4.2 Emergence 24

3.4.3 Adaptation 26

3.4.4 Fitness 29

3.4.5 Prediction 29

3.4.6 Sensing 30

3.4.7 Interaction 30

3.4.8 Stochasticity 31

3.4.9 Collectives 32

3.4.10 Observation 32

3.5 Initialization and input data 33

3.6 Submodels 33

3.7 The Neutral Model 37

4 Data Analysis 40

4.1 Introduction 40

4.2 Entropy as a Measure of Genetic Diversity 40

4.3 Fitness calculation 42

4.4 Building a Classifier for Inference 42

4.4.1 C4.5 43

4.4.2 Random Forest 44

4.4.3 JRip Rule Learner 45

4.5 Feature Selection 46

5 The genomic evolution in EcoSim and its effect on fitness 48

5.1 Introduction 48

5.2 Evolution in EcoSim versus Neutral Model 49

5.3 Emergence of New Genes 50

5.4 Building a Random Forest Classifier for Inference 51

5.5 Rule Learning Using JRip 54

5.6 Conclusion 57

6 Correlation between genetic diversity and fitness in EcoSim 59

6.1 Introduction 59

6.2 Measuring Correlation between Entropy and Fitness 60

6.3 Building Classifier for Inference 64

6.3.1 Feature Selection 64

6.4 Classification 66

6.4.1 Decision trees classification results 66

6.4.2 Random Forest Classifier for Inference 67

6.4.3 JRip Rule Learner 68

6.5 Conclusion 70

7 Behavioural and Genetic Change of Prey as a Response to Predator Removal 72

7.1 Introduction 72

7.2 Summary of EcoSim Runs 75

7.3 Predation effect on prey's behaviour 78

7.3.1 Population dynamics 78

7.3.2 Predation risk-foraging trade off 82

7.3.3 Adjustment of reproduction strategies in response to predation risk 86

7.3.4 Prey movement 88

7.4 Predator consequences on prey genomic evolution 92

7.5 Genomic classification and statistical analysis 95

7.5.1 Classifying instances belonging to case A and C 97

7.5.2 Classifying instances belonging to case B and C 97

7.5.3 Classifying instances belonging to case A and B 98

7.5.4 Extracted semantics from rules 99

7.5.5 Predator introduction 100

7.6 Conclusion 103

8 Summary, Conclusions and Future work **108**

8.1 Summary 108

8.2 Conclusions 109

8.3 Future direction 111

References **113**

Vita Auctoris **127**

List of Tables

3.1 Several physical and life history characteristics of individuals from 10 independent runs. 20

3.2 Values for user specified parameters. 34

5.1 Percentage of low fitness (LOW), high fitness (HIGH), and very high fitness (VHIGH) prey instances for 4 different runs. 52

5.2 Accuracy for training and validating with the Random Forest classifier on four runs of the simulation. 53

5.3 Accuracy percentages for training and validating with JRip after CMSS-EDA and CfsSubsetEval feature selection, for four runs of the simulation. 54

6.1 Percentage of high positive (HIGHP), high negative (HIGHN), weak correlation (WEAK CORR) between fitness and genetic diversity for window of 400 and shift of ±25 for five different runs. 66

6.2 Accuracy percentages for training and validating with the C4.5 classifier for 5 runs of the simulation. 67

6.3 Accuracy percentages for training and validating with the RF classifier for 5 runs of the simulation. 68

6.4 JRip rule learner accuracies and number of produced rules for five different runs of the simulation 69

7.1 Average energies (with std between brackets) consumed by prey while engaging in each type of action, expressed as a percent of their average energy budgets. (Note: since these are average values over the entire duration of the runs they do not add up to 1). Eating represents net energy gained from eating, including the energy spent on the eating action itself. Due to computational limitations we present here the average energy consumed for each action for four runs of case A and B prey only. *Includes all successful and failed actions. 85

7.2 Demographic characteristics as a percent (%) of population for case A and B runs. Standard deviations in brackets. *Includes successful and failed actions. **Death rate includes all causes of death (being eaten by predators (for case A prey only), energy depletion, and old age). 87

7.3 Average frequency of movement actions of all prey as a percent (%) of the total population of prey, for case A and B prey except for Speed which is expressed as number of cells. Standard deviations in brackets (std).*Includes all successful and failed attempts for each action. 90

7.4 Accuracy for training the classifier using 10 fold cross validation, accuracy of validating the classifier using the validation set, and the number of rules produced from the model 95

7.5 Results of predators introduction in different time steps 102

List of Figures

3.1 The overview and scheduling of every time step. 23

3.2 Initial FCM prey map including concepts and edges. The width of each edge represents the influence value of a concept on another. Color of an edge shows inhibitory (red) or excitatory (blue) effects. 25

3.3 An FCM for detection of foe (predator) and decision to evade with its corresponding matrix (0 for 'Foe close', 1 for 'Foe far', 2 for 'Fear' and 3 for 'Evasion') and the fuzzification and defuzzification functions. 25

3.4 The snapshot of the virtual world in one specific time step, white color represents predator species and the other colors show different prey species. 27

3.5 The snapshot of the virtual world for a specific time step of the simulation which demonstrates the pattern of grass in the world. . . 28

3.6 FCM for detection of foe (predator) - difference between perception and sensation. 31

3.7 Breeding algorithm. 36

3.8 Snap shot of the Neutral shadow world with predator and prey spatial distribution 38

5.1 Global Entropy for 10 different runs of the simulation. Top 5 curves are for EcoSim and lower 5 for Neutral Model. 50

6.1 Different prey species correlation values between entropy and fitness. x-axis represents the different time shifts. Y-axis represents the correlation values. 61

7.1 Snapshot of the world in one of the case A runs at time step 5000. The numbers of prey and predator individuals are 198554 and 27903 respectively. The colored spirals show different prey species and the white represents predators. 76

7.2 Snapshot of the world in one of the case B runs at time step 5000 showing 34254 prey individuals. 77

7.3 Snapshot of the world in one of case C runs. Figure on right represents the world at time step 15000 showing 233784 prey and the figure on left at time step 15060 where prey increased to 1040856. . 77

7.4 Total prey population. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. 79

7.5 The average amount of grass in the world. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. 79

7.6 The average prey fitness. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. 80

7.7 The average prey populations choosing reproduction action, including all successful and failed attempts. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. 80

7.8 The average prey birth ratio to population. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. 81

7.9 Grass density is the total units of grass divided by the number of prey population in the world. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone 83

7.10 The average prey foraging ratio to population. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. 84

7.11 Percentage of maximum allowed transmitted energy to offspring at birth. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. 87

7.12 The proportion of the total prey population runs that chose a movement action (escape, foraging, socialize or explore) at each time step. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. Total movement is the sum of all four of these actions, and includes all successful and failed foraging and socialization actions. 89

7.13 The proportion of all movement actions that high (case A) and low(case B) risk prey runs spent. The bars show a behavioural tradeoff between time spent foraging and time spent responding to predation risk through escape actions. 89

7.14 The average prey speed. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. 91

7.15 The average prey genetic distance from initial FCM map. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. . 92

7.16 The average prey genetic diversity (measured by entropy). Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. 93

7.17 The average number of prey species. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. 93

7.18 Distribution of genes appearing in the JRip classifier. 100

7.19 Snapshot of the world in one of the case D runs (predators are introduced to naive prey) at time step 5000. Preys show more spatial dispersal and white patches represent predators. In the first figure the numbers of prey and predator individuals are 31502 and 1008 respectively. In the second figure at time step 5030 prey decreased to 9301 and predators were 1502. At time step 5112 prey decreased to 5725 and predators were all dead. 103

Chapter 1

Introduction

1.1 Motivation

Darwin (1859) conceived the mechanism that could account for the adaptation and the diversity observed in nature. Darwin's principle of natural selection rests on a number of propositions [1]:

- The individuals of a population are not identical but vary in certain traits.
- This variation, at least partly, is heritable. Therefore, an individual shares some of these traits with its ancestors.
- Every population could potentially populate the whole world if each individual of that population realized its full reproductive potential. In reality, few (if any) individuals do, and many individuals die without reproducing at all.
- Individuals vary in their number of descendants (not only the number of children they produce, but the number of children that survive, and the offspring they leave and so forth).
- The number of an individual's descendants depends critically (but not completely) on the interactions of the traits of the individual and its environment.

Populations with these characteristics, over generations, become more adapted to their environment. With time and changing circumstances, different adaptations may become advantageous. Gradually, this mechanism gives rise to different life forms. Artificial Life (Alife) is concerned with the study of the processes and mechanisms underlying life by recreating life-like phenomena in software, hardware, and biochemicals [2]. The term 'artificial life' was coined by Langton who described Alife as "a field of study devoted to understanding life by attempting to abstract the fundamental dynamical principles underlying biological phenomena, and recreating these dynamics in other physical media-such as computers-making them accessible to new kinds of experimental manipulation and testing" [3].

An ecosystem is the complex system described by the organisms, the environment, and their physical, chemical and biological interrelationships in a given area. Ecosystem health is determined through measurable characteristics. Ecosystem health is determined through measurable characteristics. A healthy ecosystem is defined as being 'stable and sustainable'; maintaining its organization and autonomy over time and its resilience to stress [4]. So a healthy ecosystem is sustainable; that is, it has the ability to maintain its structure (organization) and function (vigor) over time in the face of external stress (resilience). A method to quantify these attributes (vigor, organization, and resilience) should be taken into consideration when modeling an ecosystem. Stability, population density, biodiversity and how energy flows through trophic levels are some measurable characteristics in evolutionary ecological models.

The metaphor of an evolving ecosystem was chosen for this study of artificial agent evolution because the evolutionary dynamics result from the modeling of individual agents and their relations to the conditions and resources, in combination with their diverse interactions with other agents in their environment. Together, these factors make up the artificial ecosystem. This view of an evolving system, where natural selection is an emergent system property, stands in contrast to traditional evolutionary algorithms where evolution is the direct result of an algorithmically explicit selection process.

1.2 Objective

Like in many disciplines; simulation modeling played a great role in studying evolutionary processes. Many ecological studies that require data of hundreds of years can be obtained by simulation modeling that produces results in a matter of a few hours or days depending on the computational cost of each system. Darwinian evolution governed by natural selection is modeled in EcoSim (an individual based ecosystem simulation). Our main challenge concerns the ability to understand the evolutionary machinery and evolved behaviour in the system. This problem is of particular importance in systems based on natural selection. The lack of an explicit objective (predefined fitness) function makes spotting and understanding the evolved behaviour of individuals a challenging task. Therefore, there must be some other way to identify and understand qualitatively novel behaviour when it emerges. The objective of the work was to capture the correlation between different emerging behaviours arising in EcoSim in order to gain deeper understanding of the model and thus natural ecosystems. Here came the integration of machine learning techniques as an analytical step for analyzing the vast amount of information produced by the simulation. The main aim of this work is to study the genomic evolution and emerging behaviours arising in EcoSim. Validating the evolutionary machinery in EcoSim by studying the evolution of possible new genes and behaviours was the first main focus. The validation step was achieved through examining how the new possible genes are capable of predicting the fitness of prey individuals and through a comparison between the genomic evolution of EcoSim and its neutral model. EcoSim is a platform that allowed the study of different ecological theories such as the relationship between species genetic diversity and fitness. The success to map this study in EcoSim acted as a validation step of the model and also a contribution to gain more insight about the correlation between genetic diversity and fitness.

The complexity of behavioural interactions in predator-prey systems has recently begun to capture trait-effects, or non-lethal effects, of predators on prey via induced behavioural changes. Non-lethal predation effects play crucial roles in shaping population and community dynamics, particularly by inducing changes to foraging, movement, and reproductive behaviours of prey. Prey exhibit tradeoffs in behaviours while minimizing predation risk (see Chapter 7 for details). EcoSim allows complex intra- and inter-specific interactions between individual evolving behavioural models called prey, as well as complex predator-prey dynamics and coevolution in a tri-trophic and spatially heterogeneous world. Another part of the study involved investigating the effects of predation risk on prey energy allocation and fitness. The semantics of the system allows comprehensive analysis of new genes and behaviours as they arise through a Darwinian evolutionary process. We asked the following questions about prey in EcoSim: are there trait-mediated effects on prey in the form of predation risk-foraging tradeoffs? How does the tradeoff affect the energy of prey and allocation of energy to reproduction? How do their reproductive strategies change in response to predation? What are the effects of predation risk on the prey population? Does the predation pressure affect prey's genomic evolution? How does prey evolve in the absence of predation pressure? What is the effect of introducing predators to nave prey?

1.3 Contributions of the thesis

- First, the evolutionary machinery in EcoSim was studied by examining the emergence of new genes and their effect on fitness. Random Forest was used to build a classifier that was able to predict the values of fitness based on the values of new developed genes. This is considered to be a validation step to ensure the validity of the behavior model and its ability to cope with changes in the environment. A feature selection step is then presented along with rule learning. These rules allow us to discover the most important features that increase fitness, and help us to understand the semantics of the behavior model.

- It has been shown how genetic evolution and diversity governs the adaptation process. We study how EcoSim's individuals adapt to their changing environment by comparing their behavior with a neutral model - a partially randomized version of EcoSim.

- Shannon entropy, which is a measure of unpredictability and disorder coming from Information theory, is used as a measure of genetic diversity. We present the difference in entropy between EcoSim and the neutral model to emphasize the adaptive characteristics of EcoSim. Furthermore, we investigate the relationship between genetic diversity and species fitness and present the correlations found between these two measures in EcoSim. Very high correlation both negative and positive between entropy and fitness was detected. In order to validate the correlation results and further understand the reasons behind these results machine learning classifiers were used to predict the correlation class variable based on training and testing sets. High accuracy for classification was seen which proves the interest of the used genetic diversity measure and its correlation with fitness. In addition, feature selection step was used to find the best features affecting the correlation values. These extracted features such as population size and spatial dispersal are similar to the factors affecting the relation between genetic diversity and fitness in community ecology. Rules were extracted and further investigated which adds more semantics to the reasons behind correlation between genetic diversity and fitness.

- EcoSim models predators and prey with a great deal of detail to their characteristics and interactions. In this study the effect of predator removal on prey's behaviour (foraging, movement and reproduction), genetic change and their capability to coevolve when predators are reintroduced in EcoSim is investigated. In addition, prey are allowed to evolve along two distinct evolutionary paths in the simulation, by either coevolving with predators or evolving in their absence. Results revealed that prey energy budgets, life history traits, allocation of energy to movements and fitness-related actions differed greatly between prey subjected to low-predation risk versus high-predation risk. High-predation risk suppressed prey foraging activity, increased movement, and decreased reproduction relative to low-risk. We used a classification algorithm to show that distinct genomes, corresponding to distinct behavioural adaptations in these prey populations, had evolved after long periods of time. As observed in many empirical studies, we found that prey alter their behaviour according to the level of predation risk. In particular, prey reduce their foraging effort when predation risk is high and instead invest more resources in antipredator behaviours. In addition, these risk-aversive behaviours negatively influence prey fitness as they reduce energy that can be allocated to reproduction. The introduction of predators to naive prey was also studied and monitored predator-prey dynamics, along with the stability of the system after this change. We show that the prey that are left to evolve for a long time without predators developed survival strategies and adaptive behaviors that were coded in their genomes, and this caused instability in the system when predators were later introduced.

1.4 Outline of thesis

- Chapter 2 reviews existing literature on evolutionary systems and the use of IBM in ecology, with a particular focus on ALife evolutionary simulations.
- Chapter 3 gives an overview of, the model used in this study, EcoSim which is a predator-prey ecosystem simulation that is capable of exhibiting long-term adaptive evolution of agent behaviour.
- Chapter 4 presents the background of the data analysis approaches used in the rest of the thesis including machine learning techniques, the deployment of entropy as a measure of genetic diversity and fitness calculation.
- Chapter 5 presents the details of studying the emergent of new genes in EcoSim and its effect on average species fitness. The comparison between genetic diversity in EcoSim and its neutral model, which is affected by continuous adaptation of the individuals to their dynamic environment, is also reviewed.
- The study of genetic diversity and its correlation with fitness is presented in Chapter 6 along with the classification and rule extraction step used to add more semantics to the correlation results.

- Chapter 7 presents the study of predators' effect on the evolution of prey's behavior and genomic structure.

Chapter 2

Background and Literature Review

2.1 Artificial life

Artificial Life, or ALife, is the research field that tries to describe and study natural life by creating artificial systems that possess some of the properties of life. The notion Artificial Life was first presented by Langton who described it by "understanding life by attempting to abstract the fundamental dynamical principles underlying biological phenomena, and recreating these dynamics in other physical media, such as computers, making them accessible to new kinds of experimental manipulation and testing." [3]. Latter on Bedaue noted Artificial Life is concerned with the study of the processes and mechanisms underlying life by recreating life-like phenomena in software, hardware, and biochemicals [2]. There are three methods to model ALife; 'soft' that uses software simulations, 'hard' which involves hardware implementation mainly in robotics and 'wet' which involves biochemistry. The first known formal model was designed by John von Neumann creating a self-reproducing, computational universal cellular automata [5]. He formalized the idea of cellular automata in order to create a theoretical model for a self-reproducing machine. He was mainly concerned with studying the evolution of complex adaptive structures motivated by the understanding of biological evolution and self-reproduction. The early use of cybernetics in information theory aiming to understand living systems was done by Wiener [6]. Due to the increase use of computer simulations, Alife has overlapped and associated with other areas in artificial intelligence such as computational intelligence, which is a nature-inspired computational methodology that addresses real world optimization problems. Although computational intelligence and Alife overlap in many aspects and share similar methodology, there is a difference in their modeling strategies. Alife is mainly concerned with gaining knowledge about living systems using computational bottom-up complex systems. On the other hand, computational intelligence is motivated by the inverse, mainly using the knowledge about living system to construct a top down centralized complex system. Whereas, computational intelligence research is essentially analytic, breaking down complex systems into basic components, ALife synthetic approach attempts to construct complex systems from elemental units. The synthetic approach is based on two concepts, emergence and adaptation.

Complex adaptive systems exhibit emergence where the behavior of the whole is more complex than the behavior of the parts [7]. Emergence is one of the characteristics of a complex system where new and coherent structures, patterns in a complex system are derived due to interactions between the elements of the system over time [8]. The characteristics of emergence were provided by Holland[7]. (a) Emergence happens in systems which compose of different interactive units that follow simple rules. (b) The interactions between the parts are nonlinear so the overall behavior cannot be predicted by summing the behaviors of the isolated parts. (c) The system functions change with the change of context which makes it difficult to predict emergent behavior. (d)The general trend of system complexity increases with increasing number of interactions. [9] defines emergence as "the origin of qualitatively new structures and functions which were not reducible to those already in exist". He classified the emergent phenomena into three different classes; computational emergence which is derived from the cellular automaton example and the mathematical theory of chaos, thermodynamic emergence which is the physicists way to emergent phenomena, and emergence relative to a model which deals with situations where observers need to change their model in order to keep up with a system's behaviour. Evolutionary emergence falls into the emergence relative to a model category. Evolutionary emergence is an essential feature in Alife as [10] noted "The essential features of computer-based Artificial Life models are: . . . There are no rules in the system that dictates global behavior. Any behavior at levels higher than the individual programs is therefore emergent. There are two types of selection that might bring such emergence. [11] referred to these as "extrinsic adaptation where evolution is governed by a specified fitness function, and intrinsic adaptation, where evolution occurs automatically as a result of dynamics of a system cause by the evolution of many interacting subsystems". When aiming to model a more open-ended evolution Alife system, intrinsic adaptation should be diploid.

Artificial evolving systems with pre-defined fitness functions, or fitness landscapes, have been well studied. GAs are biologically inspired search procedures initially developed by Holland [12] [13] [14] in the early 1960s. GAs evolve an initial random population of genomes (codings for solutions to the problem in hand) by selecting which individuals reproduced and which will be replaced. This is done by evaluating each solution's fitness function relevant to the problem and favouring the fitter solutions. A basic shortcoming of genetic algorithms and evolutionary algorithms in general is their tendency to converge. They treat evolution as an optimizer, as they reach local or global optima and eventually converge towards them. When the goal becomes building a system where autonomous agents are able to evolve and adapt in a more open-ended evolutionary dynamics, specifying in advance all the possible behaviours by optimizing an objective function is not desirable. When targeting unbounded evolution and emergence of new adaptive behavior, evolutionary algorithms (using extrinsic adaptation) should be rejected and rather a model based on natural selection (intrinsic adaptation) is more suitable. Furthermore, most existing theoretical modeling approaches rely on the genetic algorithms (GAs) model concept. These systems are optimization processes, meaning that the fate of the system is directly determined by its pre-defined fitness function with the convergence behavior

2.2 IBM

Soft Alife uses individual-based modeling (IBM) which is a bottom-up approach to simulating the interactions among individuals or groups of individuals in an attempt to create complex phenomena. IBM differs from classical equation based models (EBMs) which are typically built up from set of interrelated differential equations. Unlike EBMs, IBM consists of interacting adaptive entities which are able to capture emergent behavior and provide a greater level of useful details. The ease of modeling renders IBM as being more flexible than EBM. IBM has been used on non-computing related scientific domains such as ecological sciences (surveyed by [15]) and social sciences (surveyed by [16]).

The benefits of IBM over other modeling techniques can be captured in several points [8]: (i) agent-based models are a natural way to describe systems comprised of interacting entities; (ii) agent-based models are flexible; (iii) agent-based models capture emergent phenomena; and (iv) agent-based models provide access to a greater level of useful detail. In particular, modeling interactions between entities can be much easier in agent-based systems than in EBMs, even when one is comfortable with the concepts of partial differential equations. It is usually easy to increase the size of a simulation, adding new agents to see if interesting effects are swamped by agent numbers, or taking agents away if interesting detail is obscured. It is also possible to look at the results of simulations at different levels of detail at the level of a single agent, at the level of some specific group of agents, or at the level of all agents together. All these things are harder to manage in EBMs. In addition to their inherent naturalness and flexibility, agent-based simulations allow one to identify emergent phenomena. Emergent phenomena result from the actions and interactions of individual agents, but are not directly controlled by the individuals.

For the past decade there has been an enormous growth of use of IBM addressing different questions in ecology and evolutionary biology. Whereas classical approaches to modeling ecology often ignore individual behaviour and instead uses state- variable model that controls birth and death rates, IBM aim to "treat individuals as unique and discrete entities" (Grimm, 1999)[17] which provides for a more realistic simulation. The use of IBM in ecology and evolution has been reviewed by Grimm in 1999 and Lomnicki 1999 [18]. DeAngelis and Mooij presented another review study which focused on how the IBM field developed [15]. DeAngelis categorized the different directions along which to study individual variation in IBM into five different directions "(a) spatial variability, local interactions and movement; (b) life cycle and ontogenetic development; (c) phenotypic variability, plasticity and behavior; (d) differences in experience and learning; and (e) genetic variability and evolution." He also grouped the IBM systems into seven major study groups; movement through space, formation of patterns among individuals, foraging and population dynamics, species interactions, local competition and community dynamics, evolutionary processes, management related processes. A book by Grimm and Railsback (2005) [19] provides a set of guidelines for building, testing, and analyzing individual-based models, updated in [20]. IBM has been used in many areas in ecology including forest ecology (e.g. [21]), fisheries and marine life (e.g. [22]), conservation biology and spatial heterogeneity (e.g. [23]). Many ecological IBM systems were not designed to be general platforms that could capture different aspects in ecology and evolution but rather these models answer specific question in their narrow domain. More group of evolutionary IBMs that were designed as platforms studying evolutionary behavior, emergence, adaptation and complexity are mention below.

2.2.1 Tierra

While evolutionary computation has been studied since the 1960's, the subfield of digital evolution is much younger. The first experiments with populations of self replicating computer programs were performed in 1990 in a system called Coreworld [24], and later improved upon in Tierra [25]. Tom Ray's Tierra model is the first widely known digital evolutionary ecosystem consisting of self-replicating computer programs based on natural selection. Competition in Tierra results from finite CPU-time and memory space. Tierra is based on a virtual operating system, complete with its own, relatively robust and simple (but universal) machine language and a fixed size address space. An evolutionary run starts by seeding the empty memory space with a hand-written self-replicator program. This replicator then produces a copy of itself which is instantiated as an independent process. A small amount of stochastic behaviour is implemented for program execution, the copy process, and programs are also subject to point mutations. These mechanisms are responsible for introducing variety into the populations. If the modified programs retain their ability to replicate, and the modifications alter their probability of reproduction, Darwinian evolution can occur. A number of interesting results have been obtained from such evolutionary runs. For example, 'parasites' have appeared-short pieces of code which run another program's copying procedure in order to copy themselves. Hyper-parasites (parasites of parasites) have also been observed, along with a number of other interesting ecological phenomena. Ray demonstrated that it is possible to build an operating system in which self-replicating computer code can evolve. On the other hand, after a certain amount of time, Tierra fails to produce any new programs but only change in the number of existing ones.

Rey sparked a number of follow-up systems based on Tierra. Cosmos a Tierra-like system configured in a 2 dimensional toroidal like grid environment was used to study the role of contingency in evolution [26]. Furthermore, in Amoeba [27] the language of the digital organism along with its self replicating code is also subject to evolution. The Amoeba system, developed by Pargellis, showed the possibility of spontaneous emergence of a self-replicating program.

2.2.2 Avida

Development of AVIDA (a Tierra like system) [28] [29], in which self-replicating digital organisms consists of a circular list of instructions (its genome) and a virtual CPU evolve. In Avida, each organism lives in its own address space, unlike Tierra's shared address space. This enhancement increased the power of digital evolution as an experimental tool. AVIDA environment comprises a number of cells, each cell can contain at most one organism, and the size of an AVIDA population is bounded by the number of cells in the environment. Organisms are self-replicating, that is, the genome itself must contain the instruction to create an offspring. When an organism replicates, a cell to contain the offspring is selected from the environment and its inhabitant organism is replaced (killed and overwritten). Since digital organisms are self-replicating and compete for space, a higher merit (all else being equal) results in an organism that replicates more frequently, spreading throughout and eventually dominating the population. Hence, AVIDA satisfies the three conditions necessary for evolution to occur: replication, variation (mutation), and differential fitness (competition). Individuals in Avida do not move and in order to measure the complexity they use a fixed environment which is rarely seen in nature. This means that the system is only adapting to a preexisting environmental complexity. The processes derived from Avida and Tierra are optimization processes, similar to evolutionary algorithms, for which it has been proved that it converge toward a maximum, either local or global. Finally, as with Tierra, the complexity growth in Avida always reaches an upper bound and stops. These results with Avida do not capture the kind of continual growth in qualitative complexity or long term incremental evolution that we can observe in the biosphere. Hence, Avida and Tierra do not represent an open-ended evolution which has been defined in various ways. Nehaniv, for example, defines open-ended evolution as an unbounded increase in complexity [30]. In [31] Taylor noted about Tierra and Tierra like systems "most of these systems are only capable of producing innovations of the 'more-of-the-same' variety (e.g. more optimized code), rather than anything fundamentally new."

Avida was used to study numerous aspects of evolution [32]; issues of complexity in evolution [33] [34]. Furthermore, they investigated the emergence of complex behavior [35]. They showed that complex features do not appear suddenly but only evolve when simpler traits exist which served as a foundation upon which these complex features were built. In a recent study they showed how runaway sexual selection leads to good genes and how they should be viewed as interacting mechanisms that reinforce one another [36]. Evolving digital ecological networks was presented in [32] which models competition, parasitism and mutualism.

2.2.3 Echo

A significant advance in evolutionary IBM was presented by J. Holland in Echo [37], a generic ecosystem in which agents evolve in a resource limited environment. The world is made up of a square toroid lattice of sites which has different kinds of evolving resources encoded by a letter. Agents interact with their environment and are able to move from one site to another. They gain energy by eating and spend it on their actions such as fighting, trading and mating. Reproduction in Echo happens when an agent has replicated itself with a possible mutation when it has gained enough resources to copy its genome asexually or by sexual mating. Selection is based on the interacting agents rather than by a predefined fitness function. Emerging phenomena arise such as formation of communities and trading networks. Echo was used to study the modeling of food web complexity [38]. Echo was intended to be a general model of intrinsic adaptive system rather than modeling and answering specific questions in evolutionary biology. Due to the high abstraction level of the Echo model, the degree of fidelity to real systems is uncertain.

2.2.4 Polyworld

In PolyWorld [39], more advanced haploid agents, each controlled by an artificial neural network, with a set of primitive behaviors and learning strategies, populate a continuous environment containing number of energy sources ('food') upon which they rely on for survivor. Possible actions for agents include eat, mate, fight, move, focus and light (for vision). Agents evolve under the influence of natural selection and die when their energy is fully depleted or loose fight with another agent. An agent's genome specifies characteristics of its physiology and neural architecture which is adapted during its life through Hebbian learning. Yaeger was able to report the emergence of new population behavior such as fleeing, grazing, following and flocking. Polyworld was used to study how evolution guides complexity [40] and the passive and driven trends in the evolution of complexity [41]. Genetic clustering for the identification of species was presented was also presented in [42]. On the other hand, lack of semantics in the genomic structure (nodes) in

Polyworld, makes it difficult to reason and link together different aspects of the model. Another criticism of PolyWorld, in the context of perpetual evolutionary emergence, is that learning appears to be overwhelmingly responsible for the results. This integrated learning process adds to the computational complexity of the model. Furthermore, the high complexity of the neural networks agents limits their number making it difficult to study large ecosystem phenomena's. Geb [43] [44] is another similar artificial neural network system considered to be simpler than Polyworld as it is not trying to mimic the real world as Polyworld do. Agents which are controlled by a neural network each populate a gridded arena and compete for space with no notion of energy. There is no learning process as agents do not change during their lifetime and thus results prove it to be suited to long-term incremental artificial evolution. Geb was proven to be the first autonomous artificial system to pass the Bedau and Packard's evolutionary test [45]. According to Bedau statistics, evolutionary dynamics in Gep was proven to be unbounded [46] and thus based on intrinsic evlution. Bedau et al [45] developed a statistical measure for testing unbounded evolution.

2.2.5 Framsticks

Framsticks presented by Komosinski et al in 1999 [47] is a 3D life simulation platform addressing both research and education. The platform consists of modules that facilitates the design of various experiments in optimization, coevolution, open-ended evolution and ecosystem modeling. Agents have both mechanical structure (bodies) consisting of connected sticks and control system (brain) using artificial neural network. The neural network brain collects data from sensors and sends signals to the joints which control motion activities. The world is enriched with complex topology and a water level along with energy balls consumed by agents. Although some locomotion behaviours have evolved, the high complexity of the model did not present any different results than those obtained from much simpler evolutionary systems. This model is more concerned with the study of emerging motor behavior rather than modeling a multiple level interacting ecosystem.

2.3 Other predator-prey simulations

Some of the above mentioned systems like Polyworld and Echo model predator individuals. Other predator-prey models have also been presented focusing more on the ecological predator prey dynamics and interactions. Smith (1991) [48] uses Volterra [49]model which exhibits constant population dynamics, both in terms of oscillations in global populations as well as dynamic patchiness. The model integrated 2D spatial representation to study migration under different predation strategies. He showed that detailed movement patters in predator and prey can affect their interaction. Smith only models simple predator prey behaviour with simple genomic representation as only migration parameters are able to mutate. In [50] digital predator-prey organisms were used to study the evolution of trophic structure represented by the food web. Bell showed how different energy flow levels among organisms affect species richness and diversity. In another study [51] Lotka-Volterra equations were integrated in an IBM to examine how evolution of prey use by predators affects community stability and whether complexity of food web increases stability of the predator prey system. The results demonstrated that number of existing species decreases with the increasing complexity.

A predator-prey simulation based in a spatial collection of individual finite state machine animat agents was first presented in [52]. This model can locate hundreds of thousands of individuals evolving in a two-dimensional featureless spatial plain. Every animat carries a small set of rules that direct its microscopic behaviour and at each time-step of the simulation, each animat executes one of these rules, causing it to: move; eat; or breed. In one study the effect of introducing camouflage behaviour as an available option for predators was investigated([53]). It was shown that individuals who adopt this behaviour are relatively successful in obtaining prey and thus prolonging their lives against threat of dying of hunger. This in turn led to higher numbers of successful older predators which caused a crash in the population of prey.

In another study a time-delayed gestation period was introduced into the predator-prey selection and adaptation mechanisms ([54]). The temporal behaviour of individual animats was affected by the gestation period parameter and hence the macroscopic behaviours of the species was also affected.

Chapter 3

Individual based ecosystem simulation EcoSim

3.1 Purpose

EcoSim is an individual-based ecosystem simulation, designed by [55] to simulate agents' behavior in a dynamic, evolving ecosystem. The following description of EcoSim was previously published as a supplementary material in [56]. The agents (or individuals) of EcoSim are prey and predators acting in a simulated environment. The main purpose of EcoSim is to study ecological and evolutionary theories by constructing a complex, adaptive, and generic virtual ecosystem with behaviours and processes resembling real ecosystems. Due to the complexity of natural ecosystems and the time and funding resources required to study such theories, modeling has become crucial to the study of ecology and evolution. EcoSim is the first ecological model to use a fuzzy cognitive map (FCM) [57] to model each agent's behavior. The FCM of each agent, being coded in its genome, allows the evolution of the agent behavior through the epochs of the simulation. EcoSim as a virtual ecosystem has shown coherent behaviors of the whole simulation with the emergence of patterns observed in real ecosystems providing a general framework for the study of several specific ecological problems.

Characteristic	Predator	Prey
Max age	42 time steps (±6)	46 time steps (±18)
Min age of reproduction	8 time steps	6 time steps
Max speed	11 cells / time step	6 cells / time step
Vision distance	25 cells max	20 cells max
Level of energy at init.	1000 units	650 units
Avg. speed	1.4 cells / time step (±0.3)	1.2 cells / time step (±0.2)
Average level of energy	415 units(±82)	350 units(±57)
Avg. number of reproduction action during life	1.14 (±0.11)	1.49(±0.17)
Avg. length of life	16 time steps (±5)	12 time steps (±3)

Table 3.1: Several physical and life history characteristics of individuals from 10 independent runs.

3.2 Entities, state variables, and scales

Individuals: EcoSim has two types of individuals: predator and prey. Each individual possesses several characteristics (Table 3.1) such as: age, minimum age for breeding, speed, vision distance, level of energy, and amount of energy transmitted to the offspring. Energy is provided to the individuals by the resources (food) they find in their environment. Prey consumes grass, which is dynamic in quantity and location, whereas predator hunts for prey individuals. Each individual performs one unique action during a time step, based on its perception of the environment. Each agent possesses its own FCM that represents its genome and also its behaviors are determined by the interaction between the FCM and the environment. Thus, the FCM allows flexibility in behavioural responses to the changing environment, but since the FCM has fixed values at birth it does not model plasticity in the conventional sense (the FCM is discussed in detail below).

The energy is provided by the primary or secondary resources found in their environment. For example, prey individuals gain 250 units of energy by eating one unit of grass and predators gain 500 units of energy by eating one prey. At each time step, each agent spends energy depending on its action (e.g. breeding, eating, running) and on the complexity of its behavioral model (number of existing edges in its FCM). On average, a movement action such as escape and exploration requires 50 units of energy, a reproduction action uses 110 units of energy and the choice of no action results in a small expenditure of 18 units of energy.

Cells and virtual world: The virtual world is discrete and consists of a matrix of 1000*1000 space units called cells. Each cell represents a large space which may contain an unlimited number of individuals and/or some amount of food. The world is large enough in order to observe migration patterns, an individual moving in the same direction during its whole life cannot even cross half of the world, making large-scale migrations possible. The virtual world wraps around to remove any spatial bias. In addition, the dimensions of the world are adjustable but increasing the dimensions can increase the computation complexity of the simulation by allowing more individuals to co-exist.

Time step: Each time step involves the time needed for each agent to perceive its environment, make a decision, perform its action, as well as the time required to update the species membership, including speciation events and record relevant parameters (e.g. the quantity of available food). In terms of computational time, the speed of simulation per generation is related to the number of individuals. Recent executions of the simulation produced approximately 15 000 time steps in 35 days.

Population and Species: In average in every time step of the simulation, there are 250,000 individuals which consisting of one or more species. A species is a set of individuals with similar genome.

3.3 Process overview and scheduling

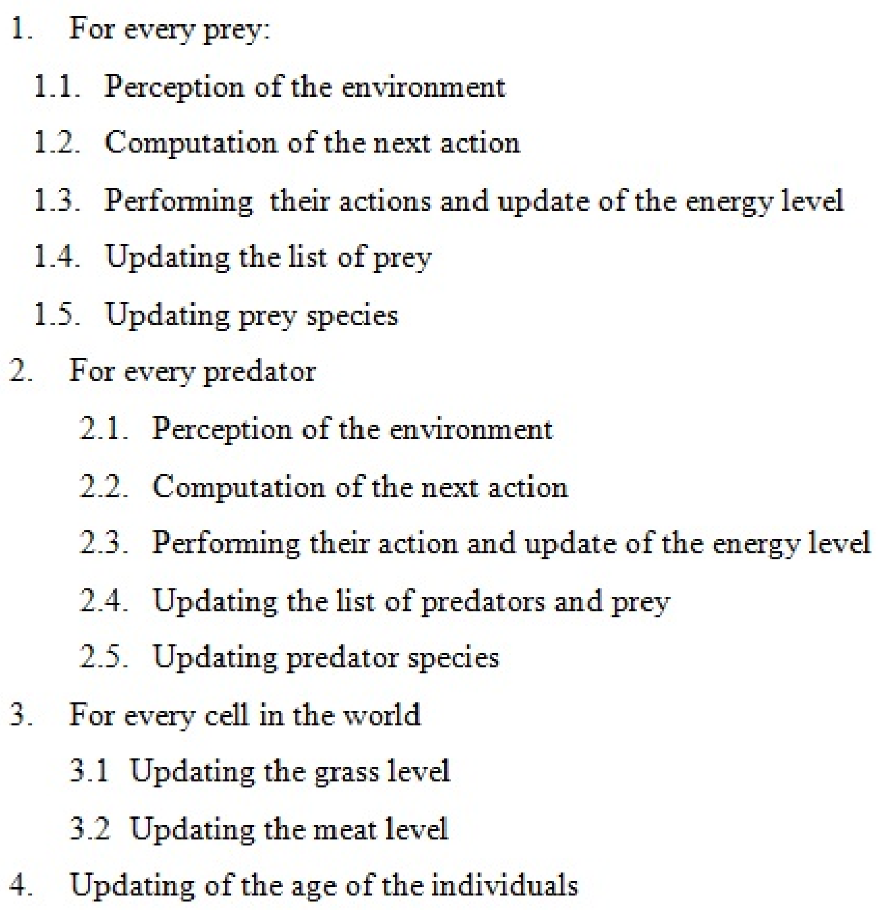
The possible actions for the prey agents are: perceive the environment to obtain information of the vicinity in terms of grass, predators, and sexual partner, evasion (escape from predator), search for food (if there is not enough grass available in the its habitat cell, prey can move to another cell to find grass), socialization (moving to the closest prey in the vicinity), exploration, resting (to save energy), eating and breeding. Predator also perceive the environment to gather information used to choose an action among: hunting (to catch a prey), search for food, socialization, exploration, resting, eating and breeding. For every individual the energy is adjusted after an action is performed at each time step. The age of every individual is also updated at each time step (age is simply the number of time steps until an individual dies). There are also two environmental processes that depend on the actions of prey and predators, the amount of grass which is consumed by prey and meat which is consumed by predators, which are also adjusted at each time step. At each time step, the value of the state variables of individuals and cells are updated. The overview and scheduling of every time step is shown in Fig.3.1.

The complexity of the simulation algorithm is mostly linear in the number of individuals. If we consider that there are N1 preys and N2 predators then the complexity of part 1 and part 2of the above algorithm, including the clustering algorithm used for speciation, will be O(N1) and O(N2) respectively ([58]). This virtual world of the simulation has 1000*1000 cells, therefore the complexity of part 3 will be O(k = 1000*1000). The complexity of part 4 will be O(N1+N2). As a result the overall complexity of the algorithm will be calculated as O(2N1+ 2N2+ k), which is O(N = 2N1 +2N2).

3.4 Design concepts

3.4.1 Basic principles

In EcoSim, a FCM is the base for describing and computing the agent behaviors. Each agent possesses a unique FCM to compute its next action. Their FCM is represented in their genome which is assigned to each individual at birth. A FCM is a directed graph containing nodes representing concepts and edges representing the influence of concepts on each other (Fig.3.2). When a new offspring is created, it is given a genome which is a combination of the genomes of its parents with some



1. For every prey:
    - 1.1. Perception of the environment
    - 1.2. Computation of the next action
    - 1.3. Performing their actions and update of the energy level
    - 1.4. Updating the list of prey
    - 1.5. Updating prey species
2. For every predator
    - 2.1. Perception of the environment
    - 2.2. Computation of the next action
    - 2.3. Performing their action and update of the energy level
    - 2.4. Updating the list of predators and prey
    - 2.5. Updating predator species
3. For every cell in the world
    - 3.1 Updating the grass level
    - 3.2 Updating the meat level
4. Updating of the age of the individuals

Figure 3.1: The overview and scheduling of every time step.

possible mutations. Formally, an FCM is a graph which contains a set of nodes C, each node Ci being a concept, and a set of edges I, each edge Iij representing the influence of the concept Ci on the concept Cj. A positive weight associated with the edge Iij corresponds to an excitation of the concept Cj from the concept Ci, whereas a negative weight is related to an inhibition (a zero value indicates that there is no influence of Ci on Cj). The influence of the concepts in the FCM can be represented in an nn matrix, L, in which Lij is the influence of the concept Ci on the concept Cj. If Lij = 0, there is no edge between Ci and Cj.

3.4.2 Emergence

In each FCM, three kinds of concepts are defined: sensitive (such as distance to foe or food, amount of energy, etc.), internal (fear, hunger, curiosity, satisfaction, etc.), and motor (evasion, socialization, exploration, breeding, etc.). The activation level of a sensitive concept is computed by performing a fuzzification of the information the individual perceives in the environment. For an internal or motor concept C, the activation level is computed by applying the defuzzification function on the weighted sum of the current activation level of all the concepts having an edge directed toward C.

Finally, the action of an individual is selected based on the maximum value of motor concepts' activation level. Activation levels of the motor concepts are used to determine the next action of the individual. For example in Fig.3.3. there are two sensitive concepts (foeClose (predator close) and foeFar (predator far)), one internal (fear), and one motor (evasion). There are also three influence edges: closeness to a foe excites fear, distance to a foe inhibits fear, and fear causes evasion. Activations of the concepts foeClose and foeFar are computed by fuzzification of the real value of the distance to the foe, and the defuzzification of the activation of evasion tells us about the speed of the evasion. The values of edges for each individual are fixed throughout his life, and are combined with another individual with possible mutation when forming a new offspring.

At the initiation of the simulation prey and predators scattered randomly all around the virtual world. Through the epochs of the simulation, distribution of

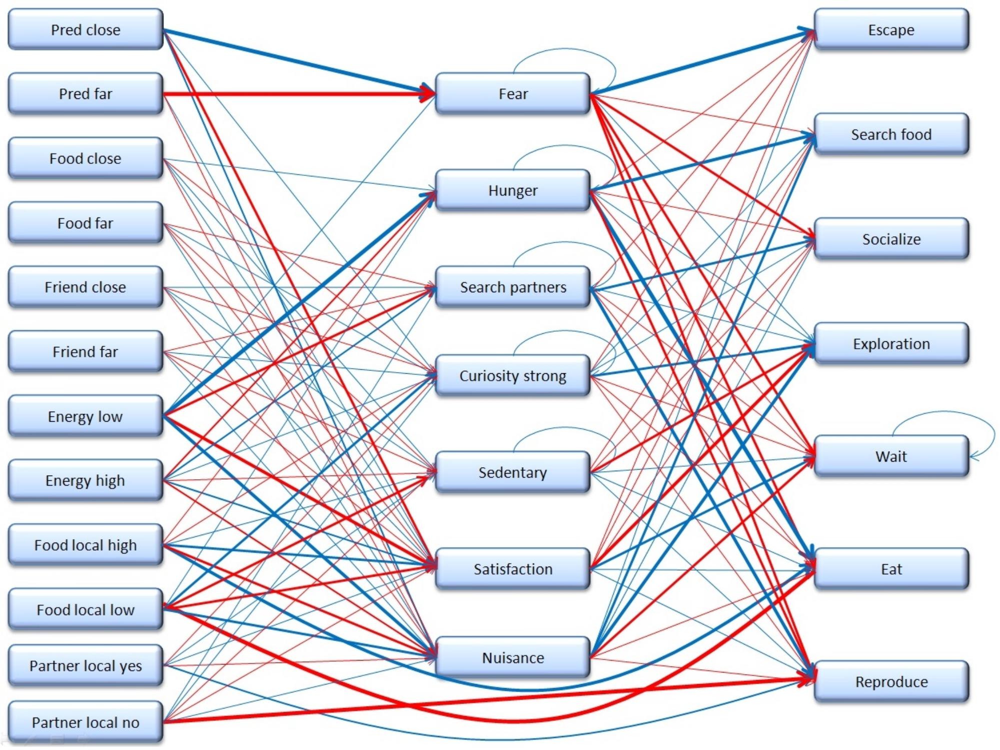


Figure 3.2: Initial FCM prey map including concepts and edges. The width of each edge represents the influence value of a concept on another. Color of an edge shows inhibitory (red) or excitatory (blue) effects.

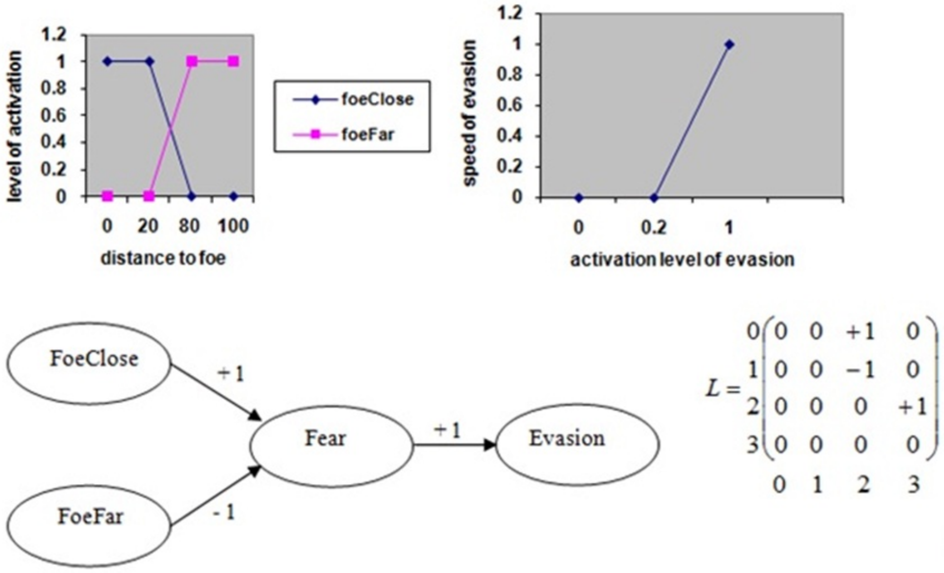


Figure 3.3: An FCM for detection of foe (predator) and decision to evade with its corresponding matrix (0 for 'Foe close', 1 for 'Foe far', 2 for 'Fear' and 3 for 'Evasion') and the fuzzification and defuzzification functions.

the individuals in the world is changed drastically based on many different factors: prey escapes from predators, individuals socialize and form groups, individuals migrate gradually to find sources of food, species emerge, etc. Fig.3.4 and Fig.3.5 show an example of a snapshot of the virtual world after thousands of time steps with emerging grouping patterns of species and grass distribution respectively. It has been shown that the data generated by EcoSim present the same kind of multifractal properties as the ones observed in real ecosystems [59]. Individuals' distribution forming spiral waves is one property of prey-predator models. The prey near the wave break has the capacity to escape from the predators sideways. A subpopulation of prey then finds itself in a region relatively free from predators. In this predator-free zone, prey starts expanding intensively and form a circular expanding region. The whole pressure process and spiral formation will be applied to this subpopulation of prey and predators again leading to the formation of the second level of spiral [60]. Because there are consecutive interactions between prey and predators during time, the same pattern repeats over and over and then self-similarity emerges in spatial distribution of individuals which is a common property of self-similar processes [61]. As can be seen in the figure individuals grouped together, and different species emerged. In addition migration phenomena can be observed, as relocation of the individuals leads to the redistribution in the population.

3.4.3 Adaptation

The genome maximal length is fixed (390 sites), where each site corresponds to an edge between two concepts of the FCM. But, as many edges have an initial value of zero, only 114 edges for prey and 107 edges for predators exist at initialization. One more gene is used to code for the amount of energy which is transmitted for the parent to their child at birth. The value of a site (gene), which is a real number, corresponds to the intensity of the influence between the two concepts. The genome of an individual is transmitted to its offspring after being combined with the one of the other parent and after the possible addition of some mutations. The behavior model of each individual is therefore unique. Step after step as more

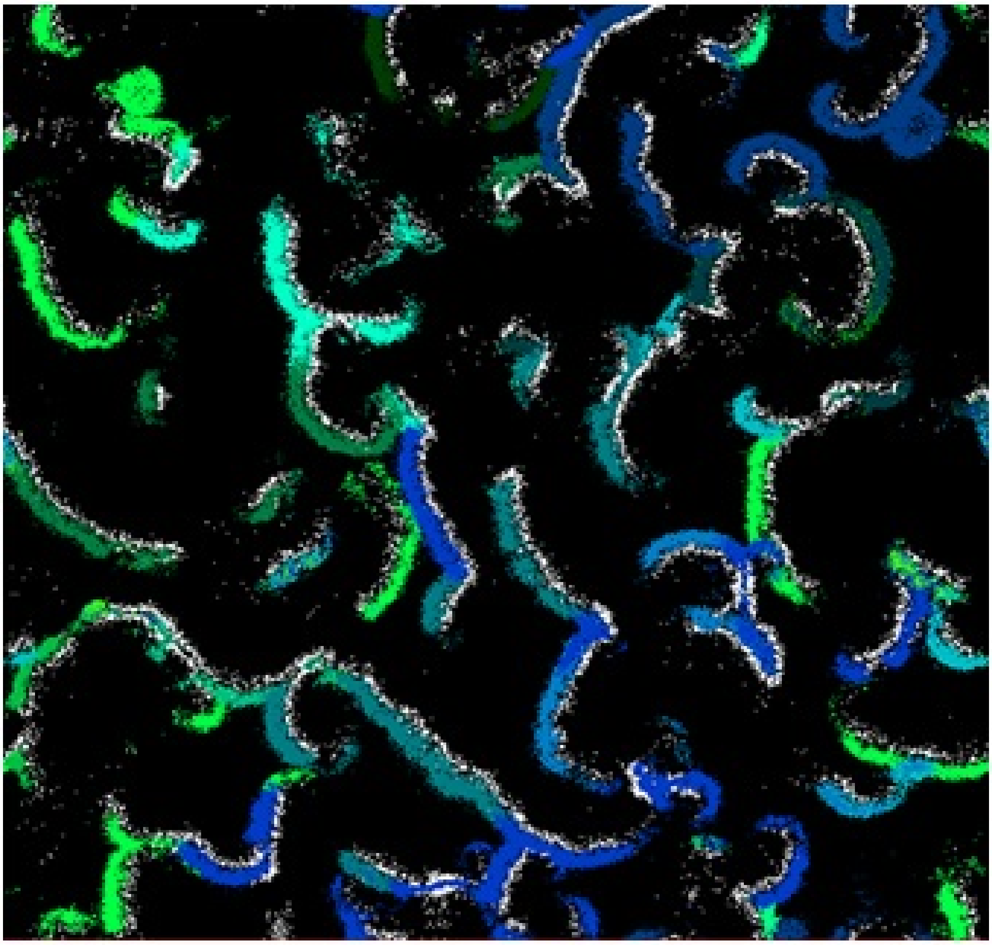


Figure 3.4: The snapshot of the virtual world in one specific time step, white color represents predator species and the other colors show different prey species.

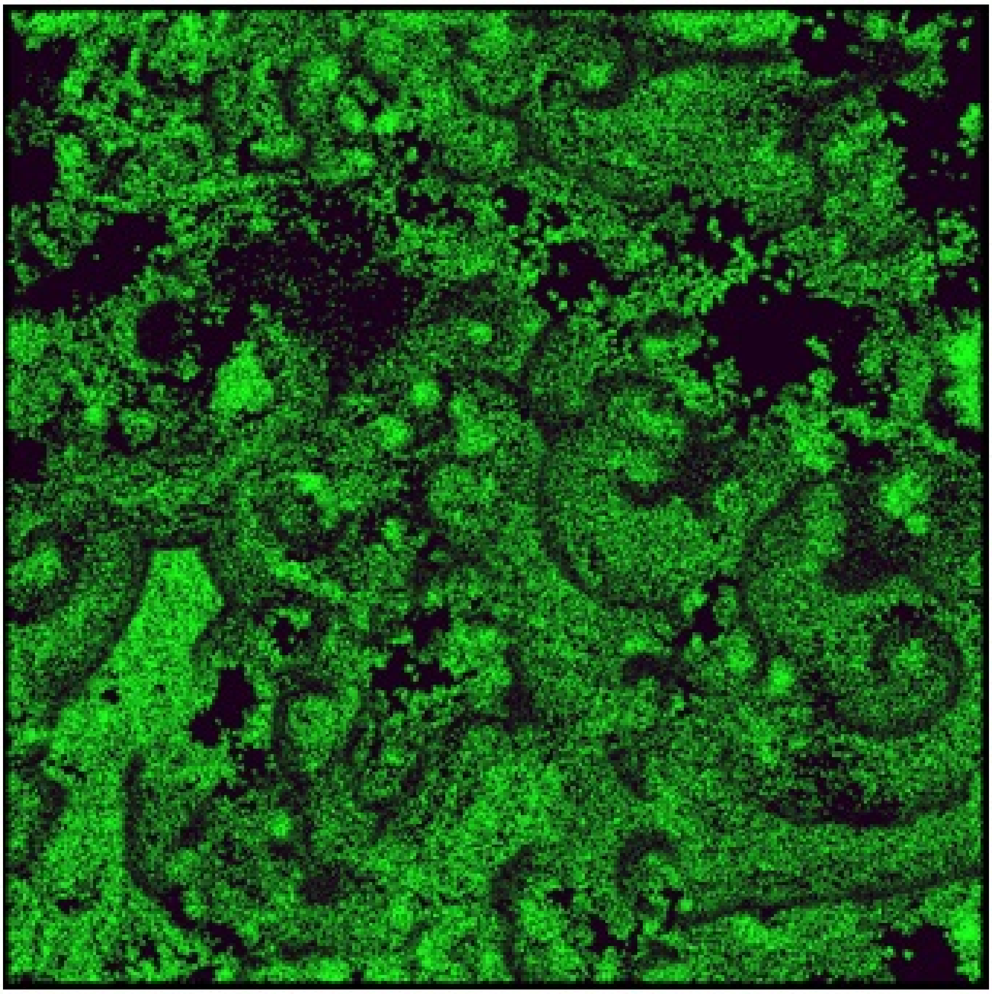


Figure 3.5: The snapshot of the virtual world for a specific time step of the simulation which demonstrates the pattern of grass in the world.

individuals are created, changes in the FCM occur due to the formation of new edges (with probability of 0.001), removal of existing edges (with probability of 0.0005) and changes in the weights associate to existing edges (with probability of 0.005). New genes may emerge from among the 265 initial edges of zero value. This emergence and disappearance of the genes in FCM is due to environmental changes and genetic drift which lead to adaptability of individuals. This emergence and disappearance of genes in the FCM, is accompanied by selective pressures due to environmental changes as well as genetic drift and can lead to adaptations over many generations.

3.4.4 Fitness

We calculate the fitness for every species as the average fitness of its individuals. The fitness of an individual is defined as the age of death of the individual plus the sum of the age of death of its direct offspring. Accordingly, the fitness value mirrors the individual's capability to survive longer and produce high number of strong adaptive offspring. There is no pre-defined explicit fitness-seeking process in the simulation but rather it is a consequence of natural selection. Individuals that are more adapt to the environment live longer, have a higher level of energy, and therefore are able to have more offspring, and can transfer them efficient genomes.

3.4.5 Prediction

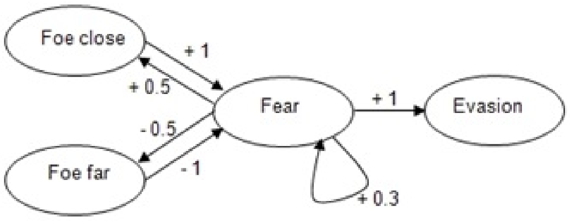
So far, there is no learning mechanism for individuals and they cannot predict the consequences of their decision. The only available information for every individual to make decision is the information coming from their perceptions at the current time step and the value of the activation level of the internal and motor concepts at the previous time steps. The activation levels of the concepts of an individual are never reset during its life. As the previous time step activation level of a concept is involved in the computation of its next activation level, this means that all previous states of an individual during its life participate in the computation of its current state. It means therefore that an individual has a basic memory of its own past that will influence its future states. As the action undertaken by an individual at a given time step depends on the current activation level of its motor concepts, the global behavior of an individual dynamically depends on a complex combination of the information it currently receives from its environment, its current internal states, and the past states it went through during its life.

3.4.6 Sensing

Every individual in EcoSim is able to sense its local environment inside of its vision range. For instance, every prey can sense the five closest foes, cells with food units and mates within the vision range, the number of grass units in its cell, and the number of possible mates in its cell. Moreover, every individual is capable of recognizing its current level of energy. It should be noticed that the FCM process enables for example, to distinguish between perception and sensation: the sensation is the real value coming from the environment, and the perception is the sensation modified by the internal states. For example, it is possible to add three edges to the previous map: one autoexcitatory edge from the concept fear to itself, one excitatory edge from fear to foeClose, and one inhibitory edge from fear to foeFar (Fig.3.6). A given real distance to the foe seems higher or lower to the individual depending on the activation level of fear. Also, the fact that the individual is frightened at time t influences the level of fear of the individual at time t + 1. This kind of mechanism gives the possibility of modeling a degree of paranoia and a degree of stress for the individual. It also allows the individual to memorize information from previous time steps: fear maintains fear. It is therefore possible to build very complex dynamic systems involving feedback and memory using an FCM, which is needed to model complex behaviors and abilities to learn from evolution.

3.4.7 Interaction

The only action that requires a coordinate decision of two individuals is reproduction. For reproduction to be successful, the two parents need to be in the same cell, to have enough energy, to choose the reproduction action and to be



$$L = \begin{pmatrix} 0 & 0 & +1 & 0 \\ 0 & 0 & -1 & 0 \\ +0.5 & -0.5 & +0.3 & +1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure 3.6: FCM for detection of foe (predator) - difference between perception and sensation.

genetically similar. The individuals cannot determine their genetic similarity with their potential partner. They try to mate and if the partner is too dissimilar, that is the dissimilarity between the two genomes is greater than a threshold (half of the speciation threshold), the reproduction fails. Predator's hunting introduces another type of interaction in the simulation. For a predator to succeed in the hunting action, its distance to the closest prey requires to be less than one. When a predator's hunting action succeeds, a new meat unit is added in the corresponding cell and the energy level of the predator is also increased by one unit of meat energy. Furthermore, there is a competition for prey and predators for food. For example, if in a given cell there is only one food unit and two agents have chosen the action of eating, the younger will act first, and so it will be the only one that can eat (in this cell) at this time step. This is a way to simulate the fact that the older help the younger to survive.

3.4.8 Stochasticity

To produce variability in the ecosystem simulation, several processes include stochasticity. For instance, at initialization time the number of grass units is randomly determined for each cell. Moreover, the maximum age of an individual is determined randomly at birth from a uniform distribution centered at a value associated with the type of agent. Stochasticity is also included in several actions of the individuals; in evasion and socialization: if there is no predator or partner respectively in the vision range of the individual, the direction of the movement would be random. Furthermore, the direction of the exploration action is always random.

3.4.9 Collectives

In EcoSim, the notion of species is implemented in a way that species emerge from the evolving population of agents. Species can become extinct if all of their members die. EcoSim implements a species concept directly related to the genotypic cluster definition [62] in which a species is a set of individuals sharing a high level of genomic similarity. In addition, in EcoSim, each species is associated with the average of the genetic characteristics of its members, called the 'species genome' or the 'species center'. The speciation mechanism implemented in EcoSim is based on the gradual divergence of individual genomes. The speciation method begins by finding the individual in a species S with the greatest distance from the species center. If this distance is greater than a predefined threshold for speciation (which is two time greater than the threshold for reproduction), a 2-means clustering is performed [58]. Otherwise, species S remains unchanged. If clustering is to be performed, two new species are created - one centered on a random individual, denoted Ir, and another centered on the individual which is the most genetically different from Ir. Subsequently, all remaining individuals in S are added to one of the two new sister species - whichever species the individual is more genetically similar. After recalculating the new centers for the two new species, the process of clustering is repeated for convergence.

3.4.10 Observation

EcoSim produces a large amount of data in every time step, including number of individuals, new and extinct species, geographical and internal characteristics of every individual, and status of the cells of the virtual world. Information regarding each individual includes position, level of energy, choice of action, specie, parents, FCM, etc. Information about the individuals and species are stored in one file with an average size of 30MB, and information for the virtual world is stored in another file with an average size of 5MB. Also there is a possibility to store all of the values of every variable in the current state of the simulation in a separate file, giving the possibility to restore the simulation from that state afterwards. The overall size of this file, which is only stored once in a while during a run of a simulation, is a few hundred MBs depending on the size of population and species. All the data is stored in a compact special format, to facilitate the storage and future analysis. There is a program which can be used to extract all the data. This program reads one file at a time and extracts all the required variables with a linear complexity for different analysis.

3.5 Initialization and input data

A parameter file is defined for EcoSim which is used to assign the values for each state variable at initial time of the simulation. These parameters are as follows: width and height of the world, initial numbers of individuals, threshold of genetic distance for prey/predator speciation, maximum age, energy, speed, vision range, and initial values of FCM for prey/predator. Any of these parameters can be changes for specific experiments and scenarios. An example of a list of most common user specified parameters for initially running the EcoSim are presented in Table 3.2.

3.6 Submodels

As mentioned earlier, each individual performs one unique action during a time step based on its perception of the environment. EcoSim iterates continuously, and each time step consists of the computation of the activation level of the concepts, the choice and application of an action for every individual. A time step also includes the update of the world: emergence and extinction of species and growth and diffusion of grass, or decay of meat. At initialization time there is no meat (dead prey) in the world and the number of grass units is randomly determined for each cell. For each cell, there is a probability, probaGrass, that the initial number of units is strictly greater than 0. In this case, the initial number is generated uniformly between 1 and maxGrass. Each unit provides a fixed amount of energy to the agent that eats it. Grass also grows in the system over time, but can become

User specified parameters	Used value
Number of Prey	12000
Number of Predators	500
Grass quantity	5790000
Prey max. age	46
Predator max. age	42
Prey max. speed	6
Predator max. speed	11
Prey energy	650
Predator energy	1000
Distance for prey vision	20
Distance for predator vision	25
Reproduction age for prey	6
Reproduction age for predator	8

Table 3.2: Values for user specified parameters.

depleted. The growth of grass in any cell depends on whether neighbouring cells have grass in them: if there is no grass in a particular cell or in any of its 8 neighbouring cells, then no grass will grow and there will be a locally empty food patch. The preys can only eat the grass, and the predators have two modes of predation: hunting and scavenging. When a predators hunting action succeeds, the killed prey is counted as 2 meat units, one consumed by the predator and one added in the corresponding cell which may be consumed by anther predator. When a predators eating action succeeds (which can be viewed as a scavenging action), one unit of meat is removed in the corresponding cell. The amount of energy is energyGrass for one grass unit when eaten by a prey and is energyMeat for one meat unit eaten by a predator. The number of grass units grows at each time step, and when a prey dies in a cell, the number of meat units in this cell increases by 2. The number of grass units in a cell decreases by 1 when a prey eats, and the number of meat units decreases by 1 when a predator eats. The number of meat units in a cell also decreases at each time step, even if no meat has been eaten in this cell.

1. Evasion (for prey only). The evasion direction is the direction opposite to the direction of the barycenter of the 5 closets foes within the vision range of the prey, with respect to the current position of the prey. If no predator is within the vision range of the prey, the direction is chosen randomly. Then the new position of the prey is computed using the speed of the prey and the direction. The current activation level of fear is divided by 2.

2. Hunting (for Predator only). The predator selects the closest cell (including its current cell) that contains at least one prey and moves towards that cell. If it reaches the corresponding cell based on its speed, the predator kills the prey, eating one unit of food and having another unit of food added to the cell. When there are several preys in the destination cell, one of them is chosen randomly. If the speed of the predator is not enough to reach the prey, it moves at its speed toward this prey. If there is no prey in the current cell and in the vicinity or it does not have enough energy to reach to a prey, hunting action is failed.

3. Search for food. The direction toward the closest food (grass or meat) within the vision range is computed. If the speed of the agent is high enough to reach the food, the agent is placed on the cell containing this food. Otherwise, the agent moves at its speed toward this food.

4. Socialization. The direction toward the closest possible mate within the vision range is computed. If the speed of the agent is high enough to reach the mate, the agent is placed on the cell containing this mate, and the current activation level of sexualNeeds is divided by 3. Otherwise, the agent moves at its speed toward this mate. If no possible mate is within the vision range of the agent, the direction is chosen randomly.

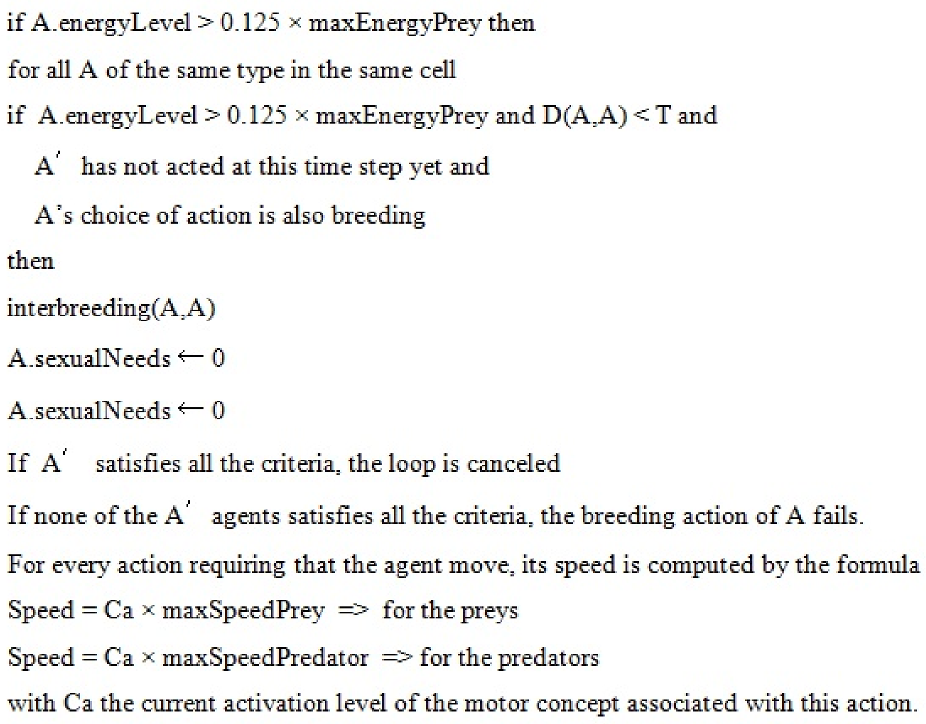
5. Exploration. The direction is computed randomly. The agent moves at its speed in this direction. The activation level of curiosity is divided by 1.5.

6. Resting. Nothing happens.

7. Eating. If the current number of grass (of meat) units is greater than 1, then this number is decreased by 1 and the preys (predators) energy level

is increased by energyGrass (energyMeat). Its activation level for hunger is divided by 4. Otherwise nothing happens.

8. Breeding. The algorithm in Fig.3.7 is applied to the agent A



```
if A.energyLevel > 0.125 × maxEnergyPrey then
for all A of the same type in the same cell
if  A.energyLevel > 0.125 × maxEnergyPrey and D(A,A) < T and
   $A'$  has not acted at this time step yet and
   A’s choice of action is also breeding
then
interbreeding(A,A)
A.sexualNeeds ← 0
A.sexualNeeds ← 0
If  $A'$  satisfies all the criteria, the loop is canceled
If none of the $A'$  agents satisfies all the criteria, the breeding action of A fails.
For every action requiring that the agent move, its speed is computed by the formula
Speed = Ca × maxSpeedPrey  =>  for the preys
Speed = Ca × maxSpeedPredator  => for the predators
with Ca the current activation level of the motor concept associated with this action.
```

Figure 3.7: Breeding algorithm.

The process of generating a new offspring (interbreeding function) consists of following steps. First, the value of birthEnergyPrey is transmitted with possible mutations from one randomly chosen parent to the offspring. Second, the edges' values are transmitted with possible mutations, and the initial energy of the offspring is computed. To model the crossover mechanism, the edges are transmitted by block from one parent to the offspring. For each concept, its incident edges' values are transmitted together from the same randomly chosen parent. Third, the maximum age of the offspring is computed. Finally, the energy level of the two parents is updated.

3.7 The Neutral Model

In order to study the genomic evolution in EcoSim, a neutral shadow of EcoSim is built. M. Bedau [45] presented the use of building a natural shadow of any evolutionary model to point out the difference in evolutionary behaviour between it and the original model. All selection processes and behaviors in the neutral shadow for the predator/prey are random, which eliminates natural selection from this model. In terms of the behavioral model of this version, all the actions such as eating, hunting (for predators), socializing, searching for food and escaping (for prey) are removed. The only two actions any individual can take are reproduction and movement. Unlike in the EcoSim, in the neutral model there is no necessity for the individuals to have genetic similarity to reproduce. Instead, in the neutral model the reproduction action is done by randomly choosing any two individuals in the world. The statistics of genetic operations (mutation rates and crossover) are the same as EcoSim. In EcoSim, individuals choose to reproduce according to their internal state, suitable environmental conditions and behavior model but not in the neutral model. To preserve population dynamics in neutral model similar to that of EcoSim, the Lotka-Volterra computational model [49] is used. This model controls the number of births and deaths at each time step. In addition, death of individuals and pairs of parents for reproduction, are randomly selected. In this way a similarity in population sizes between the neutral shadow and EcoSim is preserved. Finally, the movements in the neutral model are random, but the distribution of distances is kept the same as in EcoSim. The randomness in movement can be seen in Fig.3.8 as the formation of spatial spiral patters can no longer be observed.

The crucial property of EcoSim neutral shadow is that its evolutionary dynamics are identical to EcoSim except that neither the presence not the frequency of a genotype can be explained by its adaptive significance. This is because all selection in the neutral model is random, so no genotype has any dominance over any other. In other words, although gene states are subject to the same variation as in EcoSim, they have no evolutionary fitness consequences or effects. In addition, changes in the environment have no effect on individuals in the neutral model.

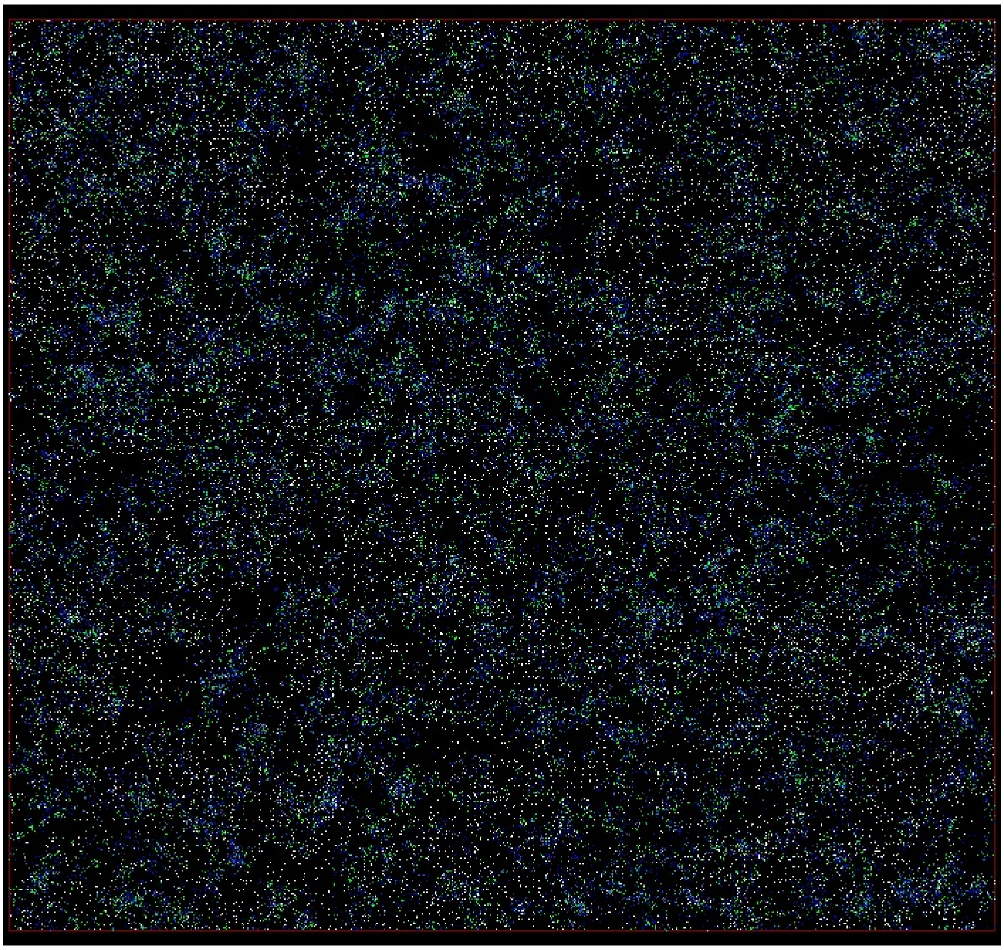


Figure 3.8: Snap shot of the Neutral shadow world with predator and prey spatial distribution

Consequently, the process of natural selection is considered to be eliminated in this neutral model.

Chapter 4

Data Analysis

4.1 Introduction

The EcoSim simulation allows capturing and storing all the characteristics of that run for latter analysis. Vast amount of information stored in (three terra bytes per simulation run), running for 30,000 time steps is produced. At each time step the current state of all individuals and the world environment characteristics are written to output files. Other information includes average characteristics for all prey and all predators such as population size, speed and age. For one simulation running for 30,000 time steps more than 30,000 files are produced. Further analysis of this information is required which is challenging to interpret manually. In addition, some calculations are done off line, that is using the output files after the simulation has stopped, such as fitness calculations. The aim to understand the different evolutionary behaviour of prey/predators encouraged us to make use of information theory and machine learning techniques.

4.2 Entropy as a Measure of Genetic Diversity

In order to measure the genetic diversity among individuals we make use of Shannon entropy. Depending on the specific problem or representation being used, ranging from biological domain to genetic programming, numerous diversity measures and methods exist. For example, Sherwin [63] has shown the efficiency of

Shannon entropy in measuring diversity in ecological community and genetics. He has also highlighted the advantages of using entropy based genetic diversity measures, and surveyed these diversity measures. A close relationship between biological concepts of Darwinian fitness and information-theoretic measures such as Shannon entropy or mutual information, was found [64]. Shannon Information theory [65] defines uncertainty (entropy) as the number of bits needed to fully specify a situation, given a set of probabilities. These probabilities can be estimated by simply counting the abundance of each genotype (site) in the population. The per-site entropy of an ensemble of sequences X, in which genotype s_i occurs with probability p_i is calculated as

$$H(X) = -\Sigma p_i \log_2(p_i) \tag{4.1}$$

where the sum goes over all different genotypes i in X. Next, the entropy content of the whole sequence (genome) is approximated by summing the per-site entropy over all sites in the sequence. This is only an approximation because it ignores interactions between sites (i.e. epistasis). There is no fixed set of genotypes but they are discreet values that change over time in the simulation. The lower the entropy, the less diverse are the genomes of a population and vice versa. There exists a limit in the values of entropy in EcoSim. When it approaches its maximum (corresponding to a uniform distribution of all genotypes) it indicates a completely uniformly distributed population close to randomness. On the other hand very low entropy (close to 0) means that there is too much similarity between individual genomes, and means that individuals need to diverge more in order to adapt to a dynamic environment. A good balance between learning from the environment (low genetic diversity) and increasing the diversity (high genetic diversity) should be met in order to ensure the well being of species. As the FCM of an individual is only written to the output files at the time step the individual is born, a dynamic data structure was used when writing a program to calculate the entropy. This dynamic data structure keeps track of birth and death of individuals so that the FCM of all individuals surviving in each time step are included in the calculation.

4.3 Fitness calculation

Fitness describes the ability of individuals to both survive and reproduce, and is equal to the proportion of the individual's genes in all the genes of the next generation. Darwinian fitness, or simply fitness, of a biological trait describes how this trait participates to the success of an organism to pass on its genes. The more likely that an individual is able to survive and live longer to reproduce, the higher is the fitness of that individual. There are mainly 2 measures of fitness, absolute fitness and relative fitness. An organism's absolute fitness is the total number of surviving offspring that an individual with a particular genotype produces during its lifetime (its lifetime reproductive success). Relative fitness of a genotype is its ability to survive and reproduce compared to other genotypes in the population. Relative fitness is the absolute fitness of an organism divided by the average fitness of the population within which that organism is found.

As EcoSim provides more flexibility and more ability in conducting more complex measurements not feasible in natural ecosystems we tune the measure of absolute fitness. We measure fitness of an individual as its age of death plus the sum of age of death of its entire direct produced offspring. Accordingly, the fitness value mirrors the individual's capability to survive longer and to produce as many strongly adaptive offspring as possible. The average species fitness is also calculated as the average fitness of its individuals. The fitness computation is only performed after termination of a run of the simulation for analysis purpose. It is therefore never used during the simulation itself, meaning that no evaluation of the fitness of the individuals is performed to select them, the natural selection process being an emerging property of the interactions between the individuals and their environment. A program has been written which traverses the output files after the termination of the simulation in order to calculate the fitness.

4.4 Building a Classifier for Inference

Classification consists of learning from exemplary rules that can later be used to classify samples into two or more predefined classes ([66]). A learning system that performs classification is known as a learner, or, more commonly, a classifier. The classifier is first fed with training data in which each sample is labeled with the correct label, or class. This data is used to train the learning algorithm, which creates models that can then be used to classify similar unclassified data When training the classifier using cross validation ([67]; [68]), a statistical method to evaluate the performance of the classifier was used. One round of cross-validation involves partitioning the data into complementary subsets, performing the analysis on one subset (the training set), and evaluating the analysis on the other subset (the testing set). For training all our models a 10-fold cross validation was used, where the data set was divided into 10 subsets, and a holdout method was repeated 10 times. Each time, one of the 10 subsets was used as the test set and the other 9 subsets were combined to form a training set. The average error across all 10 trials was computed. The advantage of this method is that it matters less how the data is divided; every data sample is included in a test set exactly once and in a training set 9 times. In addition, the data used for learning with cross-validation only represents a part of the total data as the other part has been kept to make a validation set. Instances for every run are split into two sets: train and validate which consists of unseen instances in order to evaluate the generalization capacity of the model learned on the training set.

4.4.1 C4.5

C4.5 is an algorithm which produce classifiers expressed either as decision trees or rulesets and developed by Ross Quinlan [69]. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The C4.5 constructs a very big tree by considering all attribute values and finalizes the decision rule by pruning. It uses a heuristic approach for pruning based on the statistical significance of splits At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision and splits the data set into subsets. The C4.5 algorithm then repeats the same process on these smaller subsets:

- For each attribute *a*
  Find the normalized information gain from splitting on *a*
- Let *a best* be the attribute with the highest normalized information gain which minimizes the total entropy of the sublist.
- Create a decision node that splits the sublist on *a best*
- Recurse on the sublists obtained by splitting on *a best*

The initial tree is then pruned to avoid overfitting. Pruning is carried out from the leaves to the root. For a subtree, C4.5 adds the estimated errors of the branches and compares this to the estimated error if the subtree is replaced by a leaf; if the latter is no higher than the former, the subtree is pruned. The C4.5 provides decision rules that can help in the interpretation of the classifier to some extent. Rules are generated by following the path from the root to a leaf node. For m attributes and n training instances computational complexity = O(m n log n) + O(n (log^2 n)). Some limitations of C4.5 algorithm includes possibility of over fitting and the possibility of constructing empty or insignificant branches which adds to the complexity of the tree and therefore to the number of useless rules.

4.4.2 Random Forest

Random Forest (RF) is a trademark term for an ensemble of decision trees as it grows many classification trees [70]. Through RF, decision trees are 'grown' using a random selection (but not all) of the input variables. Single, tens, hundreds, or even thousands of these trees, each with a random variable selection, may be grown for each classification problem. The final classification decision is based on a voting system of all the decision trees that have been created. Essentially, RF classifies the predictor variables as many times as the user requires, introducing an element of randomness each time, and voting for the most popular class. Unlike single decision trees which are likely to suffer from high bias (depending on how they are tuned) RF use averaging to find a natural balance between the two extremes. Since they have very few parameters to tune and can be used quite efficiently with default parameter settings (i.e. they are effectively non-parametric), Random Forests are good to use as a first cut when only little is known about the underlying model. Random Forests have higher prediction accuracy as compared to individual classification trees because the ensemble adjusts for the instability of individual trees induced by small changes in the learning sample, which would otherwise impair the prediction accuracy in test samples. RF algorithm runs fast and handles noise and missing data. It can run on large data sets with high dimensional features. On the other hand, RF are considered as black box supervised learning techniques because of its sheer size and lack of interpretability. RF internal model is barely interpretable, and thus is only defined by its inputs and outputs. It is also not easy to determine how one feature affects the prediction. Depending on the specific problem on hand, one must choose between accuracy and model readability.

4.4.3 JRip Rule Learner

JRip ([71]) implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), which is an optimized version of IREP. JRip learn rules that are easy to understand and provide informative feedback about the problem. This approach performs efficiently on large datasets, due to its low running time ((O(m log^2m)), where m in the number of training data). It is based in association rules with reduced error pruning (REP), a very common and effective technique found in decision tree algorithms. In REP for rules algorithms, the training data is split into a growing set and a pruning set. First, an initial rule set is formed that over fits the growing set, using a heuristic method. This overlarge rule set is then repeatedly simplified by applying one pruning operator among a set of possible ones. A typical pruning operator would be to delete any single condition or any single rule. At each stage of simplification, the pruning operator chosen is the one that yields the greatest reduction of error on the pruning set. Simplification ends when applying any pruning operator would increase error on the pruning set. The algorithm is divided into a building stage and an optimization stage. In the building stage the growing a rule phase and the pruning phase are repeated until reaching to stopping conditions. Growing the rules, consist of, for each rule greedily adding antecedents (or conditions) to the rule until the rule is perfect (i.e. 100% accurate) by trying every possible value of each attribute and selects the condition with the highest information gain. Then the process incrementally prunes each rule and allows the pruning of any final sequences of the antecedents. The optimization stage comes after generating the initial rule set. It generates and prunes two variants of each rule from randomized data using growing and pruning.

In comparison to other machine learning techniques (i.e Random Forests, Support Vector Machine (SVM) or Neural Networks(NN)), JRip provide a model easier to interpret and a lower number of rules. Random forests, SVM and NN are in general more efficient than JRip in terms of accuracy but on the other hand lack readability.

4.5 Feature Selection

In some of the presented experiments we use a feature selection step before applying the classification method. We used two different feature selections algorithms. The first feature selection method used, CfsSubsetEval [72], evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while having low intercorrelation are preferred. BestFirst ([73]) searches the space of feature subsets with a greedy hill-climber augmented with a backtracking facility. This was implemented under weka environment. The other method, which was previously presented by Salehi E, is a wrapper feature selection methods [74] [75] based on an estimation of distribution algorithm (EDA) called CMSS-EDA [76]. Since CMSS-EDA does not consider a small fix upper bound on the number of variables on which each variable depends, the most relevant variables using this approach were found even when there were many dependencies between them. Each subset of variables is encoded as a bitstring and the subset of variables which maximizes the AUC (Area Under ROC Curve) obtained by a Bayesian network classifier was found.

Chapter 5

The genomic evolution in EcoSim and its effect on fitness

5.1 Introduction

The emergence of complex adaptive traits and behaviors in artificial life systems requires long term evolution with continuous emergence governed by natural selection. Evolutionary adaptation can be viewed as two separate processes. The first process is the origin of new beneficial mutations. The second process is the fixation of some of those beneficial mutations by natural selection. From the genetic perspective, mutations and natural selection, through the course of evolution enforce the emergence of new traits and disappearing of others. Darwinian evolution governed by natural selection is modeled in EcoSim and the main objective of this work is to study the evolutionary process in EcoSim. Validating the evolutionary process in EcoSim requires the examining of the emergence of new genes and behaviours that contribute to the well being of individuals. First, a comparison between the evolutionary process in EcoSim and a neutral shadow of EcoSim is studied. Second, the evolutionary machinery in EcoSim was studied by examining the emergence of new genes and their effect on fitness.

5.2 Evolution in EcoSim versus Neutral Model

In order to study the genomic evolution in EcoSim, a neutral shadow of EcoSim is built (see sec. 3.7). The crucial property of EcoSim neutral shadow is that its evolutionary dynamics are identical to EcoSim except that neither the presence not the frequency of a genotype can be explained by its adaptive significance. This is because all selection in the neutral model is random, so no genotype has any dominance over any other. In other words, although gene states are subject to the same variation as in EcoSim, they have no evolutionary fitness consequences or effects. In addition, changes in the environment have no effect on individuals in the neutral model. Consequently, the process of natural selection is considered to be eliminated in this neutral model.

For this study we focus on the part of the genome that code for the behavioral model of the individual and which correspond to its FCM map and which has a maximum size of 390 sites. Every site is a real discreet number which measures the level of influence from one concept to another. Initially all prey and predator individuals are given the same values for their genome respectively. Time step after another, as more individuals are created, changes in the FCM occur due to the formation of new edges, removal of existing ones and changes in the weight associated to edges. The first couple of thousand of time steps are neglected in the calculations to overcome any misleading results due to the initial similarity between individual genomes. In each time step a value of entropy of all existing prey species is calculated, along with the entropy of the entire population of prey. The information contained within a genome determines how the organism behaves in its current environment. Thus, this information determines the capability of the organism to reproduce and transmit its genome. The environment changes from one place to another and from one time step to the next. Individuals that evolve in different parts of the world have different information about the environment they evolve in stored in their genome. Furthermore, as EcoSim models a predator-prey system, the co-evolution is also taken into consideration. The strategies (behavior) of each kind of individual (predator/prey) are continuously changing as they try to adapt to the other kind. The more the individuals try to learn the more the environment changes and the more there is still something different to learn. This fact drives the individuals to keep learning and continuously try to come up with survival strategies that helps them adapt to their changing environment. This is the reason behind the fluctuations seen in the EcoSim entropy (see sec. 4.2) curves (Fig.5.1). On the other hand the neutral model shows much more steadiness in the entropy values. Under highly random conditions and when natural selection is eliminated, the genomic structure shows neither learning nor adaptation to the surrounding environment. These results show that entropy changes through the course of evolution. The EcoSim simulation gives the chance to acquire data for thousands of generations and to study the performance of entropy as a genetic diversity measure.

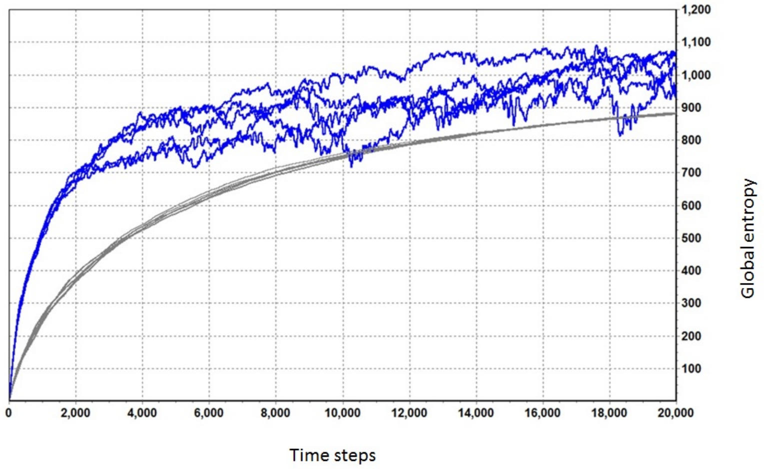


Figure 5.1: Global Entropy for 10 different runs of the simulation. Top 5 curves are for EcoSim and lower 5 for Neutral Model.

5.3 Emergence of New Genes

In a constantly changing environment individuals must continuously learn. This fact drives the individuals to evolve survival strategies that helps them adapt to their changing environment. Prey individuals die due to several reasons: reaching maximum age, lack of energy or being eaten by predators. Natural selection is not forced by limiting the number of existing species or fixing population size, but rather selection acts through the behavioral model. Individual that are not able to gain energy from food, reproduce and escape from predators will be eliminated by the evolutionary process. Thus, fitness levels are not fixed and do not always increase; rather, they vary over time. The evolutionary process of EcoSim governs the emergence of new genes and disappearing of others. By the process of natural selection only the fittest will be able to survive, and therefore the emergence of new genes is not random but adaptive, adding to the intelligence and complexity of the individuals. As the FCM is coded in the genome and represents the behavioural model, the change and evolution that happens on edges mirrors the behavioural evolution of individuals. Monitoring change in these edges values adds more understanding of the behavioural changes along the course of evolution. In order to validate EcoSim model we examine the effect, the emergence of new genes, has on individual's fitness by using classification techniques.

5.4 Building a Random Forest Classifier for Inference

Recent work in computational biology has shown an increased use of Random Forest[70], owing to its unique advantages in dealing with small sample size, high-dimensional feature space, and complex data structures [77][78]. Random forests (see sec. 4.4.2)have higher prediction accuracy as compared to individual classification trees because the ensemble adjusts for the instability of individual trees induced by small changes in the learning sample, which would otherwise impair the prediction accuracy in test samples. A new gene is a gene that had the value of zero in the initial FCM map of prey, and then mutates to and preserves a non-zero value later on in the simulation. The genes that were initially zero and then changed are monitored and extracted. The average FCM for every existing species in every time step in the simulation is calculated. From 390 possible genes there are 125 initial genes and 265 possible new genes, which can appear gradually. Each of these 265 new genes represents a feature in the classifier, making the feature space a high dimensional feature space. Each instance of the learning process consists of the set of 265 average gene values in a given species at a given time step.

Run	Percentage LOW	Percentage HIGH	Percentage VHIGH
Run 1	53%	36%	10%
Run 2	45%	39.5%	15.5%
Run 3	41%	42%	17%
Run 4	40%	46%	14%

Table 5.1: Percentage of low fitness (LOW), high fitness (HIGH), and very high fitness (VHIGH) prey instances for 4 different runs.

There are four different runs of the simulation, each having around 20000 time steps. The first couple of thousand of time steps are neglected in the calculations to overcome any misleading results due to the initial similarity between individual genomes. Consequently, most of the genes would have obtained non-zero values. Around 150000 instances are randomly extracted from every run, to build four classifiers, one for each run. The class variable to predict at a given time step is the average fitness value of the species 50 time steps later. The effect of these new genes on the individual fitness is not immediate so a 50 time steps shift was used to give time for the values of new genes to affect the fitness. 50 time steps counts as 3 to 4 generations which is enough time to observe correlation between the emerging genes and fitness values. The fitness value is discretized into three classes for the average fitness; LOW which represent values less than 85. HIGH, which is between 85 and 110, and very high VHIGH, for values higher than 110. Table 6.1 presents the percentage of instances for every class in four different runs of the simulation.

When running the RF algorithm (implemented in the weka environment [79]), 10 fold cross validation was used. Instances for every run are split into two sets: train (66% of the total data) and validate (34%) which consists of unseen instances in order to evaluate the generalization capacity of the model learned on the training set. Using cross validation and 10 classification trees, 94.7% average train accuracy of four runs with a standard deviation (std) of 0.33 and 95% validation accuracy with std of 0.3 was found (see Table 6.2). Although there are many factors affecting the fitness, it was still predicted with high accuracy knowing only the average

Run	Train Acc.	Validate Acc. on same run	Avg Test Acc. on other 3 runs
Run 1	95%	95.2%	41.5%
Run 2	94.7%	95.3%	44.2%
Run 3	94.3%	94.4%	43.4%
Run 4	94.5%	95%	44.3%

Table 5.2: Accuracy for training and validating with the Random Forest classifier on four runs of the simulation.

values of the newly developed genes. This high accuracy supports the initial assumption that the values of the new evolved genes could affect the well being of the individuals. We also tested the generality of the finding by training the classifier on one data set from one run and testing it on a data set from another run of the simulation. Although the classifier was able to learn some general rules for prediction in different runs, the average accuracy of 43.5% with std of 1.2 was not very high. One reason is that the simulation varies, from one run to another and each run has unique conditions in which survival strategies of the individuals vary. This leads to different behaviors, and thus different values of the genes that affect the fitness. These values did not evolve randomly, but were preserved by the evolutionary process thereby adding a higher level of complexity and intelligence to the individuals. It is worth noting that this increase in complexity adds an extra cost for the individual as any new gene added to this individual increase the amount of energy the individual needs to consume at any time step. Therefore, in order for new genes to be beneficial in terms of fitness, the benefit in term of advantageous behavior should be higher than the cost in energy. This supports the validity of the behavior model and demonstrates how evolutionary processes lead to adaptation and intelligence. This finding also emphasizes the role of natural selection in the simulation. Genomes that participate in the well being of its carrier host persist and survive.

	CMSS-EDA				CfsSubsetEval			
	Train Acc.	Validate Acc.	Selected Features	No. Rules	Train Acc.	Validate Acc.	Selected Features	No. Rules
Run 1	69.5%	70.9%	41	78	69.3%	70.5%	28	53
Run 2	72.2%	74.4%	35	119	73%	74.7%	41	101
Run 3	71.3%	71.5%	63	109	70.3%	71.6%	41	113
Run 4	73.7%	76.8%	47	62	73.8%	74.3%	38	55

Table 5.3: Accuracy percentages for training and validating with JRip after CMSS-EDA and CfsSubsetEval feature selection, for four runs of the simulation.

5.5 Rule Learning Using JRip

A better understanding of the semantics behind the evolution of the new genes is interesting. Which genes have a stronger influence on fitness and with which values? In order to study this phenomenon rules from the learned model predicting the fitness were extracted. The interpretability of a random forest is not as straightforward as that of an individual classification tree, where the influence of a predictor variable directly corresponds to its position in the tree. The model generated by the Random Forest can be challenging to interpret. To by-pass this limitation the JRip rule learner (see 4.4.3) was used [80] to extract more semantics from the prediction model. Moreover, in this study the accuracy obtained by JRip was reasonable and was not traded for low number of produced rules.

In order to improve the model performance and gain a deeper insight into the underlying processes affecting the results, we used a feature selection step. This pre-processing step highlights the most important genes affecting the fitness and eases the process of rule interpretation in addition to minimizing the number of rules. Two different feature selection methods are used (see sec. 4.5), CfsSubsetEval [72] using Best First searcher, and CMSS-EDA [76] and present both their results.

First, samples from the data set were used to extract features using both techniques. Then only these features were used with JRip rule extractor using separate training and validating sets with 10 fold cross validation for training the model. Different features have been selected from different runs of the simulation. This is due to the complexity of the behavior model. Survival techniques vary based on the different circumstances of each run and environment dynamics. Although the two feature selection techniques selected different features, their prediction accuracy using JRip was very similar. This shows the strong dependencies among the genes and how they collaborate with each other to adapt to their dynamic environment. Also some genes might have redundant information which could be replaced with some other set of genes. Table 6.3 shows results of JRip rule learner along with the number of rules it produces for each different run of the simulation for both CMSS-EDA and CfsSubsetEval respectively. Different IF THEN rules are learned from JRip to predict the three fitness classes. We were mainly interested in the rules that predict the very high fitness class(VHIGH) because understanding the conditions that increase the individual's species fitness is highly informative about the simulation properties. The number of rules, that predict VHIGH ranges from 2 to 25, in all runs using both feature selection methods. It should be noted that each gene can have a real positive value (for a new edge from one concept that positively influence another) or a real negative value (for a new edge from one concept that negatively influence another). Some rules with highest accuracy and hit ratio are presented. Rules are in the form of "IF edges emerge between the following concepts THEN the average fitness of the species will be very high after 50 time steps". The hit ratio of the rule is the percentage of instances following the rule.

- IF Satisfaction decreases sedentary, AND escape decreases socialize, AND search for partner decreases nuisance, AND food local low increases fear, AND satisfaction increases satisfaction, THEN fitness is VHIGH. (hit ratio=0.8, accuracy=71%)

Explanation for the previous rule is as follows. Satisfaction is an internal concept which is initially decreased by a low local (same cell as the individual) food level, or by a low energy level of the individual, or by a predator being detected in a close range (see Figure 3.2 for the initial prey map). A new edge corresponding to 'satisfaction decreasing sedentary' has evolved. The desire to escape lowers the desire to socialize because preference should be given to escaping from a close by predator. Searching for a partner decreases the nuisance, and having a low local food levels increases the internal concept of fear. 'Satisfaction increases satisfaction' is an internal feedback loop which gives persistence to the sensation of satisfaction. The combination of these new emerging genes within population tends to increase the population's fitness.

- IF curiosity decreases sedentary, AND friend close decreases exploration, AND energy high decreases exploration, AND food far increases reproduction, AND energy high increases wait, AND explore decreases explore, THEN fitness is VHIGH. (hit ratio=2.44, accuracy=85%)

Curiosity and sedentary are both internal concepts. Initially curiosity increases exploration which encourages the individual to move. Sedentary, in the initial prey map, decreases exploration. A new edge was developed between these two internal concepts, meaning that curiosity decreases sedentary, which enforce the initial semantic. Having friends close-by decreases the desire to explore and to move because having close-by friends encourages the individual to search for a local partner. Having a high level of energy decreases exploration because it might be better to reproduce and search for a partner. Having no close-by food, increases the will to reproduce instead of wasting energy by searching for distant food. Having high levels of energy increases the wait action is less obvious to interpret, but can mean that the individual has no need to move. Finally, exploration decreases the internal desire for further exploration, is also an internal feedback loop that reduce the persistence of exploration, which is mostly a random movement. These factors also increase the fitness.

- IF no local partner decreases fear, AND food local high increases wait, AND food far increases exploration, AND friend close increases eat, AND partner local yes decreases search partner, AND reproduce decreases socialize, THEN fitness is VHIGH.(hit ratio=1.53, accuracy=77%)

Fear is an internal concept that initially decreases all motion actions except escape and explore. A negative edge has been established between the perception of having no local partner and fear. If high levels of local food have been found, waiting is increased. If the food is far the need to explore increases as well. Having close-by friends increases the desire to eat and gain more energy. Local close by partners detected decreases the need of searching for partners. The desire of reproduction lowers also the socialization concept. The combination of the emergence of these factors also increases the fitness.

- IF no local partner decreases fear, AND predator close increases escape, AND socialize decreases reproduce, AND predator far increases eat, THEN fitness is VHIGH. (hit ratio=1.65, accuracy=69.4%)

Initially detecting close-by predators increases the internal concept of fear and an increased level of fear also increases the desire of escaping. A new edge emerged which directly encourages escape action if predators are detected without using internal concept. The need to socialize decreases the need to reproduce, which means that when there are no local partners (need to socialize), there is no interest to try to breed. Also a meaningful edge between increasing the eat action if predators are far was developed all leading to increasing fitness. The logical soundness of most of the produced rules shows both high semantics in the initial behavioral model of prey individuals, and the self organizing capability of our system. Some rules seems less obvious to interpret, but as the global model is a highly complex non-linear system with feedback loops, some modifications can have effects on other parts of the system and are therefore difficult to understand. The behavior model (genome) was able to evolve without any external interference besides the natural selection forces. These rules emphasize the importance of certain new genes and the strong dependencies found among them. The logical correctness of some of these rules is also a major interesting discovery.

5.6 Conclusion

The evolutionary process implemented in EcoSim has proven its effect on the behavioral model of the individuals as they adapt to a changing environment. To emphasize the capability of EcoSim to model evolutionary behavioral adaptation it was compared to a partially random version focusing on genetic diversity. It was demonstrated how entropy used to measure genetic diversity, behaves differently in both systems. The fluctuation in entropy curves for EcoSim illustrated how individuals try to learn and adapt to their environment. On the other hand the neutral model showed more steadiness in the curves due to more randomness and elimination of natural selection process. The correlation between the fitness of a population and the emergence of new behaviors was studied. The emergence of new genes was analyzed which affects the behavioral model of the agents, through the evolutionary process in EcoSim. The value of these new genes served as features for fitness prediction using Random Forest classifier implemented in weka. The high accuracy obtained from the Random Forest classifier shows the capacity of the behavior model to capture relevant information from its environment giving to the successful individual's ability to survive and to adapt to its dynamic environment. In a second step, rule learning technique was applied to extract semantics information from the prediction model. This enabled better understanding of the logical rules in the new evolved behavioral models that led to an increase in fitness. The JRip rule learner was used after a pre-processing feature selection step. The soundness of the rules obtained is very encouraging as they help to understand what kind of new behaviors can be useful in such dynamic and competitive ecosystem. As this simulation allows very long runs, it will be possible to study the dynamics of co-evolution by analyzing successive changes in the behavioral models during periods of prey adaptation to predators and vise versa.

Chapter 6

Correlation between genetic diversity and fitness in EcoSim

6.1 Introduction

Genetic change is the basis of evolution and genetic diversity serves an important role in evolution by allowing a species to adapt to a new environment and to fight off any disturbances. Genetic diversity is the variation of heritable characteristics present in a population of the same species. For a species to adapt to an ever-changing ecosystem, a significant level of variation must be present. With more variation, it is more likely that some individuals in a population will possess variations of alleles that are suited for the environment. Those individuals are more likely to survive to produce offspring bearing that allele. The population will survive for more generations because of the success of these individuals. Genetic diversity strengthens a population by increasing the likelihood that at least some of the individuals will be able to survive major disturbances, and by making the group less susceptible to inherited disorders. Many biological studies show that decreased population genetic diversity can be associated with declines in population fitness [81] [82] [83]. However, populations also adapt to their environment with the selection of the individuals with highest fitness. This driving force is opposite to the previous one as it has tendency to preferentially select high fitness alleles and consequently reduce the genetic variance in the population. This opposition of forces leads to an unstable equilibrium of genetic diversity when there is a change to the environment requiring adaptation. Because overall genetic diversity is associated with both individual fitness and population adaptive capacity, there is a need to develop an empirical quantitative understanding of the relationship between them.

6.2 Measuring Correlation between Entropy and Fitness

The virtual ecosystem EcoSim allows the study of the complex relationship between species genetic diversity and species fitness through an evolutionary process and is not limited to investigating these relationships under particular environmental conditions or at specific time periods as in most biological studies [82] [83] [84]. In EcoSim the environment is dynamic and is seen in the adaptation and genomic evolution of individuals. Thus, there are many factors affecting the genetic diversity of individuals and fitness of populations that can change over time and differs from one species to another. As we model long term evolution of many species in a dynamic environment the correlation between genetic diversity and species fitness may change over time. Shannon entropy is used as a measure of genetic diversity as presented in sections 4.2. At every time step entropy and fitness are calculated for all existing species. In order to investigate their possible correlations, first the Spearman's cross correlation [85] between genetic diversity and fitness is calculated for all prey species. The Spearman measure ranks two sets of variables and tests for a linear relationship between the variables' ranks. A perfect Spearman correlation of +1 or -1 occurs when each of the variables is a perfect monotone function of the other. The Spearman correlation coefficient is computed as follows:

$$1 - \frac{6\sum_{i=1}^{N} d_i^2}{N^3 - N} \tag{6.1}$$

where N is the number of items, and di is the distance between each population's rank of fitness and rank of diversity. A value of -1 represents negative correlation, 0.0 denotes no correlation, and 1.0 demonstrates positive correlation.

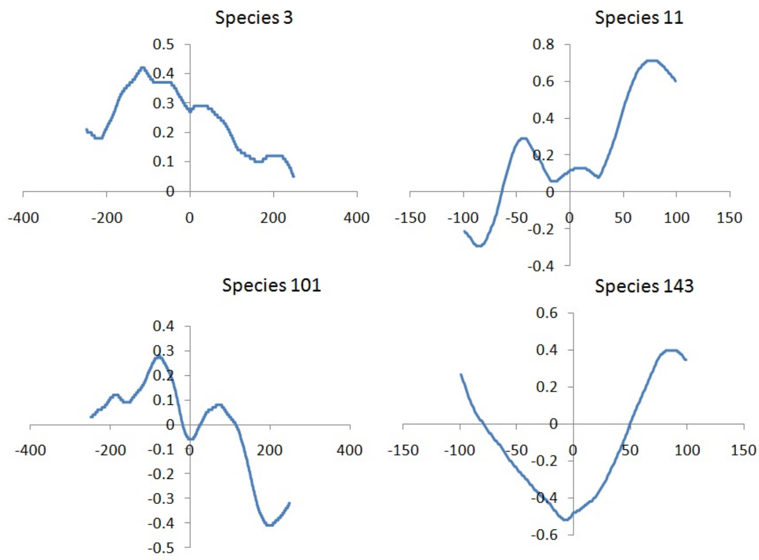


Figure 6.1: Different prey species correlation values between entropy and fitness. x-axis represents the different time shifts. Y-axis represents the correlation values.

A positive correlation indicates that either low fitness accompanies low diversity or high fitness accompanies high diversity. Alternatively, if high fitness is associated with low diversity a negative correlation is detected. In the studied evolutionary ecosystem simulation the effect of the diversity measure on fitness is not immediate. There must be a time shift between the variation in genetic diversity and its effect on fitness. Also, since the causal nature between attributes is not known in advance, the correlation in both shift directions are calculated. The Spearman correlation coefficient is computed between these two time series for every possible shift between -s and +s time steps. In essence, the entropy at time t is correlated with fitness at time t + s where s ranges from -s to +s.

Although there are many factors that might affect fitness beside genetic diversity, strong correlation between entropy and fitness for all prey species was found. The cross-correlation charts for some prey species is presented in Fig.6.1. The x-axis in these charts represents the different shifts for the time series. The y-axis represents the cross-correlation value at the corresponding shift. The figure shows not only that different species have different cross-correlation values, but also that same species correlate differently based on the time shift. Note that multiple factors affect the behaviour of species, including the dynamic environment, co-evolution and changing parameters with time. The correlation values for the same species may thus vary through the course of evolution, allowing us to investigate biologically meaningful relationships that may not be feasible by experimentation. This observation encouraged additional analyses by dividing the two time series into time frame windows and measuring correlations only within the specific time frame rather than the entire time series. In another words, these time series were split into sliding windows of 200 time steps centered at every time step within which all possible correlations are calculated with different shifts ±s. The highest correlation value (whether positive or negative) is then chosen and assigns to the species at that time step.

The results of five different runs of the simulation are presented each one containing 16,000 time steps and generating around 110,000 instances on average. Three different classes are assigned to the correlation values. Correlation with values between -0.5 and 0.5 are class WEAK CORR representing either no or weak correlation. Correlation values above 0.5 are high positive (HIGHP) and correlation values below -0.5 are high negative (HIGHN) respectively. These correlation classes are calculated for all instances (which corresponds to the set of all species at every time step) in every run and present the percentage of each class with a window of 200 and maximum shift of 25 in both directions. The averages for five runs were 26.8%, 38.4%, 34.6% for classes HIGHP, HIGHN and WEAK CORR respectively.

To better validate the calculations, variations in window and shift values were investigated. Having a window of 200 and a maximum shift of 20 in both directions gave 17%, 29.6% and 53.4% on average for five runs for HIGHP, HIGHN and WEAK CORR correlation classes respectively. Increasing the window and maximum shift to 400 and 50 was also tested. The average percentages were 23.7%, 27.5% and 48.8% for HIGHP, HIGHN and WEAK CORR classes respectively. Increasing the shift values increases the percentage of high correlation instances, as more time is needed to detect an increase in fitness after an increase in genetic diversity. Also note that increasing the window does not necessarily increase the high correlation values as some fluctuations in the entropy or fitness time series could exist. The values of shift that leads to the highest correlation values were also examined. It was found that 37.7% of instances in 5 runs obtained highest correlations from a positive shift between 10 and 25. In addition, an average of 38.7% for five runs found highest correlation in negative shift between -10 and -25. This shows that for more than 76% of the cases a window of 10 to 25 time steps was sufficient to see the effect of genetic diversity on the fitness or vice-versa. These values correspond roughly to one to three 'biological generations' (average life span of an individual) which seems a reasonable time to observe the effect of genetic variations in a population.

In order to validate the correlation results, an additional test was performed. First, both fitness and genetic diversity time series were randomized and performed the same correlation calculation. Windows of 400 with a shift of 50 were set and Spearman's cross correlation for all instances from five different simulation runs were calculated in order to compare with the original time series results. The resulting correlation values were discretized in the same way leading to 100% WEAK CORR, 0% HIGHP and 0% HIGHN. These results further validate the high correlation results obtained between genetic diversity and fitness.

The findings discussed previously of very high values for both negative and positive correlations support the claim that genetic diversity has a great influence on the well being of species. High positive correlation values mean that an increase in the genetic diversity, results in an increase in species fitness. There are many ways to interpret these results. For instance, a newly forming species in EcoSim with a small but sufficient population size would gradually increase its genetic diversity and subsequently positively correlates with its fitness. Also, these results may reflect that individuals in EcoSim adapt to their constantly changing environment. Adaptation could be mirrored by an increase in similarity of the species' FCMs (and thus a decrease in entropy) as new behaviors arise for the new environment and then diffuses throughout the population. On the other hand, negative correlations imply that a species decreases diversity, which may happen once individuals have adapted to their environment in order to reach stability. In order to further validate these results and investigate the reasons behind these correlation values, a step forward was to build a classifier that could predict the correlation values.

6.3 Building Classifier for Inference

To validate the preceding results and further investigate the reason behind these correlation values a machine learning approach to build a classifier was used. The interest of this classifier is first to see if some specific properties of a species can predict the current evolutionary behavior of the species, that is whether it can infer the correlation values based on these characteristics at a specific time. The classifier can also help to understand what factors and conditions affect the evolutionary behavior. This approach works by correctly classifying unknown instances based on a trained classifier, which demonstrates that the observed correlation is a common phenomenon that can be predicted from environmental features.

Three different classification algorithms were tested; C4.5 decision tree algorithm, Random Forests and JRip rule learner (see Chapter 4). The first step to build a classifier is to select the features that are believed to best describe the species and have direct effect on the species fitness. These features are: the number of individuals in species, the average age of individuals in species, the average speed of the individuals and their average energy level. The average number of reproduction events, average number of reproduction failing events, average activation level of reproduction and the spatial dispersal were also included. In addition, the average activation level of fear, hunger, satisfaction, nuisance, curiosity (which encourage individuals to move) were included. Finally, the entropy and fitness for each species were also among the selected features. In total there were 16 features, including the class variable correlation with the three possible values HIGHN, HIGHP and WEAK CORR. The next step was to try to select the best of these 16 features in order to both simplify the model and discover the most important features. This was accomplished using feature selection.

6.3.1 Feature Selection

To increase the quality of the classifier a wrapper feature selection method [74] [75] was used, based on an estimation of distribution algorithm (EDA) called CMSS-EDA [76]. This feature selection method is particularly efficient for problems with high level of interdependency between features. It searches for the subset of variables that maximizes AUC (Area Under ROC Curve) obtained by a Bayesian network classifier.

The best chosen features were population size, entropy, fitness, spatial dispersal, age and reproduction fail. The feature selection algorithm was used on all the five independent simulation runs and the same best features were consistently selected. These results also demonstrate the stability of the simulation that is important to be able to discover meaningful generic rules. Selection of entropy and fitness among the best features is intuitive as they are the two features being correlated and subsequently have a direct effect on the correlation class variable. Fitness and entropy values also determine the sign of the correlation results (being positive or negative). However, this particular result is not a bias in our analysis because what is being measured here is how a specific value of either entropy (or fitness), at a given time step, affects the future (or is affected by the past) correlation between fitness and entropy.

The selected features were then compared to important features used in ecological domains to study correlations between genetic diversity and fitness. For instance, one highly studied area centers on the effects of population size (which was among the selected features) on fitness. Some studies showed that population size and genetic variation are strongly positively correlated with fitness [84]. Also, loss in fitness and genetic diversity can be accompanied by a drop in population size [81]. Furthermore, positive correlation between genetic diversity, fitness, and population size was shown in [83]. Another selected feature, spatial dispersal has been shown to be an important factor for maintaining genetic diversity and fitness [83]. The last two selected features are the average age and the average reproduction fail. Considering the concept of fitness as defined in this analysis (see sec 4.3) used, it is clear that these two features have a direct effect on the fitness value because they are implicitly correlated: the higher the average age of species population the higher its fitness, and hence a lower average reproduction fail. The similarity between the best features discovered by the presented system and the most significant biological features affecting the genetic diversity and fitness is noticeable. It is important to keep in mind that these are not pre-defined or innate features of EcoSim but arise through evolution in the simulation. Finally,

Table 6.1: Percentage of high positive (HIGHP), high negative (HIGHN), weak correlation (WEAK CORR) between fitness and genetic diversity for window of 400 and shift of ±25 for five different runs.

Run	Percentage HIGHP	Percentage HIGHN	Percentage WEAK CORR	Percentage HIGH CORR
Run 1	13%	15.5%	71.5%	28.5%
Run 2	13.3%	17.6%	69.1%	30.9%
Run 3	11.3%	15.9%	72.8%	27.2%
Run 4	9.8%	11.4%	78.8%	21.2%
Run 5	11.8%	14.9%	73.3%	26.7%
Average	11.8%	14.9%	73.3%	26.7%

the biological significance of the best features selected highlights the validity of the measurements used and the finding of a strong correlation between genetic diversity and fitness in EcoSim.

6.4 Classification

6.4.1 Decision trees classification results

First, the C4.5 algorithm [69] with pruning, implemented in the WEKA [79] was used. A window of 400 is used and the shift of ±25 time steps is fixed for calculating the correlations. The reason behind that is to have all instances on the same scale making them comparable. Increasing the window to more than 400 also results in fluctuation in the fitness and entropy series. Conversely, decreasing the window tends to influence the correlations resulting in higher values. A shift value of ±25 was used based on the analysis of which shift leads to the highest correlations. Table 6.1 presents the percentages for HIGHP, HIGHN, WEAK CORR and the sum of HIGHP and HIGHN called HIGH CORR, for the five runs. The six features used to build the model are the ones selected from the feature selection process. The instances for each five runs were split into 80% for training the classifier using 10-fold cross validation and 20% for validating with C4.5 pruning model.

Table 6.2: Accuracy percentages for training and validating with the C4.5 classifier for 5 runs of the simulation.

Run	Train Acc.	Validate Acc. on same run	Avg. Validate Acc. on other 4 runs	STD Validate Acc. on other 4 runs	Number of rules
Run 1	79.3%	80.3%	60.1%	4.9	294
Run 2	74.7%	75.3%	66.8%	0.8	307
Run 3	77.2%	78.1%	63.2%	3.2	280
Run 4	80.2%	80.2%	69.1%	2.6	181
Run 5	78%	78%	66.9%	3.8	263
Average	77.9%	78.4%	65.2%	3.1	265

Table 6.2 presents training and validating accuracy on the data set from the same run. Training the classifier on the data set from one run and validating it with a data set from the other runs was also tested to infer generality of the model. The confusion matrix showed high true positive results for training and validating on the same run. The results from validation on another run showed only reasonably high true positive values when accuracies are above 65%. This is due to the fact that each run has variations in terms of attributes values and ranges and also due to possible overfitting. However, the model was able to discover some rules that can make good predictions on unclassified instances. The good classification accuracy on the test set of the same run points to the validity of the calculations of entropy as a measure of genetic diversity and its high correlation with fitness. It also shows that there exist specific conditions for the species that lead to a positive or negative correlation between fitness and genetic diversity.

6.4.2 Random Forest Classifier for Inference

The Random Forest [70] technique includes an ensemble of decision trees and incorporates feature selection and interactions within the learning process. It is nonparametric, efficient, and has high prediction accuracy for many types of data including high dimensional ones. The same 15 features were used in a feature selection step. The best six selected features were used as attributes in the RF

Table 6.3: Accuracy percentages for training and validating with the RF classifier for 5 runs of the simulation.

Run	Train Acc.	Validate Acc. on same run	Avg. Validation Acc. on other 4 runs	STD Validation Acc. on other 4 runs	Number of rules
Run 1	97.6%	97.9%	63%	4.94	1104
Run 2	96%	96.5%	68.3%	1.7	975
Run 3	96%	96.9%	66%	3.8	983
Run 4	95.7%	96.1%	67.7%	2.7	1021
Run 5	96%	96.3%	68.1%	3.3	995
Average	96.4%	96.7%	66.6%	3.2	1015

classifier. The Random Forest classifier implemented in the weka environment [79] was used. Instances for every run were split into two sets: train and validation. For training the model a 10 fold cross validation method was used. The average accuracy of five classifiers on test set, one for each simulation, was 96.7% (see Table 6.3). The high classification accuracy validates our use of entropy as a measure of genetic diversity and its high correlation with fitness, in addition to emphasizing the most important features affecting correlation results.

6.4.3 JRip Rule Learner

The model generated by the Random Forest is difficult to interpret and the high number of rules produced by C4.5 is challenging. To by-pass this limitation the JRip rule learner [80] was used to extract more semantics from the prediction model and gain more insight about the conditions affecting correlation between genetic diversity and fitness. Different IF THEN rules are learned using JRip to predict the three correlation classes using the same selected features as before. In five different runs 19 rules were discovered with an average accuracy of 76% using 10 fold cross validation see Table 6.4. Here the main interests were the rules that predict the HIGHP and HIGHN classes; and some of the produced rules are presented. The hit ratio of the rule is the percentage of instances covered by the rule.

Table 6.4: JRip rule learner accuracies and number of produced rules for five different runs of the simulation

Run	Train accuracy	Validate accuracy	No. of rules
Run 1	76%	75.6%	24
Run 2	71.7%	72%	23
Run 3	75%	75.8%	24
Run 4	79%	78.8%	7
Run 5	75.5%	76.1%	18
Average	75.4%	76%	19

- IF number of individuals is low, AND fitness is low, AND entropy is low, AND failed reproduction is high THEN correlation is HIGHP. (accuracy=55%, hit ratio=1.7)

- IF number of individuals is low, AND age is high, AND fitness is low, AND entropy is low THEN correlation is HIGHP. (accuracy=64.3%, hit ratio=0.88)

- IF fitness is low, AND age is medium, AND spatial dispersal is low THEN correlation is HIGHP. (accuracy=62%, hit ratio=0.76)

- IF number of individuals is high, AND age is high, AND entropy is high, AND spatial dispersal is high THEN correlation is HIGHN. (accuracy=82%, hit ratio=0.37)

- IF spatial dispersal is high, AND number of individuals is high, AND age is medium, AND entropy is medium, AND fitness is high THEN correlation is

HIGHN. (accuracy=75%, hit ratio=0.6)

- IF failed reproduction is low, AND entropy is high, AND number of individuals is high THEN correlation is HIGHN. (accuracy=100%, hit ratio=0.2)

The other rules discovered were similar. In general, a low number of individuals associated with a low entropy, low fitness and low spatial dispersal led to a high positive correlation between entropy and fitness. Small species tended to increase their genetic diversity which allowed them to increase their fitness and hence their survival ability. On the other hand, a high number of individuals associated with a high entropy, high fitness and high average age led to high negative correlations between entropy and fitness. Species with large population size moved towards lower genetic diversity as individuals adapted by learning common survival strategies that tended to increase their fitness. RF produced the highest accuracy but lacked readability and model understanding. The JRip rule learner gave the least accuracy but produced the least number of meaningful rules with acceptable accuracy values.

6.5 Conclusion

It has been shown how the evolutionary process implemented in EcoSim affects the behavioral model of the individuals as they adapt to a changing environment. The use of Shannon entropy as a measure of genetic diversity of an individual based evolutionary ecosystem simulation has been introduced. Very high correlation, both negative and positive, between entropy and fitness have been found, which strongly indicates how genetic diversity affects the well being of the species. In order to validate our correlation results and further understand the reasons behind these results we built different classifiers to predict the correlation class variable based on training and validating sets. We found high accuracy for classification which showed the significance of our genetic diversity measure and its correlation with fitness. In addition, we used feature selection to find the best features affecting the correlation values. We showed how these extracted features are similar to the factors affecting genetic diversity and fitness in community ecology. The similarity between results of five different runs of the simulation points to the stability of the simulation and the generality of our findings. This study demonstrated the relation between genetic diversity and fitness changes based on time and other features such as reproduction rate, population size and spatial dispersal. Three different classifying algorithms have been tested; RF, C4.5 and JRip. The RF produced the highest accuracy but lacked readability and was difficult to interpret. The rules discovered from the JRip, which seem to be biologically pertinent, gave us more understanding about the conditions affecting the correlation between genetic diversity and fitness.

Chapter 7

Behavioural and Genetic Change of Prey as a Response to Predator Removal

7.1 Introduction

Much attention has been given to the role of predation in structuring ecosystems and influencing prey behaviours ([86]; [87]; [88]); through processes occurring on ecological or microevolutionary time scales. Contemporary evolution, also known as microevolution, concerns adaptation and trait heritability observed in contemporary time (i.e. less than a few hundred generations). Although most research considers evolution to be a long-term (macroevolutionary) concern, the evidence for contemporary predator-prey evolution suggests that it should also be a short-term concern, helping to bridge the gap between ecological and evolutionary processes. Recent studies in eco-evolutionary dynamics suggests that contemporary evolution can shape ecological processes ([89]; [90]). As it is becoming increasingly clear that predators are a strong force in nature, we asked how they may shape ecosystems and influence their prey over the course of evolution.

The sudden absence of predation pressure has been considered as one of the major factors causing contemporary evolution of prey populations ([91]; [92]; [93]). In the past decade, there has been an explosion of interest in how evolution of predator-prey interactions can affect their population dynamics, such as population stability or instability ([94]), along with predator and prey behaviour ([88]). Ecologists typically treat predator and prey populations as homogeneous sets of individuals, rather than genetically diverse populations that are capable of evolution. However, predators and prey do frequently coevolve because they respond to reciprocal biotic selection which can give rise to diverse adaptations. Evolutionary change and population fluctuations can occur concurrently and interdependently, affecting predator-prey population dynamics and driving out-of-phase predator-prey cycles ([94]; [95]; [96]).

Predators do far more in shaping prey population dynamics than through lethal effects (also known as direct prey consumption) ([97]). The non-lethal consequences of predation risk, whether through predator-induced morphological changes or through behavioural compensation, are very important in determining population and community dynamics in a variety of taxa ([86]; [88]). One particularly important avenue for such non-lethal consequences to effect population level changes is through a trade-off between predation risk reduction and the benefits of foraging ([98]). Foraging-predation risk trade-offs have been used as an effective framework for understanding how non-lethal effects influence populations and communities in terrestrial and avian animals in several reviews ([88]; [99]). Studies have noted that predation risk alone may be responsible for much of the observed interactions between predators and prey ([100]), but the complexity of these interactions is just beginning to unfold as it is becoming increasingly important to synthesize our understanding of multiple interacting behaviours ([101]). By interacting with other trophic levels predators act powerfully in shaping the adaptations of organisms to their environments. Many antipredator behaviours and morphological traits are the evolutionary products of an endless antagonism between the killing instinct of their predators and the driving needs to survive and reproduce of prey ([102]; [103]). This antagonism necessarily gives rise to tradeoffs between the costs and benefits of antipredator behaviours because they evolve under energy constraints, and it is clear that organisms have evolved to choose between alternative behaviours to deal with the threat of predation under different circumstances ([104]). The evolution of antipredator, reproductive and foraging behaviours under the constraints of time, energy, and predation risk can impact not only individual fitness but also population dynamics ([96]). The abundance of organisms at different trophic levels and the behavioural complexity of their interactions within communities are two frontiers that change through evolution in predator-prey systems. Ecological modeling offers a solution for such integrated analysis. However, most models do not account for the non-lethal effects that predators can have on prey which can critically affect prey population growth and prey resource densities through cascading effects ([98]). In real ecosystems individuals engage in complex behavioural interactions between and within trophic levels which can affect energy flow. Mathematical models have been proposed to study predation risk-foraging tradeoffs, for example the Risk Allocation Hypothesis (RAH) ([105]; see [106] for a review of empirical tests of the Risk Allocation Hypothesis). Empirical work has shown that decreased foraging efficiency while under predation risk can directly limit energy transfer up the food chain by reducing both the energy acquired by prey and the energetic conversion of food to growth in intermediate consumers, thereby reducing the biomass available for predation by higher trophic levels ([107]). Thus, energy flow through an ecosystem is greatly influenced by the behavioural interactions between predators and prey. There are numerous examples of predator-prey systems in which the prey will alter their foraging behaviour, activity or reproductive behaviours due to perceived predation risk. For example,[108] was able to show that male sticklebacks adjust reproductive effort in response to predation pressure by breeding less frequently, and males were also responsive to the current probability of mating despite a background of predation risk. The authors observed that males with a higher chance of mating took more risk in finding mates than those with a lower chance of mating. Birds have been a prime study group in investigating the effects of predation on reproductive strategies. For example, [109] found that birds assess the risk of nest predation and alter their reproductive strategies accordingly. Another recent avian study demonstrated that nonlethal predation effects can reduce offspring production by 40% over the span of a breeding season, concluding that predation risk alone can affect reproductive strategies and parental effort in raising offspring, and consequently population demographics ([110]). However, it remains to be seen how these variations in predation risk and reproductive output over the lifetime of individuals will affect the evolution of this population.

The current understanding of the role of predation in nature would suggest that introducing predators into an ecosystem is likely to have a range of consequences for the prey and the ecosystem itself. These effects may vary depending on the time they are measured following reintroduction (i.e. after previously being removed, or nearing extinction) or introduction (i.e. a non-native, perhaps invasive, predator that has never existed among the prey) of a predator. Many studies have focused on relatively short-term effects of predator introductions; among these, island species and ungulates have primarily been given attention. Some studies have shown that introducing new predators to islands dramatically affects naive prey. For example, endemic island bird species that have evolved in the absence of predators are much more likely to go extinct than are exotic bird species when predators are introduced ([111]). Other studies have shown that introducing non-native predators into islands can lead to catastrophic consequences ([112]; [113]; [111] ; [114]). In this study we investigated the effect of predator removal on the general behaviour of prey, genetic change and their capability to coevolve when predators are reintroduced in EcoSim. In addition, we allowed prey to evolve along two distinct evolutionary paths in our simulation, by either coevolving with predators or evolving in their absence, and used a machine learning classification algorithm to show that distinct genomes, corresponding to distinct behavioural adaptations in these prey populations, had evolved after long periods of time.

7.2 Summary of EcoSim Runs

In order to study the effect of predation on prey evolution we set up four different scenarios. In case A, the predators and prey coexist throughout the whole simulation run so they are considered prey evolving under high predation risk. In case B, prey evolve in the complete absence of predators and thus are considered to be low predation risk prey. In case C, predators and prey start by coexisting for 15,000 time steps, and then predators are removed from the simulation leaving prey to evolve alone in the absence of all predation pressure. In case D, the simulation starts with only prey evolving alone and then we introduced predators to the naive prey. For each of these four cases 10 independent runs of the simulation were performed with the same initial conditions for 30,000 time steps (see Figures 7.1 7.2 7.3). For analysis and results, we calculated the average of these 10 runs for any parameter under examination. In case D, we investigated if the time at which predators are introduced to naive prey is a factor in their ability to survive and begin coexistence with predators. Therefore, for case D we introduced predator populations as a cohort at several different time steps using 10 different simulation runs in each test. In each simulation only prey individuals were initially present and then predator populations were introduced with the same initial ratio of predators to prey as in case A of the EcoSim (1 to 2.5), but at different time steps. Thus, prey populations had evolved and adapted their survival behavior to the absence of predators to different extents in each of these simulations. When introducing predators, we randomly chose clusters (25 cells) which were highly populated with prey.

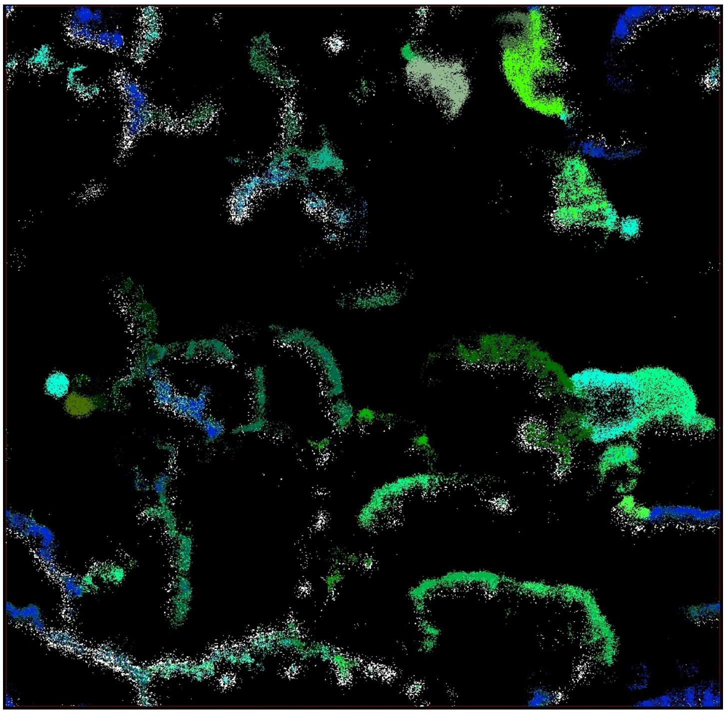


Figure 7.1: Snapshot of the world in one of the case A runs at time step 5000. The numbers of prey and predator individuals are 198554 and 27903 respectively. The colored spirals show different prey species and the white represents predators.

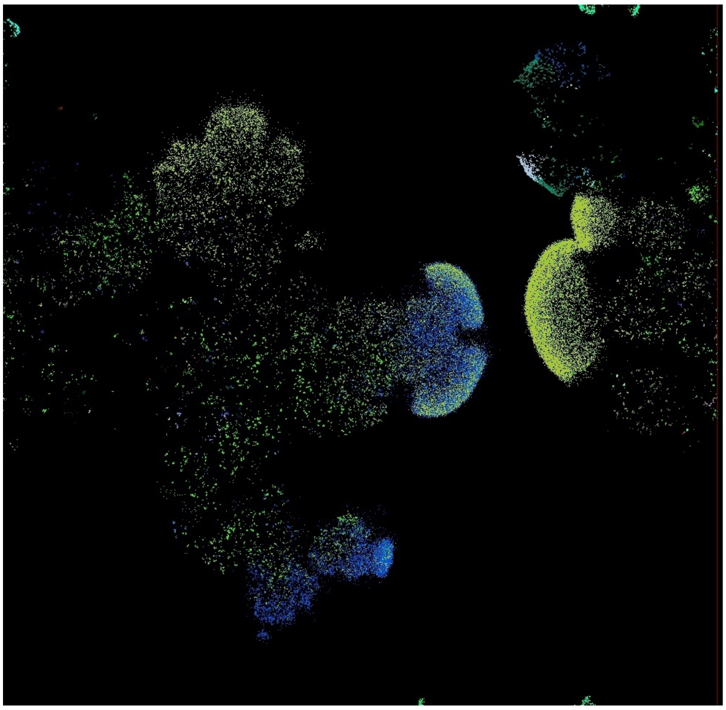


Figure 7.2: Snapshot of the world in one of the case B runs at time step 5000 showing 34254 prey individuals.

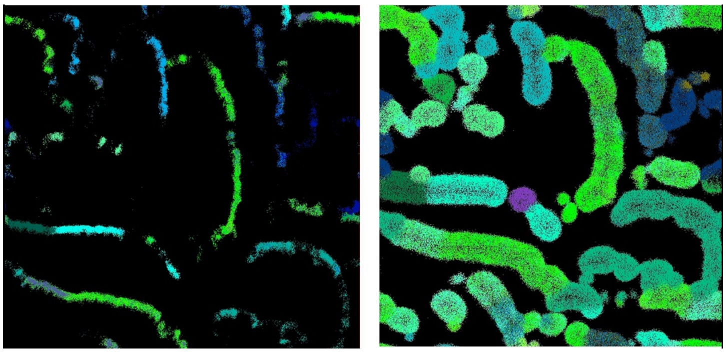


Figure 7.3: Snapshot of the world in one of case C runs. Figure on right represents the world at time step 15000 showing 233784 prey and the figure on left at time step 15060 where prey increased to 1040856.

7.3 Predation effect on prey's behaviour

7.3.1 Population dynamics

Evolution within a predator-prey system can affect their behavioural interactions and in turn population dynamics. Figure 7.4 shows the average of the prey population for 10 independent runs for each of the three different cases A, B and C. From the beginning of the simulation and for about 1,000 time steps, the prey population in case B is more than twice the prey population in case A runs. This initial spike in prey population in case B is due to the absence of predators which allows prey to feed and reproduce freely, and consequently leads to declines in grass levels (Figure 7.5) and subsequent declines in the prey population after the first 1,000 time steps. Since grass does not evolve in EcoSim, this allowed overgrazing by prey and the dramatic decline in grass until food became too scarce to sustain the high prey population. The decreased grass levels do not completely recover or reach its initial percentage as the speed of growth does not change which in return affects the prey population recovery. Particularly noteworthy is that the same phenomena is seen in case C where the sudden removal of predators caused a chaotic increase in prey population, which then led to a decrease in grass due to high consumption by prey, and a final reduction in the prey population after 1,000 times steps. The global trend in the prey population after the first 1,000 time steps is more stable and steady in case A, but is followed by a general increase during the last 28,000 time steps in case B.

The difference in prey populations between case A and case B is statistically significant (paired t-test between case A and B, p-value <0.001) (paired t-test between first and second half of case C, p-value <0.001). Prey fitness measures also mirrored the population differences by showing the same trend (Figure 7.6), where the difference is also statistically significant (paired t-test between case A and B, p-value <0.001) (paired t-test between first and second half of case C, p-value <0.001).

Both reproduction and birth rates increased in the absence of predators, as it can be seen in both case B runs and the second half of the case C runs after

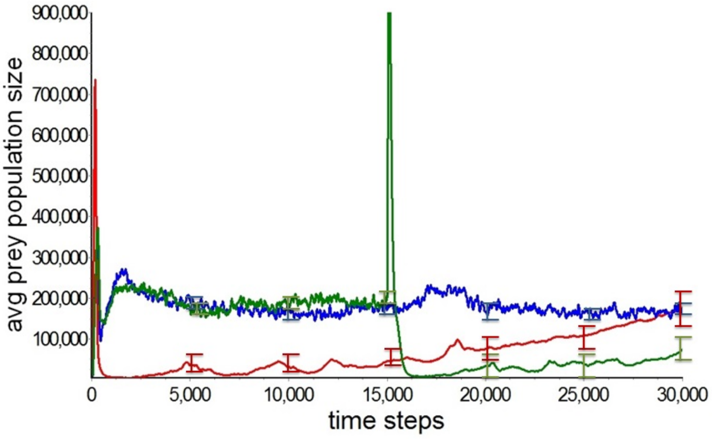


Figure 7.4: Total prey population. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone.

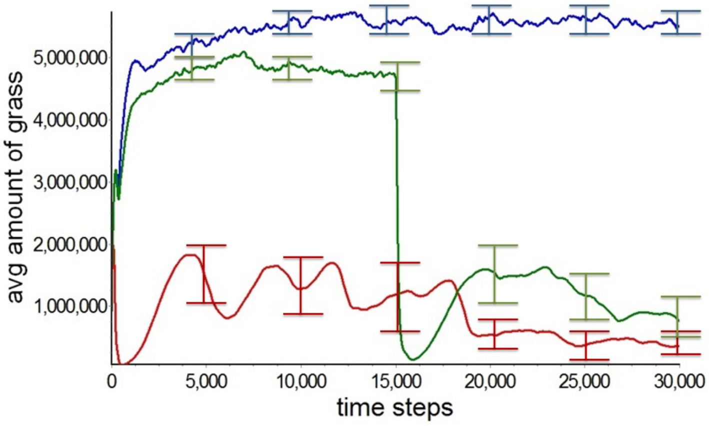


Figure 7.5: The average amount of grass in the world. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone.

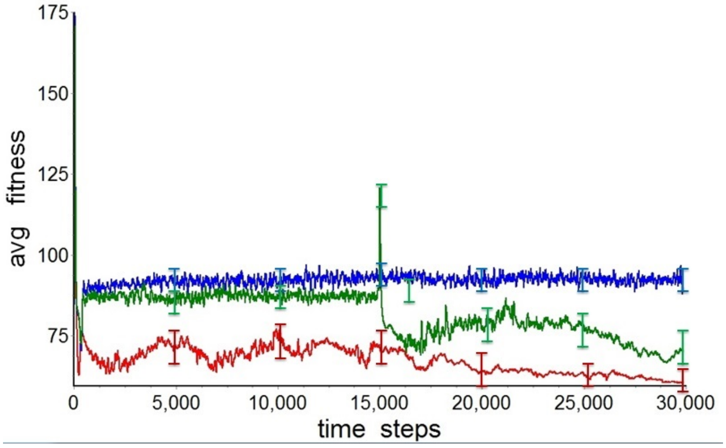


Figure 7.6: The average prey fitness. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone.

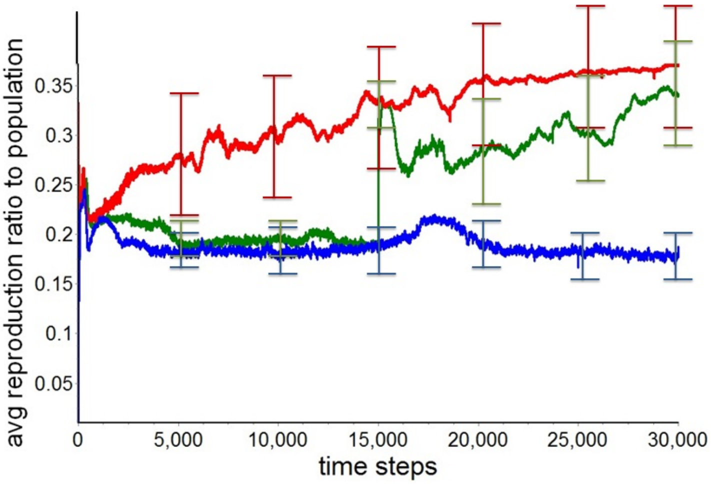


Figure 7.7: The average prey populations choosing reproduction action, including all successful and failed attempts. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone.

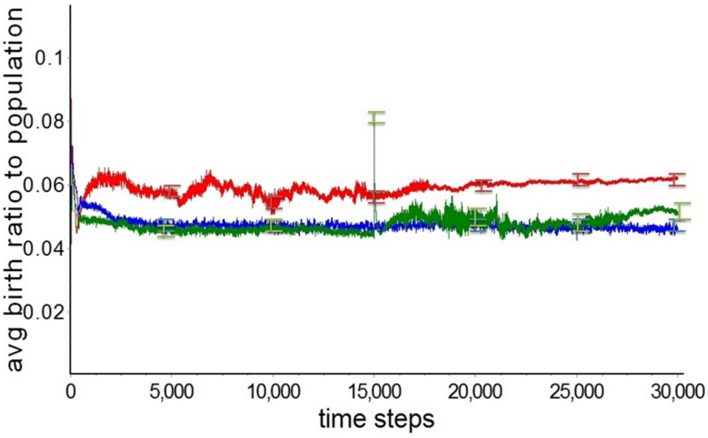


Figure 7.8: The average prey birth ratio to population. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone.

the release of predators. The difference in the ratio of individuals to the total prey population that chose the reproduction action (Figure 7.7) is statistically significant (paired t-test between case A and B , p-value <0.001) (paired t-test between first and second half of case C, p-value <0.001). The difference in the birth rate, which is the ratio of newly born prey individuals to the total prey individuals (Figure7.8) is also statistically significant (paired t-test between case A and B , p-value <0.001) (paired t-test between first and second half of case C, p-value <0.001). These results are consistent with literature reporting a reduction in reproduction levels due to predation pressure ([108]; [110]). Birth rates and death rates were significantly higher in case B runs than in case A runs (Table 7.2). While birth and death rates were relatively stable for case A prey, they increased gradually for case B prey. The consequences on population growth were very different for the two prey groups (Fig. 7.4. Increasing birth rates resulted in population growth for low-risk prey, while the high-risk prey population was relatively stable over time. This indicates that predators exerted population control on high-risk prey. Death rates are reported here as total causes of death, although EcoSim records three different causes of death (predation, low energy, and old age) for the case A condition, and two different causes for the case B condition (low energy and old age). Causes of death were similar in ranking for both prey groups: for both prey groups the most frequent cause of death was energy and the least frequent cause of death was old age. This indicates that the indirect effects of predators on prey energy budgets may have played a role in the selection of individuals, which consequently influenced population structure.

7.3.2 Predation risk-foraging trade off

Prey also showed clear differences in their foraging behaviours across cases A, B and C (Figure 7.10). These differences illustrate that prey in EcoSim exhibit a tradeoff between the foraging and predation risk, such that prey spend more time and energy on foraging when there is no predation risk compared to when there is high predation risk. Differences in foraging behaviour between case A, B and C were statistically significant (paired t-test between case A and B, p-value <0.001) (paired t-test between first and second half of case C, p-value <0.001). In addition, in the first 15,000 time steps there was a close match between case C and A in terms of foraging, but during the second half of the simulation (after predators removal in case C) there was a close match between case C and B for the next 15,000 time steps (Figure 7.10). Thus, high-risk prey foraged less frequently than low-risk prey but were more successful at eating grass. The main factor accounting for the difference in successful eating may have been the movement of prey because this affected the frequency of encountering food in the world, and therefore its consumption and availability in the world, which differed significantly between case A and case B prey as seen in Fig. 7.5 and Fig. 7.9. Predation risk-induced prey movement was in fact beneficial to the recruitment of grass because it prevented overgrazing of local food sources, as can be seen by a steady grass density for case A and steady level of grass in the world, and a dramatic decline in grass density and grass level case B. The effect of predation-sensitive foraging and overgrazing for case B prey had a snowballing effect, because if a local patch of grass becomes depleted then grass cannot regrow in a cell (i.e. when all 8 of its neighboring cells are grass-empty). As a result, grass became progressively depleted and prey became less successful in eating actions. Overall, the trends in grass abundance between the two prey groups show evidence for indirect (cascading) effects in EcoSim as predators kept grass levels stable for case A prey, while their absence led to a dramatic decline in the long run for case B prey. In EcoSim we observe both risk-averse and risk-prone prey foraging behaviours owing to the grass diffusion mechanism. Different parts of the world have different grass distribution, which allows the presence of both behaviours. In case B runs the grass level (Figure 7.5) and movement (Figure 7.12) curves show more fluctuations than case A curves. Prey adjust their behaviour from risk-aversive when food was relatively abundant in the world, to risk-prone when food became scarce.

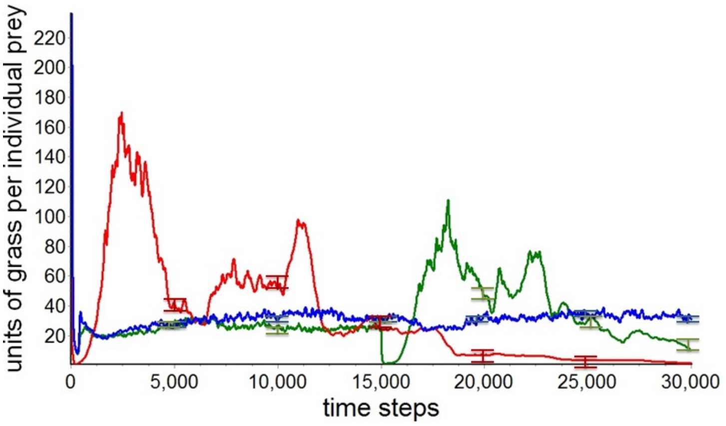


Figure 7.9: Grass density is the total units of grass divided by the number of prey population in the world. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone .

The same phenomenon has been seen in nature after the removal of predators. For example, [93] showed that removing predators from a prey population of guppies caused prey density to increase, enhance guppy trophic traits, and led to an increase in food consumption. Case C shows that even after a relatively short evolutionary time period the prey evolved adaptive behaviours to the absence of predators by allocating resources (time and energy) away from risk-aversion and to foraging. Similarly, ([93]) pointed to a specific trade-off between escape ability and foraging ability.

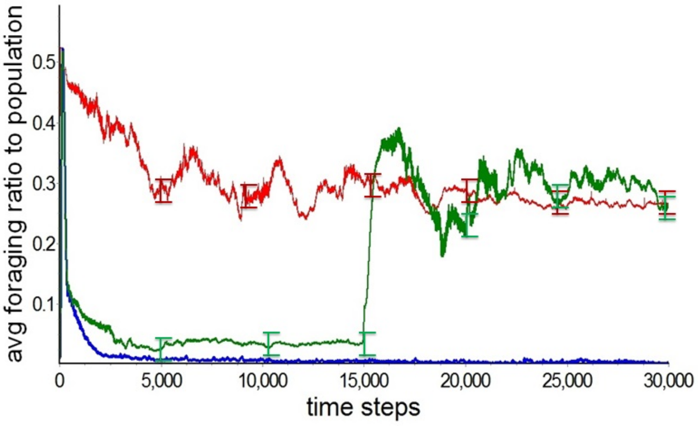


Figure 7.10: The average prey foraging ratio to population. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone.

Does the tradeoff affects the energy of prey and allocation of energy to reproduction?

The consumption of prey's energy budget is presented in (Table 7.1). On average, high-risk prey (case A) had significantly greater energy budgets of 329.33 units (std of 13.74 units) compared to 282.28 units (std of 9.33 units) for low-risk prey (case B) (paired t- test between prey energy budget in high and low risk prey runs, p-value <0.001). Furthermore, predators greatly influenced the way in which prey allocated energy to their actions, both in terms of the proportion of their energy budgets they spent on each type of action and also in the trends of energy expenditures throughout evolution. Of all movement actions foraging was the most energy costly action for low-risk prey, while explore and escape were the most energy costly for high-risk prey. These differences indicate unique behavioural adaptations to managing energy budgets.

Energy investment in reproduction also differed significantly (paired t- test between prey reproduction in case A and B runs, p-value <0.001), with high-risk prey consuming more energy for reproduction in proportion to their average energy budgets. For both case A and B the energy spent on reproduction represents the largest single portion of energy budgets. The result that high-risk prey spend more

Table 7.1: Average energies (with std between brackets) consumed by prey while engaging in each type of action, expressed as a percent of their average energy budgets. (Note: since these are average values over the entire duration of the runs they do not add up to 1). Eating represents net energy gained from eating, including the energy spent on the eating action itself. Due to computational limitations we present here the average energy consumed for each action for four runs of case A and B prey only. *Includes all successful and failed actions.

Actions	case A runs	case B runs
Average energy (units)	329.3(13.74)	282.28(9.33)
Escape	12.89(0.66)	9.76(4.29)
Foraging*	10.78(0.9)	15.79(8.36)
Socialize*	10.44(0.68)	12.86(9.12)
Explore	12.9(0.69)	9.92(4.04)
Wait	9.84(0.68)	6.13(4.17)
Eating	52.29(1.33)	66.08(10.29)
Reproduction*	50.65(1.04)	46.15(9.6)
Percentage of transmitted energy	30.75(1.68)	28.14(1.75)

energy on reproduction compared to low-risk prey, while counterintuitive, arises from the adaptation of low-risk prey to achieve higher reproduction rates. This strains their energy budgets since reproduction is very energetically expensive, and also results in a significant reduction of energy investment in offspring compared to high-risk prey (paired t-test between percentage of energy transmitted to offspring in case A and B runs, p-value <0.001).

As can be seen from the energy expenditures, evading predators through the escape action is much less energetically expensive than reproducing. Thus, the antipredator behaviour escape may be a good temporary strategy for survival from an energy perspective, but it still makes substantial decreases in overall energy so less will be available for future reproduction. This indicates that there is a tradeoff between the energy required to stay alive (by avoiding predators and foraging) and the energy required to reproduce.

7.3.3 Adjustment of reproduction strategies in response to predation risk

Previous results showed that predation risk has indirect fitness consequences through its effects on prey energy budgets and on energy allocation to offspring. Predation risk also directly influences prey fitness through its effect on reproductive behaviour (Table 7.2). Case A prey attempted reproduction much less frequently compared to case B prey, so predators effectively suppressed reproduction. Furthermore, while the reproduction action was chosen at a relatively constant frequency in case A it was chosen with increasing frequency in case B, indicating a behavioural adaptation through evolution (Fig. 7.7). Reproduction actions were also more often successful for case B, which may have occurred because reproduction success depends on genetic distance between the two prey parents; since low-risk prey moved less they were more likely encounter more genetically similar individuals and reproduce successfully.

Although low-risk prey gained more energy from more frequent successful eating actions than high-risk prey (Table 7.1), the energy budgets of prey were affected more by expenditures than by energy intake. Not only is reproduction much

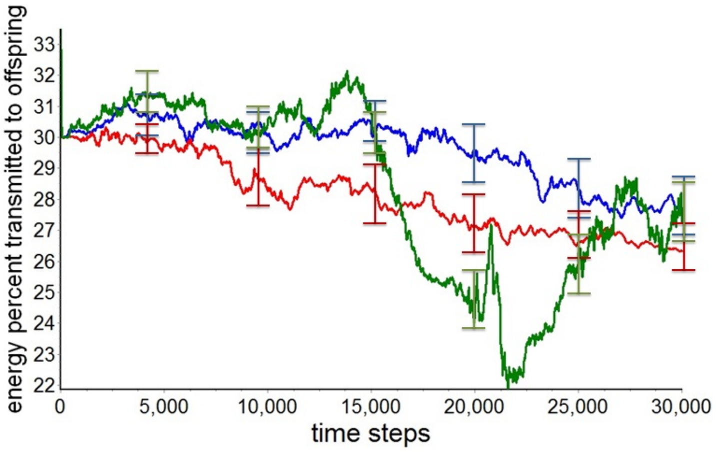


Figure 7.11: Percentage of maximum allowed transmitted energy to offspring at birth. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone.

Table 7.2: Demographic characteristics as a percent (%) of population for case A and B runs. Standard deviations in brackets. *Includes successful and failed actions. **Death rate includes all causes of death (being eaten by predators (for case A prey only), energy depletion, and old age).

Actions	case A runs	case B runs
Reproduction	18.83(6.5)	31.88(21.4)
Birth rate	4.79(1.4)	5.99(3.9)
Death rate**	4.78(1.5)	5.99(3.9)
Average age	15.3(0.7)	12.2(1.2)
Prey population	180608.2(47942)	70265.73(39228)

more energetically expensive than movement, but the birth rate was also greater for case B than case A. Higher birth rates also took an energetic toll on parental investment of energy as case B prey invested smaller and decreasing amounts of energy in their offspring throughout evolution compare to case A prey (significant paired t-test between percentage of energy transmitted to offspring in case A and B runs, p-value <0.001) (Fig. 7.11). By contrast, case A prey invested more amount of energy in their offspring throughout evolution. This shows that there is a tradeoff between reproduction rates (Fig. 7.7) and parental investment (Fig. 7.11) due to limitations in energy budgets. Empirical studies have shown situations when prey can allocate energy differently depending on the level of predation risk, whether by investing more resources in antipredator defenses or by reducing other energetically expensive activities such as reproduction ([115]; [116]), or by other means. For example, one study showed that Daphnia carinata (a small water crustacean) grow defensive crests when predation risk is high which results in less energy being invested into offspring production, as seen by their smaller eggs relative to D. carinata grown in low-risk environments ([117]). In this study Daphnia did not change food intake, so there was a clear indication that energy was allocated differently as a function of predation risk.

7.3.4 Prey movement

Prey movement was strongly affected by the presence/absence of predation risk (Table 7.3, Fig. 7.12 and Fig. 7.13). The total movement (sum of all escape, forage, socialize plus exploration actions) of prey was significantly higher throughout evolution in the simulation when predators were present (paired t-test between prey movement for case A and B runs, p-value <0.001). On the other hand, low-risk prey evolved to be progressively more sedentary as they chose the waiting action more frequently over the movement actions as seen in Fig. 7.13 and Table7.3 . Predation pressure mainly affects the proportion of the prey population that chose escape and foraging actions, and only slightly affects exploration and socialization actions; whereas escape and exploration were chosen more frequently among high-risk prey, foraging and socialization were chosen more frequently among low-risk

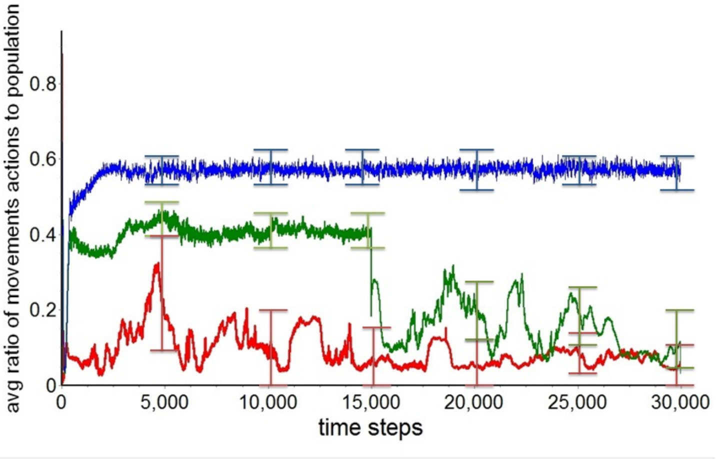


Figure 7.12: The proportion of the total prey population runs that chose a movement action (escape, foraging, socialize or explore) at each time step. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone. Total movement is the sum of all four of these actions, and includes all successful and failed foraging and socialization actions.

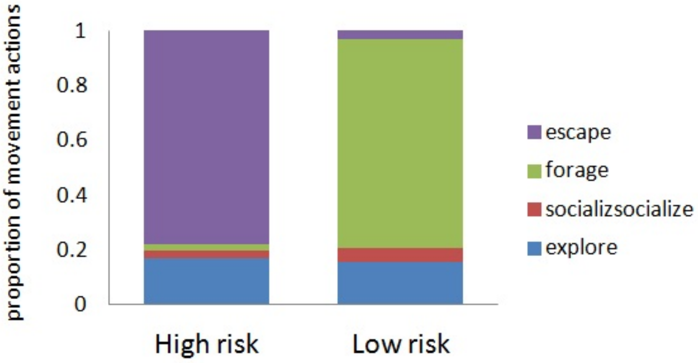


Figure 7.13: The proportion of all movement actions that high (case A) and low(case B) risk prey runs spent. The bars show a behavioural tradeoff between time spent foraging and time spent responding to predation risk through escape actions.

Table 7.3: Average frequency of movement actions of all prey as a percent (%) of the total population of prey, for case A and B prey except for Speed which is expressed as number of cells. Standard deviations in brackets (std).*Includes all successful and failed attempts for each action.

Actions	case A runs	case B runs
Escape	44.92(10.7)	1.31(2.5)
Foraging*	1.25(0.9)	29.43(17.6)
Socialize*	1.71(1)	1.97(1.7)
Explore	9.73(6.2)	6.12(8)
Total movement*	57.62(14.9)	38.84(24)
Wait	0.2(0.14)	4.6(5)
Speed in no. cells	2.75(0.1)	0.76(0.28)

prey (Fig. 7.13). All four movement actions differed significantly between case A and case B runs (p-value<0.001). This indicates that EcoSim models very strong non-lethal predatory effects on prey populations, particularly through the risk-induced escape behaviour of prey.

Thus, prey co-evolving with predators experienced higher total movement activity than prey surviving with no predators, primarily due to the escape activity induced by predators throughout the simulation. We were also interested in measuring the speed with which the prey moved and whether predation played a role in the change of speed of movement. Figure 7.14 shows that prey not only reduced their total movement activity in terms of average number of movement actions chosen, but they also reduced their speed of movement, i.e. the average number of cells moved in a time step, after predators were removed in case C. The difference is statistically significant (paired t-test between case A and B, p-value <0.001) (paired t-test between first and second half of case C, p-value <0.001). An overview of the distribution of species reveals that individuals show a strong clustering distribution with circular or spiral shapes in the presence of predators. These spiral waves are a common property of predator-prey models ([118]; [119];

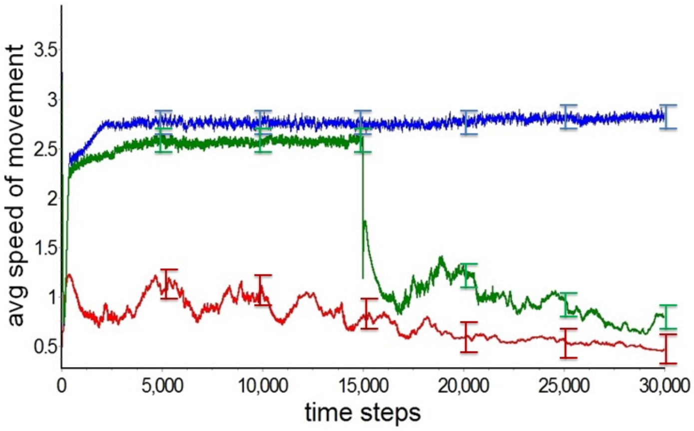


Figure 7.14: The average prey speed. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone.

[120]) and arise due to the fact that prey near the wave break have the tendency to escape from the predators by moving sideways. A subpopulation of prey then finds itself in a region relatively free from predators, and in this predator-free zone prey individuals start dispersing rapidly forming a circular expanding region ([60]). Prey are also more compact and dense in case A than case B. The difference in the average number of individuals per cell for case A (4.6 with std 0.25) and B (3 with std 0.5) was statistically significant (paired t-test between case A and B, p-value <0.001) (paired t-test between first and second half of case C, p-value <0.001). Thus, prey density per cell was greater when coevolving with predators, indicating greater clustering for case A prey than case B prey. Both small-scale movements and large-scale movements have indeed been observed to occur under high predation pressure in natural predator-prey systems ([105]). Furthermore, the type of movement changes that occur in the presence of predators may be context-specific and may vary temporally depending on the stage of the predation event. For example, increased prey movement may take the form of dispersal when they are displaced by predators, such as when insects enter streams in the presence of predators ([121]; [122]; [123]; [124]). In aquatic systems prey movement has also been observed where zooplankton dive down in the water column when predation risk is high near the surface ([125]; [126]). Overall, these examples present evidence that prey respond to predators in a risk- averse manner

by increasing movement to nearby locations that may be perceived as safe. In EcoSim, small-scale escape actions of prey coevolving with predators amounted to global changes in the spatial distribution of prey. Such large-scale movements in predator-prey systems have also been observed in nature. For example, [127] and [128] showed that migrating elk experienced reduced risk of predation when they traversed entire landscapes. These studies suggest that prey activity in wild populations may increase on large scales as a result of predation, and that prey can benefit from this with increased survival ([127]; [128]).

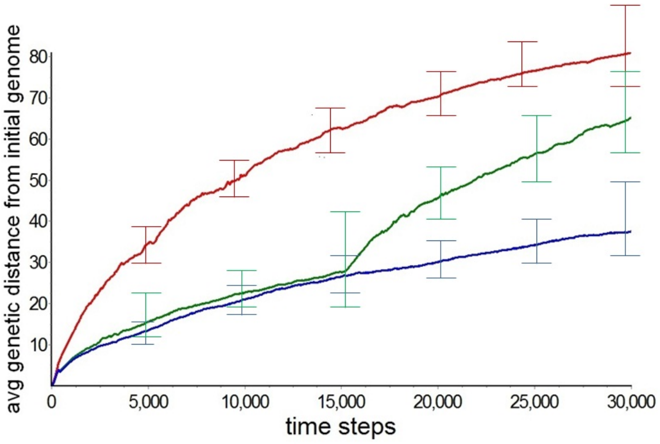


Figure 7.15: The average prey genetic distance from initial FCM map. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone.

7.4 Predator consequences on prey genomic evolution

Prey activity proved to be the critical behaviour modeled in EcoSim, affecting not only prey population dynamics but also their course of evolution. The main action affected by the presence/absence of predation risk was escape, and this effect was exaggerated since places of refuge or hiding are not meled in the EcoSim world. The urge to escape from predators and to move throughout the world decreased

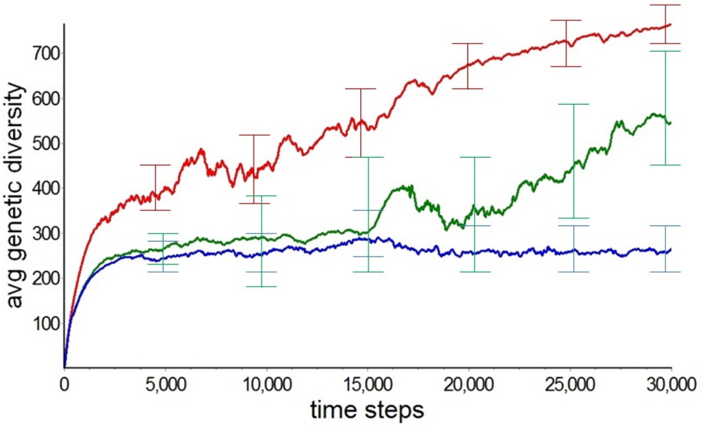


Figure 7.16: The average prey genetic diversity (measured by entropy). Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone.

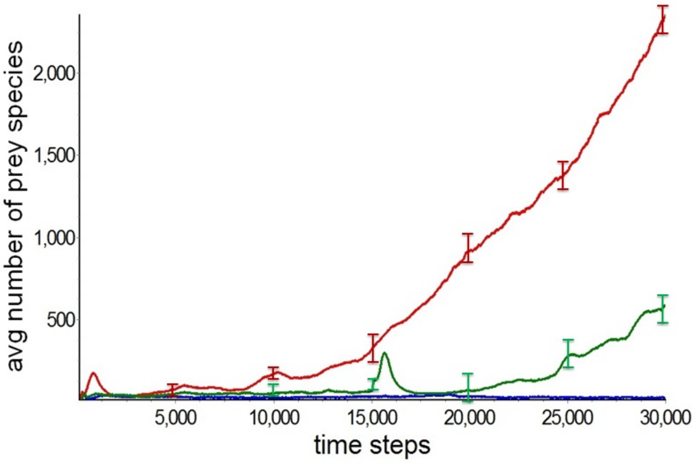


Figure 7.17: The average number of prey species. Blue line represents case A (high-risk prey), red line represents case B (low-risk prey), and green line represents case C where predators and prey co-evolve till 15,000 time steps, and then prey evolve alone.

dramatically in the absence of predators, indicating this was the major pressure on prey movement. Reduced movement by prey also had an indirect effect on gene flow, which then affected genetic diversity and in return species richness, i.e the total number of species in a given time step. Increased prey species diversity and average genetic distance from initial prey genome has previously been reported in EcoSim ([60]) and its origin proven as a consequence of introducing obstacle in the world, which restricted prey movement and reduced gene flow. In the present study we observed that the average genetic distance seen in Figure 7.15 was significantly more divergent in prey evolving in the absence of predators than prey coevolving with predators relative to the initial genomes (paired t-test between case A and B, p-value <0.001) (paired t-test between first and second half of case C, p-value <0.001).

There was a significantly higher genetic diversity (Figure 7.16) among prey in case B and the second half of case C than in case A and the first half of case C (paired t-test between case A and B, p-value <0.001) (paired t-test between first and second half of case C, p-value <0.001). The results for genetic diversity are also mirrored in a significantly greater prey species richness in case B (paired t-test between case A and B, p-value <0.001) (paired t-test between first and second half of case C, p-value <0.001), indicating that the number of prey species in the absence of predators is much higher than the number of prey species in the presence of predators, although the speciation threshold is the same and the population is smaller. This suggests that the speciation rate is directly proportional to the restriction of movement and therefore to gene flow between populations. As the population is greater in case A than B (Figure 7.17) it follows that the number of individuals per species decreases when predators were removed.

Our results show that predator removal decreased prey movement, which in return reduced the gene flow among prey individuals as occurs prior to allopatric speciation. The reduction in gene flow resulted in increased species richness, genetic diversity and average genetic distance from the initial prey genome. The predation effect on species richness has been previously studied by computer simulations using an individual-based modeling. [129] showed over a variety of parameter settings that the duration of the coexistence of two prey phenotypes dramatically

Table 7.4: Accuracy for training the classifier using 10 fold cross validation, accuracy of validating the classifier using the validation set, and the number of rules produced from the model

Classification of	Train Accuracy	Validation Accuracy	No of rules
case A & B	93.7%	93.04%	7
case A & C	93.7%	94.69%	6
case B & C	87.9%	88.4%	7

decreased as the number of predator individuals increased, suggesting that prey species diversity should decrease in the presence of predators. In addition, [60] reported an increase of species richness after introducing obstacles in the world which reduced individuals movement and restricted gene flow. Although, most empirical studies have shown that predators are better indicators of high prey species diversity, more research in this area is encouraged.

7.5 Genomic classification and statistical analysis

Having shown that predation has a behavioural effect on prey, we are interested to study if this effect extends to the genetic level. To analyze the genomes we extracted the average FCM (genome) of every existing prey species at the last 5,000 time steps of each simulation run (from time step 25,000 till 30,000) so the genetic characteristic of prey would be well developed and formed. A total of 2000 average FCM's of prey species were collected from each of case A, B and C runs. Also, since each run has unique environmental conditions which have an impact on the genome values, combining data from 10 different runs is important to prove that the results are not due to the difference between runs but rather due to the main factor we are studying the presence or absence of predators. From a total of 390 potential genes we neglected the 125 initial genes to eliminate any bias in the results. Therefore, an instance of the learning process consists of the

265 genes, each gene being an attribute in the classifier, making the feature space highly dimensional. We then combined the 20,000 FCM's coming from each of the separate 10 runs, and labeled them with a class label determining which group they belong to. Table 7.4 specifies the JRip classification accuracies along with the number of rules produced. We also calculated a one-way analysis of variance (ANOVA) test to give global insight to how these three classes (A, B and C) differ. The ANOVA test revealed a strong intraclass correlation (p-value <0.001) which indicates that the genomes of the three classes were also significantly different.

The high classification accuracy regardless of the combination of 10 different runs supports the claim that predators have an effect on prey population's genomic evolution, as it shows that the classification model is able to capture important characteristic of the genomes in each configurations that lead to accurate predictions on novel data. These results also indicate that the different conditions and environments of each run did not influence the genomes as profoundly as the presence/absence of predators; thus, we found that predation plays a strong role in genomic evolution of prey. There was also a noticeable difference in the accuracy of classifying case B and case C (slightly less) relative to the other two tests, indicating that these cases are more similar to each other than to case A, although a good classifying accuracy (88.4%) was still attained. Different IF THEN rules are learned by the JRip algorithm to predict the group any genome belongs to. The rules JRip produced also shed light on the most important genes that would differentiate between prey belonging to any of the three cases. Therefore, they add more semantics to the results and help identify the genes that contribute more to the differentiation between genomes belonging to different cases and how these genes work together. These are the 20 rules produced from the three classifiers. A negative number in a rule indicates a concept inhibiting another and corresponds to a threshold for a gene (e.g. IF energy is low → wait >-0.0946 means IF the concept 'energy' is low decreases the concept 'wait' by more than a -0.0946 * 'energy value'). A positive number in a rule indicates a concept exciting another and corresponds to a threshold for a gene (e.g. IF predator is close → wait >0.0705 means IF the concept 'predator is close' increases the concept 'wait' by more than 0.0705 * 'predator is close value'). The hit ratio is the percentage of instances the rule applies to.

7.5.1 Classifying instances belonging to case A and C

IF (predator is close → wait >0.0705) AND (energy is low → wait >-0.0946) AND (hunger → search for partner <0.0187) AND (energy is high → eat >-0.197) THEN instance belongs to case C with hit ratio = 18.53%, accuracy = 99.95%

IF (no local partner → eat >0.059) AND (local food is high → fear <0.036) AND (no local partner → eat >0.098) THEN instance belongs to case C with hit ratio = 11.07%, accuracy = 99.77%

IF (predator is far → search for partner <-0.022) AND (local food is high → socialize <0.0067) AND (local food is low → reproduce >-0.019) THEN instance belongs to case C with hit ratio = 9.64%, accuracy = 96.68%

IF (escape → satisfaction >0.0897) AND (local food is low → reproduce >-0.1093) AND (energy is high → escape >-0.0091) THEN instance belongs to case C with hit ratio = 4.52%, accuracy = 99.89%

IF (reproduce → reproduce >0.1026) THEN instance belongs to case C with hit ratio = 3.6%, accuracy= 73.06%

All other instances belong to case A with hit ratio = 52.64% and accuracy = 93.11%

7.5.2 Classifying instances belonging to case B and C

IF (predator is close → wait >0.0686) AND (energy is high → exploration <-0.0521) AND (local partner exists → hunger >0.0601) THEN instance belongs to case C with hit ratio = 7.75%, accuracy = 99.16%

IF (predator is far → search for partner <-0.0003) AND (friend is close → exploration <-0.2451) AND (local partner exists → socialize <-0.0651) AND (no local partner → eat <0.2304) THEN instance belongs to case C with hit ratio = 7.52%, accuracy = 99.93%

IF (reproduce → socialize >0.0227) AND (no local partner → escape >-0.0029) THEN instance belongs to case C with hit ratio = 16.5%, accuracy = 83.43%

IF (wait → eat <-0.0523) AND (friend is close → eat >0.0783) AND (local food is high → search for partner <0.0041) AND (friends are far → escape >-0.1403) THEN instance belongs to case C with hit ratio = 3.87%, accuracy = 99.48%

IF (nuisance → search for partner <-0.2155) AND (predator is far → eat <-0.0043) AND (socialize → wait <0.1013) THEN instance belongs to case C with hit ratio = 4.21%, accuracy = 99.13%

IF (local food is low → escape >0.088) AND (no local partner → exploration <-0.0936) AND (exploration → socialize >-0.0208) THEN instance belongs to case C with hit ratio = 4.96%, accuracy = 94.69%

All other instances belong to case B with hit ratio = 55.19% and accuracy = 85.61%

7.5.3 Classifying instances belonging to case A and B

IF (no local partner → escape <-0.085) THEN instance belongs to case B with hit ratio = 14.28%, accuracy = 99.46%

IF (predator is close → search for food <-0.0349) AND (fear → search for partner <0.5) THEN instance belongs to case B with hit ratio = 8.14%, accuracy = 100%

IF (no local partner → eat >0.0649) AND (energy is high → fear <0.6258) THEN instance belongs to case B with hit ratio = 7.43%, accuracy = 100%

IF (eat → reproduce <-0.0476) THEN instance belongs to case B with hit ratio = 6.04%, accuracy = 97.02%

IF (satisfaction → satisfaction >0.2597) AND (search for food → search for partner <0.108) THEN instance belongs to case B with hit ratio = 4.67%, accuracy = 100%

IF (reproduce → exploration >0.1221) AND (nuisance → fear <0.1044) THEN instance belongs to case B with hit ratio = 3.7%, accuracy = 99.9%

All other instances belong to case A with hit ratio = 55.72% and accuracy = 89.26%

7.5.4 Extracted semantics from rules

Studying the genes that appeared in the 20 produced rules revealed that these genes are associated with one of the three behaviours; movement, eating and reproduction. All together there were 44 genes appearing in the rules (see Figure 7.18). 42% of genes are related to concepts affecting movement such as exploration, escape, wait (not to move) and socialize; 21% of the genes in the rules are related to the eating actions (eat and search for food) and the internal concept hunger; 25% of genes are concerned with reproduction actions such as search for partner and reproduce; and the rest (12%) went to genes linked to fear and satisfaction

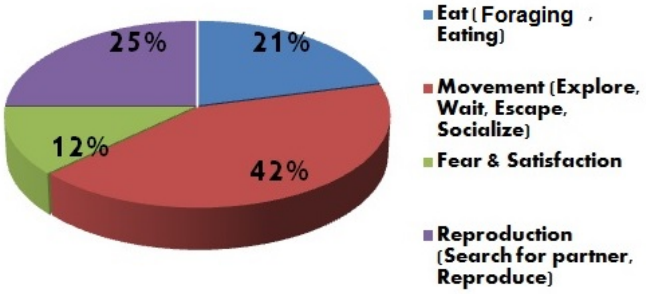


Figure 7.18: Distribution of genes appearing in the JRip classifier.

and were only found in rules associated with case A. These results indicate that movement behaviour, along with eating/foraging and reproduction, are sufficient and capable of discrimination between genomes of prey evolving with no predation pressure and prey coevolving with predators. The genes constructing the rules enforce the previous hypothesis about the effect predation has on reproduction, movement/speed and eating/foraging prey behaviours as all of these could be differentiated between the cases (see Figure 7.7, 7.12, 7.14 and 7.10). Thus, prey evolved not only distinct behaviours but also had accompanying genetic changes associated with each case.

7.5.5 Predator introduction

Although it has become increasingly important to understand the role of top predators in ecosystems due to the catastrophic effects of human activity on these species ([130]), there has been little acknowledgement of the evolutionary consequences of predator removal on prey populations, and thus minimal attention has been given to the special problems for prey that might accompany their reintroduction ([111]; [131]; [132]). Simulations offer a cost-effective method of studying long-term trends in predator-prey systems that are often too costly or impractical in field studies, and can offer important insights for biological systems. Here we show that contemporary evolution of prey in the absence of predation risk may result in those prey species being highly susceptible if predators are later introduced. The initially introduced predator population was of a consistent ratio to the prey population in all the runs, thus eliminating any bias in these results. When predators were added in case D at time step 15,000 none of the 10 replicates were able to stabilize for long and the predator population quickly died out from low energy as they consumed all the local prey (Table 7.5). Another reason for predator extinction was the highly dispersed distribution of the prey population before predators were added, which made it more difficult for predators to find food. Introducing predators at time step 5,000 resulted in one stable run out of ten which successfully maintained predators-prey coevolution as in case A. Predator introductions at even earlier time steps were affected by the prey population sizes. For example, at time step 2,500 the prey population was quite low, averaging only 4,500 individuals for 10 runs (see Figure 7.4). The low prey population adds to the difficulty of reaching stability in the system after introducing predators because food is more scarce and dispersed. Introducing predators to naive prey as early as time step 200 increased the percentage of stable runs. When we introduced predators at time step 10 there was full survival and stability of all 10 runs; however, in these runs the evolution of prey behaviour and genetic change were negligible. The success of these final 10 runs verifies that the previous results of predator introductions were not due to any instability in our simulation or choice of parameters, but rather to the genetic and behavioural changes in prey and their contemporary evolution in the absence of predators.

We found that introducing predators early in the evolutionary history of prey populations resulted in a higher chance of survival for predators than when prey have evolved for a longer period of time in the absence of predators. Our results show that the reason for such extinctions is a combination of both the evolutionary background of these well adapted prey, which accounts for their broad dispersal and decreased escape tendencies, and the short-term effects of predators on prey, which led to a high mortality that overwhelmed the capacity of the local population to reproduce and adapt to such a sudden change. It has been previously shown that the reintroduction of a predator into a site from which predators have been absent for a long time could potentially cause the local extinction of prey guppies ([133]).

Table 7.5: Results of predators introduction in different time steps

time step of pred. intro.	num of successful runs	# runs where all prey died	# runs where all pred. died	avg time steps of survival
15000	0	0	10	190
5000	1	0	9	273
2500	0	2	8	134
200	4	2	4	76
10	10	0	0	n/a

These results also show that the amount of time required for these evolutionary changes to occur is relatively brief, on the order of 4-20 generations. [132] studied the consequences of wolves and brown bear recolonization in Scandanavia and North America after an absence of 50-130 years and showed that this led to the substantial increase in the mortality of both young and adult moose prey. The impact of predator reintroduction on prey populations is highlighted in studies like that of [131] who showed that as much as 20% of the elk population living in the headwaters of the Madison River in Yellowstone National Park was killed following reintroduction of wolves in just one winter season ([134]). However, our results suggest that predators are also at risk and caution against focusing on prey behaviour alone, a point previously made in [135]. Furthermore, it is well established that predator-prey interactions can lead to dynamic population cycles ([95]; [96]; [94]), and our results consistently show that predator populations were affected by the decline in local prey (in addition to their spatial distribution), which ultimately led to predator extinction. Taken together, these results suggest that stability in the predator-prey model could be achieved if predators are either introduced gradually into prey populations, allowing prey to adaptively evolve to face predation pressure by adopting greater antipredator responses (escape actions), or by introducing predators early enough in the evolutionary history of prey so that they do not first become maladapted to the absence of predators.

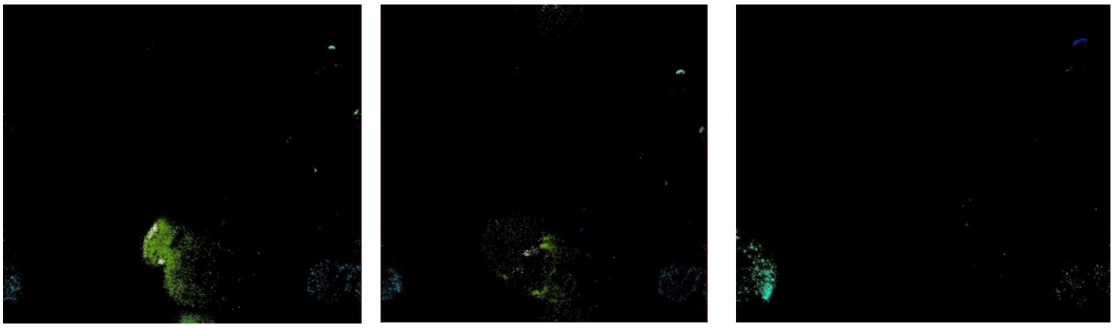


Figure 7.19: Snapshot of the world in one of the case D runs (predators are introduced to naive prey) at time step 5000. Preys show more spatial dispersal and white patches represent predators. In the first figure the numbers of prey and predator individuals are 31502 and 1008 respectively. In the second figure at time step 5030 prey decreased to 9301 and predators were 1502. At time step 5112 prey decreased to 5725 and predators were all dead.

7.6 Conclusion

The novelty of EcoSim is based on modeling in depth the complex tri tophic interaction between different components of an evolutionary dynamic ecosystem. The used genomic model with integrated semantics helps us study the effect predation has on the genomic evolution of prey, a study which has not been given attention before. We were able to show that predators play a large role in stabilizing predator-prey systems and exert a great influence on prey evolution. By comparing prey activity across various types of predator-prey systems (cases A-D) we found that predators significantly influenced prey reproduction, total movement and foraging behaviours. Specifically, predators suppressed reproduction and birth rates, increased total movement, and decreased foraging activity of prey under predation risk. The population level consequences of these behavioural changes were observed as significant demographic differences, with the population being initially greater and more stable for the predator-prey coevolution system, and initially smaller but steadily increasing for the prey alone system.

We found that the energy of prey in EcoSim is affected in several ways which differ between high- and low-risk prey groups. First, we found that energy budgets differed between prey groups: energy was greater for high-risk prey than for low-risk prey. Second, we found that prey allocated energy to survival and fitness differently between prey groups: high-risk prey engaged in more movement actions and had lower birth rates, yet they spent proportionally more energy on reproduction compared to low-risk prey because they engaged in fewer of energetically costly reproductions (Fig. 7.7. Conversely, low-risk prey engaged in fewer movement actions but had greater birth rates, resulting in a smaller energy budget and spending proportionally less energy on reproduction. Third, we found that maternal investment in offspring (in terms of offspring energy at birth) differed greatly: high-risk prey invested a constant amount of energy in their offspring, while low-risk prey invested a decreasing amount of energy in their offspring, which may reflect a tradeoff with the higher birth rate since reproduction is very energetically expensive on parents. These results also indicate that prey evolved life history characteristics that were not present at the start of the simulation. Low-risk prey tended to act as 'energy maximizers'([136]; [137]): their fitness benefits from foraging increased with foraging effort, so they foraged as much as possible to meet energetic needs for continued reproduction while spending less time doing non-foraging activities. By contrast, high-risk prey tended to behave as 'time minimizers' ([136]; [137]): they spent more time doing activities other than foraging such as escaping and socializing, and their energy benefits to fitness was fixed, as seen by a relatively constant reproduction rate and near constant parental investment of energy.

Our results suggest that the importance of movement for prey evolution in EcoSim cannot be underestimated. The different spatial distributions of prey seen in the presence/absence of predators results from their behaviour model which does not impose movement, but facilitates much more movement when prey experience predation risk through their only antipredator behaviour, escaping. It is important to note that while many prey species seek refuge when they perceive high risk in nature, and hence reduce their movement, in the present study our model did not allow for this type of antipredator response but will be considered in the future.

EcoSim provides an unparalleled study environment for such analysis. We found energy budgets for prey in the high- and low-risk part of this study were greatly affected by energy expenditures, specifically due to energetically expensive reproduction actions, and in part due to movement activity. Previous studies (e.g. [138]) may not have considered the combined effects of predation on foraging success and fitness, but our study suggests that when considering multiple behaviours that are mediated by predation risk the net effect on energy budgets may not be equivalent across different risk regimes, and instead they are context-specific and can sway in the direction of the more energetically costly event. In EcoSim this means that predators actually offset the costs of reproduction for high-risk prey by causing them to engage in fewer energetically expensive reproductions compared to low-risk prey. We find this in nature when some organisms postpone reproduction during periods of high predation risk ([139]; [140]), a strategy which may be beneficial in the long run under variable risk conditions ([108]).

The time scale of our study differs from the time scale considered in behavioural ecology. EcoSim extends beyond ecological time scales by allowing prey to adaptively evolve to different predation regimes, giving rise to distinct gene pools of prey. To date there have been a few long-term studies spanning from a few years to a few decades that have tracked prey population changes in response to predators, particularly with snowshoe hare in the Yukon and elk in Yellowstone National Park ([141]; [142]), but the majority of studies look at much shorter times that range from minute-by-minute responses of prey to predators, to responses spanning days, weeks and months. Although ecological and evolutionary processes can simultaneously affect the outcomes of predator-prey systems, few studies have considered the evolutionary consequences of adaptations to predation risk ([96]). In EcoSim, higher activity in the short run by choosing escape actions more frequently resulted in higher fitness and steady population size in the long-run. Escape activity is commonly seen in nature, particularly when there are no refuges to hide. However, a short-term strategy of predator avoidance may not always be optimal in the long-run. This was demonstrated clearly in [143] who showed that when prey used a recurrent vigilance strategy coupled with low foraging activity in high-risk situations this led to long-term decreases in fitness, most likely caused by high risk of starvation and reduced energy intake. The strategy for risk reduction that is most beneficial in the short run depends on multiple factors, and there is no simple relationship to the long-term effects.

Many authors stress the importance of more integrative work in the study of predator-prey systems. Behavioural ecologists are particularly adamant about the need to consider multiple compensatory behaviours when determining the fitness consequences of antipredation behaviour ([101]). Such a research agenda would work best if the full range of organisms' compensatory behaviours were understood in terms of their energetic costs and benefits, but this has not been done ([144]). Furthermore, since energy budgets directly impact fitness through the allocation of energy to reproduction and to offspring, essentially all behaviours that affect energy intake or expenditure can be thought of as beneficial or harmful to fitness. Thus, behaviours should evolve to optimize energy budgets for allocation of energy to fitness, even under conditions of variable predation. Yet, among all the current models of prey behaviour, for example the Risk Allocation Hypothesis ([105]) there are limitations in accounting for the trait-mediated effects of predators on prey foraging, activity and reproduction in predicting community-level outcomes. Other ecologists have stressed that empiricists and theoreticians must work together to build models that make accurate predictions of evolutionary phenomenon based on small-scale and short-term ecological data, particularly by acknowledging trait-mediated effects ([144]). The complexity of ecosystems and behaviours of organisms present real problems for ecologists looking to understand how nature works, but by integrating information from multiple fields and multi-dimensional studies this may be possible. EcoSim offers a flexible environment with many levels of complexity where large-scale evolutionary phenomenon can be examined. We have shown that it can be applied to study how prey behave in response to predators and how this behaviour affects their energy budgets and fitness, and we have seen that it generates ecologically accurate results comparable to many empirical studies. As we have made theoretical predictions for the evolutionary outcomes of predator-prey interactions in terms of energy and fitness, the next step would be to compare these results to more long-term observational studies in nature.

A key advantage of our study was the ability to investigate not only prey behavioural changes associated predation risk, but also to peer inside their genomes to study genomic changes associated with predation risk. Ecologists have to deal with noisy, high-dimensional data that are often non-linear and do not meet the assumptions of conventional statistical procedures. To overcome this problem, we used JRip, a supervised rule learner classifier, to classify unknown average prey species genomes as belonging to case A, B or C. JRip successfully classified average genomes with high accuracy and statistical significance, proving that these prey genomes were indeed distinct. JRip also produced meaningful rules that added more semantics to the classification process by emphasizing the significance of three key traits in classifying genomes: movement, reproduction and foraging, which were also significantly different in our results across the different cases. Simulations offer a cost-effective method of studying long-term trends in predator-prey systems that are often too costly or impractical in field studies, and can offer important insights for biological systems. Our results suggest that survival of predators following their introduction is sensitive to the evolutionary state of their prey. We have shown that removal of predators causes contemporary evolution of prey after time intervals as short as 200 time steps (about 15 generations). The length of time after initial predator removal and the strength of predation pressure following predator reintroduction will impact not only preys' ability to adapt but also the predator's survival

Chapter 8

Summary, Conclusions and Future work

8.1 Summary

I first reviewed the existing literature on simulation modeling used in Alife in Chapter 2. The use of individual base modeling in ecology and how it differs from equation based modeling was also presented. The chapter also gave an overview of several existing computational ecosystems. In the following chapter 3 a detailed overview of the used evolutionary ecosystem simulation was presented. The chapter describes the different entities, state variables and scale of the system. A process overview along with the design concept is also given. A neutral model of EcoSim, which acts as the randomized version is also presented. Chapter 4 gives a brief description about the data analysis tools used throughout the thesis including different classification algorithms, feature selection and rule extraction. The use of Shannon entropy as a measure of genetic diversity along with the used fitness metric was presented.

In Chapter 5 the genomic evolution of EcoSim was examined by comparing it to the genomic evolution of EcoSim's neutral model. Classification was used to successfully predict the average species fitness giving only the new genes with high accuracy. A feature selection step was performed which shed a light on the most important genes affecting fitness and well being of species. The JRip rule learner was used which produced meaningful rules that added more semantics to the results. In chapter 6 the use to Shannon entropy as a measure of genetic diversity was presented. The genetic diversity of all existed species along with their fitness was calculated. The main aim was to test the biological hypothesis that genetic diversity affects species fitness. The Spearman's cross correlation was used to measure the effect genetic diversity has on fitness. In a further step these correlation results were used as a class variable that was predicted by a built classifier. The JRip rule learner was used to extract rules that added more understanding of the conditions leading a positive or negative correlation between genetic diversity and fitness. These rule were studied which concluded that large species tend to decrease their diversity in order to increase their fitness. On the other hand, relatively smaller species correlate positively between genetic diversity and fitness.

The ecological effects of predator removal and its consequence on prey behaviour have been widely investigated in Chapter 7. We investigated the effects of predation risk on prey energy allocation and fitness. We investigated foraging-predation risk tradeoffs and the energy-fitness consequences on prey at an unprecedented scale of population size, evolutionary time, and behavioural complexity. Results revealed that prey energy budgets, life history traits, allocation of energy to movements and fitness-related actions differed greatly between prey subjected to low-predation risk versus high-predation risk. High-predation risk suppressed prey foraging activity, increased movement, and decreased reproduction relative to low-risk. A classification algorithm was used to demonstrate the difference between genomes belonging to prey coevolving with predators and prey evolving in the absence of predation pressure which is a new direction is studying predator prey interactions.

8.2 Conclusions

We chose to use a simulation modeling complex ecosystem that was designed to address broad theoretical questions in ecology and evolutionary biology. The novelty of this simulation comes from the fact that it models in great details the interaction between hundreds of thousands of behaviourally and genetically distinct individuals in a tri-trophic dynamic complex ecosystem over large evolutionary time scale allowing for Darwinian natural selection to emerge. Our novel individual-based behavioural model of evolution allows us to approach pertinent biological questions through a system with higher complexity than has previously been published in literature. The integrated semantics of the genomic model helps in reasoning and analyzing the causes and factors behind any evolved new behaviour. As we do not have a predefined fitness function, all the emerged new behaviours and genes are all due to the evolutionary machinery of EcoSim governed by individuals adaptation and natural selection. From the genetic perspective, mutations and natural selection, through the course of evolution in EcoSim enforce the emergence of new traits and disappearing of others. Thoughts out studies we were interested in validating the evolutionary machinery of EcoSim, gain more understanding about the individuals evolving behaviour, and investigate some ecological theories.

Due to the enormous amount of produced data in each simulation run, detecting and interpreting emerging behaviours in EcoSim was a challenging task. Machine learning techniques were used throughout the thesis to assist in the analysis. To our knowledge this is the first time machine learning techniques have been used to analyze results obtained from an evolutionary ecosystem simulation. The thesis validated the evolutionary machinery of EcoSim through several studies. A comparison between the genetic behaviour of EcoSim and its neutral model demonstrated the difference between emerging behaviours based on natural selection and randomness. In another study the possible new genes which arise in EcoSim were able to predict the average species fitness with high accuracy which demonstrates the important linkage between the genes values, and therefore the behavioral model, and the fitness of individuals. This is a strong validation that the evolutionary model is governed by natural selection even though no fitness functions exist to give preference on a behaviour over another. The predefined genes were not included in this study to eliminate any bias in the results. A further rule extraction step was taken to understand which of these new genes correlates more with the fitness and well being of species. As hypothesized that interacting entities of EcoSim were sufficient to model natural ecosystems we were able to observe same emerging phenomena's as those arising in natural ecosystems. The high correlation between average species genetic diversity and average species fitness found in EcoSim, is also another major concept in ecology which emerged in our model. The rules extracted from the classification model emphasized the most important features affecting the correlation values. These rules added more semantics to our model and highlights important factors that can further assist biologists in their studies.

In another study EcoSim was used to examine the dynamic relation between predators and prey which is also an active study in biological literature. Here we tested the role of predator removal on the contemporary evolution of prey traits such as movement, reproduction and foraging. We showed that the contemporary evolution of prey behaviour owing to predator removal is also accompanied by prey genetic change. We employed machine learning methods, now recognized as holding great promise for the advancement of our understanding and prediction of ecological phenomena, showing that the genetic changes in prey are due to predator removal. Moreover, we showed that predator introductions to naive prey might be destabilizing if prey have evolved and adapted to the absence of predators which is a significant result and concern for conservative biology. Our results suggest that both predator introductions and predator removal from an ecosystem have widespread effects on the survival and evolution of prey by altering their genomes and behaviour, even after relatively short time intervals. The different results found in the predator prey dynamics study was supported by similar results found in biological literature which further proves the validity of our predator prey model and add emphasis on the high capabilities of EcoSim in modeling and studying many ecological phenomena.

8.3 Future direction

EcoSim offers a flexible environment with many levels of complexity where large-scale evolutionary phenomenon can be examined. We have shown that it can be applied to study how prey behave in response to predators and how this behaviour affects their energy budgets and genomic evolution, and we have seen that it generates ecologically accurate results comparable to many empirical studies. In continuation in this direction we would like to extend EcoSim in order to integrate alternative antipredator behaviours into the system. Currently the prey's only antipredator reaction is fleeing in the opposite direction of the predator. Studying hiding and camouflage is of great interest to us. Predator-induced escape actions by prey are dependent on the prey's ability to accurately detect the predators and on the distance between the prey and the predator. More generally, prey usually have several behavioural options for how to respond to prey when they know they are at risk of predation. They have to make a decision whether to hide and reduce activity, or run away and escape activity, and their choice depends on many factors including availability of nearby refuges, their hunger state and the quality of the food at their current location. A future step would be to give the prey in EcoSim the option of hiding or escaping, but the hiding action should probably be used more frequently than escape, and their activity should also decrease, in order for our result to match most of the literature.

References

[1] M. Begon, C. R. Townsend, J. L. Harper, Ecology: from individuals to ecosystems, Wiley-Blackwell, 2009.

[2] M. A. Bedau, Artificial life: organization, adaptation and complexity from the bottom up, Trends in cognitive sciences 7 (11) (2003) 505–512.

[3] C. G. Langton, C. Taylor, J. Farmer, S. Rasmussen, Artificial life ii: Proceedings of the workshop on artificial life, february 1990, sante fe, new mexico. redwood city, calif (1992).

[4] R. Costanza, Toward an operational definition of ecosystem health, Ecosystem health: New goals for environmental management (1992) 239–256.

[5] J. Von Neumann, A. W. Burks, et al., Theory of self-reproducing automata.

[6] N. Wiener, Cybernetics; or control and communication in the animal and the machine.

[7] J. Holland, Emergence: from chaos to order (1998).

[8] E. Bonabeau, Agent-based modeling: Methods and techniques for simulating human systems, Proceedings of the National Academy of Sciences of the United States of America 99 (Suppl 3) (2002) 7280–7287.

[9] P. Cariani, Emergence and artificial life, Artificial Life II 10 (1991) 775–798.

[10] C. G. Langton, et al., Artificial life, Addison-Wesley Publishing Company Redwood City, CA, 1989.

[11] N. H. Packard, Intrinsic adaptation in a simple model for evolution, Artificial life 141.

[12] J. Holland, Outline for a logical theory of adaptive systems, Journal of the ACM (JACM) 9 (3) (1962) 297–314.

[13] J. Holland, Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence, U Michigan Press, 1975.

[14] J. Holland, Escaping brittleness: the possibilities of general purpose learning algorithms applied to parallel rule-based system, Machine learning (1986) 593–623.

[15] D. L. DeAngelis, W. M. Mooij, Individual-based modeling of ecological and evolutionary processes, Annual Review of Ecology, Evolution, and Systematics (2005) 147–168.

[16] M. Niazi, A. Hussain, Agent-based computing from multi-agent systems to agent-based models: a visual survey, Scientometrics 89 (2) (2011) 479–499.

[17] V. Grimm, Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future?, Ecological modelling 115 (2) (1999) 129–148.

[18] A. Łomnicki, Individual-based models and the individual-based approach to population ecology, Ecological modelling 115 (2) (1999) 191–198.

[19] V. Grimm, S. F. Railsback, Individual-based modeling and ecology, Princeton university press, 2005.

[20] S. F. Railsback, V. Grimm, Agent-based and individual-based modeling: a practical introduction, Princeton University Press, 2011.

[21] H. H. Shugart, et al., A theory of forest dynamics. The ecological implications of forest succession models., Springer-Verlag, 1984.

[22] K. A. Rose, E. S. Rutherford, D. S. McDermot, J. L. Forney, E. L. Mills, Individual-based model of yellow perch and walleye populations in oneida lake, Ecological Monographs 69 (2) (1999) 127–154.

[23] B. H. Letcher, J. A. Priddy, J. R. Walters, L. B. Crowder, An individual-based, spatially-explicit simulation model of the population dynamics of the endangered red-cockaded woodpecker,¡ i¿ picoides borealis¡/i¿, Biological Conservation 86 (1) (1998) 1–14.

[24] S. Rasmussen, C. Knudsen, R. Feldberg, M. Hindsholm, The coreworld: Emergence and evolution of cooperative structures in a computational chemistry, Physica D: Nonlinear Phenomena 42 (1) (1990) 111–134.

[25] T. Ray, An approach to the synthesis of life, Artificial Life II (1991) 371–408.

[26] T. Taylor, J. Hallam, Replaying the tape: An investigation into the role of contingency in evolution, Proceedings of Artificial Life VI, Los Angeles (1998) 256–265.

[27] A. N. Pargellis, Digital life behavior in the amoeba world, Artificial Life 7 (1) (2001) 63–75.

[28] C. Adami, C. T. Brown, Evolutionary learning in the 2d artificial life system avida, in: Artificial life IV, Vol. 1194, Cambridge, MA: MIT Press, 1994, pp. 377–381.

[29] C. Ofria, C. Wilke, Avida: a software platform for research in computational evolutionary biology, Artificial Life 10 (2004) 191–229.

[30] C. Nehaniv, Measuring evolvability as the rate of complexity increase (2000) 55–57.

[31] T. Taylor, Creativity in evolution: Individuals, interactions and environments, Creative evolutionary systems (2001) 79–108.

[32] M. A. Fortuna, L. Zaman, A. P. Wagner, C. Ofria, Evolving digital ecological networks, PLoS computational biology 9 (3) (2013) e1002928.

[33] C. Adami, C. Ofria, T. C. Collier, Evolution of biological complexity, Proceedings of the National Academy of Sciences 97 (9) (2000) 4463–4468.

[34] C. Adami, Sequence complexity in darwinian evolution, Complexity 8 (2) (2002) 49–56.

[35] R. Lenski, C. Ofria, R. Pennock, C. Adami, The evolutionary origin of complex features, Nature 423 (2003) 139–144.

[36] C. H. Chandler, C. Ofria, I. Dworkin, Runaway sexual selection leads to good genes, Evolution 67 (1) (2013) 110–119.

[37] J. Holland, The echo model, Proposal for a Research Program in Adaptive Computation, Santa Fe Institute.

[38] O. J. Schmitz, G. Booth, Modelling food web complexity: the consequences of individual-based, spatially explicit behavioural ecology on trophic interactions, Evolutionary Ecology 11 (4) (1997) 379–398.

[39] L. Yaeger, Computational genetics, physiology, metabolism, neural systems, learning, vision, and behavior or polyworld: Life, Artificial Life III (1992) 263–298.

[40] L. S. Yaeger, How evolution guides complexity, HFSP journal 3 (5) (2009) 328–339.

[41] L. Yaeger, V. Griffith, O. Sporns, Passive and driven trends in the evolution of complexity, arXiv preprint arXiv:1112.4906.

[42] J. Murdock, L. S. Yaeger, Genetic clustering for the identification of species, in: Proceedings of the 13th annual conference companion on Genetic and evolutionary computation, ACM, 2011, pp. 29–30.

[43] A. Channon, R. Damper, Evolving novel behaviors via natural selection, Proceedings of Artificial Life VI, Los Angeles (1998) 384–388.

[44] A. Channon, R. Damper, Towards the evolutionary emergence of increasingly complex advantageous behaviours, Internaltion Journal of Systems Science 7 (31) (2000) 843–860.

[45] M. A. Bedau, E. Snyder, N. H. Packard, A classification of longterm evolutionary dynamics, in: Proc. of Art. Life VI, MIT Press, 1998, pp. 228–237.

[46] A. Channon, Passing the alife test: activity statistics classify evolution in geb as unbounded, in: 6th European Conference on Advances in Artificial Life, Springer-Verlag, 2001, pp. 417–426.

[47] M. Komosiński, S. Ulatowski, Framsticks: Towards a simulation of a nature-like world, creatures and evolution, in: Advances in Artificial Life, Springer, 1999, pp. 261–265.

[48] M. Smith, Using massively-parallel supercomputers to model stochastic spatial predator-prey systems, Ecological Modelling 58 (1) (1991) 347–367.

[49] V. Volterra, Variations and fluctuations of the number of individulas in animal species living together, Animal Ecology 3 (1931) 409–448.

[50] G. Bell, The evolution of trophic structure, Heredity 99 (5) (2007) 494–505.

[51] W. Yamaguchi, M. Kondoh, M. Kawata, Effects of evolutionary changes in prey use on the relationship between food web complexity and stability, Population ecology 53 (1) (2011) 59–72.

[52] C. Scogings, K. Hawick, H. James, Tools and techniques for optimisation of microscopic artificial life simulation models, in: Proceedings of the Sixth IASTED International Conference on Modelling, Simulation, and Optimization, Gabarone, Botswana, 2006, pp. 90–95.

[53] C. J. Scogings, K. A. Hawick, Modelling predator camouflage behaviour and tradeoffs in an agent-based animat model, in: Proc. IASTED International Conference on Modelling and Simulation (MS2013), no. CSTN-184, WorldComp, 2013, pp. 802–032.

[54] C. J. Scogings, K. A. Hawick, Introducing a gestation period of time-delayed benefit into an animat-based artificial life model, in: Proc. 12th IASTED Int. Conf. on Artificial Intelligence and Applications(AIA13), IASTED, 2013, pp. 43–50.

[55] R. Gras, D. Devaurs, A. Wozniak, A. Aspinall, An individual-based evolving predator-prey ecosystem simulation using fuzzy cognitive map as behavior model, Artificial Life 15 (4) (2009) 423–463.

[56] M. Mashayekhi, B. MacPherson, R. Gras, A machine learning approach to investigate the reasons behind species extinction, Ecological Informatics 20 (2014) 58–66.

[57] B. Kosko, Fuzzy cognitive maps, Int. Jornal of Man-Machine Studies (1986) 65–75.

[58] A. Aspinal, R. Gras, K-means clustering as a speciation method within an individual-based evolving predator-prey ecosystem simulation, Proc. of the Active Media Technology, LNCS (2010) 318–329.

[59] L. Seuront, F. Schmitt, Y. Lagadeuc, D. Schertzer, S. Lovejoy, S. Frontier, Multifractal analysis of phytoplankton biomass and temperature in the ocean, Geophysical Research Letters 23 (24) (1996) 3591–3594.

[60] A. Golestani, R. Gras, M. Cristescu, Speciation with gene flow in a heterogeneous virtual world: can physical obstacles accelerate speciation?, Proceedings of the Royal Society B: Biological Sciences 279 (1740) (2012) 3055–3064.

[61] V. Biktashev, J. Brindley, A. Holden, M. Tsyganov, Pursuit-evasion predator-prey waves in two spatial dimensions, Chaos: An Interdisciplinary Journal of Nonlinear Science 14 (4) (2004) 988–994.

[62] J. Mallet, A species definition for the modern synthesis, Trends in Ecology & Evolution 10 (7) (1995) 294–299.

[63] W. B. Sherwin, Entropy and information approaches to genetic diversity and its expression: Genomic geography, Entropy 12 (2010) 1765–1798.

[64] C. Bergstrom, M. Lachmann, Shannon information and biological fitness, in: In IEEE Information Theory Workshop, 2004, pp. 50–54.

[65] C. Shannon, A mathematical theory of communication, Bell Systems Technical Journal (1948) 379–423.

[66] C. Kampichler, R. Wieland, S. Calmé, H. Weissenberger, S. Arriaga-Weiss, Classification in conservation biology: a comparison of five machine-learning methods, Ecological Informatics 5 (6) (2010) 441–450.

[67] S. C. Larson, The shrinkage of the coefficient of multiple correlation., Journal of Educational Psychology 22 (1) (1931) 45.

[68] F. Mosteller, D. L. Wallace, Inference in an authorship problem, Journal of the American Statistical Association 58 (302) (1963) 275–309.

[69] R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, 1993.

[70] L. Breiman, Random forests, Machine Learning 45 (2001) 5–32.

[71] W. W. Cohen, Y. Singer, A simple, fast, and effective rule learner, in: Proceedings of the National Conference on Artificial Intelligence, JOHN WILEY & SONS LTD, 1999, pp. 335–342.

[72] M. A. Hall, Correlation-based feature subset selection for machine learning, Ph.D. thesis, University of Waikato, Hamilton, New Zealand (1998).

[73] J. Pearl, Heuristics: intelligent search strategies for computer problem solving.

[74] Q. Yang, E. Salehi, R. Gras, Using feature selection approaches to find the dependent features, in: 10th International Conf. on Artificial Intelligence and Soft Computing, LNAI, 2010, pp. 487–494.

[75] Y. Saeys, I. Inza, P. Larraaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (19) (2007) 2507–2517.

[76] E. Salehi, R. Gras, Efficient eda for large optimization problem via constraining the search space of models, in: GECCO'11, ACM, 2011, pp. 73–74.

[77] D. R. Cutler, T. C. E. Jr, K. H. Beard, A. Cutler, K. Hess, J. Gibson, J. Lawler, Random forests for classification in ecology, Ecology 88 (2007) 2783–2792.

[78] R. Diaz-Uriarte, S. A. de Andres, Gene selection and classification of microarray data using random forest, BMC Bioinformatics 7 (1) (2006) 3.

[79] I. Witten, E. Frank, Data Mining- Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, USA, 2000.

[80] W. Cohen, Fast effective rule induction, in: 12th International Conference on Machine Learning, 1995, pp. 115–123.

[81] D. Reed, R. Frankham, Correlation between fitness and genetic diversity, Conserv Biology 17 (2003) 230–237.

[82] J. Markert, D. Champlin, R. Gutjahr-Gobell, J. Grear, A. Kuhn, T. McGreevy, A. Roth, M. Bagley, D. Nacci, Population genetic diversity and fitness in multiple environments, BMC Evolutionary Biology 10 (2010) 1471–2148–10–205.

[83] S. Vandewoestijne, N. Schtickzelle, M. Baguette, Positive correlation between genetic diversity and fitness in a large, well-connected metapopulation, BMC Biology 6 (2008) 1741–7007–6–46.

[84] J. Oostermeijer, M. van Eijck, J. den Nijs, Offspring fitness in relation to population size and genetic variation in the rare perennial plant species gentiana pneumonanthe (gentianceae), Oecologia 97 (1994) 289?296.

[85] S. Siegel, Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, New York, 1956.

[86] S. L. Lima, Nonlethal effects in the ecology of predator-prey interactions, Bioscience (1998) 25–34.

[87] E. E. Werner, S. D. Peacor, A review of trait-mediated indirect interactions in ecological communities, Ecology 84 (5) (2003) 1083–1100.

[88] W. Cresswell, Non-lethal effects of predation in birds, Ibis 150 (1) (2008) 3–17.

[89] G. Fussmann, M. Loreau, P. Abrams, Eco-evolutionary dynamics of communities and ecosystems, Functional Ecology 21 (3) (2007) 465–477.

[90] D. M. Post, E. P. Palkovacs, Eco-evolutionary feedbacks in community and ecosystem ecology: interactions between the ecological theatre and the evolutionary play, Philosophical Transactions of the Royal Society B: Biological Sciences 364 (1523) (2009) 1629–1640.

[91] M. L. Pace, J. J. Cole, S. R. Carpenter, J. F. Kitchell, Trophic cascades revealed in diverse ecosystems, Trends in Ecology & Evolution 14 (12) (1999) 483–488.

[92] D. Reznick, M. J. Butler IV, H. Rodd, Life-history evolution in guppies. vii. the comparative ecology of high-and low-predation environments, The American Naturalist 157 (2) (2001) 126–140.

[93] E. P. Palkovacs, B. A. Wasserman, M. T. Kinnison, Eco-evolutionary trophic dynamics: loss of top predators drives trophic evolution and ecology of prey, PloS one 6 (4) (2011) e18879.

[94] M. H. Cortez, S. P. Ellner, Understanding rapid evolution in predator-prey interactions using the theory of fast-slow dynamical systems, The American Naturalist 176 (5) (2010) E109–E127.

[95] P. A. Abrams, The evolution of predator-prey interactions: theory and evidence, Annual Review of Ecology and Systematics (2000) 79–105.

[96] T. Yoshida, L. E. Jones, S. P. Ellner, G. F. Fussmann, N. G. Hairston, Rapid evolution drives ecological dynamics in a predator–prey system, Nature 424 (6946) (2003) 303–306.

[97] S. Creel, D. Christianson, Relationships between direct predation and risk effects, Trends in Ecology & Evolution 23 (4) (2008) 194–201.

[98] E. L. Preisser, D. I. Bolnick, The many faces of fear: comparing the pathways and impacts of nonconsumptive predator effects on prey populations, PLoS One 3 (6) (2008) e2465.

[99] J. L. Verdolin, Meta-analysis of foraging and predation risk trade-offs in terrestrial systems, Behavioral Ecology and Sociobiology 60 (4) (2006) 457–464.

[100] E. L. Preisser, D. I. Bolnick, M. F. Benard, Scared to death? the effects of intimidation and consumption in predator-prey interactions, Ecology 86 (2) (2005) 501–509.

[101] J. Lind, W. Cresswell, Determining the fitness consequences of antipredation behavior, Behavioral Ecology 16 (5) (2005) 945–956.

[102] M. Edmunds, Defence in animals: a survey of anti-predator defences, Longman Harlow, 1974.

[103] P. H. Harvey, P. J. Greenwood, Anti-predator defence strategies: some evolutionary problems, Behavioural ecology: an evolutionary approach. Blackwell, Oxford (1978) 129–151.

[104] S. L. Lima, L. M. Dill, Behavioral decisions made under the risk of predation: a review and prospectus, Canadian Journal of Zoology 68 (4) (1990) 619–640.

[105] S. L. Lima, P. A. Bednekoff, Temporal variation in danger drives antipredator behavior: the predation risk allocation hypothesis, The American Naturalist 153 (6) (1999) 649–659.

[106] M. C. Ferrari, A. Sih, D. P. Chivers, The paradox of risk allocation: a review and prospectus, Animal Behaviour 78 (3) (2009) 579–585.

[107] G. C. Trussell, P. J. Ewanchuk, C. M. Matassa, The fear of being eaten reduces energy transfer in a simple food chain, Ecology 87 (12) (2006) 2979–2984.

[108] U. Candolin, Reproduction under predation risk and the trade–off between current and future reproduction in the threespine stickleback, Proceedings of the Royal Society of London. Series B: Biological Sciences 265 (1402) (1998) 1171–1175.

[109] J. Fontaine, T. Martin, Parent birds assess nest predation risk and adjust their reproductive strategies, Ecology Letters 9 (4) (2006) 428–434.

[110] L. Y. Zanette, A. F. White, M. C. Allen, M. Clinchy, Perceived predation risk reduces the number of offspring songbirds produce per year, Science 334 (6061) (2011) 1398–1401.

[111] T. M. Blackburn, P. Cassey, R. P. Duncan, K. L. Evans, K. J. Gaston, Avian extinction and mammalian introductions on oceanic islands, Science 305 (5692) (2004) 1955–1958.

[112] F. Courchamp, M. Langlais, G. Sugihara, Cats protecting birds: modelling the mesopredator release effect, Journal of Animal Ecology 68 (2) (2001) 282–292.

[113] J. Estes, K. Crooks, R. Holt, Ecological role of predators, Encyclopedia of biodiversity 4 (2001) 280–1.

[114] D. Croll, J. Maron, J. Estes, E. Danner, G. Byrd, Introduced predators transform subarctic islands from grassland to tundra, Science 307 (5717) (2005) 1959–1961.

[115] D. F. Fraser, J. F. Gilliam, Nonlethal impacts of predator invasion: facultative suppression of growth and reproduction, Ecology 73 (3) (1992) 959–970.

[116] B. D. Wisenden, Female convict cichlids adjust gonadal investment in current reproduction in response to relative risk of brood predation, Canadian journal of zoology 71 (2) (1993) 252–256.

[117] M. J. Barry, The costs of crest induction for daphnia carinata, Oecologia 97 (2) (1994) 278–288.

[118] N. F. Otani, A. Mo, S. Mannava, F. H. Fenton, E. M. Cherry, S. Luther, R. F. Gilmour Jr, Characterization of multiple spiral wave dynamics as a stochastic predator-prey system, Physical Review E 78 (2) (2008) 021913.

[119] C. J. Krebs, M. S. Gaines, B. L. Keller, J. H. Myers, R. H. Tamarin, Population cycles in small rodents., Science (New York, NY) 179 (4068) (1973) 35.

[120] J. Bascompte, R. V. Solé, et al., Spatiotemporal patterns in nature.

[121] G. E. Forrester, Influences of predatory fish on the drift dispersal and local density of stream insects, Ecology (1994) 1208–1218.

[122] A. Sih, D. E. Wooster, Prey behavior, prey dispersal, and predator impacts on stream prey, Ecology 75 (5) (1994) 1199–1207.

[123] D. Wooster, A. Sih, A review of the drift and activity responses of stream prey to predator presence, Oikos (1995) 3–8.

[124] S. J. McCauley, L. Rowe, Notonecta exhibit threat-sensitive, predator-induced dispersal, Biology letters 6 (4) (2010) 449–452.

[125] S. Bollens, B. Frost, Predator-induced diet vertical migration in a planktonic copepod, Journal of Plankton Research 11 (5) (1989) 1047–1065.

[126] P. Dawidowicz, C. J. Loose, Metabolic costs during predator-induced diel vertical migration of daphnia., Limnology and Oceanography 37 (8) (1992) 1589–1595.

[127] M. Hebblewhite, E. H. Merrill, Multiscale wolf predation risk for elk: does migration reduce risk?, Oecologia 152 (2) (2007) 377–387.

[128] M. Hebblewhite, E. H. Merrill, Trade-offs between predation risk and forage differ between migrant strategies in a migratory ungulate, Ecology 90 (12) (2009) 3445–3454.

[129] S. Merilaita, Frequency-dependent predation and maintenance of prey polymorphism, Journal of evolutionary biology 19 (6) (2006) 2022–2030.

[130] D. R. Strong, K. T. Frank, Human involvement in food webs*, Annual Review of Environment and Resources 35 (2010) 1–23.

[131] R. A. Garrott, J. A. Gude, E. J. Bergman, C. Gower, P. White, K. L. Hamlin, Generalizing wolf effects across the greater yellowstone area: a cautionary note, Wildlife Society Bulletin 33 (4) (2005) 1245–1255.

[132] J. Berger, J. E. Swenson, I.-L. Persson, Recolonizing carnivores and naive prey: conservation lessons from pleistocene extinctions, Science 291 (5506) (2001) 1036–1039.

[133] D. N. Reznick, C. K. Ghalambor, K. Crooks, Experimental studies of evolution in guppies: a model for understanding the evolutionary consequences of predator removal in natural communities, Molecular Ecology 17 (1) (2008) 97–107.

[134] P. White, R. Garrott, Yellowstones ungulates after wolves–expectations, realizations, and predictions, Biological Conservation 125 (2) (2005) 141–152.

[135] S. L. Lima, Putting predators back into behavioral predator–prey interactions, Trends in Ecology & Evolution 17 (2) (2002) 70–75.

[136] T. W. Schoener, Optimal size and specialization in constant and fluctuating environments: an energy-time approach., in: Brookhaven symposia in biology, Vol. 22, 1969, p. 103.

[137] T. W. Schoener, Theory of feeding strategies, Annual review of ecology and systematics 2 (1971) 369–404.

[138] G. C. Trussell, C. M. Matassa, B. Luttbeg, The effects of variable predation risk on foraging and growth: Less risk is not necessarily better, Ecology 92 (9) (2011) 1799–1806.

[139] C. Magnhagen, Reproduction under predation risk in the sand goby, pomatoschistus minutes, and the black goby, gobius niger: the effect of age and longevity, Behavioral Ecology and Sociobiology 26 (5) (1990) 331–335.

[140] H. Ylönen, Vole cycles and antipredatory behaviour, Trends in ecology & evolution 9 (11) (1994) 426–430.

[141] C. Krebs, S. Boutin, R. Boonstra, A. Sinclair, J. Smith, M. Dale, R. Turkington, Impact of food and predation on the snowshoe hare cycle, AAAS, 1995.

[142] S. Creel, J. Winnie Jr, B. Maxwell, K. Hamlin, M. Creel, Elk alter habitat selection as an antipredator response to wolves, Ecology 86 (12) (2005) 3387–3397.

[143] M. Watson, N. J. Aebischer, W. Cresswell, Vigilance and fitness in grey partridges perdix perdix: the effects of group size and foraging-vigilance trade-offs on predation mortality, Journal of Animal Ecology 76 (2) (2007) 211–221.

[144] B. Bolker, M. Holyoak, V. Krivan, L. Rowe, O. Schmitz, Connecting theoretical and empirical studies of trait-mediated interactions, Ecology 84 (5) (2003) 1101–1114.

Vita Auctoris

Marwa Fouad Khater was born in 1976 in Cairo, Egypt. She obtained her bachelor degree in Computer Science from Kuwait University in Kuwait in 1998. She obtained her MSc degree in Computer Science from Ain Shams University in Cairo, Egypt in 2003. She joined University of Windsor and obtained her PhD degree in Computer Science in 2014.