Clayton School of Information Technology
Monash University

Literature Review — Semester 1, 2014

Simulated Evolution of Non-Regular Strategies for Repeated Games

Bradon Thomas Hall

22051635

June 13, 2014

Supervisor: Julian Garcia

Contents

The mechanisms that allow for the evolution of cooperation in competitive, survival-of-the-fittest scenarios can be investigated using game theory and computational simulations. The repeated Prisoners Dilemma game provides a simple way of modelling the competitive environment, with strategies competing at playing this game. I discuss literature describing Direct Reciprocity and Assortment, the two mechanisms for the evolution of cooperation relevant to my project. I focus on the impact that the choice of how to simulate strategies has on what strategies evolve for the repeated Prisoner's Dilemma. Based on this review of literature, I suggest extending existing work using novel representations that allow for more complex strategies to evolve.

1 Introduction

Cooperation is a behaviour observed in many environments; perhaps the most obvious example is in biology. Darwin's Theory of Evolution- in which individual fitness determines whose genes are passed on to the next generation- seems at first glance to fail to explain these cooperative behaviours. Other members of the population would gain an advantage by exploiting these cooperators, yet by some mechanism cooperation emerges and persists. Darwin suggested that mechanisms such as kinship and reciprocity are necessary for cooperation to succeed. These and other mechanisms have been investigated, but the problem of cooperation is still not fully understood (Pennisi; 2005).

Evolutionary Game Theory (EGT) provides a means to investigate the evolution of cooperation via mathematics and computational simulations, complementing observational investigation in relevant fields like biology (Dawkins; 1989). This review focuses on EGT research involving the Prisoner's Dilemma (PD); a simple but effective means to model cooperation.

The Prisoner's Dilemma refers to the following scenario: two prisoners each have the choice of two options; they keep quiet (cooperate, C), or they can spill the beans (defect, D) and testify against their partner. If both cooperate, each serve a year in jail (receive payoff R, Reward). If both betray their partner, they serve 2 years (receive payoff P, Punishment). If one testifies and one keeps quiet, the prisoner who testifies walks free (receive payoff T, Temptation), and his/her partner serves 5 years (receive payoff S, Sucker).

	Cooperate	**Defect**
Cooperate	R = 4, **R = 4**	S = 1, **T = 5**
Defect	T = 5, **S = 1**	P = 2, **P = 2**

Table 1: Payoff Matrix for the Prisoner's Dilemma, with example values.

The specific times served (payoffs) do not define the game; the game is a Prisoner's Dilemma if the best outcome for a player is to betray their partner, the worst outcome is to be betrayed, and both players testifying is worse than both players staying silent. The hierarchy of payoffs is given by:

$$T > R > P > S \tag{1}$$

The best overall outcome (minimum sum of both sentences) occurs if both choose to keep quiet; however, they are better off individually if they testify against their partner. If their partner has chosen to keep quiet, they can walk free by snitching. If their partner is testifying, they can reduce their sentence by doing so too. They can only gain by switching to defecting, but this is not the optimal outcome- hence the dilemma.

In EGT payoffs determine the fitness of an individual. This fitness determines what strategies are selected for in the next generation. The aim of the technique is to predict what strategies will perform well under evolutionary competition. In the standard one-shot Prisoner's Dilemma cooperation is never successful- to obtain cooperation, some mechanism(s) are required.

Nowak (2006) discusses mechanisms for the evolution cooperation; the most familiar concept is likely to be kinship. There is an advantage for helping out a related individual- you share many genes, and increasing their likelihood of reproduction increases the likelihood of passing on those shared genes (Hamilton; 1964). The concept is not limited to biology, and here I consider it in general- the idea that related strategies can benefit from helping each other.

The idea of Group Selection is to extend the concept of kinship to a group level- cooperating for the benefit of the group. Models of group and kin selection are mathematically equivalent, and mediated by Assortment (Traulsen; 2010). Assortment will be used to describe these mechanisms. Assortment refers to population structure, and in particular the likelihood of meeting another 'alike' player.

Consider an extreme case. A population of 100 contains 90 members that always defect (ALLD) and 10 members that always cooperate (ALLC). In a situation where every member meets every other member with equal probability the defectors exploit the cooperators, gaining a higher payoff, and the cooperators inevitably lose. In a situation where cooperators band together, and only interact with each other, they will do well. Since mutual cooperation has a higher payoff than mutual defection in the Prisoner's Dilemma, the cluster of cooperators gains the higher payoff/fitness. This simple example demonstrates that Assortment can be a basis for cooperation (Bergstrom; 2003; Eshel and Cavalli-Sforza; 1982).

Direct Reciprocity is another mechanism, summed up fairly adequately by 'you scratch my back, I'll scratch yours'. The game is played multiple times between two players. If a member of a population learns that another member is willing to cooperate, they may establish such behaviour (Fogel; 1993).

Sharing the fact that you are willing to cooperate (by initiating cooperation) is costly- the other member might not be willing to cooperate, and so you are exploited. Or the other member may goad you into cooperative behaviour, with the intention of exploiting you. In general, in order for Direct Reciprocity to be useful, the reward for mutual cooperation must be high enough in comparison to mutual defection, being exploited must not be too costly, and for establishing cooperation to be worth the potential price, there must be a sufficient likelihood of you interacting with a player again (Axelrod; 1984).

Indirect Reciprocity is when behaviour is mediated by reputations (Axelrod; 1997). For example, members of a population observe that a particular individual never cooperates, therefore attempting cooperation with such an individual would be pointless. Conversely, another member is seen to always cooperate. If you have access to this information, and there will be no repercussions for your actions, exploiting them would maximise your payoff. Adding reputation, and a risk that players will punish you for exploiting other players provides a potential incentive for cooperation. How likely reputation is to spread determines the success of this mechanism in enabling cooperation (Nowak and Sigmund; 1998).

Finally, Network Reciprocity can be used to model a scenario where chances of meeting other players are determined by a spatial structure. Links in a graph between two players increase probability of meeting again, and clusters of cooperators can form. Networks may essentially be random graphs, or may have some underlying cause for links between players. This may represent spatial factors, or a collection of social links (Ohtsuki et al.; 2006).

In this review I will focus on Direct Reciprocity and Assortment, the two mechanisms relevant for my project. In Section 2, Direct Reciprocity will be discussed with example literature. Section 3 will discuss Assortment, and when the two mechanisms are combined. I will discuss the importance of representations in Section 4, and in Section 5 describe how my project fills existing gaps the literature.

2 Direct Reciprocity

Direct Reciprocity has been widely studied, with both analysis using game theory and simulations. When the game is repeated, the possible strategies are no longer simply Cooperate or Defect. Now, the players choose what to do over a series of games. In the repeated Prisoner's Dilemma, infinite strategies exist.

If the game is repeated a fixed number of times, no cooperation is expected- if all possible strategies are permitted. If both players know the number of rounds to be played, the result of the one-shot Prisoner's Dilemma is expected during the last round: mutual defection. Backward induction leads to a fully defecting game (Aumann; 1995). This can be avoided by having probabilistic length games. Instead of playing a fixed number of times, strategies play each other once, then with probability δ they play again, resulting in a probabilistically determined game length.

Playing the Dilemma results in a payoff/fitness which is used to determine both how a player performed, and to select players to reproduce or survive. All payoffs are positive (to avoid negative fitness), so if payoffs from each individual game were simply summed, there would have been an advantage to playing more rounds. To account for this, the rounds can be summed with a discount weight:

$$\Pi_{AB} = (1-\delta)\sum_{i=0}^{N} \delta^i \pi^i_{AB} \tag{2}$$

Where Π gives fitness after a sequence of games, and π^i_{AB} is the payoff A recieves playing B in the ith game between them.

2.1 Axelrod's Tournaments

Axelrod and Hamilton (1981) investigated the behaviour of strategies using computer tournaments, with repeated Prisoner's Dilemma games played between strategies submitted by people to the tournaments. This study compared fitness (payoff of games) of submitted strategies playing each other. Strategies play each other probabilistically; the game is not a fixed size, instead a continuation probability determines if they continue to play. Out of the 14 submitted strategies, Tit-For-Tat (TFT) performed best. This is a simple strategy that reciprocates the other player's behaviour, after initially cooperating. Initially it cooperates, then it repeats the other player's last move. When playing a strategy that always cooperates (ALLC), it performs identical to ALLC. When playing a strategy that always defects (ALLD), it loses on the first turn, then performs identical to ALLD- given a long enough number of games played, the strategies have similar payoffs.

Axelrod and Hamilton identified three criteria to determine if a strategy can be successful. The strategy must be robust- it must perform fairly well against a wide variety of strategies. TFT does so, including against the a population composed of the simple ALL* examples, but only if there is a sufficient chance of the game repeating. Secondly, the strategy must be stable- a mutant strategy must not be able to invade in the event the strategy becomes established. For example, ALLC is not stable against ALLD. A single ALLD in a population of ALLC will gain better payoffs (fitness) and invade. Lastly, the strategy must be initially viable- it must be able to gain a foothold.

The Tournaments provide insight into the role Direct Reciprocity can play as a mechanism for the success of cooperation. A desirable extension is to allow evolution of the strategies, rather than just allowing the example strategies to compete. From the outcome of the tournaments one might initially suspect Tit-For-Tat might emerge and always dominate. However, Boyd and Lorberbaum (1987) showed this might not be the case; Tit-For-Tat is not (and indeed no strategy is) evolutionary stable in the repeated dilemma game (Boyd and Lorberbaum; 1987; van Veelen and García; 2010).

2.2 Evolving Strategies in Simulations

Axelrod (1987) used a simple Genetic Algorithm approach to study the evolution of strategies for the Prisoner's Dilemma (Back et al.; 1997). A half-memory approach is one in which the opponent's sequence of Cooperate (C) or Defect (D) moves are remembered. A full memory approach remembers its own moves, or equivalently the results (both C is the same as Reward, R, both D is the same as Punish, P, etc). A string of {C,D} represents each strategy. A Full-Memory approach was used in Axelrod (1987).

Based on previous moves, a location in that string is looked up, and states the next move to perform. The first 6 bits determines the initial strategy, then there is not a history of 3 moves. It gives a 'history' to base decisions on, when no actual history is available. 6 bits are required since both the opponents and their own moves need to be encoded. To encode the strategy, 64 bits are used. Three moves are remembered, with 4 possibilities- so 4^3. A total of 70 bits were used by Axelrod's simulation, allowing for moves to be determined based on the last 3 results. That is, it is a full-memory simulation of length 3. This allows for any of 2^{70} different strategies to appear- analysis using a matrix of payoff of every strategy against every other strategy would be a square matrix with 2^{70} elements on each side. This is why explicit analysis selects notable examples from the population, and demonstrates why simulation will be useful.

Strategies were evolved using Genetic Algorithms. Strategies are generated in a population, and then they compete with 8 strategies from the original tournaments to determine their fitness. A new population is determined by reproducing from the population, with fitness determining what strategies reproduce. They then compete with the 8 representatives again. Reproduction involves a small chance of random mutation, and a crossover from each parent. The string representing the strategy is the chromosome, and the new child is produced from parts of each parent's chromosome. A population of 20 strategies competed in games 151 moves long, competing with 8 other members of the population in each round, for 50 generations. Technology of the time limited the size of simulations; today we have the luxury of being able to throw far more computing power at the problem, allowing for much longer (and with larger population) simulations. The initial population was random. The strategies that evolved mostly resembled Tit-For-Tat, and Axelrod identified general patterns for strategies. They continued cooperating after mutual cooperation had been established, they responded to opponent defection with defection, they returned to cooperation after it was restored by the opponent following a defection, and when stuck in a pattern of mutual defection they do not attempt to escape and allow themselves to be exploited.

Strategies that performed better than Tit-For-Tat did emerge. These developed ways to exploit one of the 8 representatives, having a higher fitness against that, at a cost of a lower fitness against some of the other 8 representatives. If the net result is a gain over Tit-For-Tat, the strategy performs better. This only applies to the environment of this simulation. In an environment where the competing strategies is not fixed, for example when the competition is picked from the population, those exploited strategies might be expected to die off and remove the advantage the exploiter gains.

Axelrod also performed simulations in which reproduction only involved the chromosome of one parent; reproduction was asexual, involving no crossover, so new strategies only appeared from random mutations. Again, Tit-For-Tat-like strategies evolved. However, strategies that exploited one of the 8 representatives did not appear as often. Reproduction mechanism affects results of this simulation- in this case, crossover explored the strategies differently to only mutating, and as will be discussed further how strategies mutate from one to another is important.

2.3 Fogel's Automata

Fogel (1993) studied the evolution of strategies using Finite State Automata (FSA). These are abstract computing machines with a number of states, and a number of transitions (Sipser; 2006). Each state can be labelled an accept state, and one state is labelled the start state. Based on some input sequence, transitions are followed from one state to the next. When the machine has read all input, the final state may be an accept or reject state. So, it can compute whether a sequence meets or does not meet a criterion.

The sequence of {C,D} choices (the history) in a game of Prisoner's Dilemma lends itself to this method; view the strategy as a language defined by a FSA. When a history is a word in that language (when its FSA ends in an accept state) the strategy cooperates, when it is not it defects. Fogel used FSA with a maximum of 8 states, to simplify analysis of evolved strategies. The strategies were full-memory, knowing both their opponent's history of moves and their own. A population of size 50 to 1000 was used, 151 games were played for each interaction, and there were at most 200 generations. Number of generations and repetitions of experiments were limited by hardware. As in Axelrod, new strategies are created by mutations of parents, and what strategies reproduce are decided by fitness.

Results were similar to those found by (Axelrod; 1997)- strategies developed that would establish mutual cooperation if possible. Fogel suggests initial sequences of symbols form patterns, which other machines can recognise to establish cooperation- a 'handshake'. Interestingly, size of the population was not observed to affect the evolution of cooperative behaviour, nor the time taken for it to evolve. The best machines produced in the 50 population example did have lower fitness compared to best machines from larger populations, though any effect of population size appeared to diminish after sufficient population size was reached.

When payoff for exploiting the opponent is increased, it is expected that mutual cooperation should decrease. For example, if the payoff for alternating Exploiter (payoff T), Exploited (payoff S) is larger than continuous mutual cooperation, this behaviour should be selected for. Indeed, transitions to long-term mutual cooperation is not seen in the structure of the best performing strategies in these circumstances. Fogel identified the impact of the payoffs in the PD matrix; when mutual cooperation results in the highest overall payoff it can be expected to evolve, when the payoff from alternating defection and cooperation exceeds the payoff for cooperating, mutual cooperation does not evolve. When either alternating or mutually cooperating has equal reward, initial population state determines the outcome.

This approach limits the strategies that can evolve to those representable by Automata with 8 nodes. It also uses fixed length games; with an expanded strategy space, cooperation would not be expected with this simulation.

2.4 Miller's Automata

Miller (1996) used FSA, of a size of 16 states, to perform Prisoner's Dilemma simulations. Each FSA is represented by a string of bits. To define each state, 9 bits are used. One bit defines what to do if the history input to the machine ends at the state- or, equivalently defines if it is an accept state. Four bits define what state to transition to in response to a cooperation, and the final 4 define what to do in response to a defection (16 states, so

4 bits are required to identify the target state). Four bits at the start of the string define the initial state, followed by 16 sets of 9 bits, one set for each state. Since this is a total length of 148 bits, 2^{148} strategies are possible (ignoring the fact many will be actually equivalent, just different expressions of the same strategy).

Evolution follows a similar process to previously described research. Based on performance in the game (playing every player including itself in a sequence of 150 games), some strategies in the population are selected for reproduction. From a population of 30, 20 are selected from to reproduce. These 20 also survive to the next generation. Reproduction involves a crossover and a mutation process. In crossover, a sequence from both parents is taken and combined. In mutation, a bit is flipped. The 10 individuals that result from crossover are added to the population, replacing the 10 least fit. Initial populations were random.

Simulations were run with two levels of noise included. For both noise levels, the number of states reduced from the initial value. Number of states is calculated from the minimal representation of the FSA- not the actual equivalent FSA that is in the simulation. More noise resulted in fewer states. One interpretation of this result is that establishing a pattern of behaviour between individuals requires a simple ‘message’ to communicate if the environment is noisy. In the simulation, defection was reciprocated at a high rate, and cooperation was reciprocated at a lower rate- but still reciprocated. Higher noise also negatively impacted rates of cooperation.

Miller’s work provides a clear technique for defining a FSA, a simple mutation method, and various features of FSA that may be worth studying- like the number of states, and number of terminal states (both transitions at this state are to itself). However, in limiting the number of states, the strategy space is limited. Games are also fixed length; if the number of states were unbounded, this simulation would eventually collapse to a population of full defectors (Aumann; 1995). Probabilistic length games would avoid this issue.

A summary of the works regarding Direct Reciprocity examined in this review is shown in Table 2. The representations column indicates the means by which individual players are modelled. A strategy describes how the individual player will play the game, in response to a history of previous movements (a sequence of {C,D}, Cooperate and Defect). The representation chosen determines the types of strategies that can evolve, the number of strategies, and the set of strategies they can represent. The importance of the representation will be developed later in this review. In the repeated Prisoner’s Dilemma, there are an infinite number of strategies. An individual representation will allow a subset of these strategies- they will have a strategy space. The strategy space is listed, along with how the games are repeated (fixed length, probabilistic length), and number of generations simulated. In the notes column, I indicate that Axelrod and Hamilton (1981) is a study of how strategies perform at the repeated Prisoner’s Dilemma- with no exploration of strategies aside from those added initially. I indicate van Veelen et al. (2012) used both repetition of the game, and Assortment of the population.

3 Structured Populations

Bergstrom (2003) investigated the role Assortment can play in circumstances in which Direct Reciprocity does not enable cooperation- when games are one shot. In previously discussed research, the probability of encountering an individual in a population was uniform (the population is well-mixed). Bergstrom (2003) instead considers a scenario where the probability of encountering a cooperator, given the individual is a cooperator, is p. The probability of encountering a defector, given the individual is a defector, is q. The population consists of just cooperators and defectors. The likelihood of meeting an alike

Paper	**Represent.**	**Strat. Space**	**Repetition**	**Gen**	**Notes**
Axelrod and Hamilton (1981)	Fixed, 1/2-Mem	Submitted	Prob.	Not Evolutionary	
Axelrod (1987)	Lookup, Full-Mem	Length 3	Fixed 151	50	
Fogel (1993)	FSA, Full-Mem	Regular, 8	Fixed 151	200	
Miller (1996)	FSA, 1/2-Mem	Regular, 16	Fixed 150	50	
van Veelen et al. (2012)	FSA, 1/2-Mem	Regular, ∞	Prob.	5*10^5	Structure
García and Traulsen (2012)	Lookup, 1/2-Mem	Length 1	Prob.	$2 * 10^8$	
Imhof et al. (2005)	Fixed, 1/2-Mem	3 Strategies	Analytical		

Table 2: Selection of Direct Reciprocity Papers

player is therefore increased (for p, q > 0). The assortivity index is a function of q and p, where x denotes the proportion of the population that are cooperators:

$$a(x) = p(x) - q(x)$$

They also found the difference (D) between the payoff of the cooperator and the payoff of the defector. In this equation, b is the benefit recieved from someone cooperating with you, and c is the cost you pay to cooperate (so R=b-c, S=-c, T=b, P=0 in the PD game matrix).

$$D(x) = a(x)b - c \tag{3}$$

They find when the reward for cooperation times the assortivity index is greater than the cost for helping, cooperators will do better than defectors. When it is less than, defectors will do better. This is Hamilton's rule- which indicates whether an individual shares enough genes for helping them to be an increase in fitness for the potential helper- but with an Assortment parameter replacing relatedness.

Consider the payoffs in Table 3. Take a simple Assortment scenario, with a single-shot game. With probability r, a player plays with an 'alike' player, without sampling from the population. With probability $1 - r$, the player plays with a player sampled from the population. The population consists of two strategies: ALLC, and ALLD.

	Cooperate	**Defect**
Cooperate	R = 4, **R = 4**	S = 0, **T = 5**
Defect	T = 5, **S = 0**	P = 1, **P = 1**

Table 3: Payoff Matrix for the Prisoner's Dilemma, with example values

When sampling from the population (or equivalently, when there is no structure), the payoffs Π are:

$$\Pi_{ALLC} = \frac{N_{ALLC} - 1}{N_{TOTAL} - 1} * (R = 4) + (S = 0)$$

$$\Pi_{ALLD} = \frac{N_{ALLC}}{N_{TOTAL} - 1} * (T = 5) + \frac{N_{ALLD} - 1}{N_{TOTAL} - 1} * (P = 1)$$

Since the Temptation payoff is always greater than the Reward payoff and the Punishment payoff is always greater than the Sucker's payoff, in an unstructured population ALLD has a higher expected payoff.

When structure is added, the expected payoff is:

$$\Pi^{(r)}_{ALLC} = (1 - r)\Pi_{ALLC} + r * (R = 4)$$
$$\Pi^{(r)}_{ALLD} = (1 - r)\Pi_{ALLD} + r * (P = 1)$$

Substituting the example game, with 50 ALLC players and 50 ALLD players:

$$\Pi^{(r)}_{ALLC} = (1-r)\frac{4*49}{99} + r*(4) = \frac{196}{99} + \frac{200}{99}r$$
$$\Pi^{(r)}_{ALLD} = (1-r)\frac{299}{99} + r*(1) = \frac{299}{99} - \frac{200}{99}r$$

So if $r > 103/400$, ALLC has a higher fitness. That is, in a sufficiently structured population, cooperation can become favorable even in the one shot version.

Assortment is known to enable the success of cooperative behaviour (Bergstrom; 2003). Next, I will discuss when games are both repeated, and there is Assortment of the population.

The behaviour that evolves at with various levels of probability of a game continuing and population structure was discussed in Axelrod and Hamilton (1981). It was investigated in detail with a simple model with both simulation and analysis by van Veelen et al. (2012). The method I will use in my project is based on this paper. Both general analytic results were found, and simulation results that may be representation-specific (FSA with half-memory were used in the simulations in this paper).

The Prisoner's Dilemma game played was defined as follows:

	Cooperate	**Defect**
Cooperate	R = b-c, **b-c**	-c, **b**
Defect	b, **-c**	0, **0**

Table 4: Payoff Matrix for van Veelen et. al.

The analysis focused on identifying behaviour as both continuation probability (so, number of times the game is played) and Assortment are varied. Assortment is the likelihood of meeting an alike player: there is r chance that the other player is the same as the current player, and $1-r$ chance that they other player is selected at random from the population (Eshel and Cavalli-Sforza; 1982). Several regions of predictable behaviour exist in a graph of continuation probability vs assortment. This graph is reproduced in Figure 1. The graph is for b/c=2, different ratios will change the curves separating regions- but the results will be qualitatively the same.

In Region I, ALLD is an equilibrium- determined by comparing the payoff of ALLD to a mutant that cooperates at least once. It is not the only equilibrium in the region, but equilibria strategies will play Defect against themselves and ALLD. In this region, cooperation is not expected; no cooperating strategy is an equilibrium.

In Region II, a variety of strategies are equilibria; for example, ALLD and TFT. So, both extremes of behaviour may be seen in a population under these circumstances.

In Region III, there are again a variety of strategies that are equilibria, but ALLD and other complete defectors are no longer equilibria. Cooperative behaviour of varying types and degrees is expected to be common most of the time (equilibria can be escaped, for a short time).

In Region IV, ALLC is an equilibrium. All other equilibria always cooperate when playing against ALLC or themselves. Fully cooperative behaviour is expected to dominate. More regions are defined, but these are the ones of primary concern.

These general regions are expected to hold regardless of representation (I and II are particularly well defined, and not expected to vary). The simulations van Veelen et. al. conducted were of FSA with an unbounded number of states and half-memory (of unbounded length). The simulation is initialized with simple individuals: ALLD. Every member plays one other member of the population for a number of rounds determined stochastically, based on the continuation probability. With probability r a member will

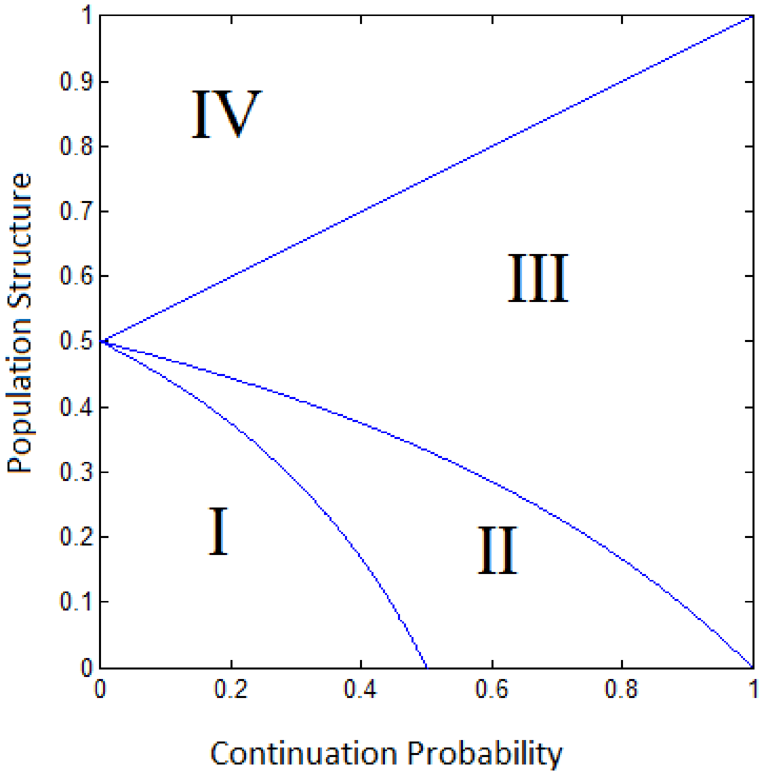

Figure 1: Regions of behaviour for varied continuation probability and structure

play against an identical strategy. With probability $1 - r$ a member will play against another player from the population.

This payoff was used to calculate the probability for the individual to reproduce into the next generation. Stochastically, members of the population are selected for the new generation, so unfit individuals tend to be removed. During reproduction, mutation could also occur, resulting in new strategies.

Results of the simulation were consistent with the analysis. A colourmap indicating the amount of cooperation in the simulation for varied continuation probabilities and assortment parameters was produced, with behaviour matching that predicted in each region. In general, both increased assortment and increased continuation probability increases cooperation (but there are exceptions).

Indirect invasions are observed in the simulation (García and Traulsen; 2012). Indirect invasions are when a new strategy enters the population not by performing well against the currently dominant strategy, but by 'springboarding' off another strategy. For example, consider that TFT is resistant to ALLD, and will only get exploited on the first move. ALLC is a neutral mutant of TFT (it plays against TFT the same way TFT plays against itself), and in a finite population of mostly TFT, ALLC can become more common by neutral drift. ALLD can exploit ALLC, and so when this neutral drift occurs, ALLD can use it to invade the population. This was summarised as: "unconditional cooperation is therefore cooperation's worst enemy" (van Veelen et al.; 2012). Indirect invasions were observed establishing cooperative behaviour in a defecting population too.

The results of the analysis should be valid regardless of representation. The simulation built upon previous work, investigating the interplay between reciprocity and population structure, and using unbounded rather than bounded FSAs allowing an infinite strategy space. However, the set of all possible strategies for the Prisoner's Dilemma is larger than this space.

Other representations have their own strategy space. The choice of representation will exclude some, and enable some that were not possible in other representations. Representation also affects the chance of a strategy appearing, or the route by which it appears, since the mutation process may differ.

The model described in van Veelen et al. (2012) was chosen as the basis of my project. Both varied assortment and varied chance of a game continuing are modelled. Broad behaviour (such as whether strategies that always cooperate can dominate, or can be somewhat successful) can be predicted for ranges of assortment and continuation probability. This provides bounds for what behaviour is expected. Detailed behaviour can be discovered through simulation. The impact of variations in the simulation- such as changing how strategies are represented- can be compared quantitatively. By using both assortment and continuation probability, conditions not explored in previous literature (Region III for example in Figure 1) can be investigated.

4 Expanding the Strategy Space

Different representations affect the evolution of cooperation - as noted in section 2.2. There are a number of ways to change the representation and allow for a larger strategy space. In a lookup table, the number of previous moves a player remembers when making a decision changes the range of strategies that can be represented. In a FSA, the maximum number of states limits the strategies that can be represented; however even an unbounded FSA is limited. A FSA can determine if a sequence is in a regular language defined by the FSA. So, its strategy space is limited to regular strategies.

Changing the model also allows for different strategies. I have focused on simulations with half-memory strategies. That is, each agent has access to the history of its opponent's moves. The simulation could be changed, and allow each agent access to its own and its opponent's history- it could have full-memory. An example of a strategy that can evolve in full-memory simulations but not half-memory simulations is Win-Stay, Lose-Shift. If it successfully cooperates with or exploits the opponent, it continues with the successful strategy. If it loses, for example gets exploited, it tries the other move. To accomplish this it of course needs the result; knowing the result {R,T,S,P} is equivalent to knowing the history of both players.

4.1 Representations

Small strategy spaces can be useful; in Axelrod's original Tournaments (Axelrod and Hamilton; 1981), a set of submitted strategies competed, with interesting results. A space consisting of 3 strategies- ALLD, ALLC and TFT- was used by Imhof et al. (2005) to examine evolutionary cycles when noise is present. Useful as small strategy spaces are, it is not certain that the strategies that perform well in a small space will be successful when the strategy space is expanded.

The limitation of FSA representations to regular languages (and further limitation if the number of states is bounded) has been mentioned. Does an evolutionarily successful strategy that cannot be represented by an FSA exist (ie. a non-regular strategy)? I'm not sure, and this is a major aspect I intend to investigate, but a strategy that is very successful at playing the Prisoner's Dilemma in stochastic games (where noise can cause some moves to be performed opposite to what a strategy intended) certainly exists. Press and Dyson (2012) described a class of Zero Determinant (ZD) strategies. These strategies move so as to set the opponent's score, or to set the ratio between the opponent's score and the ZD player's score. The best a strategy can achieve against these is the score or ratio the ZD is attempting to set. This is an example of a non-regular strategy that is successful at playing the game against other strategies (ZD are non-regular in general, but regular examples exist; TFT is a ZD strategy). This does not mean it would become the most common strategy in an evolutionary environment for a long time. Adami and Hintze (2013) showed that coercive ZD strategies will not be stable- it may invade a population, but the advantage of coercion is short-lived.

One extreme example of a representation with a larger strategy space is Turing Machines (Sipser; 2006). This would include all strategies described by FSA, but additionally all computable strategies. Any such simulation would be very complex, and encounter problems such as what to do about machines that never halt. A suitable 'middle-ground' may provide some insight to how the larger strategy space alters the evolution of strategies. One solution would be to move a single step up the Chomsky hierarchy; Push-Down Automata are more powerful than Finite Automata Machines, representing all regular languages plus all context-free languages.

The size of the strategy space is not the only factor to consider; how that space is explored (via mutations) matters (García and Traulsen; 2012).

4.2 The Impact of Mutations

The impact of mutations on how the strategy space might be explored can be demonstrated with a simple lookup table example and a simple FSA example (Fogel; 1993; García and Traulsen; 2012). Consider a lookup table of memory 1. It is 3 bits in length- it needs to know what move to perform initially when there is no history (1 bit), and what do do in response to a cooperate or a defect (2 bits). ALLD is represented by (111). Suspicious Tit-For-Tat (STFT), which defects initially then reciprocates, is represented by (101). They are one mutation apart- flip the middle bit. The simplest FSA representation is a single node, in a non-accept state, with both C and D transitions pointing to itself. The simplest STFT representation is two nodes, an extra transition, and two move transitions (Figure 2). Start on a non-accept node, stay on it if the opponent defects, move to an accept node if the opponent cooperates, and so forth. The actual distance between ALLD and STFT depends on how the FSA mutations are done- but the mutation distance is longer.

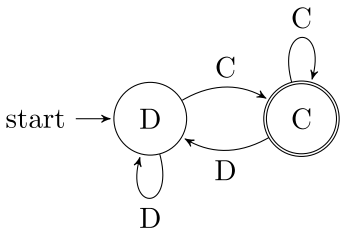

Figure 2: Suspicious Tit-For-Tat Represented by a Finite State Automaton

García and Traulsen (2012) explored the impact of differing mutations distances or probabilities (probability of mutating to a particular strategy should be dependent on how many mutations are needed). A lookup table to length 3 was used, and games were repeated, without structure. In analysing a small set of strategies (such as the 8 in this case), one approach is to assume uniform mutation probabilities. For n strategies, when a mutation occurs it mutates to any individual strategy at the same rate (1/n). So if a uniform mutation model is used, ALLD to STFT has the same probability as ALLD (111) to ALLC (000), which requires 3 bit-flip mutations in a lookup table representation, and ALLD to TFT, which requires 2 bit-flip mutations.

Both simulation and theoretical analysis comparing uniform and bitwise mutations showed that the frequency of cooperative strategies was significantly lower- assumptions made about mutations, or the mutation methods used affect the results of evolutionary simulations.

It is important to consider these findings for a number of reasons. It highlights the need to consider assumptions made when mutating strategies. Uniform probability is not a good assumption depending on the model context, how to mutate a FSA or PDA may also be important. For example, a single mutation may be to add a node that connects to itself. This might have different results to a single mutation being to add a node, that connects to random other nodes in the Automaton. If mutations are viewed not from a node and transition perspective, but with a representation as a string of bits like in Miller (1996), mutations could be performed just by flipping bits- a variety of mutation methods are possible, each exploring strategies differently.

Aside from the possibilities for mutations and how they are evaluated (eg. does it resemble something observed in nature? Is it simple?), if there is a difference in results using different representations in simulations, the role the mutation process plays in that difference must be considered.

5 Conclusion

The various mechanisms for cooperation to evolve that were discussed (Direct Reciprocity and Assortment) have been fairly well explored individually (Imhof et al.; 2005; Nowak; 2006). When there is sufficient Assortment, the cooperators have the advantage (Bergstrom; 2003). When the game is repeated enough times (with probabilistic length), cooperators have the advantage (Axelrod; 1984).

Developments in research on the evolution of cooperation when these factors are combined has not been explored to the same depth. One aspect that has not been examined is the impact representations with larger strategy spaces will have on the evolution of cooperation.

What are the possible outcomes of expanding the strategy space in a repeated, structured simulation? I would expect that the analysis of van Veelen et al. (2012) would hold; the reasoning was not representation-specific. Strategies in those regions will follow the predicted behaviour, but the strategies that appear may vary when a different representation is used. Recent research, such as García and Traulsen (2012) indicates the importance of representations and mutations. In the case of switching to a Push Down Automata representation, one possibility is that the results are identical. PDA have states and transitions like FSA, but add a stack that each transition can access, reading and removing the item at the top of the stack and pushing to the top of the stack. A PDA can simulate all FSA, plus additional strategies (Context-Free)(Sipser; 2006). They are capable of things a FSA is not; for example rudimentary counting with history of any length (eg. push X times to the stack if a Defect is read, pull Y times if Cooperate is read and the stack is not empty. If the stack is empty when all history has been read, Cooperate). If the best strategies inside context-free strategy space are also regular (if those capabilities are not advantageous), it may have the same overall outcomes.

Another possible outcome is that strategies tend to be more or less cooperative than in the original simulation; that the extra allowed strategies find ways to exploit strategies more, or to establish cooperation more. The behaviour within the regions described by van Veelen et al. (2012) may display different patterns; perhaps in one region it tends towards cooperation more than the FSA simulation did, and in another it tends towards defection more.

We know enough to make some predictions about the behaviour of agents playing the Repeated Prisoner's Dilemma if the representations are changed, so as to allow a larger strategy space. We know that no strategy is the best strategy (all strategies can be invaded). How the dynamics change when larger strategy spaces are allowed is not fully understood, and is the focus of my research.

6 References

Adami, C. and Hintze, A. (2013). Evolutionary instability of zero-determinant strategies demonstrates that winning is not everything, *Nature communications* **4**(2193).

Aumann, R. J. (1995). Backward induction and common knowledge of rationality, *Games and Economic Behavior* **8**(1): 6–19.

Axelrod, R. (1984). *The Evolution of Cooperation*, Basic Books, New York.

Axelrod, R. (1987). The evolution of strategies in the iterated prisoner's dilemma, *Genetic Algorithms and Simulated Annealing* pp. 32–41.

Axelrod, R. (1997). *The Complexity of Cooperation*, Princeton University Press, Princeton.

Axelrod, R. and Hamilton, W. D. (1981). The evolution of cooperation, *Science* **211**: 1390–1396.

Back, T., Fogel, D. B. and Michalewicz, Z. (1997). *Handbook of evolutionary computation*, IOP Publishing Ltd.

Bergstrom, T. C. (2003). The algebra of assortative encounters and the evolution of cooperation, *International Game Theory Review* **5**(03): 211–228.

Boyd, R. and Lorberbaum, J. P. (1987). No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game, *Nature* **327**: 58–59.

Dawkins, R. (1989). *The selfish gene*, Oxford University Press, New York.

Eshel, I. and Cavalli-Sforza, L. L. (1982). Assortment of encounters and evolution of cooperativeness, *Proceedings of the National Academy of Sciences USA* **79**: 1331 – 1335.

Fogel, D. B. (1993). Evolving behaviors in the iterated prisoner's dilemma, *Evolutionary Computation* **1**(1): 77–97.

García, J. and Traulsen, A. (2012). The structure of mutations and the evolution of cooperation, *PLoS One* **7**: e35287.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. i, *Journal of theoretical biology* **7**(1): 1–16.

Imhof, L. A., Fudenberg, D. and Nowak, M. A. (2005). Evolutionary cycles of cooperation and defection., *Proceedings of the National Academy of Sciences USA* **102**: 10797–10800.

Miller, J. H. (1996). The coevolution of automata in the repeated prisoner's dilemma, *Journal of Economic Behavior & Organization* **29**(1): 87–112.

Nowak, M. A. (2006). Five rules for the evolution of cooperation, *Science* **314**(5805): 1560–1563.

Nowak, M. A. and Sigmund, K. (1998). The dynamics of indirect reciprocity, *Journal of theoretical Biology* **194**(4): 561–574.

Ohtsuki, H., Hauert, C., Lieberman, E. and Nowak, M. A. (2006). A simple rule for the evolution of cooperation on graphs and social networks, *Nature* **441**(7092): 502–505.

Pennisi, E. (2005). How did cooperative behavior evolve?, *Science* **309**(5731): 93–93.

Press, W. H. and Dyson, F. J. (2012). Iterated prisoners dilemma contains strategies that dominate any evolutionary opponent, *Proceedings of the National Academy of Sciences* **109**(26): 10409–10413.

Sipser, M. (2006). *Introduction to the Theory of Computation, Second Edition*, Course Technology, Boston.

Traulsen, A. (2010). Mathematics of kin-and group-selection: Formally equivalent?, *Evolution* **64**(2): 316–323.

van Veelen, M. and García, J. (2010). In and out of equilibrium: Evolution of strategies in repeated games with discounting, *Technical report*, Tinbergen Institute Discussion Paper.

van Veelen, M., García, J., Rand, D. G. and Nowak, M. A. (2012). Direct reciprocity in structured populations, *Proceedings of the National Academy of Sciences USA* **109**: 9929–9934.