Performance of Strategies for the Iterated Prisoner's Dilemma in a Natural Environment

Jasper van Tilburg

BSc Thesis, Delft University of Technology

August 22, 2018

Abstract

In nature, all species have their own behaviors and strategies for survival. Some species survive and reproduce, while others become extinct. This paper proposes a model to simulate these strategies and test their performance. Natural behavior is represented as strategies for the iterated Prisoner's Dilemma (IPD). Agents wielding one of ten common IPD strategies are deployed in a natural spatial environment with biologically realistic conditions, where they continuously play Prisoner's Dilemma games. If the payoffs are well enough, agents are able to reproduce. The harshness of the environment is determined by three factors. The cost of living directly controls the climate and age limitation and energy limitation affect an agent's ability to reproduce. Another influencing factor is evolution, which gives agents the option to adopt different strategies in later stages. Harsh environments are defined by high costs of living, high reproduction costs and low life expectancy. Results show that cooperative strategies are more likely to survive and reproduce in harsh environments. Moreover, evolution is in the advantage of cooperative strategies, because many unsuccessful defectors evolve into cooperators.

1 Introduction

The Prisoner's Dilemma (PD) is a founding concept in game theory used in numerous studies. It is a cooperative game where players have the choice to either cooperate or defect and receive payoffs depending on the outcome. It is a non-zero-sum game, meaning that one player's gain does not necessarily balance the other player's loss. For mutual cooperation both players receive reward R. For mutual defection both players receive a punishing payoff P. If a player defects while the opponent cooperates the player receives a temptation payoff T and the opponent receives a sucker payoff S. A classical PD game is defined when $T > R > P > S$ and $2R > T + S$.

The Iterated Prisoner's Dilemma (IPD) is an extended version of PD, where the game is played repeatedly and the player is able to remember previous choices of its opponent. IPD provides an abstract framework for investigating the evolution of cooperation. Multi-agent simulations using IPD are very useful in studying cooperation levels and social cohesion in nature and society [1].

In a single Prisoner's Dilemma game defection is the preferred strategy. Regardless of the opponent's move, defecting receives the best payoff. In the iterated Prisoner's Dilemma however, opponents repeatedly play against each other and cooperating will pay off better on the long term, as proved by Axelrod [2]. He was the first to organize a tournament of a N-step iterated Prisoner's Dilemma with a fixed N, where participants entered their strategies. Overall, successful strategies were the ones that allowed for cooperation. The winning strategy, sent by professor Anatol Rappoport, is called tit for tat (TFT) which is based on attempt of cooperation and later on reciprocity - simply copying the opponent's last move. Axelrod concluded that in order to be successful, strategies should be nice, forgiving and simple. Nice means cooperation on the first move and forgiving means reciprocity even after being defected.

2 Related Work

Spatial evolutionary games are played in two-dimensional or three-dimensional spatial structures and usually restrict interactions between players to local neighborhoods. Numerous researchers have designed models to study behavior of IPD strategies in spatial environments. Nowak and May were pioneers in this field of research. They showed indefinite persistence between cooperators and defectors in spatial environments. In their model, each patch-owner plays PD games with its immediate neighbours. At the start of the next generation players switch strategy if this proves to be more successful [3].

Schweitzer and Mach performed similar experiments, but instead of using the strategies of unconditional cooperation and defection they used a 3-bit binary string representation of strategies, introduced by Nowak and Sigmund [4]. The first bit represents the initial choice of the agent. The second and third bit represent its choice when the opponent cooperates and defects, respectively. They found that all players eventually adopt TFT, provided that the games were played for enough iterations against each opponent [5].

Lindgren and Nordahl used a variation of binary string representation of strategies. Longer binary strings enable agents to choose their next move based on a bigger memory of previous moves. Different types of mutation allowed the agents to evolve their strategies over time into more complex ones. Successful strategies were binary strings evolved from the simple binary strings 11 and 01 representing unconditional cooperation and TFT, respectively [6].

Wilensky built a model to explore the implications of some complex strategies as well. He deployed agents in a spatial environment with periodic bounds. The agents moved randomly through space and repeatedly played PD games against the same opponents [7].

Table 1: Overview of related models

	Strategies	Evolution	Mutation	Reproduction	Aging
Nowak & May [3]	ALLC, ALLD	x			
Schweitzer & Mach [4]	3-bit binary strings	x			
Lindgren and Nordahl [6]	n-bit binary strings	x	x		
Netlogo [7]	ALLC, ALLD, RAND, TFT, GRIM				
Smaldino [8]	ALLC, ALLD			x	
Proposed model	ALLC, ALLD, RAND, gTFT, sTFT, TFTT, TTFT, GRIM, Pavlov	x		x	x

Finally, Smaldino studied cooperation in harsh environments and the emergence of spatial patterns as they occur in all types of natural systems, living and non-living. He discussed the strategies of unconditional cooperation and defection and the coexistence between the two strategies in a natural model. In specific, he experimented with factors that determined the environment's harshness. The cost of living and sucker payoff S played an important role in the harshness. He found that the larger the cost of living and the value for S were, the more successful the cooperators. Accordingly, low costs of living and sucker payoffs served defectors [8].

Table 1 shows an overview of models related to this model and the different features they posses. The features and strategies are discussed into detail in section 3 and section 4.

3 Contribution

This paper reports an extension of Smaldino's study and model. Whereas Smaldino discussed only the two unconditional strategies, this paper talks about several complex strategies for IPD. It is a research regarding the question: How do strategies for IPD perform in a biologically realistic environment? Multi-agent simulations were used to experiment with several factors influencing the strategies in the environment. These experiments try to show how complex IPD strategies would perform if they were deployed in nature. The simulations were run on an extension of Smaldino's model. In this model the agents can be compared to individuals in nature. These individuals wield survival strategies which can be cooperative (flocks) or defective (predators). The model features several variables that influence realistic factors found in nature. Harshness of the climate is determined in the model by a cost of living, age limitation and energy limitation. The cost of living k directly controls the climate. The higher k, the harder for agents to survive. Age limit h affects an agent's ability to reproduce. The sooner it will die, the sooner it has to reproduce to be successful. Thus, low age limits make the environment harsher. Finally the reproduction threshold and reproduction cost are in close relation with the energy limit. Agents are able to reproduce if their energy level is two-third of the maximum with a cost of one-third of the maximum. The higher the energy limit, the longer it takes to become strong enough to reproduce. Another realistic feature of the model is evolution. In nature individuals accommodate to harsh environments by small changes in their physique, habits and behavior. Evolution is reflected in the model by the option to change strategies in later stages. The model is described in detail in section 5.

To experiment with the performance of the strategies some hypotheses were drafted. The following hypotheses are based on conclusions of Axelrod and Smaldino:

1. Successful strategies are nice and forgiving.
2. Harsh environments are in the advantage of cooperators.
3. Evolution is in the advantage of cooperators.

Note that Axelrod concluded that strategies should be simple as well. This is left out in hypothesis 1 as this paper only discusses rather simple strategies and does not go into detail about complex ones.

4 Strategies

In 2004 and 2005 similar competitions to the one organized by Axelrod were held. Jurišić et al. did a review of strategies submitted in all three tournaments and strategies that have emerged in between. They state that most new successful strategies are based on the principle of TFT [9].

In their review Jurišić et al. classify nine strategies as default types shown in Table 2. Furthermore, they mention the winning strategy of the tournament in 2005, called Adaptive Pavlov (APavlov). APavlov tries to recognize the opponent's strategy, categorize it in one of the nine strategies (RAND if unknown) and respond in an optimal way [10]. Note that Pavlov can be either played with a cooperation in the first move (Cooperative Pavlov or PavlovC) or a defection (Defective Pavlov or PavlovD).

All IPD strategies can be further categorized in subsets. First, there is the set of fixed strategies containing: always cooperating (ALLC), always defecting (ALLD) and random (RAND). The actions of the other strategies are more complicated as they depend on their opponent's behavior. Second, there is a separation between nice strategies and non-nice strategies. Nice means that they always start with a cooperating move. For example, TFT and grim trigger (GRIM) are nice strategies and suspicious TFT (STFT) and PavlovD are not. Finally, strategies have some level of forgiveness. For example, tit for two tats (TFTT) is more forgiving than TFT, which is in its turn more forgiving than two tits for tat (TTFT). GRIM is the sternest strategy.

The model described in section 5 will contain the default strategies from Table 2, including both PavlovC and PavlovD.

Table 2: Default Types of Strategies

Designation	Description
ALLC	Strategy always plays cooperation
ALLD	Strategy always plays defection
RAND	Strategy has a 50% probability to play cooperation or defection
GRIM	It starts with cooperation, but after the first defection of its opponent continues with defection
TFT	It starts with cooperation and then it copies the moves of the opponent
STFT	As TFT but starts with defection
TFTT	As TFT but defects after two consecutive defections
TTFT	As TFT but for each defection retaliates with two defections
Pavlov	Action results are divided into 2 groups, positive actions are T and R and negative actions are P and S - if the result of previous action belonged to the first group, action is repeated and if the result was in the second group, then the action was changed, it is also called win-stay, lose shift

5 Model Description

As explained in section 3, the model used in this paper is based on the work of Smaldino [8]. In his model however, he only used unconditional cooperation and defection, i.e. only the strategies ALLC and ALLD. The model proposed in this paper features multiple complex strategies. Netlogo was used to build this model [11].

Agents are deployed on unique random cells in a 100 x 100 square lattice with periodic boundary conditions. They are initialized with an energy level h drawn from a uniform distribution between 1 and 50. Once the simulation starts, agents try to find an opponent on one of its 8 neighboring cells (Moore neighborhood). If successful the agents play a single game of the Prisoner's Dilemma against each other and the resulting payoffs add up to their current energy level. If no opponent is found the agent randomly moves up a single cell in its Moore neighborhood. Payoffs are determined in Table 3.

Table 3: Payoff Matrix

	Opponent's move	
Player's move	Cooperate	Defect
Cooperate	Player: 3 Opponent: 3	Player: 0 Opponent: 5
Defect	Player: 5 Opponent: 0	Player: 1 Opponent: 1

Harshness of the environment is directly controlled by the cost of living k. Each game cycle this value is deducted from the energy level of every agent. If its energy level falls below zero, the agent dies and is removed from the environment. On the other hand the agents have an energy limit. It is important that $k > 1$, otherwise defectors can survive on their own, because $P = 1$. Furthermore, agents have a reproduction threshold of two-third of the energy limit. If an agent obtains an energy level above this threshold the agent will try to reproduce and hatch an agent with the same strategy in its Moore neighborhood. Reproduction will only be successful if there is at least one unoccupied neighboring cell. The child agent will be provided with an energy level of one-third of the energy limit, deducted from its parent agent.

The model allows for evolution of strategies. Agents are able to imitate strategies from neighbors that outperform their own. The model provides a variable threshold for evolution q. If the sum of an agents energy level and q is lower than the energy level of one of the agents with a different strategy in its Moore neighborhood, it will adopt the strategy of that agent.

6 Results

Experiments were conducted with an initial population of one-tenth of the lattice capacity, i.e. an equal distribution of 100 agents per strategy. The variables to be tested were cost of living k, age limit m, energy limit h and evolution threshold q. Default values were chosen as follows: $k = 1.1$, $m = 0$ (no age limit), $h = 150$ and $q = 0$ (no evolution). As explained in section 5, k should be larger than 1 to avoid survival by mutual defection. Still k is chosen not much higher than 1 so the environment will not be too harsh by default. Values for age limit, energy limit and evolution are determined by Smaldino's model, i.e. no aging, no evolution and an energy cap of 150.

Results were determined by the average of 20 runs per simulation. The deviations in results were significantly small at this number of runs. Every run kept going until the populations of strategies stabilized, i.e. when deviations in the relative number of agents

per strategy stayed within a margin of 0.1% of the total population for 50 consecutive generations. Simulations for experimenting with an age limit were set on a fixed number of generations, because agents are constantly dying and replaced, which never leads to an exact stable state. These simulations ran for 2000 generations. After this rather large number of generations populations still changed, but stopped growing or shrinking, making it stable enough for good results. The measurement for the success of strategies is the number of agents counted in the environment in the last generation.

First simulations were run using the default values for k, m, h and q. Figure 1 shows the average agent count per strategy over time. After approximately 350 generations ($t = 350$) the complete environment was occupied with agents. The populations stabilized when the agent count hit 10 000. After a few generations differences in growing rates arise and no populations pass each other. At $t = 350$ a clear separation can be made between the nice strategies and the non-nice. The lower four strategies start with a defection (RAND with a 50% chance, thus can not be called nice) and the upper six start with cooperation. Furthermore, there seems to be an advantage for rather unforgiving strategies. GRIM ends up as largest population followed by TTFT, TFT and TFTT. This is the order from sternest to most forgiving as described in section 4.

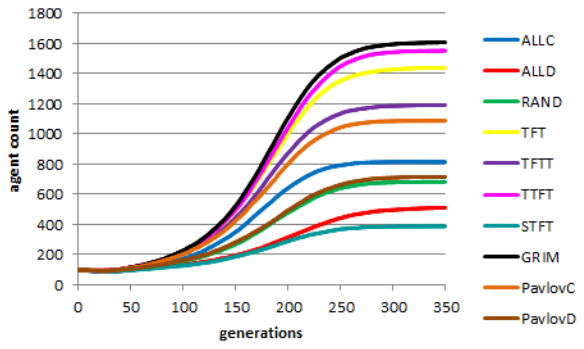

Figure 1: Agent count per strategy over time without evolution

6.1 Evolution

By evolution the imitation of neighboring strategies is meant. When an agent has sufficiently less energy than its strongest neighbor, it will adopt the strategy of this neighbor. The experiment on the effect of evolution was conducted on two values for the evolution threshold q: $q = 0$ (no evolution) and $q = 20$. When $q < 20$ agents kept evolving too often and never reached a stable composition and when $q > 20$ differences in populations become negligible.

Simulations with evolution took about 50 generations more to stabilize than without evolution. Evolution kept doing its work even after all 10 000 cells were occupied. Results shown in figure 2 are similar to the previous experiment without evolution. The order of strategies in the agent count is nearly the same. In this graph however, the strategies are widerspread. The difference in agent count between the best strategy (GRIM) and worst strategy (STFT) has nearly doubled compared to runs without evolution. At approximately $t = 200$ some populations stagnate and start shrinking. At this point many agents with losing strategies realize their strategy is dying out and they switch to a neighboring winning strategy. Most of these shrinking strategies are not nice. Interestingly PavlovC is shrinking as well, even though it starts with cooperation. The decrease in population is sufficient enough to drop below the agent count of ALLC.

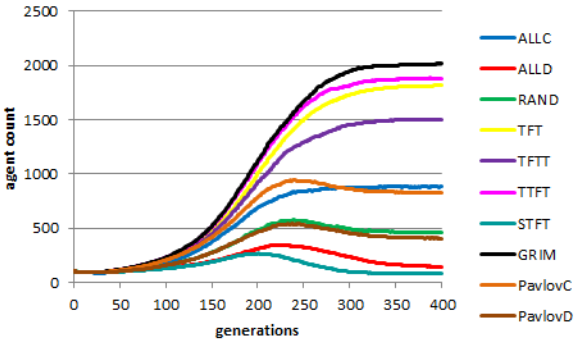

Figure 2: Agent count per strategy over time with evolution

6.2 Cost of Living

The harshness of the environment is directly influenced by the cost of living. Each generation, cost of living k is deducted from each agent's energy level. The higher k, the harsher the environment. The experiment was conducted on 11 values for k, from 1.0 to 2.0 with steps of 0.1. Below 1, the cost of living will not affect any agent. Above 2, it will become hard for any agent to survive.

Figure 3 shows the results of simulation runs on the cost of living. For $k = 1$, populations stabilized at approximately $t = 350$. It took longer for larger values of k. For $k = 2$, stabilization took about 1300 generations. The final average agent count per strategy is shown in the graph for $1 \leq k \leq 2$. Again we see a gap between nice strategies and non-nice ones.

When the cost of living is increased there is a clear decrease in the agent count for non-nice strategies. Moreover, non-nice strategies become extinct when the environment is sufficiently harsh. Nice strategies take in their place, which results in a slight increase in agent count. Interestingly, PavlovC does not seem to be affected by the cost of living at all. For both $k = 1$ and $k = 2$, the final agent count for PavlovC is just above 1000. ALLC even outperforms PavlovC in harsh environments.

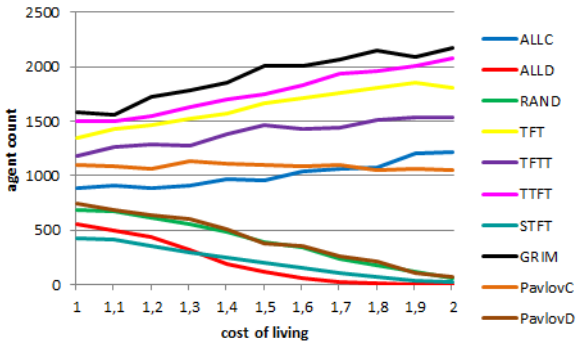

Figure 3: Final agent count with a varying cost of living

6.3 Age Limit

In nature, all individuals are affected by aging. How old one becomes differs per species and per individual. For research purposes the age limit is generalized. Every agent has the exact same age limit. The experiment on the effect of age limitation m was conducted on 10 values for m, from 100 to 500 with steps of 50 and $m = 0$, being no age limit. The value $m = 50$ was excluded from the experiment. For this age limit, none of the agents survived because they died of age before they could reproduce.

In figure 4 it can be seen that an age limit is in the advantage of nice strategies. A short life time results in extinction of non-nice strategies. After an agent reaches its limit, it dies and makes place for a strong strategy waiting to reproduce. Again PavlovC is the exception. Instead of having an advantage, low age limits even affect the strategy negatively.

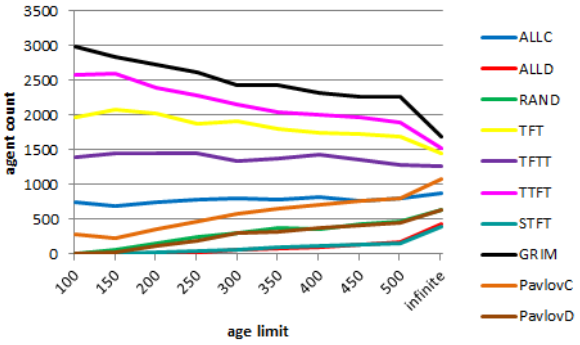

Figure 4: Final agent count with a varying age limit

6.4 Energy limit

In the model, reproduction becomes more expensive as the energy limit rises. Thus, increasing the energy limit will make reproduction conditions worse. Similar to nature, agents have to be strong and healthy enough to reproduce in harsh conditions. The experiment on the effect of h was conducted on 10 values for h, from 50 to 500 with steps of 50.

Figure 5 shows the results of the experiment. For $h = 500$, the gap between nice and non-nice strategies is clearly visible. The lower the energy limit, the more this gap fades. For $h = 50$, non-nice strategies RAND and PavlovD even perform the same as ALLC. The top four strategies count approximately 200 agents more when the energy limit goes from the default value ($h = 150$) to 500. The lower four strategies drop about 150 in agent count in this interval. Surprisingly, PavlovC and ALLC are not affected by the change in energy limit. The agent count deviates less than 50 in the interval $50 < h < 500$.

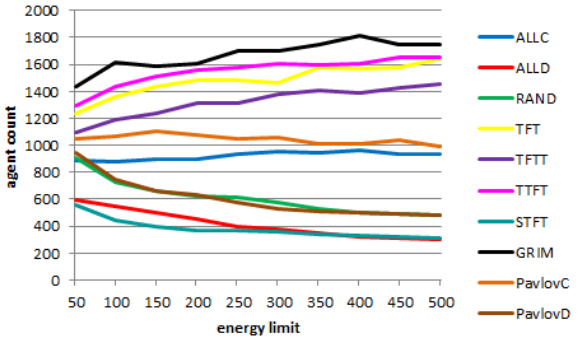

Figure 5: Final agent count with a varying energy limit

6.5 One-on-one Results

Previous experiments were conducted with all strategies at once. In the environment, 1 000 agents of ten different strategies compete for the highest agent count. One-on-one competitions were held as well. Again, the initial population consisted of one-tenth of the capacity, i.e. 500 agents of two different strategies. All strategies played against every other strategy ten times. The number of runs is less than previous experiments because deviations were smaller in one-on-one simulations. The resulting average percentage of the total population was used for the measurement of performance.

Results were quite similar to previous experiments. The order in success of strategies is quite the same. Only PavlovD has an unexpected boost in rankings. It switched places with ALLC. Its success is due to the fact that PavlovD overrules forgiving strategies. In one-on-one competitions PavlovD reaches nearly 100 percent of the population against ALLC and a little less for TFTT and TFT. It is outperformed by TTFT and GRIM, but these strategies do not inhibit PavlovD in competing against other strategies as they did in experiments with all strategies mixed in the environment. Figure 6 shows the exact results.

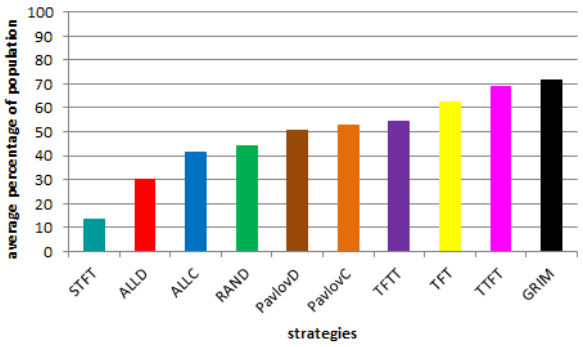

Figure 6: Average percentage of the total population per strategy in one-on-one competitions

7 Conclusion

In this paper, the performance of IPD strategies in nature was investigated. An extension of Smaldino's model was proposed which is able to deploy ten default strategies for IPD in a biologically realistic environment.

Success of strategies was measured by the final agent count. The order of the strategies was mostly the same in all experiments. The six nice strategies outperformed the four remaining non-nice strategies. GRIM, TTFT, TFT and TFTT stayed quite close to each other, followed by ALLC and PavlovC. Those two strategies alternated fifth and sixth place dependent on the settings of the environment. RAND, PavlovD, ALLD and STFT always ended up in the final four places. The order of the first four strategies is the exact order from sternest to most forgiving. Overall, nice but stern strategies are most successful. This conclusion both confirms and contradicts hypothesis 1. Axelrod concluded that strategies should be nice and forgiving. Results verify that strategies should be nice, i.e. start with cooperation. A strategy that starts with defection in its first move performs drastically worse than its twin strategy that starts with cooperation. Examples of this huge difference are TFT and STFT and PavlovC and PavlovD. On the other hand results contradict that strategies should be forgiving. TFT was the winner of Axelrod's tournament. TFT is quite forgiving, but is outperformed in this model by GRIM, which is not forgiving at all. This disagreement may be due to the lack of various strategies. Success of forgiving strategies can be tested better with more different strategies and above all with the presence of stochastic strategies. RAND is the only stochastic strategy in this model. TFT will perform much better with stochastic strategies, because after a defection of its opponent TFT will later on attempt to reestablish cooperation. GRIM would never cooperate again which results in lower payoffs. It should also be stated that this model relies heavily on self cooperation, i.e. cooperation with agents of the same strategy. This is what makes TFT so successful and STFT not. Two TFT agents keep cooperating, while two STFT agents keep defecting each other, due to the first move.

Hypothesis 2 can be tested with the the cost of living and the ability to reproduce, which is affected by age limitation and energy limitation. All conducted experiments partially confirm the hypothesis that harsh environments are in the advantage of cooperators. High costs of living, high energy limits and low age limits make the environment harsh and result in the success of cooperative strategies and the failure of defective strategies, given that the cooperative strategies are nice.

The same goes for hypothesis 3. The difference between performance with evolution and without are significant. Evolution gives cooperative strategies a boost and inhibits defective strategies, again, given that they are nice. After approximately 200 generations defective strategies start adopting nice, cooperative strategies.

An interesting aspect of the results is the performance of PavlovC. PavlovC is cooperative, nice and

forgiving. Therefore, according to the hypotheses, PavlovC should have an advantage in harsh environments and an advantage with evolution. However, it has neither of these advantages. In evolutionary environments PavlovC agents switch to different strategies resulting in a decrease of nearly 300 agents compared to runs without evolution. Moreover, the cost of living and energy limit does not seem to affect the agent count of PavlovC at all and an age limit even harms the strategy. These findings can not easily be explained and should be tested more extensively with more varying strategies.

During experiments clustering and forming of patterns emerged. It was decided that this was out of the scope of this research. For future works, clusters and patterns within IPD strategies could be an interesting topic. Furthermore, the model can be extended with more strategies in the future. Only in Axelrod's first tournament 223 different strategies were submitted. A large part of these strategies are stochastic. The participation of stochastic strategies could change results, especially in the forgiveness of strategies.

Nature counts millions of species. Each of them living in their own climate and having their own strategy of survival. This strategy is profitable against some species and disadvantageous against others. But which strategy has the best chance of survival? By simulating natural behavior in IPD games a lot can be learned about survival and extinction of species in nature.

References

[1] R. Chiong and M. Kirley. Effects of Iterated Interactions in Multiplayer Spatial Evolutionary Games. *IEEE Transactions on Evolutionary Computation*, 16(4):537–555, 2012.

[2] M. Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.

[3] M.A. Nowak and R.M. May. Evolutionary Games and Spatial Chaos. *Nature*, 359(6398):826–829, 1992.

[4] F. Schweitzer and R. Mach. Distribution of Strategies in a Spatial Multi-Agent game. In *Proceedings from the 8th International Workshop on Game Theoretic and Decision Theoretic Agents (GTDT)*, pages 46–54, Hakodate, Hokkaido, Japan, 2006.

[5] M.A. Nowak and K. Sigmund. Tit for Tat in Heterogenous Populations. *Nature*, 355(6398):250–253, 1992.

[6] K. Lindgren and M. G. Nordahl. Evolutionary Dynamics of Spatial Games. *Physica D: Nonlinear Phenomena*, 75(1-3):292–309, 1994.

[7] U. Wilensky. NetLogo PD N-Person Iterated model. http://ccl.northwestern.edu/netlogo/models/PDN-PersonIterated, 2002.

[8] Smaldino P.E. Cooperation in Harsh Environments and the Emergence of Spatial Patterns. *Chaos, Solitons Fractals*, 56:6–12, 2013.

[9] M. Jurišić, D. Kermek, and M. Konecki. A Review of Iterated Prisoner's Dilemma Strategies. In *MIPRO 2012 - 35th International Convention on Information and Communication Technology, Electronics and Microelectronics - Proceedings*, volume 4, pages 1093–1097, 2012.

[10] J. Li. How to Design a Strategy to Win an IPD Tournament. *The Iterated Prisoner's Dilemma: 20 Years on*, 4:89–104, 2007.

[11] U. Wilensky. NetLogo. http://ccl.northwestern.edu/netlogo/, 1999.