

Masters Thesis Report

School: Courant Institute of Mathematical Sciences, NYU, New York

Department: Computer Science

Inavamsi B Enaganti (ibe214@nyu.edu)

Thesis Advisor : Prof Bhubaneshwar Mishra

Thesis Reader : Prof Bari Anasse

Title : Cooperation and Deception in multi-agent signaling

Keywords: Signaling Games, Cooperation, Deceptive Signaling, Iterated Prisoner's Dilemma, Evolutionary Game Theory, Dove-Hawk Game, Predator-Prey Model, Biomimicry, Artificial Intelligence, Machine Learning

Honor Code: All code and work done in this thesis is completely original.

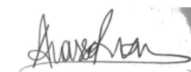
Code : <https://github.com/inavamsi/Thesis>

Thesis Reader 1
Prof Bhubaneshwar Mishra



5-12-20

Thesis Reader 2
Prof Bari Anasse



6-3-20

Contents

Contents 2

Brief 3

Introduction 4

Basic Theory and Definitions 5

Agent 5

Signaling Game 5

Prisoner's Dilemma (PD) 6

Iterated Prisoner's Dilemma (IPD) 6

Stochastic Iterated Prisoner's Dilemma 7

Some standard agents 7

Literature Survey 8

The Game 9

Replicator Equations 9

Equilibrium 9

The Game 12

Memory Agents 12

Agent Estimation and Optimal action in Noisy environment 15

Brief on Approaches 17

Tournaments and Replication without Mutation (Approach 1) 18

Tournaments 18

Replicator Equations 20

Asexual Replication with mutation (Approach 2.1) 21

Simulation 21

Invasion 22

Other Parameters 25

Mutation Rate and effects 27

Genetic Algorithms (Approach 2.2) 29

Everything GA 30

Network(Approach 3) 31

A Naive Network 31

Perceptron memory 34

Costly signaling 35

Comparing Agents 36

Recommender-Verifier (Approach 4) 38

Global Reputation 38

Community Utility 39

Deceptive Signaling between Predator and Prey 41

Mullers mathematical model revised 42

Dynamics of a Predator-Prey model in the presence of Mimicry 44

Generalized Lotka Volterra equations for multi predator – multi prey model 44

Mimicry in a simple two predator- two prey model 45

Generalizing mimicry in a two predator – two prey model 50

Some interesting scenarios in Bio-mimicry 54

Model 55

Ecosystem 55

Prey 56

Predator 57

Signal Drift 59

Interesting Phenomena in 1-Dimensional signaling 61

Pooling Equilibrium 61

Polymorphic Mimicry 62

Mimicry can be bad to both species 63

Interesting Phenomena in 2-Dimensional signaling 65

Drift in 2-Dimensions 65

Switching directions 66

Conclusion 68

Future Work 68

References 69

Brief

‘Through cooperation deception arises. To beat deception, cooperation is required.’

We aim to study cooperation and deception in a system with multiple agents through utility and signaling. We start with the classic standard for cooperation namely the ‘Prisoner’s Dilemma’ and then we move on to the ‘Iterated Prisoner’s Dilemma’ which we treat as an iterated version of a signaling Game. This is because the previous actions of an agent are a signal to the opponent about the agent’s type. We then move on to Bio-mimicry and deception where we study dynamics and interesting phenomena that arise due to signaling between Predator and Prey. Cooperation and deception are two sides of a coin and it is imperative to understand both of them as we develop better and more efficient Artificial Intelligence systems.

Introduction

In today's world where humanity has barely scratched the surface of Artificial Intelligence(AI), there is a need for efficient mechanisms that can incentivize, analyze, and verify the evolution of AI. There will be drastic impacts on the structure of society with the evolution of AI. AI will be constantly breaking new frontiers with great impacts on human life. From simple ecosystems to complex interactions we can optimize conditions of survival and evolution.

In the eternal war to strive for optimality and relevance, the world can be envisioned a huge system of interactions which can be modeled as games. When each agent competes for optimality we do not see an optimal system but an evolving system that can take any random turn. We see many fascinating and sometimes counter-intuitive developments.

As nature has already shown and has beautifully been put by Richard Dawkins in 'The Selfish gene', agents tend to be selfish. Through the passage of time surprisingly this selfish world gives rise to cooperation as we see the emergence of altruistic traits. Consider two different species. Species A where every agent is selfish and acts based on personal utility. Species B with a sense of community where agents share their resources. During times of prosperity because of the presence of excess resources, we see that through natural selection members of species A become stronger over time. But during special times like the case of resources being grouped in scarce regions we see that only a small percentage of species A will find resources and survive. On the contrary in harsh situations even if a smaller fraction of species B finds the same resources because of sharing, a significantly larger amount of species B will survive. All in all, species B has an evolutionary advantage on the whole despite members of species A having an evolutionary advantage on the individual level.

This sense of cooperation and trust has been ingrained in humans to a pretty deep extent. We learn to trust Banks, Government, and our fellow citizens from small transactions to large policies and even when it comes to situations where lives are at stake. Due to the ability of reproduction and the emergent phenomena of natural selection, this sense of cooperation has been genetically hardcoded in us and is one of the frontrunner qualities in helping us build civilization and advance. But, in the case of AI, these agents are not limited by concepts of survival and death, and thus to learn cooperation is a very difficult task. With a lack of complete information and also due to the presence of huge amounts of randomness in this world, it is very hard to encode the notion of cooperation. So to create this urge for cooperation we need to understand structures and encodings for cooperation to occur. We further study this with the help of the iconic Prisoner's Dilemma which is the standard for cooperation. We specifically consider a more natural and relevant version called the Iterated Prisoner's Dilemma.

In the above system, there is an innate requirement of trust to form a community. An agent will only want to form a community with those they can trust and identify. This results in the emergence of signals to identify other agents. A signal is a message from Player A to player B indicating or conveying information about their type. As these signals arise there will be malicious agents that find it in their best interest to be deceptive. It is highly improbable that there is no deception at all. Deception is a very natural phenomenon that emerges just like cooperation. We thus study emergent 'cooperation' and with it 'deception' as they are two

sides of the same coin. This involves creating better mechanisms and systems that enable cooperation while understanding the dynamics in the presence of deception. This thesis aims to understand cooperation and deception through the dynamics of a few simple games. This will help us create incentivize and verify rapidly changing technology. We also cover a wide range of signaling dynamics from Iterative signaling(IPD) to emergent phenomena of learning agents in the presence of evolving signals.

Some examples are as follows. Consider two competitors attacking each other's websites through bots. We can understand the dynamics of such situations and figure out the optimal strategy for each company. This has been shown in the 'Dynamics of Predator-Prey model in the presence of mimicry'. Another example is that of fake currency and spambots. How to create effective mechanisms that are more robust and resilient to attacks can be studied through signaling games and understanding of equilibria that arise. We can also understand strange emergent phenomena that cannot be explained through conventional means. We also look at some sub-optimal signaling conventions because of the existence of a pooling equilibrium.

Basic Theory and Definitions

Agent

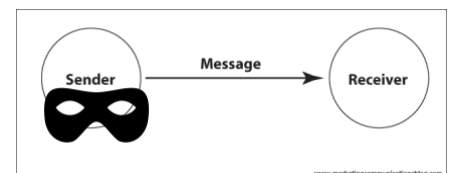
We will constantly refer to agents through the thesis. A very general agent can be considered to consists of the following attributes.

- Body : this can refer to architecture or genetic makeup that can evolve. Moore's Law can be considered to affect the body of an agent.
- Mind : Consists of the learning and action making system. This makes decisions and updates beliefs.
- Memory : Store history and experiences. Accessed by the mind.
- Actions : This consists of taking actions and sending signals
- Senses : Perception that receives signals.

Signaling Game

A Signaling Game plays out as follows

1. Sender(S) has a type t .
2. Sender(S) send a signal/message to Receiver(R).
3. Receiver(R) receives the message.
4. Receiver(R) takes action based on his/her belief about Sender(S).
5. Receiver(R) gets a payoff/reward based on his action and type of Sender(S)



We now look at some common Bayesian equilibrium in Signaling Games.

Pooling Equilibrium : Senders with different types all choose the same message. This means that the sender's message does not give any information to the receiver, so the receiver's beliefs are not updated after seeing the message.

Separating Equilibrium : senders with different types always choose different messages. This means that the sender's message always reveals the sender's type, so the receiver's beliefs become deterministic after seeing the message.

Hybrid Equilibrium : Sender randomizes between Pooling and Separating

After a signaling convention is set. A malicious agent will realize that it is better to send the same signal as a Good agent. But technology might not allow so. There is also a cost attributed to every signal. We shall explore this in later sections.

Prisoner's Dilemma (PD)

Consider the standard game of **Prisoner's Dilemma**

Two criminals A and B are imprisoned. Each prisoner is given the opportunity either to betray the other by testifying that the other committed the crime, or to cooperate with the other by remaining silent. The possible outcomes are:

- If A and B each betray the other, each of them serves two years in prison
- If A betrays B but B remains silent, A will be set free and B will serve three years in prison (and vice versa)
- If A and B both remain silent, both of them will serve only one year in prison (on the lesser charge).

More generally it is represented by a matrix as follows.

Table 1 : Payoff Matrix for Prisoner's Dilemma

	Cooperate	Defect
Cooperate	(R,R)	(S,T)
Defect	(T,S)	(P,P)

The payoffs R, S, T, P denote the Reward for mutual cooperation, Sucker's Payoff, Temptation to defect, Punishment for mutual defection. The following conditions hold.

$T > R > P > S$

The 'dilemma' faced by the prisoner's here is that, whatever the other does, it is better off confessing than remaining silent. Thus resulting in a worse off situation than the case where both remained silent.

Defection is the dominant strategy. The Nash Equilibrium is mutual defection.

Iterated Prisoner's Dilemma (IPD)

We now consider the **Iterated Prisoner's Dilemma**

If two players play prisoner's dilemma more than once in succession and they remember previous actions of their opponent and change their strategy accordingly, the game is called iterated prisoner's dilemma. This game is essentially an iterated a signaling game. Each player knows the move played by the opponent in the previous turns. This history of moves played essentially is a signal from the opponent. This gives information about the type of the opponent. Additionally, we enforce that **$2R > S + T$** . This is to prevent alternating cooperation and defection giving a greater reward than mutual cooperation.

If the game is repeated a finite number of times. The optimal strategy in the last move is to mutually Defect. Inductively we can see that the Nash Equilibrium is to Defect every round. For cooperation to emerge the number of rounds must be unknown.

Stochastic Iterated Prisoner's Dilemma

In an encounter between player X and player Y , X 's strategy is specified by a set of probabilities P of cooperating with Y . P is a function of the outcomes of their previous encounters or some subset thereof.

If P is a function of only their most recent n encounters, it is called a "memory- n " strategy. A memory-1 strategy is then specified by four cooperation probabilities: $P = \{P_{CC}, P_{CD}, P_{DC}, P_{DD}\}$, where P_{ab} is the probability that X will cooperate in the present encounter given that the previous encounter was characterized by (ab) where X played a and Y played b in the previous turn. In our experiments we shall further delve into this.

Some standard agents

Always Defects(ALLD)

ALLD always defects irrespective of what the opponent does.

As a memory-1 strategy it can be written as $[0,0,0,0]$.

Always Cooperates(ALLC)

ALLC always cooperates irrespective of what the opponent does.

As a memory-1 strategy it can be written as $[1,1,1,1]$.

Tit-for-Tat(TFT)

TFT always cooperates on the first move and mimics the previous move of the opponent for the remaining moves.

As a memory-1 strategy it can be written as $[1,0,1,0]$.

Random

This agent cooperates with probability p and defects with probability $1-p$.

As a memory-1 strategy it can be written as $[p,p,p,p]$.

Some other famous variations

GRIM : Cooperates till the opponent defects, then onwards always defects.

TFTT : Tit-for-two-Tat, Retaliates with defection only after two consecutive defections

TTFT : Retaliates with two defections for every defection of opponent

STFT : TFT that starts with a defection.

PAVLOV : Win-stay lose-shift strategy

Literature Survey

The Evolution of Cooperation | Book authored by Robert Axelrod[14]

The book is the most cited work in Political Science literature. It develops the basis and requirements for cooperation.

The Iterated prisoner's Dilemma 20 years on | Book[1]

This book is a collection of cutting edge advancements in Iterated Prisoners Dilemma Tournaments and Games in the early 21st century.

Axelrod's tournament | Tournament conducted annually

In the 1980's Axelrod ran an IPD tournament inviting agents for the same. We can see a multitude of agents from basic rules to learning and signaling agents. A tournament is run every year and multiple new agents are listed.

Iterated Prisoner's Dilemma and Evolutionary game Theory | Paper[2]

Describes successful strategies for a tournament as Nice, Retaliating, Forgiving and Clear. Also describes a whole slew of strategies. Gives a basic introduction to generating strategies through Genetic Algorithms and also describes evolutionary and replicator dynamics of certain strategies.

Learning IPD through Co-evolution | Paper[3]

Introduces the concept of co-evolution of multiple agents. Multiple agents undergo evolution together and various parameters like noise is tested. Introduction of agents to N-Player IPD.

How to Design a Strategy to Win an IPD Tournament | Paper[4]

Describes a method here an agent can learn to detect an opponent through the first few moves and respond accordingly.

Evolving continuous behaviors in the Iterated Prisoners' Dilemma | Paper[10]

Cooperation through Evolution. Use Finite Automata and Feed Forward Perceptron.

Multiagent Reinforcement Learning in the Iterated Prisoners' Dilemma | Paper[7]

Use Q-Learning to develop an agent that can play against Tit-for-Tat and another Q-Learner.

Invincible Strategies for Iterated Prisoner's Dilemma | Paper[8]

Use the stationary distribution of Memory-1 strategies to define invincible strategies.

No Strategy can win in the repeated Prisoners Dilemma | Paper[6]

Describes invasion and how every strategy can be overturned by the correct set of mutants

Modelling the commons as a game with vector payoffs | Paper[16]

It is a novel attempt to introduce a concept of vector payoff so as to implement Cobb Douglas Utility that not only maximizes personal interest but also the community interests. The result was that a all defect Nash Equilibrium was no longer the Nash Equilibrium. Instead it was a partially cooperative strategy that was the new Nash Equilibrium.

The Game

Replicator Equations

Replicator Equations is a model where the fitness function is determined by the population distribution instead of being fixed.

The vector $\mathbf{x} = (x_1, \dots, x_n)$ denotes the distribution of types in the population.

The payoff matrix A encodes all the fitness information. This is because type x_i will interact with type x_j in proportion to its proportion in the population, and in these interactions the payoff will be A_{ij} .

Although populations are finite, the discrete case is computationally intensive. Given a large enough sample it will be the same in the limit.

The rate of change of a type of population is

$$\frac{dx_i}{dt} = x_i((Ax)_i - \mathbf{x}^T A \mathbf{x})$$

Where $(Ax)_i$ is the expected payoff and $\mathbf{x}^T A \mathbf{x}$ is the average fitness of the entire population.

The change in ratio of proportions with time can be shown to be :

$$\frac{d}{dt} \left(\frac{x_i}{x_j} \right) = \frac{x_i}{x_j} ((Ax)_i - (Ax)_j)$$

This shows that the change in ratio is entirely driven by the difference in fitness between types.

Equilibrium

We consider a dynamic system where a group of agents are constantly interacting and replicating with respect to the payoff matrix A . We reach a non-trivial equilibrium condition when the fitness of each type is exactly the same. Thus effectively the rate of change of the type of population is zero. Trivially the rate of change also becomes zero when one population dies out and the other population takes full control.

We consider a world of agents with days and nights alternating. Each day food keeps spawning and the agents go looking for food. Every piece of food is contested by two agents randomly chosen from the population. Every agent of a particular species/type does the exact same thing when they encounter another agent.

Every night all agents consume one unit of food. If an agent does not get 1 unit of food it dies. If it is lucky and gets more than one unit it replicates perfectly with the excess food at the cost of 1 unit each.

We now look at the following worlds with replicating agents.

Case 1

There are two types of agents, Dove and Hawk in the population. The Dove always cooperates(shares food) while the Hawk always defects(fights for the food).

For any interaction between two agents we consider a standard game of PD with the payoff matrix as follows. ($k > 0$)

Table 2 : Payoff Matrix for Dove Hawk Game Case 1

	C	D
C	1/2, 1/2	0, 1
D	1, 0	$\frac{1}{2} - k, \frac{1}{2} - k$

Let d, h denote the proportion of population of Doves and Hawks respective

Expected payoff of a Dove = $d\left(\frac{1}{2}\right) + h(0)$

Expected payoff of a Hawk = $d(1) + h\left(\frac{1}{2} - k\right)$

At equilibrium the expected payoff is the same, thus:

$$d\left(\frac{1}{2}\right) + h(0) = d(1) + h\left(\frac{1}{2} - k\right)$$

$$d = 1 - \frac{1}{2k} \quad h = \frac{1}{2k}$$

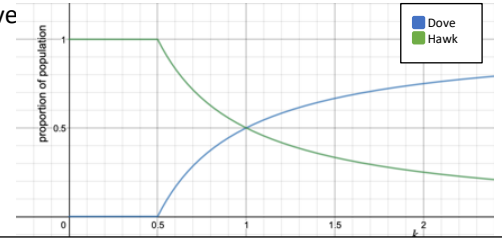


Figure 1: Plot of population vs k at equilibrium

Figure 1 shows the Proportion of Dove and Hawk during equilibrium.

When $k=1$, at equilibrium the fitness of Doves and Hawks are both

$$x^T A x = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \frac{1}{4}$$

Case 2

There are two types of agents, Dove and Hawk in the population. The Dove always cooperates(shares food) while the Hawk always defects(fights for the food).

For any interaction between two agents we consider a standard game of PD with the payoff matrix as follows. ($k_2 > 0, k_1 \geq 0$)

Table 3 : Payoff Matrix for Dove Hawk Game Case 2

	C	D
C	1/2, 1/2	$k_1, 1 - k_1$
D	$1 - k_1, k_1$	$\frac{1}{2} - k_2, \frac{1}{2} - k_2$

Solving similarly as above, during equilibrium we have

$$d = \frac{2k_2 + 2k_1 - 1}{2k_2} \quad h = \frac{1 - 2k_1}{2k_2}$$

Consider Figure 2. In the region below the line $k_2 + k_1 < \frac{1}{2}$ during equilibrium the Dove population dies out and only Hawks remain.

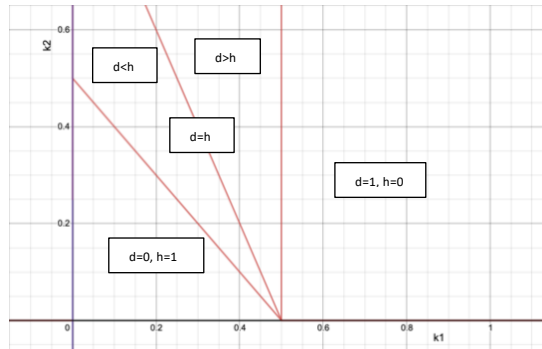


Figure 2: Plot of population ratios at equilibrium for given k_1 and k_2

Case 3

We consider three types of agents namely Dove, Hawk and a Crow. The Dove always cooperates (shares food), the Hawk always defects (fights for the food) and the Crow cooperates with Doves and other Crows but defects with Hawks.

Every interaction is given by the payoff matrix Table 1.

Let d, h, c be the proportion of population of Doves, hawks and Crows respectively.

$$d + h + c = 1$$

$$\text{Expected Payoff for Dove } D = d\left(\frac{1}{2}\right) + h(k_1) + c\left(\frac{1}{2}\right)$$

$$\text{Expected Payoff for Hawk } H = d(1 - k_1) + h\left(\frac{1}{2} - k_2\right) + c\left(\frac{1}{2} - k_2\right)$$

$$\text{Expected Payoff for Crow } C = d\left(\frac{1}{2}\right) + h\left(\frac{1}{2} - k_2\right) + c\left(\frac{1}{2}\right)$$

At equilibrium the expected payoff for all agents is the same.

- $D=C \Rightarrow k_2 + k_1 = 1/2$
- $N=H \Rightarrow d(1/2 - k_1) = c(k_2)$
- $D=H \Rightarrow d\left(\frac{1}{2} - k_1\right) = h(k_2 + k_1 - 1/2) + c(k_2)$

Solving we get the following strict conditions for equilibrium.

$$\begin{aligned} d &= c \\ k_2 + k_1 &= 1/2 \end{aligned}$$

So in a world with $k_2 + k_1 = \frac{1}{2}$, we reach an equilibrium with Doves and Crows at the same population. But

$k_2 + k_1 = \frac{1}{2}$ implies that Sucker payoff = Punishment Payoff. So technically there is no longer incentive to cheat. Hence we can conclude that no equilibrium where all three species survive exists. Figure 3 shows the equilibrium region given the ratios of Doves and Crows. The ratio of Hawks is precisely $h = 1 - d - c$.

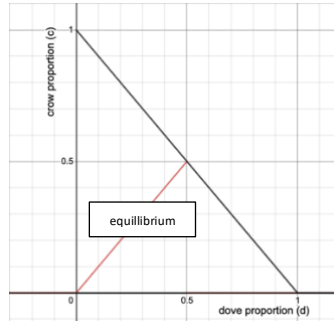


Figure 3: Plot of equilibrium region given population ratio of Doves and Crows

The payoff matrix for the replicator equation will be

$$A = \begin{pmatrix} \frac{1}{2} & k_1 & \frac{1}{2} \\ 1 - k_1 & \frac{1}{2} - k_2 & \frac{1}{2} - k_2 \\ \frac{1}{2} & \frac{1}{2} - k_2 & \frac{1}{2} \end{pmatrix}$$

When k_1 is 0 as in Case 1 \Rightarrow Doves and Crows increase at the same rate till Hawks die out.

At equilibrium fitness of every agent type is : $x + (1 - 2x)(k_1)$

$$A = \begin{pmatrix} \frac{1}{2} & k_1 & \frac{1}{2} \\ 1 - k_1 & k_1 & k_1 \\ \frac{1}{2} & k_1 & \frac{1}{2} \end{pmatrix} \quad x^T = \begin{pmatrix} x \\ 1 - 2x \\ x \end{pmatrix}$$

The Game

In the above cases if we consider the PD to be an IPD instead. This implying that every interaction of two agents is an IPD as opposed to a standard PD. It can be noticed that the Dove-Crow-Hawk interaction in a PD is very similar to a ALLC-TFT-ALLD interaction in an IPD in the limit.

We can now define our system which is similar to the world we saw with Doves, Hawks and Crows.

We have a population of agents P each having a Strategy, a Consumption rate (c) and a way to replicate with an associated cost (r).

The world continuously alternates between Day and Night.

Every day M boxes spawn. Each box contains N units of food.

Every box is contested by two randomly drawn agents from P . This is done through a game of IPD where each round corresponds to one unit of food.

Finally once all the boxes have been distributed, it is Night time.

Each night every agent has 3 possible actions which are as follows

- Death : If $c > \text{Total food collected}$
- Reproduce and Survive : If $c+r < \text{Total food collected}$
- Just survive : $c < \text{Total food collected} < c+r$

Each day an agent can participate in multiple contests for food if it is lucky to be chosen. All unused food is stored to be used the next day.

In short the system in consideration is a group of agents where interactions are defined as IPD's. The agents learn and evolve to form an interesting dynamic system.

Memory Agents

The agent used throughout the paper will be a memory agent that is limited by the amount of history that can be accessed.

An **$M(a,b)$** agent is one that remembers the last a moves it made and the last b moves it's opponent made during an IPD.

Consider an IPD of n rounds between agent $M_x(a_x, b_x)$ and $M_y(a_y, b_y)$.

Let the moves of $M_x(a_x, b_x)$ be x_1, x_2, \dots, x_n and the moves of $M_y(a_y, b_y)$ be y_1, y_2, \dots, y_n .

Then in the i 'th iteration of the game.

The move/strategy of M_x will be a function of $(x_{i-a_x-1}, \dots, x_{i-1}, y_{i-b_x-1}, \dots, y_{i-1})$

The move/strategy of M_y will be a function of $(y_{i-a_y-1}, \dots, y_{i-1}, x_{i-b_y-1}, \dots, x_{i-1})$

A **Stochastic $M(a,b)$** or **$SM(a,b)$** encodes an array of probabilities of cooperating for different possible situations of historical moves. It's strategy consists of an array of length 2^{a+b} , each cell corresponding to one of the possible historical states.

If we assume an error-free world, a game between stochastic agents devolves to a simple Markov chain of historical states. The stationary distribution gives us the expected payoff in the limit. One drawback is that these agents cannot learn strategies like GRIM(Cooperate till opponent Defects, then always Defect) which require either full history or a dedicated

variable to store special information. This shall be further explored in later sections with the introduction of extra memory to keep track of special features.

Example

An M(1,1) agent will use only the moves of the previous turn to make the next move.

A Stochastic M(1,1) agent encodes its strategy into a vector (p_1, p_2, p_3, p_4) .

Here, p_1 denotes the probability of cooperating when 'CC' has been played in the previous move. Similarly, p_2 denotes the probability of cooperating when 'CD' has been played, and p_3 denotes the probability of cooperating when 'DC' has been played, and p_4 denotes the probability of cooperating when 'DD' has been played.

With the SM(1,1) encoding of ['CC', 'CD', 'DC', 'DD']:

An ALLC(Always cooperate) agent will have an SM(1,1) encoding of (1,1,1,1).

An ALLD(Always defect) agent will have an SM(1,1) encoding of (0,0,0,0).

A TFT(Tit-for-Tat) agent will have an SM(1,1) encoding of (1,0,1,0).

A PAVLOV(Win-Stay Lose-Shift) agent will have an SM(1,1) encoding of (1,0,0,1)

A Random agent will have an SM(1,1) encoding of (0.5,0.5,0.5,0.5)

During an IPD between two SM(1,1) agents namely P and Q the states form a Markov chain. Let the strategy vectors of the agents be (p_1, p_2, p_3, p_4) and (q_1, q_2, q_3, q_4) . The possible states are 'CC', 'CD', 'DC', 'DD' where the first literal corresponds to P's move and the second literal in the state corresponds to Q's move.

Suppose P and Q played C and C respectively in the (i-1)'th round of the IPD. Then the probability of P, Q playing D, D the next round is $(1 - p_1)$ or $\overline{p_1}$ and $(1 - q_1)$ or $\overline{q_1}$ respectively. Since the moves of P and Q are independent of each other the probability of going to state 'DD' from 'CC' is $(1 - p_1)(1 - q_1)$ or $\overline{p_1}\overline{q_1}$.

The transition matrix thus becomes the following.

$$\begin{pmatrix} p_1 q_1 & p_1 \overline{q_1} & \overline{p_1} q_1 & \overline{p_1} \overline{q_1} \\ p_2 q_2 & p_2 \overline{q_2} & \overline{p_2} q_2 & \overline{p_2} \overline{q_2} \\ p_3 q_3 & p_3 \overline{q_3} & \overline{p_3} q_3 & \overline{p_3} \overline{q_3} \\ p_4 q_4 & p_4 \overline{q_4} & \overline{p_4} q_4 & \overline{p_4} \overline{q_4} \end{pmatrix}$$

Example

Code link : <https://github.com/inavamsi/Thesis/tree/master/Markov%20Chain>

We consider two simple SM(0,1) agents.

SM(0,1) corresponds to an encoding of ['C', 'D']. Just the last move of the opponent.

This is less powerful than SM(1,1) as standard strategies like PAVLOV cannot be encoded. Nevertheless, we can encode the other standard agents like the following.

ALLD corresponds to (0,0) encoding.

ALLC corresponds to (1,1)

TFT corresponds to (1,0) encoding

Given the encoding of two agents $P = [p_C, p_D]$ and $Q = [q_C, q_D]$. The corresponding Markov chain can be encoded as transitions of space $S = \{s_{CC}, s_{CD}, s_{DC}, s_{DD}\}$ where s_{ij} denotes a state where the previous moves made by the agents P and Q are 'i' and 'j' respectively.

The probability of a transition between two states is given by.

$$\begin{aligned} p(s_{CC}|s_{ij}) &= p_i q_j & p(s_{CD}|s_{ij}) &= p_i \bar{q}_j \\ p(s_{DC}|s_{ij}) &= \bar{p}_i q_j & p(s_{DD}|s_{ij}) &= \bar{p}_i \bar{q}_j \end{aligned}$$

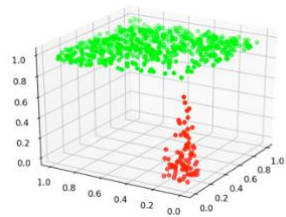
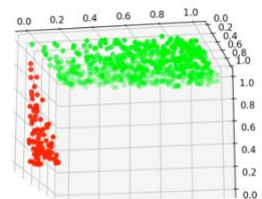


Figure 4 : Regions of Cooperation(Green) and Defection(Red).

Figure 4 shows the regions of high cooperation and defection between two SM(0,1) agents, who are encoded on the x,z and y,z planes respectively. Green corresponds to a stationary distribution where 'CC' is over 0.8 and Red corresponds to a stationary distribution where 'DD' is over 0.8. Although it seems that cooperation is much more likely than defection. It should be noted that there exist other agents that behave with other tendencies. Green region contains all ALLC-ALLC and ALLC-TFT integrations while the red region contains ALLD-ALLD and ALLD-TFT interactions.



If we consider error where with ε probability the decision of an agent is the opposite of what was intended. The transition matrix changes as follows.

$$p(s_{CC}|s_{ij}) = p_i q_j \bar{\varepsilon}^2 + p_i \bar{q}_j \varepsilon \bar{\varepsilon} + \bar{p}_i q_j \varepsilon \bar{\varepsilon} + \bar{p}_i \bar{q}_j \varepsilon^2$$

Example

Consider an IPD between two SM(0,2) agents.

P and Q encode strategies as (p_1, p_2, p_3, p_4) and (q_1, q_2, q_3, q_4) respectively. Note that the states here correspond to the last two moves of the opponent.

So SM(0,2) corresponds to an encoding of ['CC', 'CD', 'DC', 'DD'].

Standard Agents are encoded as following.

ALLC encodes as (1,1,1,1).

ALLD encodes as (0,0,0,0).

TFT encodes as (1,1,0,0).

TTFT encodes as (1,0,0,0).

TFTT encodes as (1,1,1,0).

Although this encoding is more powerful in some aspects compared to SM(1,1) it cannot encode agents like PAVLOV which require knowing one's own history.

A state will be represented as ('CC', 'DC') where the first two literals belong to encoding of P and the last two literals belong to player Q . Thus the first(last) two literals correspond to the last two moves of player $Q(P)$. Note that a state ('CC', 'DC') can possibly only got to one of four states which are ('CC', 'DC'), ('CC', 'CD'), ('CD', 'DC'), ('CD', 'CD').

A transition of states in a Markov Chain of states will be as follows. Let P_i and Q_i denote the moves of PA and Q in round I of the IPD.

$(Q_{i-2}Q_{i-1}, P_{i-2}P_{i-1}) \rightarrow (Q_{i-1}Q_i, P_{i-1}P_i)$ with probability $(Q_{i-2}Q_{i-1} \rightarrow Q_i) (P_{i-2}P_{i-1} \rightarrow P_i)$

So, the state ('CC', 'CD') will go to the state ('CC', 'DD') with probability $p_1 q_2$.

Agent Estimation and Optimal action in Noisy environment

Trying to optimize payoff in a signaling game goes hand in hand with estimating the type of the opponent. In an IPD the type is determined by strategy. We use Bayes rule to estimate the optimal move given that we are not a stochastic agent (Agent A).

Consider a population of Random, ALLC, ALLD and TFT with population proportion r, c, d, t respectively.

Agent A plays an IPD of 'n' rounds with a randomly sampled agent B. The moves are $(a_1, b_1) \dots (a_n, b_n)$ where a_i and b_i are either 'C' or 'D'. Let $\varepsilon < \frac{1}{2}$ be the error probability which is the probability of 'C' getting switched 'D' or vice-versa.

The prior probability of types(strategies) is

$$P(B = \text{Random}) = r, P(B = \text{ALLC}) = c, P(B = \text{ALLD}) = d, P(B = \text{TFT}) = t$$

Let $\delta(x, y) = 1$ if $x = y$ and 0 otherwise

After j moves,

$$P(a_1 b_1 \dots a_j b_j | B = \text{Random}) = \frac{1}{2^j}$$

$$P(a_1 b_1 \dots a_j b_j | B = \text{ALLC}) = \prod_{i=1}^j (1 - \varepsilon)^{\delta(a_i, "C")} \varepsilon^{\delta(a_i, "D")}$$

$$P(a_1 b_1 \dots a_j b_j | B = \text{ALLD}) = \prod_{i=1}^j (1 - \varepsilon)^{\delta(a_i, "D")} \varepsilon^{\delta(a_i, "C")}$$

$$P(a_1 b_1 \dots a_j b_j | B = \text{TFT}) = \prod_{i=2}^j (1 - \varepsilon)^{\delta(a_{i-1}, b_i)} \varepsilon^{1 - \delta(a_{i-1}, b_i)}$$

Using Bayes Rule we can calculate,

$$P(B = \text{agent} | a_1 b_1 \dots a_j b_j) = \frac{P(a_1 b_1 \dots a_j b_j | B = \text{agent}) P(B = \text{agent})}{P(a_1 b_1 \dots a_j b_j)}$$

We can ignore the common denominator and consider the ratio of numerators.

Now after j rounds, the probability of B playing 'C' in the following round can be represented as the following.

$$P(b_{j+1} = 'C' | a_1 b_1 \dots a_j b_j) = \sum P(B = \text{agent} | a_1 b_1 \dots a_j b_j) P('C' | B = \text{agent})$$

Where by definition,

$$\begin{aligned} P('C' | B = \text{ALLC}) &= 1 \\ P('C' | B = \text{ALLD}) &= 0 \\ P('C' | B = \text{Random}) &= \frac{1}{2} \\ P('C' | B = \text{TFT}) &= (1 - \varepsilon)^{\delta(b_j, "C")} \varepsilon^{1 - \delta(b_j, "C")} \end{aligned}$$

But, A cannot play optimally by maximizing reward for round $n+1$. Based on the remaining number of moves $n - j$.

Since we assumed $\varepsilon < \frac{1}{2}$

Let $R(agent)$ denote the expected payoff if played optimally in the remaining $n - j$ moves if B(opponent) is agent. Let $T(a, b)$ denote the payoff due to move a and b. We use $E(a, b)$ to denote the Expected payoff due to A wanting to play a and B wanting to play b..

$$E(a, b) = T(a, b)(1 - \varepsilon)^2 + T(\bar{a}, b)\varepsilon(1 - \varepsilon) + T(a, \bar{b})(1 - \varepsilon)\varepsilon + T(\bar{a}, \bar{b})\varepsilon^2$$

If B(opponent) is one of ALLC, Random, ALLD agent then it would be optimal to Defect(D) till the end. The expected payoffs if played optimally till the end of the game would be as follows.

$$\begin{aligned} \frac{1}{n}R(B = ALLC) &= E('D', 'C') \\ &= T('C', 'C')\varepsilon(1 - \varepsilon) + T('C', 'D')\varepsilon^2 + T('D', 'C')(1 - \varepsilon)^2 \\ &\quad + T('D', 'D')(1 - \varepsilon)\varepsilon \\ \frac{1}{n}R(B = ALLD) &= E('D', 'D') \\ \frac{1}{n}R(B = Random) &= T('C', 'C')\frac{\varepsilon}{2} + T('C', 'D')\frac{\varepsilon}{2} + T('D', 'C')\frac{(1 - \varepsilon)}{2} + T('D', 'D')\frac{(1 - \varepsilon)}{2} \end{aligned}$$

The optimal move against TFT is a lot more tricky to estimate. But if we assume $T(a, b)$ to follow all the rules of IPD payoff then the optimal move would be to 'C' till $n - j - 1$ move and then 'D' on the last move. This can be proved inductively.

On the last move A will get $R(B = ALLC)$ if 'C' was played in the second last move or A will get $R(B = ALLD)$ if 'D' was played in the second last move.

$$\begin{aligned} R(B = TFT|b_j = 'C') &= E('C', 'C') + (1 - \varepsilon)(n - j - 2)(T('C', 'C')(1 - \varepsilon) + T('C', 'D')(\varepsilon)) + \\ &\quad (\varepsilon)(n - j - 1)(T('D', 'D')(1 - \varepsilon) + T('D', 'C')(\varepsilon)) + R(ALLC) \\ R(B = TFT|b_j = 'D') &= E('C', 'D') + (1 - \varepsilon)(n - j - 2)(T('C', 'C')(1 - \varepsilon) + T('C', 'D')(\varepsilon)) + \\ &\quad (\varepsilon)(n - j - 1)(T('D', 'D')(1 - \varepsilon) + T('D', 'C')(\varepsilon)) + R(ALLC) \end{aligned}$$

$$\begin{aligned} R(B = TFT) &= (1 - \varepsilon)\delta(b_j, 'C')R(B = TFT|b_j = 'C') \\ &\quad + (\varepsilon)\delta(b_j, 'D')R(B = TFT|b_j = 'D') \end{aligned}$$

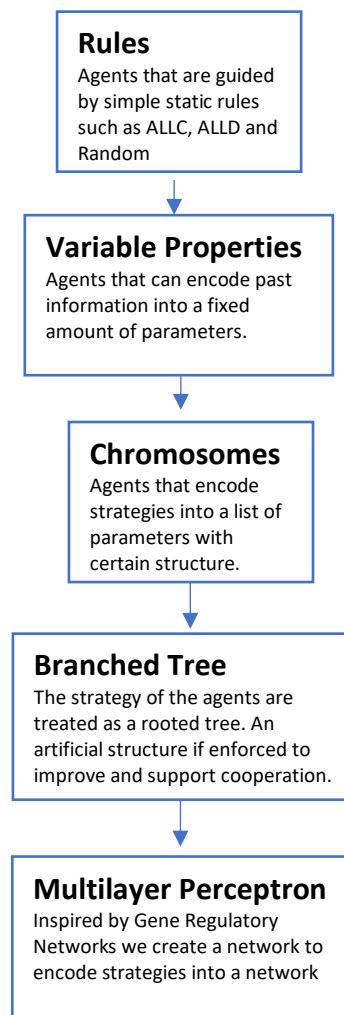
On move $j + 1$, to maximize expected total payoff we choose 'C' if $P(B = TFT|a_1b_1 \dots a_jb_j)R(B = TFT) > P(B = ALLC|a_1b_1 \dots a_jb_j)R(B = ALLC) + P(B = ALLD|a_1b_1 \dots a_jb_j)R(B = ALLD) + P(B = Random|a_1b_1 \dots a_jb_j)R(B = Random)$. Otherwise we choose 'D'.

This is the general framework to choose optimal action through Bayesian estimation in a multi-agent iterated signaling game with error. In the case of an unbounded game we add a discounting parameter γ to discount future reward.

Brief on Approaches

The intention is to study the emergence of cooperation and thus understand the underlying structure of the system in cooperation. We go through a series of experiments where we continually evolve and improve the structure of agents so they can cooperate better.

The general timeline of the experiments is as follows. The gradual process of improving the structure of an agent and system is shown as simple chart below.



Tournaments and Replication without Mutation (Approach 1)

Tournaments

A tournament in a non-replication system where a group of starter agents contest to determine the final winner. Just like Axelrod's Tournament as shown in the Literature Review Section.

Despite not helping us evolve new agents or understand cooperation, tournaments allow us to grasp basic structures of interaction.

We start with a standard Round Robin tournament to see how different agents fare when pitted in an open field. Each player plays every other player in each round. For the sake of understanding we simply look at some standard agents.

Code : <https://github.com/inavamsi/Thesis/tree/master/Tournament>

The above code outputs the winner of all possible tournaments given a list of agents. Here are a few results.

```
TFT TFFT STFT PAVLOV ALLC → STFT
ALLD Random GRIM TFFT STFT PAVLOV ALLC → ALLD
Random GRIM TFFT STFT PAVLOV ALLC → PAVLOV
ALLD GRIM TFFT STFT PAVLOV ALLC → ALLD
GRIM TFFT STFT PAVLOV ALLC → PAVLOV
ALLD Random TFFT STFT PAVLOV ALLC → ALLD
Random TFFT STFT PAVLOV ALLC → Random
ALLD TFFT STFT PAVLOV ALLC → ALLD
TFFT STFT PAVLOV ALLC → STFT
ALLD Random GRIM TFFT TFFT STFT PAVLOV ALLC → TFFT
Random GRIM TFFT TFFT STFT PAVLOV ALLC → TFFT
```

An interesting thing to note is that the existence of a certain agent causes a significant influence on the dynamic between other agents. A stable population cannot be disturbed by an agent but with the help of a third agent an invasion can occur. This will be properly discussed in the next section with the help of Evolutionary Stable Strategies.

A classic example of cooperation is as follows. Individually ALLD always wins but with the existence of two cooperating agents in the population, the cooperating agents are better off.

```
ALLD TFFT -> ALLD
ALLD TFT -> ALLD
ALLD TFT TFFT -> TFFT
```

As an additional experiment this data was trained on a 3 layer Neural network.

Code: https://github.com/inavamsi/Thesis/blob/master/tournament_nn.py

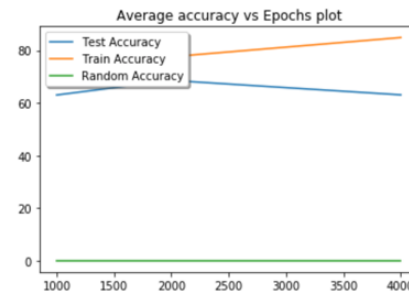


Figure 5 : Plot of Accuracy vs Epochs

The best accuracy was around 70% as plotted in Figure 5.

In most tournaments, certain dominant species win most of the time. This results in a lot of similar outputs, possibly leading to this very high success rate.

Replicator Equations

As described in the section 'The Game' we have a set of agents. These agents interact and do one of three things every night – Live, Die, Replicate.

Code: <https://github.com/inavamsi/Thesis/tree/master/Replication-Standard%20Agents%20No%20Mutation>

From table 1, when $k=1$, Dove and Hawk population will be the same. This is a simulation is shown in Figure 6. It can also be easily represented as a predator prey model.

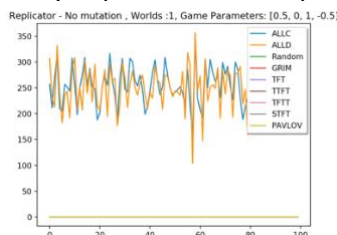


Figure 6 : Dove Hawk Population timeline

A replication game of multiple agents will result in the death of undesirables and a cooperative equilibrium. For example Figure 7 shows a population with all agents.

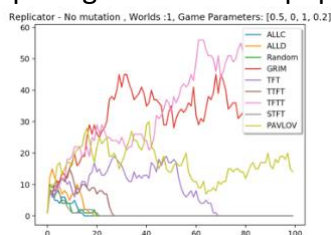


Figure 7 : Population timeline 8 agents

But this is one of the multiple stochastic routes and not repeatable due to a discrete population and quite a bit of randomness. We should be observing the same in the limit. The perfect way is to solve a system of replicator equations as we did before. We could also increase the sample size and take an average over multiple worlds.

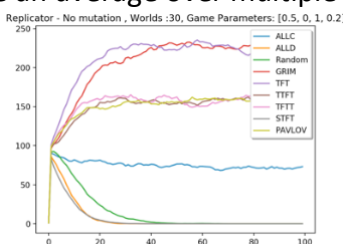


Figure 8 : Population timeline 8 agents averaged over 30 worlds

Figure 8 shows a larger population averaged over 30 worlds. We can clearly see that this is very accurate representation as there is a very stable equilibrium. Agents that defect without provocation are directly pruned out. Once these agents are gone there are no defection because of the lack of provocation and the remaining population stay in equilibrium.

A simple calculation will yield the same results. It is trivial to show the equilibria after Day 40. For the other days a direct replicator equation will give us the rate of change of population at every instant which can be verified easily. The tiny perturbations we see in the living species is just noise during random selection. In the limit these will be straight horizontal lines.

Asexual Replication with mutation (Approach 2.1)

Agents

We will use SM agents that were defined earlier. The agents will replicate as before with the addition of mutation. We consider Gaussian noise $N(0, \sigma)$ of standard deviation σ .

At this stage we do not consider a mutation in the amount of memory. But this will be dealt with later on. Consider the parent $SM_p(a,b)$ agent and kid $SM_k(a,b)$ agent with the encoding $(p_1, p_2 \dots p_n)$ and $(k_1, k_2 \dots k_n)$ respectively, where $n = 2^{a+b}$. Then the following relation holds.

$$k_i = \max(0, \min(1, p_i + N(0, \sigma))) \quad \forall i \in \{1 \dots n\}$$

Note that k_i represents the probability of cooperating given a certain state and thus is bounded by $[0,1]$. This relation represents the mutation from parent to child.

Simulation

We shall now see a few examples of evolution of the simple $SM(0,1)$ agents.

The encoding of an $SM(0,1)$ agent is of the form $[C', D']$ which refers to the probability of cooperating when the opponent's last move was C or D respectively.

If we plot these agents (Figure 10) with the first probability of the y axis and the second one on the x axis, we can observe the evolution of agents over time.

As shown in Figure 9, the top right corner will correspond to agents behaving very much like ALLC while the bottom left corresponds to agents which behave very similarly to ALLD. Agents in the center are Random agents and the top left corner corresponds to TFT agents.

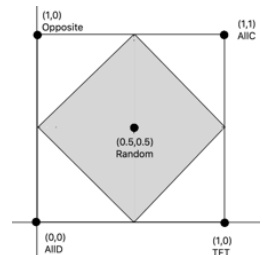


Figure 9 : Regions of similarity for different agents

We shall see a simple simulation with a bunch of randomly generated agents (not Random).

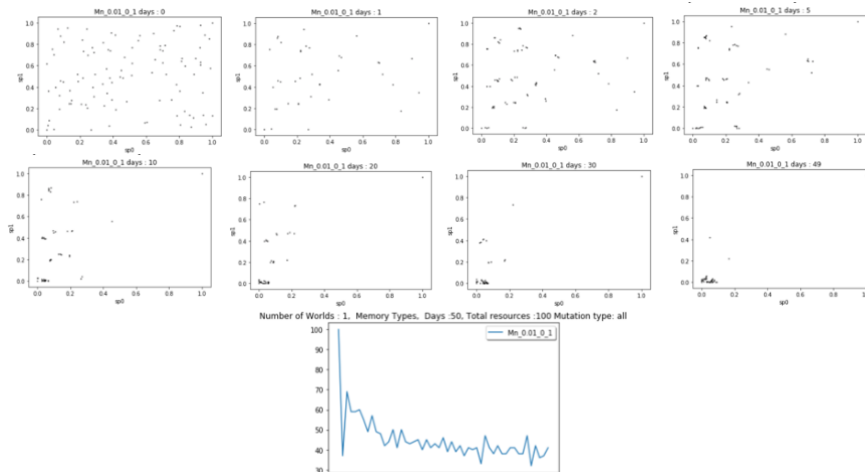


Figure 10 : The plot of randomly initialized $SM(0,1)$ agents over time.

The final result is basically a bunch of agents very similar to ALLD. This is a classic example of the overwhelming power of ALLD agents. But this does not imply that all evolution results in ALLD.

With special structure, the final agents do not always devolve to a state of complete defection. A standard case is starting with just agents that are strictly retaliatory. Implying that the probability of cooperating when opponent defects is 0. Such SM(0,1) agents can be encoded as (p,0). We shall see what happens in Figure 11.

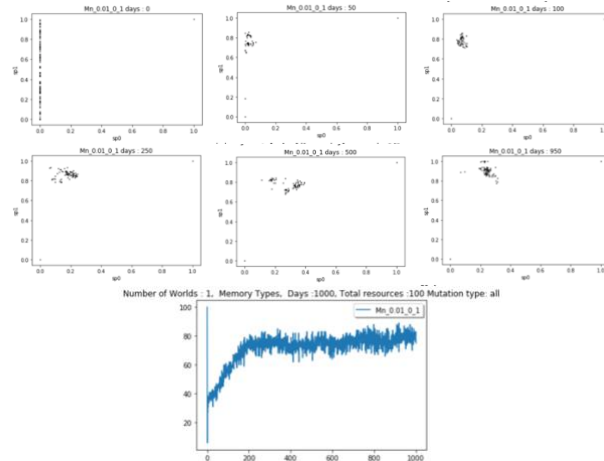


Figure 11 : The plot of randomly initialized retaliating SM(0,1) agents over time.

Although taking over 900 generations we can surprisingly see that these agents not only tend towards TFT and corporate, but after reaching that state the system tends towards a less retaliatory state as a whole. It is also obviously clear as to why the total population in the second case is higher than the first case despite both situations being given the same amount of resources.

These experiments are repeatable and despite not giving the exact same result every-time, in the limit the results tend to be the same. In this implementation changing the number of worlds tends the system towards the limit at a cost of high computation.

Analyzing these structures will help us understand and better develop mechanisms and the necessary structure for cooperation. We shall see how this happens in the later Sections.

Invasion

Parrondo's Paradox : 'A combination of losing strategies can become a winning strategy'.

An evolutionarily stable strategy (ESS) is a strategy (or set of strategies) which, if adopted by a population in a given environment, is impenetrable, meaning that it cannot be invaded by any alternative strategy (or strategies) that are initially rare.

(S, S) is a Nash equilibrium in a two player game if and only if this is true for both players and for all $T \neq S$:

$$E(S,S) \geq E(T,S)$$

Maynard Smith and Price specify two conditions for a strategy S to be an ESS. For all $T \neq S$, either

$$E(S,S) > E(T,S), \text{ or } [E(S,S) = E(T,S) \text{ and } E(S,T) > E(T,T)]$$

There is also an alternative, stronger definition of ESS, due to Thomas. This definition requires that for all $T \neq S$

$$E(S,S) \geq E(T,S), \text{ and } E(S,T) > E(T,T)$$

The two definitions are not equivalent. Considering the Maynard Price definition it is evidently seen that $E(\text{ALLD}, \text{ALLD}) > E(\text{ALLD}, \text{TFT})$ because of the first move and it thus ALLD is ESS but with respect to Thomas definition $E(\text{ALLD}, \text{TFT}) \not\geq E(\text{TFT}, \text{TFT})$ and it thus ALLD is not ESS.

The second definition makes more sense in our case.

Figure 12 shows how TFT invades a population of ALLD

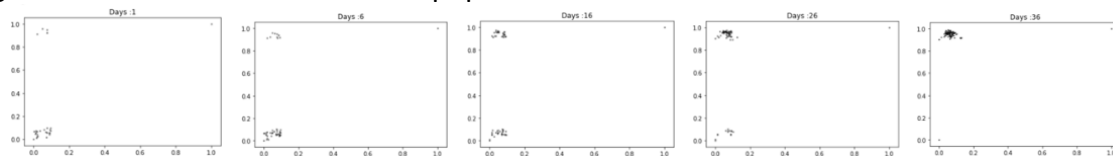


Figure 12 : Timeline of TFT invading ALLD

What we saw was a case of direct invasion, where TFT was directly able to invade ALLD.

A population of TFT can never be invaded directly by ALLD agents. But if there are enough ALLC agents also present the ALLD agents can invade through the ALLC agents. This is called indirect invasion.

Figure 13 shows the variations due to starting configuration.

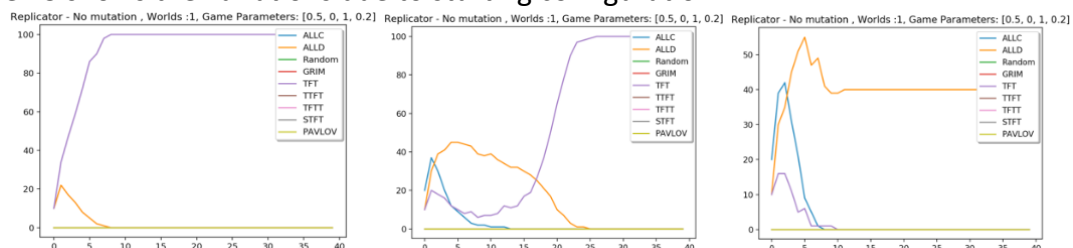


Figure 13 : Timeline of agents with different ratios of starting population of ALLD, TFT, ALLC.

In the first case, with only TFT and ALLD agents, there is no chance of ALLD agents invading. But with the introduction of ALLC agents it all changes and ALLD can indirectly invade a population of TFT agents. Although the probability of ALLD agents invading in this specific situation is low, it is still possible.

It can be assumed that the TFT agent is most evolutionary stable. This is not the case as seen below. We see a population drift to more non cooperative zones. As shown in Figure 14, this was the case given a large enough time frame.

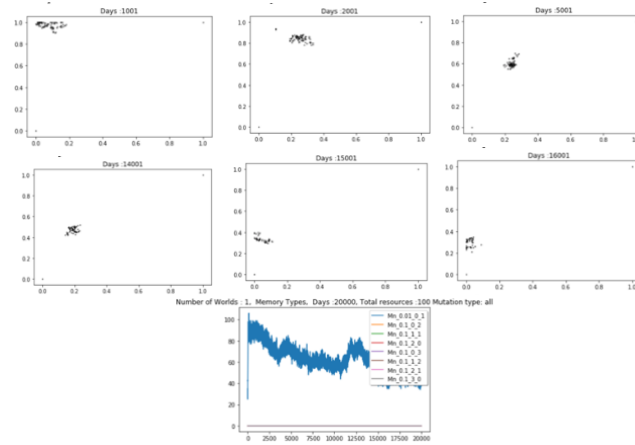


Figure 14 : Drift of TFT agents to ALLD

Even in the case of multiple mutation rates, we notice that there is an optimal mutation rate which gives best survival chances. But in the following case the optimal agents most of the time evolved to ALLD agents when given random starting conditions. As shown in Figure 15, we can see that the chances of cooperation are pretty grim as the required amount of mutation rate for the jump to cooperative agents is pretty high. And the high mutation rate agents die out really fast due to the inadequacy of proper adaptation.

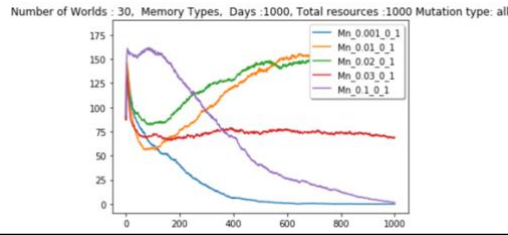


Figure 15 : Timeline of different stochastic memory(SM) agents

Also in subsequent sections there will be interesting cases of indirect invasions where ALLD can invade TFTT with the help of other agents that that arise due to mutation. Here is an example that will be discussed in the next section.

In figure 16 we consider a population of SM(0,1) agents. It can be seen that TFTT(Yellow) gets periodically invaded by ALLD(Blue) with the help of other rare agents present in the population.

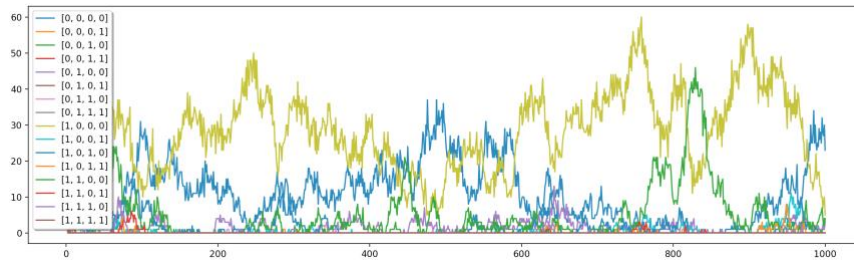


Figure 16 : Timeline of SM(0,1) agents closest to corresponding one of 16

Other Parameters

We can ask the following questions.

- When to replicate?
- Should I share resources with my kid?
- What if we create new parameter that can capture more information like discounted history?
- What should the mutation rate be?

When to replicate ?

One obvious way to replicate is immediately when resources are available, but sometimes it would be wiser to wait and hold on to resources. Then replicate once a certain amount is present. There is a tradeoff between ensuring one survival or replicating to ensure increased advantage. This is shown in the example below.

Waiting to gather more resources does not penalize an agent that does better early on but in the case of cooperating agents the number of agents of the same species matters. As seen before, A lone TFT does not have any survival advantage but as more TFT's are added the advantage becomes apparent.

Example

Code: https://github.com/inavamsi/Thesis/tree/master/Replication_with_mutation

SM(1,1) agent with mutation rate $N(0,0.1)$.

Blue is ALLD, Yellow is TTFT and Green is TFT, Red is TFTT. In figure 17, we can see regions of cooperation arising with ALLD and the TTFT (less forgiving than TFT) being most dominant.

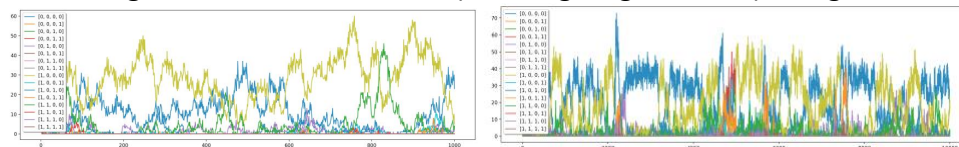


Figure 17 : Timeline of SM(1,1) agents closest to corresponding one of 16 vectors with immediate

When we introduce an added security of at least one day of savings before reproduction the result as shown in Figure 18, does not look very different.

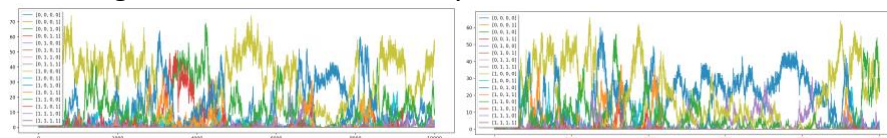


Figure 18 : Timeline of SM(1,1) agents closest to corresponding one of 16 vectors with reproduction after savings.

The difference becomes apparent when we increase the amount of savings. In Figure 19 are four timelines for 0,3, 5 and 7 times savings. There is more stability and the evolution is more gradual. This is a sort of an elitist approach where only the best of the best reproduce.

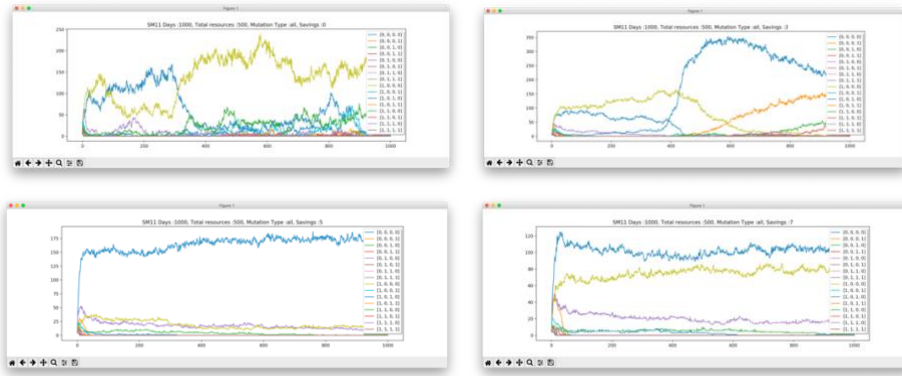


Figure 19 : Timeline of SM(0,1) agents closest to corresponding one of 16 vectors with different amount of savings before

Splitting resources with kids.

A new problem arises as to how much resources should be given to the child so that the child has time to mature to adapt to the surroundings, but the tradeoff is that most children are going to be suboptimal and sharing resources will be a waste. All these parameters can be added as chromosomes(parameters) to find optimal agents. A more interesting way is to model this as a Stochastic Multi armed Bandit problem where the agent learns over time which process to proceed with. But that also has many challenges and is not in the scope of this paper.

Discounted History

Since an agent with higher memory should do better in the long run, we introduce a parameter which stores the discounted history of the agent and the opponent. A discount value of 0 translates to an SM(1,1).

And it can be noticed that it is preferred to have a low level of discounted history in the long run. This is clear as the ecosystem on the whole is constantly changing . It is interesting to see that optimal discount value is very close to an SM(1,1) agent although an SM(1,1) agent performs subpar when compared to it.

Code: [https://github.com/inavamsi/Thesis/tree/master/Replication with mutation](https://github.com/inavamsi/Thesis/tree/master/Replication%20with%20mutation)

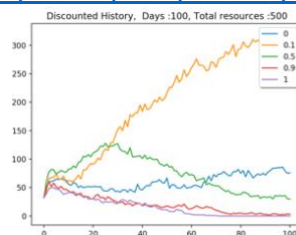


Figure 20 : Timeline of SM(1,1) agents for different discounted

Mutation Rate and effects

We have considered a Gaussian distribution with mean zero for error during mutation.

Mutation rate is again a tradeoff between converging faster to an optimal solution vs escaping local maxima. The simplest way out is to use Genetic algorithms and this is discussed in the next section.

When the experiment was repeated with multiple rate of mutations for SM(1,1) agents a very peculiar phenomena can be observed. There is no deterministic result even when the sample space and size were increased. Also there is a sudden emergence of cooperation and as a critical point is reached. It is evident that cooperation takes place as the total population size can only be sustained with cooperation given the available number of resources. This shows that mutation rate instills a sense of community. All species with the same mutation rate reach a dynamic equilibrium.

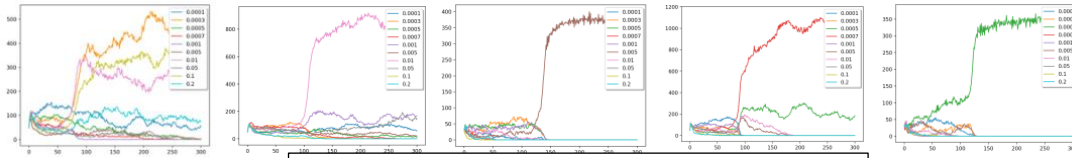


Figure 21 : Timeline of SM(1,1) agents with different mutation

This method seems to be perfect giving a good structure for cooperation and a lot of unexplored space like discounted history and mutation rate to explore. But there is a large problem. A large anomaly is noticed as shown below.

Code: <https://github.com/inavamsi/Thesis/tree/master/Memory>

As seen in Figure 22, we consider a huge population size. Counterintuitively, SM(2,0) agents come out on top followed by SM(0,1) agents.

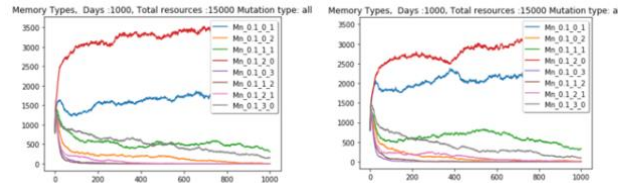


Figure 22 : Timeline of SM(m,n) agents with different amounts of memory.

This was also the case when we considered in a smaller number of species, namely SM(2,0) comes out on top over SM(1,1). In Figure 23, we average over 200 worlds and notice that this is a recurrent phenomenon.

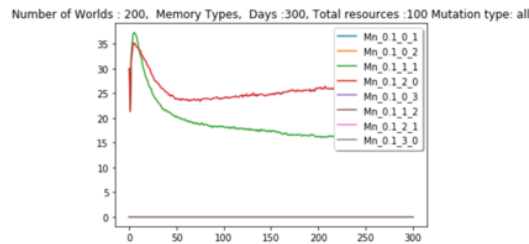


Figure 23 : Timeline of SM(2,0) and SM(1,1) agents with different amounts of memory averaged over 200 worlds.

Agents that know their own history do better than agents than agents that know the opponent. This is because agents that only know themselves can easily devolve to ALLD type

strategies while other agents require a lot more time to prune their huge number of parameters. This is a critical problem and requires a better approach which is discussed further sections.

Note : We have considered a Gaussian distribution with mean zero during mutation. To get a sense of the dynamics, it is perfectly fine to use the gaussian distribution but if we want to model existing occurrences and predict the future we have to realize that most mutations are suboptimal and some extreme mutations die out thus we might have to look at distributions with a fatter or thinner tail. One would be Student-T distribution.

Genetic Algorithms (Approach 2.2)

The previous approach does pretty well but is riddled with flaws. So now we explore the system with from the view of Genetic algorithms where reproduction is a result of more than one agent.

There are multiple ways to explore genetic algorithms. The following is one of the ways.

Chromosome: The strategy encoding remains the same for the SM agents.

Mating : Instead of replicating, we choose two agents and take the bitwise average of their encodings with a certain amount of mutation.

Death : Every round we kill of the a percentage of the least fit agents. Each day we reset all scores back to 0 as the entire ecosystem has changed.

Code: <https://github.com/inavamsi/Thesis/tree/master/GA>

As shown in Figure 24, we notice that the results and also underlying structure are very different from the case of mutation with replication.

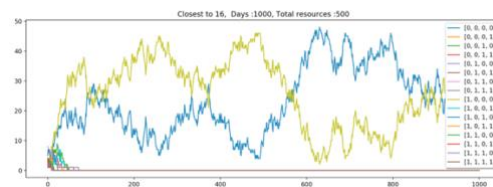


Figure 24 : Population timeline with reproduction by mating of two individuals.

There is a lot less variation as compared to the previous cases. Most agents die out and only ALLD and TTFT survive and dance.

But in the case of discounted history as shown in Figure 25, we see a very different structure and the lack of importance to discounting. There are no associations that can be drawn.

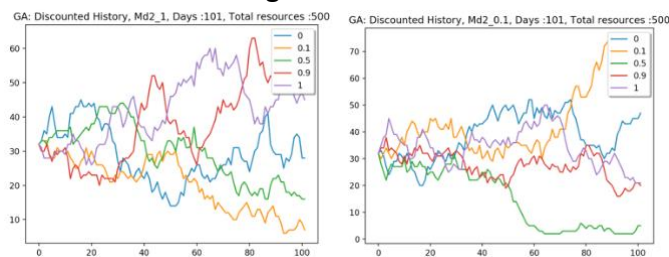


Figure 25 : Population timeline with discounted history.

Coming to the most important point. We will have to see if the anomaly of ‘knowing myself is better than knowing opponent’ has been rectified. The timeline is plotted in Figure 26.



Figure 26 : Population timeline different agents with different sizes of memory history.

We face the same problem as before, there are too many parameters for effective fitting and thus more complex agents get pruned really quickly. The SM(0,3) agents die out really fast and do not have time to adapt.

Everything GA

This has not been fully implemented.

The idea is to create a complete Genetic algorithm setup where agents can evolve to accommodate more memory. Also this structure will have chromosomes for every parameter form rate of mutation discounted history. This is to perfectly emulate nature as we evolved from a bunch of simple molecules.

The reason for failure of both approaches is the fundamental lack of structure. There is a requirement of structure in the strategy. To solve this we take inspiration from gene regulatory networks and this is shown in the following section.

Network(Approach 3)

What we learnt was the need for structure in our encoding. The naïve approach would be to create associations between parameters of our existing encoding. It also has to be robust and to a certain degree immune to noise. An obvious result would imply that each parameter should be associated more than once so that a letdown on one association will not hamper the result greatly.

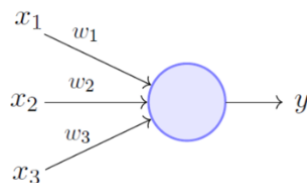
A Naive Network

The first canonical approach to convert a string to more structured setting is directly weighting the results.

So we define a model as follows. The first layer is an input string of 1's(C) and 0's(D) representing the concatenated history that is accessible by the agent. We create another layer which is created by a weighted sum of the previous layer. We continue till we get a single result. This is the probability of cooperating or playing C.

This can be thought of as perceptron without bias. The weights are normalized and bounded in the range [0,1].

Consider an $M(a,b)$ agent. Let $n = a+b$, then we define an $RM_{[1]}(a,b)$ agent as follows.



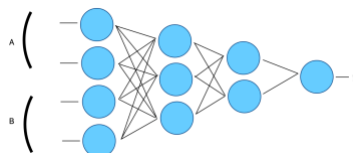
$$y = \sum_{i \in \{1 \dots n\}} x_i w_i$$

$$x_i = \begin{cases} 1, & \text{if C was played in corresponding round by corresponding player} \\ 0, & \text{if D was played in corresponding round by corresponding player} \end{cases}$$

$$0 \leq w_i \leq 1, \sum_{i \in \{1 \dots n\}} w_i = 1$$

So an $RM_L(a,b)$ agent takes a string of length $a+b$ as input and has $length(L)$ layers with each layer i having $L[i]$ nodes, where $L[i]$ corresponds to the i 'th layer.

Example $RM_{[3,2]}(2,2)$ where y is probability of playing C.



Why would such a weak model be chosen when we could directly use a more powerful perceptron model. It is as follows.

- Given an $RM_L(a,b)$ agent.

The total amount of weights is $n * L(1) + \sum_{i \in \{1 \dots len(L)-1\}} L(i) * L(i + 1) + L(length(L))$.

Although not by much it has a lower number of parameters than a similar multi-layer perceptron.

- An all 'C' input will result in a 'C' output because the weights are normalized and bounded by [0,1]. Thus allowing us to prune out ALLD and irrational agents. A one layer $RM_{\square}(0,1)$ agent is effectively a TFT agent. This kind of forces a level of cooperation by structure itself.

Code: https://github.com/inavamsi/Thesis/tree/master/Tree_no_bias

One improvement is that among $M(0,1)$, $M(0,2)$ and $M(0,3)$ agents. The more powerful $RM_L(0,3)$ agents do better as required, which was not the case in previous models. In Figure 27 we consider a variety of $L = \{[2],[3],[2,2],[3,2]\}$ so as to forgo bias.

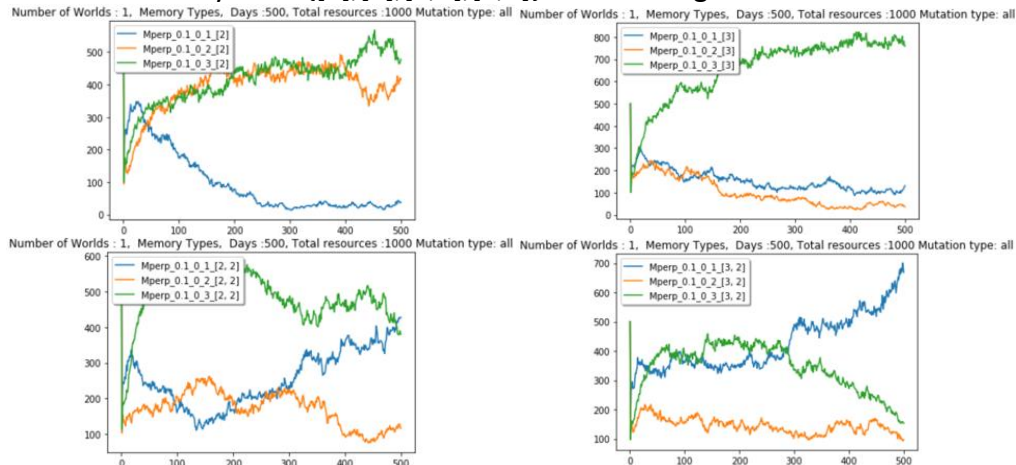


Figure 27 : Population timeline with different RM agents.

It is no surprise that for $L=[3,2]$ the $RM(0,1)$ model wins as clearly there are too many parameters to prune and the mutation rate is not enough. Here the $RM_L(0,1)$ is a TFT and doing better is a very good sign as compared to previous approaches as the agent easily can devolve to an ALLD agent.

This seems like a better model but it has its drawbacks. This model does not have to ability to adapt to the required extent. Also we face the same problem as the predecessors where an agent knowing itself is better off. This is shown in Figure 28.

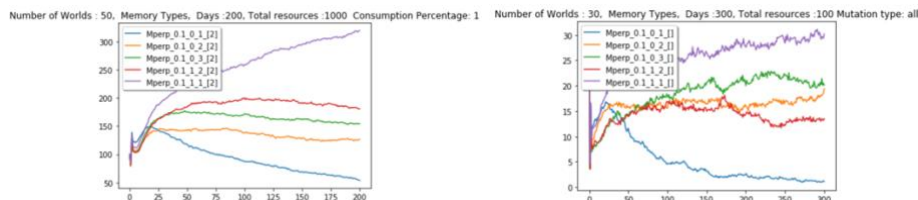


Figure 28 : Population timeline with different RM agents.

The agent is not robust enough so less powerful agents dominate. As clearly seen an $M(1,1)$ agent dominates in the presence of $M(1,2)$ and $M(0,3)$ agents.

We see the anomaly more clearly when we consider an population of $RM_L(2,0)$, $RM_L(1,1)$ and $RM_L(0,2)$, agents. On Figure 29, we consider a variety of $L = \{[2],[3],[2,2],[3,2]\}$ so as to forgo bias. In all cases we see the $RM_L(2,0)$ dominating and surviving at the end.

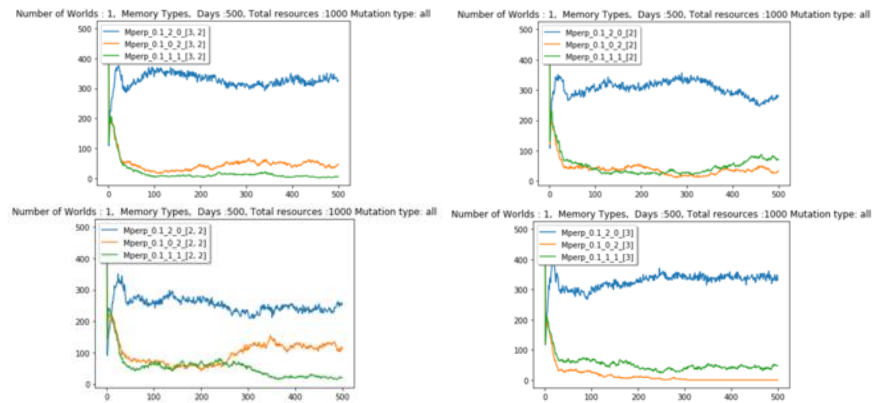
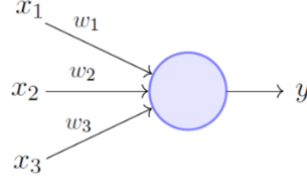


Figure 29 : Population timeline with different RM agents.

Perceptron memory

Taking inspiration from gene regulatory networks we consider a perceptron model for the strategy encoding where a single layer perceptron is defined as below. We denote this agent as $PM_L(a,b)$.

We have an additional parameter called bias on every node. Instead of a weighted sum we have a threshold function, which is 1 only if the weighted sum is greater than the bias. This is exactly how gene regulatory networks work. There is a required amount of threshold for activation.



$$y = \begin{cases} 1, & b + \sum_{i \in \{1 \dots n\}} x_i w_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$x_i = \begin{cases} 1, & \text{if } C \text{ was played in corresponding round by corresponding player} \\ 0, & \text{if } D \text{ was played in corresponding round by corresponding player} \end{cases}$$

$$-1 \leq b \leq 1, -1 \leq w_i \leq 1, \sum_{i \in \{1 \dots n\}} w_i = 1$$

Code : <https://github.com/inavamsi/Thesis/tree/master/Perceptron>

Compared the previous model to the weights which can now be anywhere from $[-1,1]$ and are not normalized. Although there is not a forced sense of cooperation we can see this model performing as required as shown in Figure 30.

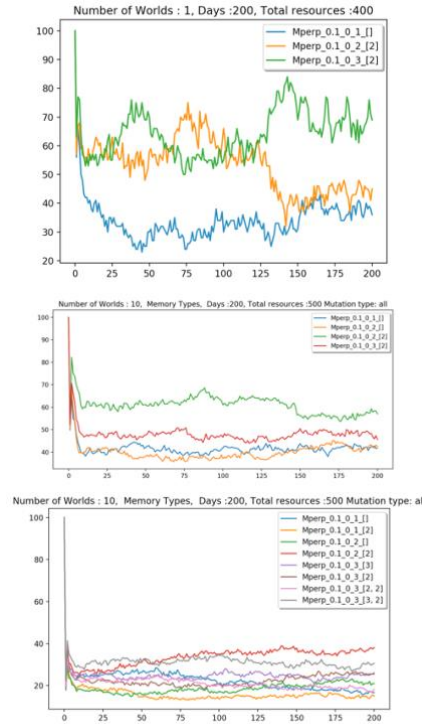


Figure 30 : Population timeline with different PM agents over multiple worlds.

Also the anomaly of ‘knowing self-history being better than opponent history’ is not seen(Figure 31) in the perceptron agent. It is not evident as to the more optimal one between an PM(0,2) agent or an PM(1,1) agent as there is a lot of context involved. But definitely an PM(2,0) should not be optimal or else cooperation is not possible.

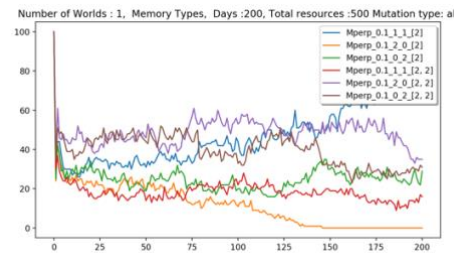


Figure 31 : Population timeline with PM agents with different sizes of history.

Costly signaling

We assign a cost for more complex structures which results in more complex signals.

One every important aspect is the increased amount of energy for an agent that remembers more and also has to do more computation(because of more complicated strategy encoding) for deciding a move.

We have to determine limits of penalty that can be given to agents for the above reasons. The limits up to which the equilibrium is maintained and no anomalies are present. What can also be called as the cost of being more complex. The brain was initially a huge vulnerable organ that took a huge portion of the energy and thus was initially a big liability in a harsh survivor world. This cost paid off in the long run. So the system should place importance on energy but also has to be flexible enough to allow long term changes which have high chances of being suppressed initially.

We consider a percentage increase in energy as to how much memory is being used for the encoding. A 1% energy cost implies n% increase in energy consumption where n is the number of nodes. In figure 32, we can see clearly that more complex agents do better but as energy consumption increases smaller agents obviously do better. This is a proper tradeoff between consumption, complexity, overfitting and rate of learning.

Code: <https://github.com/inavamsi/Thesis/tree/master/Energy%20Consumption>

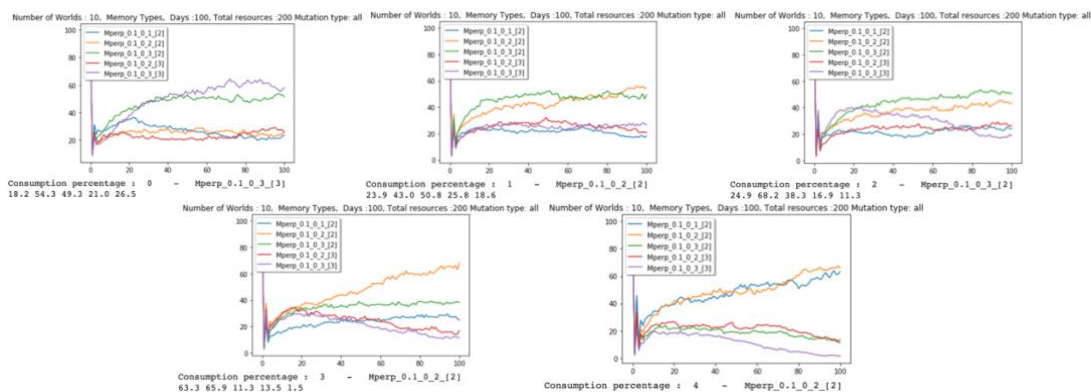


Figure 32 : Population timeline with PM agents with different costs of signaling.

Comparing Agents

We now have a huge variety of agents. We would like to see how complexity affects the ability of an agent to be a cooperative agent.

A standard measure of cooperation ability of an agent can be done with a measure of how well it can behave as a simple TFT agent.

Let us consider a $PM_{[]} (0,1)$ agent.

It has one weight and one bias. So given input c its output will be $\begin{cases} 1, & c \cdot w + b > 0 \\ 0, & \text{otherwise} \end{cases}$

To behave as a TFT agent, on input 1(C) output should be 1(C) and on input 0(D) output should be (D). This gives us two equations.

$$w + b > 0$$

$$b > 0$$

For both equations to hold true we need to have the following conditions.

$$w > 0$$

$$b > 0$$

$$|w| > |b|$$

Probability of a $PM_{[]} (0,1)$ agent to behave like a TFT = $P(w > 0) \cdot P(b > 0) \cdot P(|w| > |b|)$
 $= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

So in a randomly initialized population of $PM_{[]} (0,1)$ agents, $\frac{1}{8}^{\text{th}}$ of them will behave like TFT.

Let us consider a $PM_{[]} (1,1)$ agent. (Same for $PM_{[]} (0,2)$)

It has two weights and one bias.

To behave like a TFT the following conditions have to hold.

$$w_1 + w_2 + b > 0$$

$$w_2 + b > 0$$

$$w_1 + b \leq 0$$

$$b \leq 0$$

This translates to a probability ~ 0.041 .

Code: <https://github.com/inavamsi/Thesis/tree/master/Comparing%20Agents>

Table 4.1 : Comparing cooperation capability of different PM agents

Layer List L	Probability of being like TFT		
	$PM_L(0,1)$	$PM_L(1,1)$	$PM_L(0,2)$
[]	0.125	0.0416	0.0424
[2]	0.0498	0.0159	0.0155
[3]	0.062	0.01858	0.01892
[2,2]	0.0113	0.00391	0.00361
[3,2]	0.0141	0.00456	0.00482
[4]	0.0714	0.0215	0.0210
[3,3]	0.0146	0.0042	0.0043
[3,2,2]	0.378	0.00118	0.00106
[2,2,2]	0.283	0.000863	0.000853

For an SM agent the probability of acting like a TFT agent is 0. This is because only exact encodings will result in a specific agent. To do a proper comparison we consider a threshold of how close the agents are to a TFT agent. The results are shown in table 4.2

Table 4.2 : Comparing cooperation capability of different agents given threshold

Agent	Probability of being like TFT		
	Threshold =100%	Threshold =99%	Threshold =90%
SM(0,1)	0	0.00025	0.02715
SM(1,1)	0	0.00001	0.01849
SM(0,2)	0	0.00001	0.0488
RM_[2](0,1)	1	1	1
RM_[2](1,1)	0.0208	0.0273	0.705
RM_[2](0,2)	0.01386	0.05761	0.9976
PM_[2](0,1)	0.0486	0.04977	0.05
PM_[2](1,1)	0.01629	0.0162	0.032
PM_[2](0,2)	0.01627	0.01719	0.04593

These numbers clearly show how much more effective RM and PM agents are compared to SM agents.

Although these numbers for 100% threshold looks small, we have to realize that the number of types of cooperative agents blow up. Keeping track of all these types will give us a more significant amount. For example an SM(0,2) agent can encode both TFTT and TTFT agents while SM(0,1) cannot.

Recommender-Verifier (Approach 4)

What we have done so far is effectively pruning and supporting agents of undesired and desired qualities respectively by adding rules and regulations. This can be done more effectively by using a Verifier who constrict the search space. We constantly introduce new agents with the help of mutations, this can be carried out by a recommender who keeps recommending new agents and enlarge the search space.

This is effectively a Recommender-Verifier system where Recommenders keep adding new agents and verifiers keep pruning the undesirable ones. As the system grows bad recommenders and verifiers are also pruned out of the system.

Global Reputation

We consider a set group of standard agents with a global oversight. The global spectator updates the rankings of players by viewing each match and updating global reputation of corresponding players using some inherent rules.

Every agent can access the global reputation and decide to play or not play. If they choose not to play they have to pay a penalty while the other agent gets some compensation.

Consider the following experiment,

Code: <https://github.com/inavamsi/Thesis/tree/master/Reputation>

We take a population of standard agents. We let them play around for a few days till the reputation stabilizes to an equilibrium. We notice distinct reputation for each agent using different verifiers.

For example when we use a population of [ALLC, ALLD, GRIM, TFT, TFTT, TTFT, STFT]. We encode 4 rules for the global spectator, namely ["Naive Cooperation", "Strict", "Good Behavior", "Retaliating"]. Each rule supports a specific type of behavior in the agents.

This can be used to find the properties of agents that have complex encodings like an agent with Perceptron encoding of strategy. A rule encoding of [0,0,0,1] supports all retaliating agents like STFT and ALLD and will give them a high reputation as shown in Figure 33.

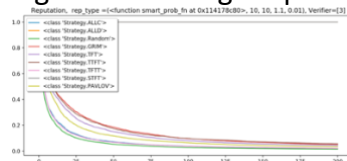


Figure 33 : Plot of Reputation against time

If rule encoding of [0, 0, 0.42, 0.58] results in a beautiful separation of agents into 3 groups as shown in Figure 34.

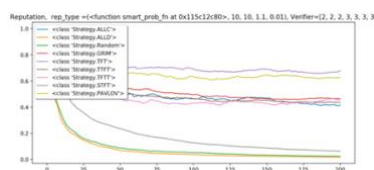


Figure 34 : Plot of Reputation against time

Community Utility

We have tried multiple methods to enforce a notion of community and cooperation. A simple method is to encode a notion of community by introducing community utility. This method entails using the Cobb-Douglas utility function to incentivize cooperation in one-shot non iterative Games.[9]

Consider the following two player Game 1 represented in Table 5.

Table 5: Scalar Payoff Matrix for Game 1

	Cooperate	Defect
Cooperate	2,2	1,3
Defect	3,1	1,1

The one-shot game Nash Equilibrium is (Defect, Defect). Any finitely bounded repeated version of this game also results in the same Nash Equilibrium as one can just backtrack from the last step. The unbounded repeated game has two Nash Equilibria, Alternating between (Cooperate, Defect)-(Defect, Cooperate) or just (Cooperate, Cooperate).

The paper ‘Modelling the commons as a game with vector payoffs’[9], encodes a notion of community using the Cobb-Douglas utility function and treating reward as a vector payoff of your reward and your opponent’s reward. The resulting payoff matrix is shown in Table 6.

Table 6: Modified Vector Payoff matrix of Table 5

	Cooperate	Defect
Cooperate	[2,2],[2,2]	[1,3],[3,1]
Defect	[3,1],[1,3]	[1,1],[1,1]

Cobb-Douglas Utility : $u(v_1, v_2) = k \cdot v_1^\alpha \cdot v_2^{1-\alpha}$

Where

- k is a constant
- v_1 is the reward
- v_2 is the reputation multiplier
- $\alpha \in (0,1)$

Considering the following utility function $u(v_1, v_2) = v_1^{0.6} \cdot v_2^{0.4}$, the Table 6 payoff matrix becomes Table 7 payoff matrix.

Table 7: Modified Payoff matrix of Table 6 using Cobb-Douglas Utility function

	Cooperate	Defect
Cooperate	2,2	1.552, 1.933
Defect	1.933, 1.552	1,1

The Nash Equilibrium becomes (Cooperate, Cooperate).

Now, consider the following two player game with payoff matrix represented by Table 8.

Table 8: Payoff Matrix for Game 2

	Cooperate	Defect
Cooperate	1,1	0.9,1.1
Defect	1.1,0.9	0.9,0.9

We can modify the scalar payoff into a vector payoff that encodes long-term benefit, community, and reputation. Let us consider a reputation multiplier and modify the game. Suppose Cooperating increases reputation by 20% while defecting decreases reputation by 20% we encode the payoff as shown in Table 9.

Table 9: Modified Vector Payoff matrix of Table 8

	Cooperate	Defect
Cooperate	[1,1.2],[1,1.2],	[0.9,1.2],[1.1,0.8]
Defect	[1.1,0.8],[0.9,1.2]	[0.9,1.2], [0.9,1.2]

Considering the following utility function $u(v_1, v_2) = v_1^{0.7} \cdot v_2^{0.3}$, the Table 9 vector payoff matrix becomes the Table 10 Payoff matrix.

Table 10: Modified Payoff matrix of Table 9 using Cobb-Douglas Utility function

	Cooperate	Defect
Cooperate	1.037, 1.037	0.95, 1.032
Defect	1.032, 0.95	0.95, 0.95

It can be clearly seen that the Nash Equilibrium is again (Cooperate, Cooperate). We can thus induce a notion of community.

This is yet another way in which we can induce cooperative behavior.

Deceptive Signaling between Predator and Prey

We first explore a refinement of Muller's math model for two prey unfavorable to the predator. Two additional papers have been added in the following sections which describes deceptive signaling through mimicry in a predator prey model.

Revised Muller's mathematical model

Muller describes a simple model describing the ratio in which two mimicking prey species are consumed before the predator learns that they are unfavorable. We look at it from a different perspective.

Dynamics of a Predator-Prey model in the presence of Mimicry

This paper describes the dynamics of a predator prey system in the presence of mimicry. We have introduced a novel framework using utility to describe mimicry.

Some Interesting Phenomena in Biomimicry

This paper describes Interesting hypothetical mathematically feasible scenarios that are but may or may not be biologically feasible. This paper extends over the shortcomings of the first paper and introduced prey evolution and multiarmed bandit predators.

Mullers mathematical model revised

Muller's Mathematical model

Müller presumed that the predators had to attack n unprofitable prey in a summer to experience and learn their warning coloration. Calling a_1 and a_2 the total numbers of two unprofitable prey species, then, Müller argued, if the species are completely unlike they each lose n individuals.

However, if they resemble each other, then species 1 loses $\frac{a_1 n}{a_1 + a_2}$ individuals, and species 2 loses $\frac{a_2 n}{a_1 + a_2}$ individuals. Species 1 therefore gains $n - \frac{a_1 n}{a_1 + a_2} = \frac{a_2 n}{a_1 + a_2}$ and species 2 similarly gains $\frac{a_1 n}{a_1 + a_2}$ in absolute numbers of individuals not killed.

The proportional gain compared to the total population of species 1 is $g_1 = \frac{a_2 n}{a_1(a_1 + a_2)}$ and similarly for species 2 is $g_2 = \frac{a_1 n}{a_2(a_1 + a_2)}$, giving the per head fitness gain of the mimicry when the predators have been fully educated.

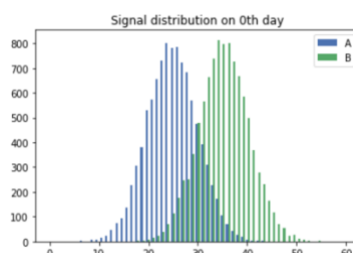
Revised Model

We consider a signaling game between the predator and prey. Every prey sends a signal and the predator receives the signal. Based on prior knowledge, the predator chooses to consume the prey. On consumption the predator updates knowledge on the respective signal. We consider a signaling spectrum $S = [0, L] \subset R$. Every prey sends a signal in the range $[0, L]$. If we combine all the signals from a particular species it forms the signal distribution of the species.

We consider two species namely A_1 and A_2 with population sizes a_1 and a_2 .

As stated in Muller's mathematical model, both species are unfavorable to the predator. We assume that the predator needs to consume n prey from a range of length t ($[s_0, s_0 + t] \subset S$ where $s_0 \in [0, L - t]$) signals to learn that the range is unfavorable. This implies that the total number of prey that will be consumed to learn that signal in the range $[0, L]$ is unfavorable is $\frac{L \cdot n}{t}$.

Every prey of a species does not produce the exact same signal as there is bound to be variation due to mutation. Let the signal distribution functions of the signals of species A_1 and A_2 be f_1 and f_2 respectively. We consider the two species to be mimicking if there is an overlap between the distributions f_1 and f_2 . See figure for reference.



Given that a predator requires to kill n individuals before it learns to avoid a signaling range of length t .

To find expected number of deaths for a specific species we do the following.

$$\text{Fraction of prey of species } A_1 \text{ consumed at a given signal} = \frac{P(p \in A_1 | s)}{P(p \in A_1 | s) + P(p \in A_2 | s)} = \frac{f_1(s)}{f_1(s) + f_2(s)}$$

Let total number of prey of species A_1 and A_2 consumed be n_1 and n_2 respectively.

Total number of prey consumed in length ds is $\frac{ds \cdot n}{t}$.

$$n_1 = \int_0^L \frac{f_1(s)}{f_1(s) + f_2(s)} \frac{n}{t} ds, \quad n_2 = \int_0^L \frac{f_2(s)}{f_2(s) + f_1(s)} \frac{n}{t} ds, \quad n_1 + n_2 = 1$$

Note: Real signal distributions are discrete due to innate thresholds and gene regulatory networks. This is an approximation in the limit given a large number of prey over a large feasible signaling range.

Dynamics of a Predator-Prey model in the presence of Mimicry

Abstract

This paper presents a novel idea of understanding toxicity and nutritional value of a prey to a predator with utility. Furthermore, we introduce the notion of prey mimicry in a multi predator - multi prey system. We identify and analyze the stability of equilibrium that arise as a result of the one or more predators inability to distinguish between certain prey.

Keywords: Mimicry, Predator-Prey, Lotka-Volterra Equations, Equilibrium, Stability, Toxic prey

INTRODUCTION

The dynamics between predator and prey is a fascinating subject. This subject has been extensively studied and formalized. We introduce the notion of prey toxicity and mimicry and study the resulting dynamics. This can easily be extended to other fields and sciences. A classic example would be the dynamics behind counterfeit currency which is essentially a toxic mimic to real currency.

We assume a simple system with multiple predators and prey. Each prey species can either be toxic or nutritious to a predator. A rational predator will always consume the nutritious prey and avoid the toxic prey. The predator needs to be able to distinguish between the different prey species.

To distinguish between prey of different species, the predator uses the signal sent by the prey. The signal can be anything from color, smell, sound or even a dance pattern. So a red toxic species can easily be distinguished from a blue nutritious species. But in the case where both prey send the same signal to a predator, they cannot be distinguished by the predator. The predator essentially treats them as a single prey and this can be considered as mimicry.

Interestingly, mimicry depends on the point of view. For example, two prey can both be Red in color but produce different sounds. So a predator that hunts using just visual signals will not be able to distinguish between the two. Whereas the predator the hunts using acoustic signals will be able to distinguish between the two prey. Furthermore, a prey can be toxic to certain predators but nutritious to others. The toxicity and nutritional values is encoded as utility of the prey to the predator. We specifically explore the dynamics of this example in further sections.

The concept of mimicry becomes trivial without the presence of toxicity. To sum up, we aim to study the interesting dynamics of a predator prey system with the presence of mimicry and toxicity.

Generalized Lotka-Volterra equations for multi predator – multi prey model

In general for a multi predator-multi prey model we consider the following assumptions.

1. The prey population finds ample food at all times.

2. The food supply of the predator population depends entirely on the size of the prey population.
3. The rate of change of population is proportional to its size.
4. During the process, the environment does not change in favor of one species, and genetic adaptation is inconsequential.
5. Predators have limitless appetite.
6. In the absence of prey the predators exponentially decay.
7. There is no competition between the prey.
8. Every prey species has a fixed utility for every predator species.

We consider n prey where each prey is denoted by x_i , $i \in \{1 \dots n\}$. We consider m predators where each predator is denoted by y_j , $j \in \{1 \dots m\}$.

$$\frac{dx_i}{dt} = \alpha_i x_i - \sum_{j=1}^m \epsilon_{ij} \beta_{ij} x_i y_j, \quad \forall i \in \{1 \dots n\}$$

$$\frac{dy_j}{dt} = \sum_{i=1}^n \epsilon_{ij} \beta_{ij} u_{ij} x_i y_j - \delta y_j, \quad \forall j \in \{1 \dots m\}$$

- $\epsilon_{ij} \in \{0, 1\}$ is an indicator variable indicating whether x_i is a prey for predator y_j .
- $\alpha_i > 0$ corresponds to growth rate of prey x_i
- $\beta_{ij} > 0$ denotes the number of interactions or rate of predation between prey x_i and predator y_j
- u_{ij} denote the utility of prey x_i to predator y_j . This can be scaled as required
- $\delta_j > 0$ denote the loss/death rate of predator y_j

If $u_{ij} > 0$ then prey x_i is nutritious to predator y_j and is beneficial. If $u_{ij} < 0$ then prey x_i is toxic to predator y_j and is fatal.

If a predator y_j can distinguish between a toxic x_t and a nutritious x_n . Then $\epsilon_{nj} > 0$, while $\epsilon_{tj} = 0$. Optimally, the predator should choose to not consume the toxic prey. The interesting case is when the nutritious prey x_n mimics toxic prey x_t . If a prey x_i is toxic to predator x_j then we consider that $\epsilon_{ij} = 0$. This gets interesting when the predator cannot distinguish between a toxic prey and a nutritious prey. This is also known as Batesian mimicry.

Mimicry in a simple two predator- two prey model

We consider two prey A and B. We also consider two predators

P and Q. We assume the same exponential growth rate for both prey ($\alpha = \alpha_A = \alpha_B$), the same rate of interactions for every pair of predator and prey

($\beta = \beta_{AP} = \beta_{AQ} = \beta_{BP} = \beta_{BQ}$) and the same exponential loss/death rate for both predators ($\delta = \delta_P = \delta_Q$)

$$\begin{aligned}\frac{dA}{dt} &= \alpha A - \epsilon_{AP}\beta_{AP}AP - \epsilon_{AQ}\beta_{AQ}AQ \\ \frac{dB}{dt} &= \alpha B - \epsilon_{BP}\beta_{BP}BP - \epsilon_{BQ}\beta_{BQ}BQ \\ \frac{dP}{dt} &= \epsilon_{AP}\beta_{AP}u_{AP}AP + \epsilon_{BP}\beta_{BP}u_{BP}BP - \delta P \\ \frac{dQ}{dt} &= \epsilon_{AQ}\beta_{AQ}u_{AQ}AQ + \epsilon_{BQ}\beta_{BQ}u_{BQ}BQ - \delta Q\end{aligned}$$

A. Single Predator - Two Prey

Consider an ecosystem with only one type of predator P ($Q = 0$). A is nutritious to predator P ($u_{AP} > 0$) while prey B is toxic to predator P ($u_{BP} < 0$).

Case 1: If A and B can be distinguished by the predator P.

$$\epsilon_{AP} = 1, \epsilon_{BP} = 0$$

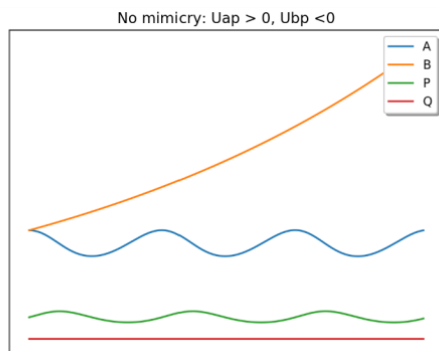


Fig. 1. Prey B grows exponentially while A and P follow the standard single predator - single prey model.

Case 2: If A mimics toxic B and cannot be distinguished by the predator P.

$$\epsilon_{AP} = 1, \epsilon_{BP} = 1$$

Case 2.1: If $u_{AP}A + u_{BP}B \geq 0$

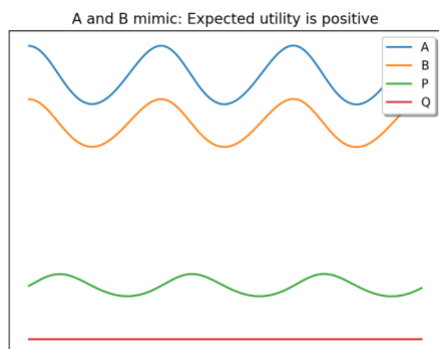


Fig. 2. Effectively a one predator - one prey model.

This can be treated as single predator single prey model with the utility of the single prey being the expected utility which is $(u_{AP}A + u_{BP}B)/(A+B)$.

As the predator cannot distinguish between the two prey species it sees a single signal which can be attributed to an expected reward or utility.

Given that the growth rate of A and B is the same, if $\beta_{AP} \neq \beta_{BP}$ then the prey species with higher rate of interaction goes extinct. This is a standard result of one predator - two prey model.

Case 2.2: If $u_{AP}A + u_{BP}B < 0$

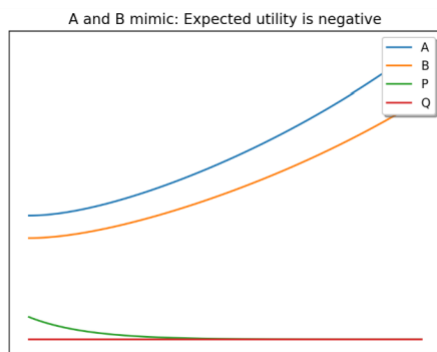


Fig. 3. Predator dies out and then prey grows exponentially.

The rate of change of predators is negative. Thus both predators decay and die out while the prey under no predation grow exponentially.

Equilibrium In an equilibrium or a fixed point of the system all the derivatives are zero.

$$\frac{dA}{dt} = \alpha A - \beta AP = 0$$

$$\frac{dB}{dt} = \alpha B - \beta BP = 0$$

$$\frac{dP}{dt} = \beta u_{AP}A + \beta u_{BP}B - \delta P = 0$$

The fixed points are $(A, B, P) = (0, 0, 0)$ and $(x, y, \frac{\alpha}{\beta})$ such that $u_{AP}x + u_{BP}y = \delta/\beta$. To check for stability of the fixed points we check whether the real part of all eigen values of the Jacobian are negative.

$$J(A,B,P) = \begin{pmatrix} \alpha - \beta P & 0 & -\beta A \\ 0 & \alpha - \beta P & -\beta B \\ \beta u_{AP}P & \beta u_{BP}P & \beta u_{AP}A + \beta u_{BP}B - \delta \end{pmatrix}$$

$$J(0,0,0) = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & -\delta \end{pmatrix}$$

α is an eigenvalue and is positive hence this is not a stable fixed point. This shows that $(0,0,0,0)$ is not stable even in the generalized version.

$$J(x, y, \frac{\alpha}{\beta}) = \begin{pmatrix} 0 & 0 & -\beta x \\ 0 & 0 & -\beta y \\ \alpha u_{AP} & \alpha u_{BP} & 0 \end{pmatrix}$$

We see that $\lambda^3 = \lambda(-\alpha\delta)$. One of the eigenvalues is 0 and thus the equilibrium is non-hyperbolic and thus not robust.[13]

B. Two Predator - Two Prey

We look at a few examples of mimicry in this model.

- Prey A and B are distinctly identifiable by both the predators P and Q.
- Prey A and B can be distinguished by predator P but not by Q (or vice versa).
- Prey A and B can't be distinguished by either predator P or Q.

When a predator y_j can distinctly identify a prey x_i and the corresponding utility u_{ij} is negative then clearly $\epsilon_{ij} = 0$.

In the trivial case where both prey have positive utility to both predators. The model follows a standard two predator-two prey model. In the case that both prey have negative utility to both predators. Both predators die out.

Note: Below are a few cases in each condition. Other cases are a result of change in labels. Consider all epsilon values to be 1 by default.

The indicator variables can be interpreted as following. 1. If $\epsilon_{ij} = 0$ and $\epsilon_{i'j} = 0$ then predator y_j will go extinct. The final model is essentially a one predator - one prey model.

2. If $\epsilon_{ij} = 0$ and $\epsilon_{ij} = 0$ then prey x_i will grow exponentially.

The final model is essentially a one predator - one prey model.

3. If $\epsilon_{ij} = 0$ and $\epsilon_{i'j'} = 0$ and $i \neq i'$ and $j \neq j'$, then the model is essentially two independent one predator - one prey model.

4. If predator y_j cannot distinguish between x_i and $x_{i'}$ then

$$\epsilon_{ij} = \epsilon_{i'j} = 0 \text{ if } E_u \leq 0$$

$$\epsilon_{ij} = \epsilon_{i'j} = 1 \text{ if } E_u > 0$$

Where $E_u = (u_{ij}x_i + u_{i'j}x_{i'}) / (x_i + x_{i'})$ is the expected utility predator y_j gets from the mimicry ring.

Table 1. Prey A and B are distinctly identifiable by both the predators P and Q.

U_{AP}	U_{BP}	U_{AQ}	U_{BQ}	Indicator
+	+	+	-	$\epsilon_{BQ} = 0$
+	+	-	-	$\epsilon_{AQ} = 0, \epsilon_{BQ} = 0$
+	-	+	-	$\epsilon_{BP} = 0, \epsilon_{BQ} = 0$
+	-	-	+	$\epsilon_{AQ} = 0, \epsilon_{BP} = 0$

+	-	-	-	$\epsilon_{AQ} = 0, \epsilon_{BQ} = 0$
---	---	---	---	--

Table 2. Prey A and B can be distinguished by predator P but not by Q (or vice versa).

U_{AP}	U_{BP}	U_{AQ}	U_{BQ}	Indicator
+	+	+	-	Depends on expected utility
+	+	-	-	$\epsilon_{AQ} = 0, \epsilon_{BQ} = 0$
+	-	+	-	Not fixed
+	-	-	+	Not fixed
+	-	-	-	$\epsilon_{BP} = 0, \epsilon_{AQ} = 0, \epsilon_{BQ} = 0$

Table 3. Prey A and B cannot be distinguished by either predator P or Q.

U_{AP}	U_{BP}	U_{AQ}	U_{BQ}	Indicator
+	+	+	-	Depends on expected utility
+	+	-	-	$\epsilon_{AQ} = 0, \epsilon_{BQ} = 0$
+	-	+	-	Depends on expected utility
+	-	-	-	$\epsilon_{AQ} = 0, \epsilon_{BQ} = 0$

C. Special Cases of Two Predator - Two Prey

The interesting situation occurs in Table 2 case 3 and 4 labelled as 'Not fixed'.

We take a look at case 3 defined as follows. Prey A and B can be distinguished by predator P but not by Q.

In case 3, the utilities are as follows. $u_{AP} > 0, u_{BP} < 0, u_{AQ} > 0, u_{BQ} < 0$.

In case 4, the utilities are as follows. $u_{AP} > 0, u_{BP} < 0, u_{AQ} < 0, u_{BQ} > 0$

Predator P can distinguish between prey A and B and thus will never choose to consume B as utility is negative. On the other hand predator Q has no choice but to consume whatever prey it encounters to survive.

We assume the following e 's. It trivially reduces in other cases as seen before. $\epsilon_{AP} = 1, \epsilon_{BP} = 0, \epsilon_{AQ} = 1, \epsilon_{BQ} = 1$

In a equilibrium or a fixed point of the system all the derivatives are zero. The trivial fixed point is $(A, B, P, Q) = (0, 0, 0, 0)$. The other fixed point is when $P = 0$ and thus exactly the same as Case 2 of one predator - two prey model mentioned before.

It is impossible for both predator and prey species to all be alive in an equilibrium, this is because A will be predated more than B as long as P is non zero.

Generalizing mimicry in a two predator – two prey model

$$\frac{dA}{dt} = \alpha_A A - \epsilon_{AP} \beta_{AP} A P - \epsilon_{AQ} \beta_{AQ} A Q$$

$$\frac{dB}{dt} = \alpha_B B - \epsilon_{BP} \beta_{BP} B P - \epsilon_{BQ} \beta_{BQ} B Q$$

$$\frac{dP}{dt} = \epsilon_{AP} \beta_{AP} u_{AP} A P + \epsilon_{BP} \beta_{BP} u_{BP} B P - \delta_P P$$

$$\frac{dQ}{dt} = \epsilon_{AQ} \beta_{AQ} u_{AQ} A Q + \epsilon_{BQ} \beta_{BQ} u_{BQ} B Q - \delta_Q Q$$

Equilibrium There are multiple fixed points or equilibrium but we want to focus on the equilibrium where none of the prey and none of the predators go extinct.

There is only one fixed point at which all $A, B, P, Q \neq 0$. The fixed point is (a, b, p, q) where,

- $a = \frac{\delta_P \epsilon_{BQ} \beta_{BQ} u_{BQ} - \delta_Q \epsilon_{BP} \beta_{BP} u_{BP}}{\epsilon_{BQ} \beta_{BQ} u_{BQ} \epsilon_{AP} \beta_{AP} u_{AP} - \epsilon_{AQ} \beta_{AQ} u_{AQ} \epsilon_{BP} \beta_{BP} u_{BP}}$
- $b = \frac{\delta_P \epsilon_{AQ} \beta_{AQ} u_{AQ} - \delta_Q \epsilon_{AP} \beta_{AP} u_{AP}}{\epsilon_{AQ} \beta_{AQ} u_{AQ} \epsilon_{BP} \beta_{BP} u_{BP} - \epsilon_{BQ} \beta_{BQ} u_{BQ} \epsilon_{AP} \beta_{AP} u_{AP}}$
- $p = \frac{\alpha_A \epsilon_{BQ} \beta_{BQ} - \alpha_B \epsilon_{AQ} \beta_{AQ}}{\epsilon_{BQ} \beta_{BQ} \epsilon_{AP} \beta_{AP} - \epsilon_{AQ} \beta_{AQ} \epsilon_{BP} \beta_{BP}}$
- $q = \frac{\alpha_A \epsilon_{BP} \beta_{BP} - \alpha_B \epsilon_{AP} \beta_{AP}}{\epsilon_{BP} \beta_{BP} \epsilon_{AQ} \beta_{AQ} - \epsilon_{AP} \beta_{AP} \epsilon_{BQ} \beta_{BQ}}$

The Jacobian will be as follows.

$$J(A, B, P, Q) = \begin{pmatrix} 0 & 0 & -\epsilon_{AP} \beta_{AP} & -\epsilon_{AQ} \beta_{AQ} \\ 0 & 0 & -\epsilon_{BP} \beta_{BP} & -\epsilon_{BQ} \beta_{BQ} \\ \epsilon_{AP} \beta_{AP} u_{AP} & \epsilon_{BP} \beta_{BP} u_{BP} & 0 & 0 \\ \epsilon_{AQ} \beta_{AQ} u_{AQ} & \epsilon_{BQ} \beta_{BQ} u_{BQ} & 0 & 0 \end{pmatrix}$$

We consider the special cases we saw in the previous section. All other cases devolve to simpler models.

Prey A and B can be distinguished by predator P but not by Q. The utilities are as follows. We consider static indicator variables. Thus $\epsilon_{AP} = 1$, $\epsilon_{BP} = 0$, $\epsilon_{AQ} = 1$, $\epsilon_{BQ} = 1$ The fixed point becomes (a, b, p, q) where

- $a = \frac{\delta_P}{\beta_{AP} u_{AP}}$
- $b = \frac{\delta_Q \beta_{AP} u_{AP} - \delta_P \beta_{AQ} u_{AQ}}{\beta_{BQ} u_{BQ} \beta_{AP} u_{AP}}$
- $p = \frac{\alpha_A \beta_{BQ} - \alpha_B \beta_{AQ}}{\beta_{BQ} \beta_{AP}}$
- $q = \frac{\alpha_B}{\beta_{BQ}}$

Special Case 1 $u_{AP} > 0, u_{BP} < 0, u_{AQ} > 0, u_{BQ} < 0$
Conditions for equilibrium

- $\alpha_A \beta_{BQ} > \alpha_B \beta_{AQ}$
- $\delta_P \beta_{AQ} u_{AQ} > \delta_Q \beta_{AP} u_{AP}$

Special Case 2 $u_{AP} > 0, u_{BP} < 0, u_{AQ} < 0, u_{BQ} > 0$
Conditions for equilibrium

- $\alpha_A \beta_{BQ} > \alpha_B \beta_{AQ}$

For both special cases the Jacobian is as follows.

$$J(A,B,P,Q) = \begin{pmatrix} 0 & 0 & \beta_{AP} & \beta_{AQ} \\ 0 & 0 & 0 & \beta_{BQ} \\ \beta_{AP} u_{AP} & 0 & 0 & 0 \\ \beta_{AQ} u_{AQ} & \beta_{BQ} u_{BQ} & 0 & 0 \end{pmatrix}$$

Proof of claim : There exists a positive eigenvalue for the below matrix if $u_{AP} u_{AQ} u_{BQ} < 0$.

$$J(A,B,P,Q) = \begin{pmatrix} 0 & 0 & \beta_{AP} & \beta_{AQ} \\ 0 & 0 & 0 & \beta_{BQ} \\ \beta_{AP} u_{AP} & 0 & 0 & 0 \\ \beta_{AQ} u_{AQ} & \beta_{BQ} u_{BQ} & 0 & 0 \end{pmatrix}$$

The characteristic equation is

$$\lambda^4 + \lambda^2(\beta_{BQ}^2 u_{BQ} + \beta_{AQ}^2 u_{AQ} + \beta_{AP}^2 u_{AP}) + \beta_{BQ} u_{BQ} \beta_{AQ} u_{AQ} \beta_{AP} u_{AP} = 0$$

This is of the form $ax^2 + bx + c = 0$ where

- $x = \lambda^2$
- $a = 1$
- $b = (\beta_{BQ}^2 u_{BQ} + \beta_{AQ}^2 u_{AQ} + \beta_{AP}^2 u_{AP})$
- $c = \beta_{BQ} u_{BQ} \beta_{AQ} u_{AQ} \beta_{AP} u_{AP}$

$$\lambda^2 = x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$\lambda = \pm \sqrt{\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}}$$

We consider the following root. $\lambda = \sqrt{\frac{-b + \sqrt{b^2 - 4ac}}{2a}}$

Since $u_{AP} u_{AQ} u_{BQ} < 0$ it follows that $4ac < 0$. Thus $\sqrt{b^2 - 4ac} > -b$. Which results in $\lambda = \frac{\pm \sqrt{b^2 - 4ac}}{2a}$ which is a positive number.

Thus we have found the required positive eigenvalue(λ).

There exists a positive eigenvalue for the Jacobian matrix if $u_{AP} u_{AQ} u_{BQ} < 0$. Thus this is an unstable equilibrium.

We can conclude that there is no stable equilibrium in a two predator - two prey model where each predator has one toxic and one non-toxic prey when one predator can distinguish between the two while the other cannot.

Example: We consider an equilibrium/fixed point. By perturbing the parameters we see that it is not a stable equilibrium. This has been shown in Fig.4 and Fig5.

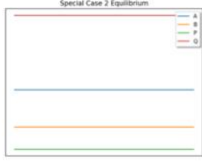


Fig. 4. Special Case 2: Equilibrium Fixed point according to $\alpha_A \beta_{BQ} > \alpha_B \beta_{BQ}$.

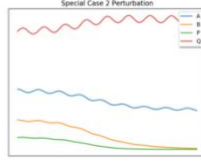


Fig. 5. Special Case 2: Unstable equilibrium with small perturbation given $\alpha_A \beta_{BQ} > \alpha_B \beta_{BQ}$.

Section Conclusion

We have presented a formal framework for analyzing dynamics of the signaling game between predator and prey using utility.

We have also shown that there is no stable equilibrium in a two predator-two prey system when one predator can distinguish between the two prey while the other cannot. With the additional condition that at least one prey is toxic to each predator.

Discussion

We have assumed a non-evolving system where the predator does not continually learn and evolve as well, although at a slower pace. A possible solution to the learning capability of a predator is a dynamic indicator variable which has been discussed below.

This will be discussed in a following paper which treats signals and utility as related quantities and allows for their evolution through the natural selection of the predator.

A. Generalizing to m predator n prey mimicry

When a predator y_j can distinctly identify a prey x_i and the corresponding utility u_{ij} is negative then clearly $\beta_{ij} = 0$. On the other hand if a predator y_j cannot distinguish between some prey species x_s , $s \in S \subseteq \{1 \dots n\}$. The predator views this chunk as a single unit and the utility will be the expected utility from the entire chunk. The expected utility is $E[u_{y_j}(x_s)] = \sum_{s \in S} \beta_{sj} u_{sj} x_s$.

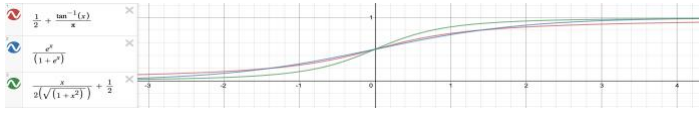
$$\text{Then } \forall s \in S \text{ we have, } \begin{cases} \epsilon_{sj} = 0 & \text{if } E[u_{y_j}(x_s)] \leq 0 \\ \epsilon_{sj} = 1 & \text{if } E[u_{y_j}(x_s)] > 0 \end{cases}$$

The problem starts here as the expected utility is dynamically changing with time as other predators interact with specific subsets of the population.

B. Dynamic Indicator Variable $\epsilon \in [0, 1]$

We consider a dynamic epsilon value that continuously changes with the expected utility. This will effectively allow the predator to be rational and choose a signal only if it is getting a positive utility from it. And similar to natural predators as utility approaches 0, the indicator variable approaches 0. For this we use an activation function

We look at some standard activation functions.



But these functions still have significant indicators when utility is negative. Thus we can scale and translate these functions to behave as shown in Fig. 7.

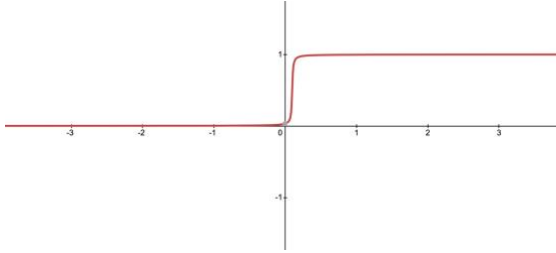


Fig. 7. The function is $f(x) = \frac{1}{2} + \frac{\tan^{-1}(100(x-0.1))}{\pi}$

As utility becomes negative the predator will drastically reduce the amount of prey consumed. We consider the small value to be the exploration parameter to constantly check the utility it would receive. Thus for Mimicry our indicator variable will look like the following if A and B are not distinguishable by Q.

$$\epsilon_{AQ} = \epsilon_{BQ} = \frac{1}{2} + \frac{\tan^{-1}(k - \frac{u_{AQ}A + u_{BQ}B}{A+B} - x_0)}{\pi}$$

C. Example of practical use for Artificial Intelligence

Two competing companies A and B have people visiting their websites. Each company invests money into advertising/services to those that visit their website. Both companies create bots to bring down their competitors. It is beneficial if the recipient was human, but resources wasted if the recipient was a bot.

A bot which is the malicious agent will try to solve the captcha to pass off as a human. The websites will get bad utility from a bot and thus create better captcha which are harder for bots to crack. This resembles a coevolutionary chase between predator and prey. The bot is trying to mimic the human introducing the element of mimicry. This perfectly mirrors two predators(Companies) consuming prey(visitors). Some prey are toxic(bots) and some nutritious(humans). Their ability to distinguish between the signals(Captcha/Turing test) changes with time as they learn.

If we know the dynamics one can easily analyze the amount of money a company must spend on creating bots for competitors or how much one must invest into a captcha system. We specifically study the dynamics of such predator-prey models in the following part.

Some interesting scenarios in Bio-mimicry

Abstract

This paper presents a novel idea of looking at bio-mimicry from the perspective of utility. Every prey has a signal and utility to the predator. Multi-armed bandit predators with zero prior information are introduced into the ecosystem. As a result of exploration and exploitation the predators naturally select the prey that result in the evolution of those signals. This produces a variety of interesting phenomena which are presented.

Introduction

Standard predator - prey models treat predators and prey as non-evolving, non-learning entities. In reality, over time both predators and prey start a co-evolutionary chase which is fueled by continual evolution and learning. A simple example is when the predators consume the slower prey which are the easiest to catch. The remaining prey population is faster on average and succeeding generations will be faster. As a result of quicker prey the slower predators are pruned out which over time results in faster predators.

We focus on situations where prey can evolve to be more toxic which should result in more resistant predators. The scope of this paper is the cases where the lifespan of a predator covers a significantly large number of generations of prey. This will result in the predator not being able to evolve fast enough to gain immunity to the toxic prey. So naturally predators have a learning mechanism and thus learn and continually adapt over time. When a predator eats a prey they update their inner beliefs about the prey based on the effect the prey has on the predator. To accomplish this we treat the predator as a learning agent who uses reinforcement learning to update their policy of choosing to eat or not eat a prey from the same species when encountering it again in the future.

One lacking requirement is the ability of the predator to distinguish between multiple prey so that they can update the respective belief. To distinguish they use the signal corresponding to the prey. A signal can be visual, acoustic or even a smell. A red bird will essentially be sending a different visual signal from a blue bird.

Breaking it down, the predators receive a signal and based on their prior beliefs, they choose to consume the prey that sent the signal. It gets interesting when two different species evolve to send the same signal. Effectively the predator cannot distinguish between the two and will thus share a common belief with respect to the common signal.

We also see a fascinating variety of phenomena arising due to difference in value the predator has to the different species. We effectively model this using utility. The utility of a toxic prey to a predator will be negative. Furthermore the utility function encodes preference and is exactly what we need to model the behavior of evolving predators and prey. In regards to two mimicking species that share a common signal, the utility the predator associates to the signal will be the expected utility from the two species in question.

To keep it simple, we analyze the following simple system. We consider a signaling game with multiple species of Predator and Prey. The prey (Lifespan of Mayfly ≈ 24 hours) has a relatively small lifespan compared to the predator (Lifespan of Common Toad ≈ 12 years). This

way the prey can undergo mutations in its signal and adapt while the predator can only learn about the prey but cannot change genetically in the short period. As the signals of the prey mutate and evolve there will be convergence and divergence of signals of different species. We try to understand the conditions and extent of mimicry with respect to the utility value of different prey species to the predator. The predators will all be multi-armed bandit agents. We look at a few interesting phenomena that arise as a mathematical consequence of our assumptions.

Consider the following example where this can be applied in the field of Artificial Intelligence. Two competing companies A and B have people visiting their websites. Each company invests money into advertising/services to those that visit their website. Both companies create bots to bring down their competitors. It is beneficial if the recipient was human, but resources wasted if the recipient was a bot.

A bot which is the malicious agent will try to solve the captcha to pass off as a human. The websites will get bad utility from a bot and thus create better captcha which are harder for bots to crack. This resembles a coevolutionary chase between predator and prey. The bot is trying to mimic the human introducing the element of mimicry. This perfectly mirrors two predators(Companies) consuming prey(visitors). Some prey are toxic(bots) and some nutritious(humans). Their ability to distinguish between the signals(Captcha/Turing test) changes with time as they learn.

If we know the dynamics one can easily analyze the amount of money a company must spend on creating bots for competitors or how much one must invest into a captcha system. We specifically study the dynamics of such predator-prey models in the following part.

New work

Bio-mimicry has scarcely been interpreted using a formal mathematical framework of utility. Another novel idea is using multi-armed bandit predators that can continuously learn and update the non-stationary distribution of signals.

Most of the following phenomena explained below are mathematically feasible scenarios. Some like 'Mimicry can be bad to both species' are original and hypothetical scenarios that may or may not have biological feasibility.

Model

Ecosystem

The ecosystem consists of multiple prey and multiple predators. On a given day a predator encounters some number of prey. Each prey gives out a respective signal. Based on the predators prior knowledge it chooses to consume a prey of a specific signal. After eating the prey, based on the utility it got from the prey it updates its information about the signal. This continues over time leading to a variety of phenomena.

It is to be noted that the prey mutate and evolve over each generation but the predator whose lifespan is significantly larger than the prey continuously learns and update sits

knowledge. As naturally seen, the predator comes across a randomly sampled set of prey and has to choose one prey of them.

There are two types of ecosystems we can focus on. In the first the prey also compete with each other for resources. In the second one the prey have a reproduction rate and can indefinitely grow if there is no predation. Both ecosystems have very distinct properties and showcase very different phenomena under similar conditions. We mainly focus on the indefinite growth in the examples seen in this paper. signaling range.

Prey

Every prey has the following properties which will be defined below.

- Gene Vector which encodes toxicity, evasion rate, etc.
- Signal
- Mutation rate
- Reproduction rate

Species Prey divided into species. Every prey of a specific species share common traits such as reproduction rate, mutation rate and genetic constraints. In the paper each species will be represented by a different color.

Gene vector, Signal and Utility to Predator Every prey has some genetic values which we can encode as a gene vector. The gene vector contains is bound by the genetic constraints of the prey species. A natural analogy is the set of chromosomes. The set of genes in the gene vector determines a variety of things from how the prey looks to its speed of evasion from the predator. It essentially is analogous to a change in chemical composition which changes the taste, toxicity and nutritional value to the predator. To sum it up, the gene vector determines the signal the prey sends and also the utility a predator receives from consuming the respective prey.

We consider a signaling range bounded in the range $[L,R]$. Let g be the signal function that takes gene vector as input and returns signal.

Example 1: Consider two prey species A and B. Let the gene vector of A be $[2,1]$ and the gene vector of B be $[1,2]$. Suppose $g([x,y]) = x+y$ is the signaling mechanism of both species, then prey both A and B give out the same signal 3. Given a predator P with utility function $f([x,y]) = x^2 - y$, it can be seen that the utility of A to the predator is 3 while the utility of B to the predator is -2 which implies that B is toxic to the predator P. P cannot distinguish between the A and B as it receives the same signal 3.

Example2: Consider two prey species A and B. Let the gene vector of A be $[2,1]$ and the gene vector of B be $[1,2]$. Suppose $g([x,y]) = 2x + y$ is the signaling mechanism of the species A and $g([x,y]) = x+y$ is the signaling mechanism of the species B. Then prey of A signal 5 while B signals 3. Given a predator P with utility function $f([x,y]) = x^2 - y$, it can be seen that the utility of A to the predator is 3 while the utility of B to the predator is -2 which implies that B

is toxic to the predator P. P can clearly distinguish between the two prey and will learn that signal 5 corresponds to a good prey while signal 3 corresponds to bad prey.

Since the signal is the only thing visible to the predator, we look specifically at parts of the gene vector affected by the signal. For the sake of simplicity we assume that the signaling function(g) to be a injective function. This implies that the gene vector can be inferred from the signal. As the signaling function is injective, we can assume that the utility to the predator which depends on gene vector depends on the signal. We thus use the notation $u_{AP}(s)$ which refers to the utility of prey A with signal s to the predator P.

When we say constant utility we imply that the mutations in the gene vector result in changes in signal but not in utility for the predator. An example would be that the change in the first gene could make a prey more red in color but the first gene does not affect the toxicity or taste of the prey to predator in any way and thus does not influence utility.

If utility to predator is proportional to gene vector which in turn is proportional to signal, then utility will increase with increasing signal. Signal can be appropriately scaled along with mutation rate.

Mutation Every prey species has a natural system that has evolved with some constraints. There is a cost attributed to most changes. Some changes affect reproduction rate, some affect color, some affect speed or toxicity and other evasion. This is a result of mutations in the gene vector. Effectively mutations in the gene vector result in new signals and utilities.

Mutations can be Gaussian, which implies that the most of the mutations will be concentrated around the point of origin with a very small probability of large mutations. We consider a simplified version where mutations are restricted to the current signal and the neighboring two signals. Given mutation rate p , then the probability of signal s mutating to $s-1$ or $s+1$ is p , while the probability of staying at s is $1 - 2p$.

For example, if we consider $u(s) = s$ then it means that a mutation in the gene vector will result in a change in s and a proportional change in utility.

Signaling distribution Every prey has a signal. So a prey species will have a signaling distribution which is a collection of signals from all the prey of the species. As the species are naturally selected the signal distribution changes shape. this will be described in detail below.

Predator

A predator has the following properties.

- Prior Knowledge
- Learning system and related parameters
- Signal perception function
- Utility function
- Consumption rate

Species We also have multiple species of predators each with their own learning systems, utility and perception.

Utility The utility function takes in the gene vector of prey as input and return utility to predator for consuming prey as output. As defined in the Prey section, we use the notation $u_{AP}(s)$ which refers to the utility of prey A with signal s to the predator P.

Signal from predator perspective There are multiple prey species, each with their own signal distribution. The predator only sees the combined distribution of signals.

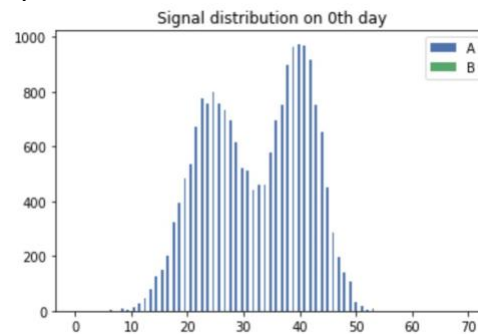


Fig 1: This is the signal the Predator sees.

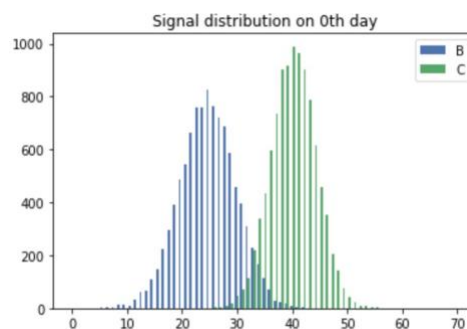


Fig 2: This is the actual signal distribution of two separate prey species.

Signal perception function In higher dimensions, predators see a projection of the signal onto a smaller space. For example given a 2 dimensional signal $s = (s_1, s_2)$ of a prey species A, a predator P that sees only the first dimension will not be able to distinguish (10,3) and (10,20) but it will be able to distinguish between (10,3) and (5,3). To encode this property we use a signal perception function. A real world example is that off color blind predators. Another example is a species that can has a weak smell and weak eyesight and thus distinguishes a prey as a function of both.

So a signal perception function $f(s_1, s_2) = s_1 + s_2$ will not be able to distinguish between (10,20) and (25,5). This essentially reduces a 2 dimensional signal to a one dimensional signal. This is sometimes more efficient due to the cost of evolving more complex senses or a more robust learning system.

Consumption Rate This refers to the number of prey the predator needs to consume at a given day. This can be easily modified to required amount of utility per day.

Learning and Prior Knowledge We assume that a new predator with no prior knowledge is introduced into a new environment where it has to learn the utilities from the signal.

Due to evolution, the signal distributions is non-stationary and utility corresponding to a signal keeps changing. Thus the predator needs to forget over time and thus needs a discounting parameter. Clearly neighboring signals will have similar utilities for simple functions. Thus the predator needs to update its learning from a specific signal to surrounding signals as well. Finally we need a predator that explores for new opportunities but also exploits as is naturally seen.

UCB Bandit We consider a standard discounted UCB agent with a bandit for every signal in the range $[L,R]$. During update, the bandits of surrounding signals are also updated. The predator chooses the signal s that maximizes the following value.

$$\mu_s + \sqrt{\left(\frac{\alpha \log(n)}{2n_s}\right)}$$

Where,

- μ_s is the discounted average utility gotten so far from signal s and its neighbors.
- α is a scaling constant
- n is the total number of prey consumed so far
- n_s corresponds to the number of prey of signal s consumed so far

Traditional UCB where we update the corresponding arm is really slow to learn and not robust enough in the presence of discounting. To solve this we have modified the update rule to update all neighbors in a gaussian manner. Closer neighbors are highly affected by update and farther neighbors have a negligible effect.

Signal Drift

As predators start exploring different signals they learn the corresponding utility over time and start exploiting signals of high utility. This is essentially natural selection and results in the drift of the signal distribution for the particular prey species. But there are various conditions that result in the magnitude and direction of the drift. We look at the major ones.

There are two main types of signal drift. Signal drift of a prey species due to a signaling range corresponding to a lower utility. The other one being signal drift due to a signaling range with lower expected utility. Depending on the ecosystem one may dominate over the other producing interesting phenomena.

Signal Drift due to lower self-utility

Consider a prey species A with constant utility to predator P , meaning that the utility does not change with signal. For example a prey could change its voice frequency but its utility to the predator might not change at all. This will result in the predator uniformly sampling the prey as it prefers every signal exactly the same.

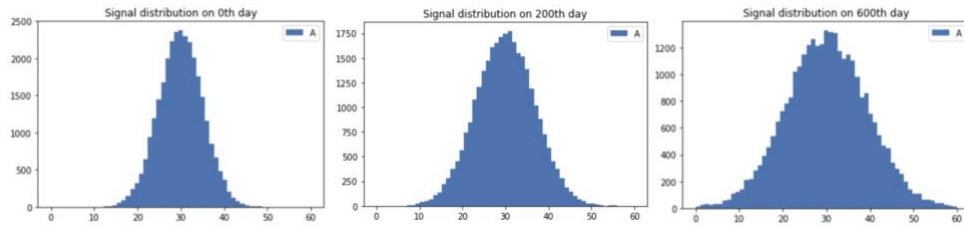


Fig 3: Signal distribution over a 600 generations plotted at different time intervals. The distribution gets flatter and fatter over time. In this specific example, the rate of consumption by predators is exactly equal to the reproduction rate so the total number of prey at any given generation is approximately 30000.

We model utility as a function of signal. Given a signal distribution for the prey species, the prey species mutate and slowly change the signaling distribution. The predator will choose signals that induce a higher utility in the prey. This results in the movement of the signaling distribution in the other direction. An example is shown of such a drift is shown below. We consider a prey with utility proportional to the signal resulting in a left shift because the utility on the left will be lower.

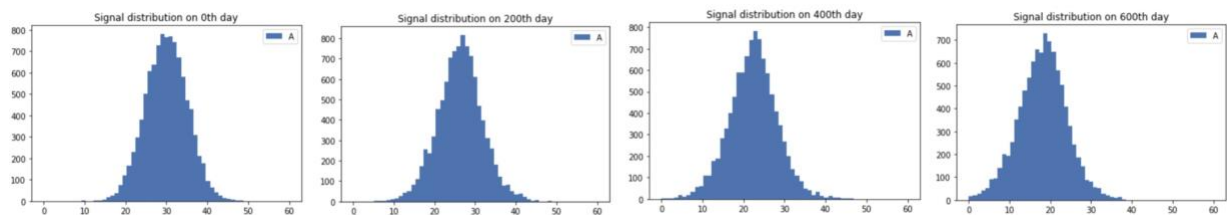


Fig 4: Signal distribution over a 600 generations plotted at different time intervals. We see that the signal distribution shifts left as right has higher utility.

An uniform movement in one direction specifically happens in increasing or decreasing functions. If the utility which is a function of signal is not monotonic (increasing or decreasing) then the prey population will greedily descend in the direction of lower utility. An example has been shown with a non-monotonic utility function.

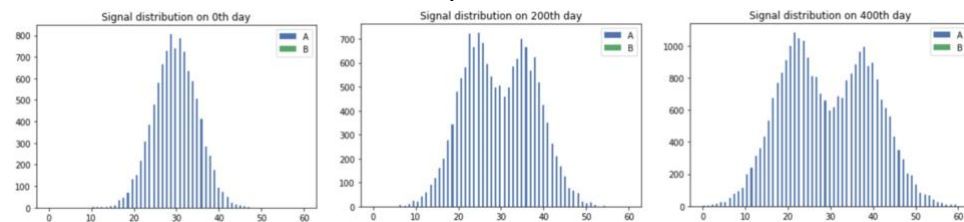


Fig 5: Signal distribution over a 600 generations plotted at different time intervals. The utility function is $u(s) = k - (s - 30)^2$ see that the signal distribution splits at the point $s=30$.

Signal drift due to lower expected utility

Consider two prey species A, B and a Predator P.

A(Blue) :Constant high utility to P B(Green) :Constant low utility to P

Over generations, due to exploiting of prey with highest utility by predator we see a natural shift in the signaling range of the population.

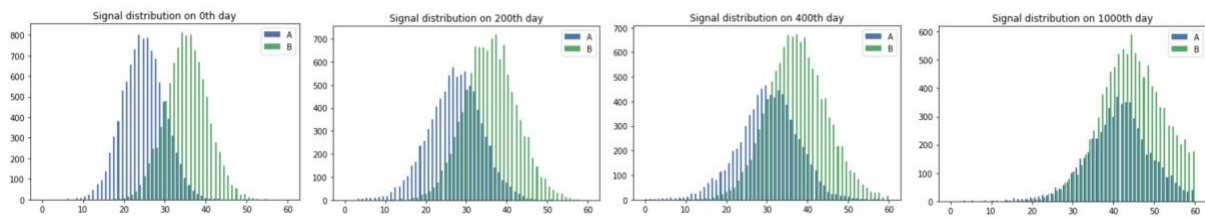


Fig 6: Signal distribution over a 1000 generations plotted at different time intervals.

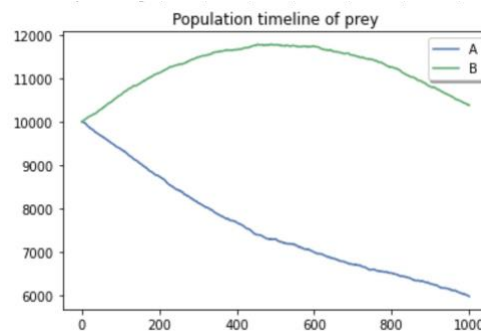


Fig 7: Population time series over the 1000 generations during drift. In this specific example, the total consumption of predator is higher than the growth rate of prey species combined. This results in a decline of total population over time.

The predator preys on the signaling range with maximum utility. Thus signals that are produced by only the species with high utility will be the prime targets for the predator. On exploitation the prey species with high utility will naturally evolve to signaling range that is shared with a prey species of lower utility. The predator will learn that the common signaling range and associate it with the expected utility. Over generations, a drift in the signal is achieved.

In the same notion, a prey species with low utility will drift away from prey species with higher utility. But this will be slower than the drift in case of a prey species with high utility.

Interesting Phenomena in 1-Dimensional signaling

We consider a one dimensional gene vector for the prey which corresponds to a one dimensional signal. All species can signal on a band from L to R.

Pooling Equilibrium

This phenomena occurs when multiple prey species pool together which results in the predator not being able to distinguish between any of the species. This is an equilibrium as no species can be naturally selected and thus there is no induced drift.

Consider 3 prey species A, B, C and Predator P.

A (Blue): constant high utility to P

B (Green): constant low utility to P

C (Red): constant medium utility to p

$$u_{BP}(s) < u_{CP}(s) < u_{AP}(s)$$

$$\alpha_B = \alpha_C = \alpha_A$$

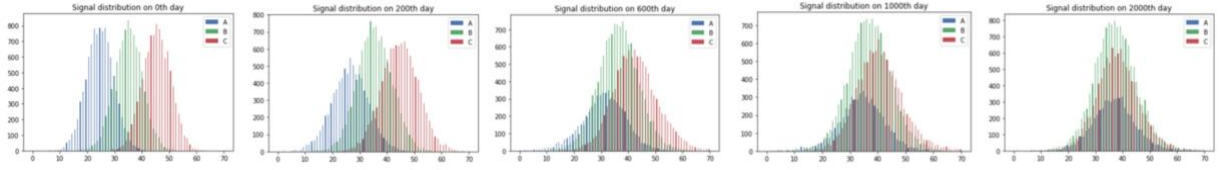


Fig 8: Signal distribution over a 2000 generations plotted at different time intervals.

The pooling equilibrium is stable. This is because any change in the signal distribution of a single species will change the distribution of expected utility. This effectively will trigger a drift of all prey species such that the equilibrium is maintained.

Polymorphic Mimicry

This phenomena is the splitting of a single peaked signal distribution of a prey species to a multipeak distribution. This can easily be seen in the case of non monotonic functions. Interestingly it is possible to generate the same phenomena using only simple linear monotonic functions.

We consider a situation where prey choose a trade-off between reproduction rate and toxicity. Higher toxicity implies a lower reproduction rate but also a lower utility to the predator.

Consider 3 prey species A, B, C and Predator P.

A (Blue): Toxicity is high, utility to predator P is low and reproduction rate is low

B (Green): Toxicity is low, utility to predator P is high and reproduction rate is high

C (Red): Toxicity is moderate, utility to predator P is moderate and reproduction rate is moderate

$$u_{BP}(s) > u_{CP}(s) > u_{AP}(s)$$

$$\alpha_B < \alpha_C < \alpha_A$$

In the scenario where prey species B would have gone extinct. Instead we observe that in the presence of mimicry not only does prey species B thrive at the end but also its signal splits into two peaks.

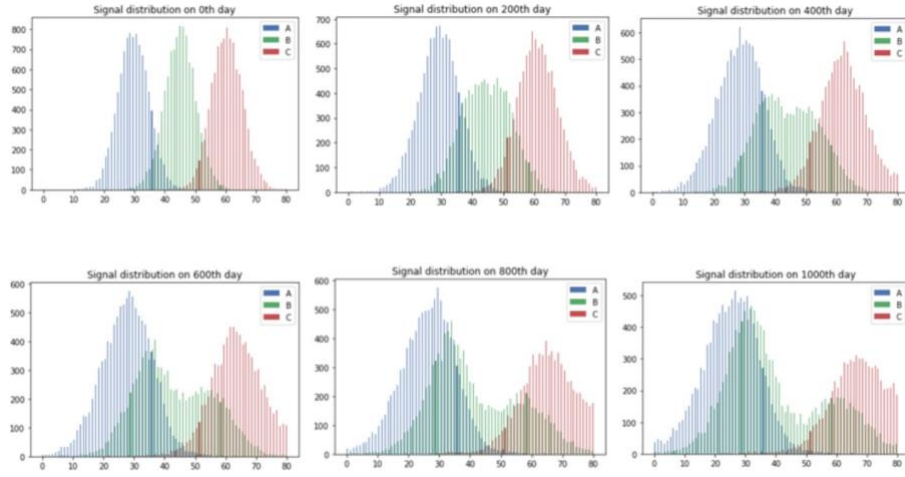


Fig 9: Signal distribution over a 1000 generations plotted at a period of every 200 generations.

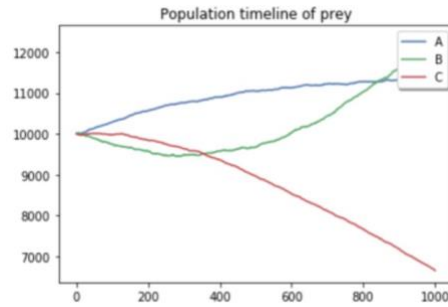


Fig 10: Population time series of prey over 1000 generations

Mimicry can be bad to both species

We show a situation where the mimicry of A and B results in mutual extinction. In the control ecosystem where the other does not exist, they actually thrive.

Consider 3 prey species A, B, C and Predator P.

A (Blue): Utility to predator P is negatively correlated to signal : ($u_{AP}(s) = a_0 - a_1s$)

B (Green): Utility to predator is constant and rate of mutation is insignificant : ($u_{BP}(s) = b_0$)

C (Red): Utility to predator is correlated to signal : ($u_{CP}(s) = c_0 + c_1s$)

Since utility of A is negatively correlated with signal it drifts towards the right(R) because the predator exploits A of lower signal. Since utility of C is correlated with signal it drifts towards the left(L) because the predator selects species C of higher signal. Since, utility of B is constant there is no drift but instead just a flattening of the signal distribution.

Case 1: An ecosystem with only prey species A and B A is naturally selected towards a signal of lower utility and thus drifts right(R). Initially the utility of A is higher and thus the corresponding signals are exploited but as A drifts right its utility decreases till a tipping point where its utility equals the utility of the agents of species B. As a result we see an initial drop in population for A followed by recovery.

The turning point is when a critical chunk of species A drifts past to signal higher than $s = \frac{a_0 - b_0}{a_1}$. Thus effectively resulting in a lower utility for the predator. A rational predator will make the choice to choose signals corresponding to B compared to A.

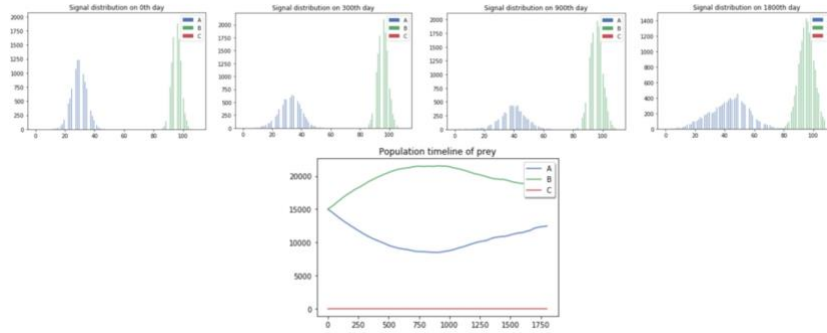


Fig 11: Signal distribution over a 1800 generations.

Case 2: An ecosystem with only prey species B and C This is very similar to the previous case. C is naturally selected towards a signal of lower utility and thus drifts left(L). Initially the utility of C is higher and thus the corresponding signals are exploited but as C drifts left its utility decreases till a tipping point where its utility equals the utility of the agents of species B. As a result we see an initial drop in population for C followed by recovery. The turning point is when a critical chunk of species A drifts past to signal lower than $s = \frac{c_0 - b_0}{c_1}$.

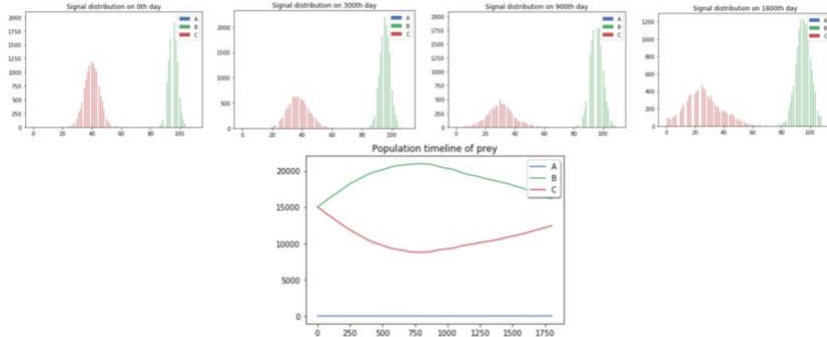


Fig 12: Signal distribution over a 1800 generations.

Case 3: An ecosystem with all prey species A,B,C As a result of the pooling equilibrium formed by the signals of A and C. The predator P cannot distinguish between the signals of A and C and thus cannot naturally select them.

At the point of pooling, if the expected utility of the signals in the pooling equilibrium is lesser than the utility of the signals of prey species B. This leads to exploitation and thus eventually extinction.

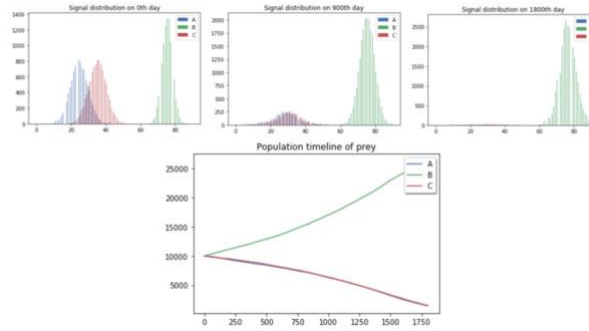


Fig 13: Signal distribution over a 1800 generations.

We consider a simple case where $u_{AP}(s) = 3 - 0.05s$, $u_{BP}(s) = 1.3$, $u_C(s) = 0.05s$ and initial distributions mean at points 25 for A and 35 for C. $s = 30$ will be the mean of the pooling equilibrium between A and C after drifting. This is because initially both A and C have the same utility and are thus selected together. The expected utility of any signal in this pooling equilibrium will be $3/2$. This is greater than the utility of prey B and thus the pooling equilibrium will be exploited till extinction.

Above was a very specific example, but can be easily generalized to show that a pooling equilibrium can form and thus stop natural selection towards a lower utility, effectively leading to extinction. This is mainly because of the lack of freedom in the 1-Dimensional signaling space.

Oscillating Mimicry

Oscillating mimicry where signals of two species is constantly oscillating, is not possible when we consider a one dimensional gene vector, identity signal function(g) and a fixed predator population. This is due the formation of a pooling equilibrium.

Interesting Phenomena in 2-Dimensional signaling

2-Dimensional signaling opens up a whole new world phenomena. In 1-Dimensional signaling due to a lack of freedom we are restricted to certain type of phenomena. here we present just a basic idea of how far we can go.

Drift in 2-Dimensions

Drift in higher dimensions is similar to 1-dimensional case except with more freedom. The drift is determined by the Signal perception function, utility function of predator and genetic vector of prey.

Consider a prey species A and a predator P. A has a genetic constraint that it cannot reproduce if its genes are too different from the rest. This stops the signaling distribution from spreading out in all directions. The predator P can see signals only along the line $s_1 - s_2 = k$ because of the signal perception function $f(s_1, s_2) = s_1 + s_2$. Furthermore a utility function $u_{AP}(s_1, s_2) = s_1 + s_2$ results in a diagonally left-down drift for the prey species A.

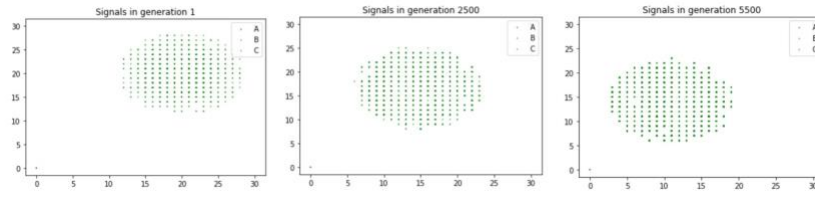


Fig 14: Signal distribution over a 5500 generations plotted at different times.

If both Red and Blue are toxic, then by continuously introducing new predators into the system we observe that Red and Blue converge towards a common signal. This is also known as Mullerian mimicry. The reason is shown in the previous section ‘Revised Mullerian Mathematical Model’.

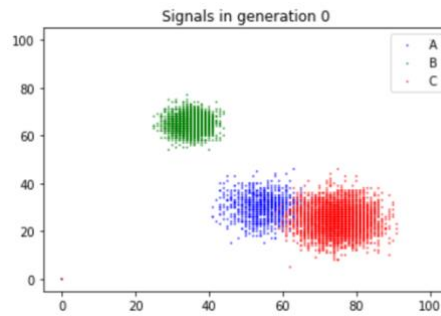


Fig 15: Mullerian mimicry between Red and Blue after continuously introducing new predators. The prey in the common region are feasted on lower.

Switching directions

In the presence of two prey species and two predator species, we can create very interesting phenomena which involve changes in drift direction. This is not possible in 1-Dimensional signaling due to the creation of a Pooling Equilibrium.

Consider two prey species A, B and two predator species P and Q. Let the predators and prey have the following functions.

- P signal perception function : $f(s_1, s_2) = s_1$. This means that P can distinguish only the x coordinate of the signal.
- Q signal perception function : $f(s_1, s_2) = s_2$. this means that Q can only distinguish the y coordinate of the signal.
- The signaling function of A as a function of the gene vector : $g(v_1, v_2) = (v_2, -v_1)$
- The signaling function of B as a function of the gene vector : $g(v_1, v_2) = (-v_1, -v_1)$
- The utility function of P as a function of gene vector : $u(v_1, v_2) \propto v_1 + v_2$
- The utility function of Q as a function of gene vector : $u(v_1, v_2) \propto v_1 - v_2$

The above functions result in the following utility functions.

- $U_{AP}(S_1, S_2) \propto S_1 - S_2$
- $U_{AQ}(S_1, S_2) \propto -S_1 - S_2$
- $U_{BP}(S_1, S_2) \propto -S_1 - S_2$
- $U_{BQ}(S_1, S_2) \propto -S_1 + S_2$

Starting at the appropriate positions with correct scaling parameters such that A is diagonally left-up of B. If initially $U_{AP} > U_{BP}$ and $U_{BQ} > U_{AQ}$, then A will drift left while B will drift down. After a point when $U_{AP} = U_{BP}$ and $U_{BQ} = U_{AQ}$, the predators start gaining interest in the other prey which they were not predating originally. Slowly the predators reduce the predating of their original prey and begin predating the other prey. This will result in A drifting upward and B drifting right. So A initially starts drifting left, followed gradual change in direction towards a drift upward. Similarly B initially drift downward and gradually shifts to drifting right.

Section Concluding Remarks

In most of the phenomena we have focused on using only linear functions. Although this may not be the case, it only shows that a simple set of linear rules can result in complex mechanisms like 'The Game of Life'. If we introduce higher degree or more complex functions we will be able to construct even more fascinating phenomena for bio-mimicry. Furthermore if we increase the dimension of signaling and the nature of genetic constraints, we can explore an even wider variety of phenomena.

This paper gives us a broad framework to understanding the existence of multiple mathematically feasible phenomena. This can easily be extended to a the dynamics between multiple agents that can systematically be described as Learning predators and evolving prey.

A standard example where this can be used is in understanding the introduction of Artificial intelligence or learning systems into new ecosystems. Another example is that of fake currency which mimics real currency. If we design a set of learning bots that learn to remove fake currency from a system. The bots learn over time and utility is precisely that of removing fake currency and not real currency. Fake currency can be modelled as prey. If fake currency is left unchecked it will be misused and there will be more malicious agents pumping it into the economy without the fear of getting caught. Furthermore too much fake currency also does not work and this represents prey-prey competition. If we identify enough fake currency we can narrow down the source which is similar to lowering reproduction in prey due to a smaller population.

Multiple systems can effectively be modeled using this framework. We can test out a plethora of interesting phenomena and potentially learn to stop them if required, with minimal changes.

Conclusion

The thesis started with understanding the dynamics of AI cooperation and deception through signaling games. This is required to better understand the changes AI will bring into society. We have thus presented certain basic games and their dynamics as we so that we can help build a more effective and efficient tomorrow.

Throughout the thesis, we have presented quite a few novel ideas which solve or help us better understand the dynamics behind cooperation and deception. We have looked at cooperation through replicator equations, invasive strategies, and evolutionarily stable strategies. We have created simple frameworks like optimal play in a noisy environment of standard IPD agents, to slightly more complicated frameworks like predator-prey multi-dimensional signaling. We have tried many approaches from reputation, genetic algorithms, perceptron memory, community utility, each with their benefits. We have introduced deception through biomimicry, learning predators, and evolving prey. This lays a foundation that can be translated into multiple real-world problems we face in the new age of AI.

As mentioned multiple times, cooperation and deception go hand in hand. It is imperative to understand these dynamics if we want a more effective future in the presence of changing technology.

Future Work

1. Harsanyi and Aumann have described hierarchies of belief and common knowledge. We aim to extend signaling games to an inductive setup and understand the dynamics of said system.
2. Use generators to generate visual pattern signals. Then use predators equipped with convolutional neural nets to receive visual signals. Thus, study Mullerian mimicry of butterflies as studied by muller using reinforcement agents that can see(Computer Vision).
3. Dynamics of intra-specie signaling to share information and communicate.
4. Dynamics of changing system where Predators and Prey both evolve and learn.

References

1. The Iterated Prisoners' Dilemma : 20 years on
Graham Kendall, Xin Yao, Siang Yew Chong
2. Chong, Siang Yew & Humble, Jan & Kendall, Graham & Li, Jiawei & Yao, Xin. (2007). Iterated Prisoner's Dilemma and Evolutionary Game Theory. 10.1142/9789812770684_0002.
3. Chong, Siang Yew & Humble, Jan & Kendall, Graham & Li, Jiawei & Yao, Xin. (2007). Learning IP D Strategies Through Co-evolution. 10.1142/9789812770684_0003.
4. Axelrod, Robert. *The Evolution of Cooperation*. Basic Books, 2006.
5. Li, Jiawei. (2007). How to Design a Strategy to Win an IPD Tournament. 10.1142/9789812770684_0004.
6. García, J., & van Veelen, M. (2018). No Strategy Can Win in the Repeated Prisoner's Dilemma: Linking Game Theory and Computer Simulations. *Frontiers In Robotics And AI*, 5. doi: 10.3389/frobt.2018.00102
7. Sandholm, Tuomas & Crites, Robert. (1996). Multiagent Reinforcement Learning in the Iterated Prisoner's Dilemma. *Bio Systems*. 37. 147-66. 10.1016/0303-2647(95)01551-5.
8. Wang, Shiheng & Lin, Fangzhen. (2017). Invincible Strategies of Iterated Prisoner's Dilemma.
9. Casey, W., Kellner, A., Memarmoshrefi, P., Morales, J., & Mishra, B. (2018). Deception, identity, and security. *Communications Of The ACM*, 62(1), 85-93. doi: 10.1145/3190836
10. García, J., & van Veelen, M. (2018). No Strategy Can Win in the Repeated Prisoner's Dilemma: Linking Game Theory and Computer Simulations. *Frontiers In Robotics And AI*, 5. doi: 10.3389/frobt.2018.00102
11. Wei Chu, Lihong Li, Lev Reyzin, Robert E. Schapire (2011), "Contextual bandits with linear payoff functions", *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*: 208–214
12. Müller, Fritz (1879). "Ituna and Thyridia; a remarkable case of mimicry in butterflies. (R. Meldola translation)". *Proclamations of the Entomological Society of London*. 1879: 20–29.
13. Equilibrium : <http://www.scholarpedia.org/article/Equilibrium>
14. Predator-Prey model : http://www.scholarpedia.org/article/Predator-prey_model
15. Axelrod, Robert (1984), *The Evolution of Cooperation*, Basic Books, ISBN 0-465-02122-0
16. Lejano, Raul & Ingram, Helen. (2012). Modeling the commons as a game with vector payoffs. *Journal of Theoretical Politics*. 24. 66-89. 10.1177/0951629811423120.