Quasi-Stable States in the Iterated-Prisoner's Dilemma

Philip T. Mueller

Fittest Bits, Niwot, CO
phil.mueller@fittestbits.com

Abstract

This paper describes the states of a heterogenous population of agents playing the Iterated-Prisoner's Dilemma. The interactions of the agents are governed by five interaction processes which range from highly localized interactions to complete mixing; while the evolution of the agents is governed by five adaptive processes which range from local processes to global processes. For certain combinations of interaction processes, adaptive processes and control parameters, the populations alternate between periods of cooperation and defection while spending relatively little time in between. In addition, even at high rates of mutation, the population does not degenerate into random play.

Introduction

The Prisoner's Dilemma is a two player simultaneous non-zero sum game. Each player can chose to cooperate or defect with the payoffs for the various combinations shown in Table 1.

Payoff table (player1, player2)		Player 2	
		cooperate	defect
Player 1	cooperate	R=3, R=3	S=0, T=5
	defect	T=5, S=0	P=1, P=1

Table 1: Prisoner's Dilemma payoff table. T is the temptation to defect; S is the suckers payoff; R is the reward for mutual cooperation; and P is the punishment for mutual defection.

The dilemma occurs when $T > R > P > S$ and $2R > T + S$. For those constraints, the two players are collectively better off if they both cooperate, but individually will do at least as well as the other player if they defect.

When two players play the Prisoner's Dilemma multiple times, it is called the Iterated-Prisoner's Dilemma (IPD). The IPD has been used extensively over the past 40 years to study cooperation and conflict in economic, social, biological and political settings (Axelrod, 1984; Boyd and Lorberbaum, 1987; Poundstone, 1992; Nowak and Sigmund, 1995; Hoffman and Waring, 1996; O'Riordan, 2001). In addition, the IPD has been used to study the evolution of strategies (Axelrod, 1997). Changing the relative payoffs has also been studied (Angeline, 1994; Yoshida et al., 1998; Delahaye and Mathieu, 1996).

Cyclical and irregularly oscillating populations have been studied (Nowak and Sigmund, 1989; Nowak and Sigmund, 1993). These experiments were based on populations containing only a few unique strategies and at each time step the proportion of each strategy was adjusted based on that strategy's performance. However, in the experiments here, the strategies themselves evolve.

In addition, the experiments in this paper focus on errors in the adaptive process; while others have focused on errors in the play of the IPD (Molander, 1985; Bendor, 1993; Wu and Axelrod, 1995).

The interaction processes of these experiments are derived from the fact that most human interactions are based on location, social labeling or organizational roles that make some interactions more likely than others. In addition, the adaptive processes of these experiments are derived from the fact that humans learn rapidly from comparisons between their own performance and the performance of others. The variation in interaction processes and adaptive processes from local to global allow us to see how these parameters affect the stability of the population. As we shall see, the combination of local interactions with global adaptations leads to unstable behavior.

In this paper, we are interested in the dynamics of these instabilities and not the cases which have a single evolutionarily stable strategy.

Experiments

Each experiment consists of a population of 256 agents implementing a stochastic strategy for playing the IPD, an interaction process determining how opponents are chosen, and an adaptive process determining how the agents evolve over time. There are five interaction processes which move from fixed localized interactions to complete mixing in small steps; and five adaptive processes, one global, two local and two control.

The stochastic strategy for playing the IPD is made up of a 3-tuple of real (64-bit floating point) values (i, p, q) (Nowak and Sigmund, 1992). The value i is the probability of cooperating on the first move; p is the probability of cooperating if the other player cooperated on the previous move; and q is the probability of defecting if the other player defected on the previous move. Thus, TIT-FOR-TAT (TFT) (Axelrod, 1984) would be represented by (1.0, 1.0, 0.0); always defect (ALLD) by (0.0, 0.0, 0.0); and always cooperate (ALLC) by (1.0, 1.0, 1.0). These three values are the "genes" that will be acted on by the adaptive processes described below.

The initial strategies are uniformly distributed across the (p,q) space, with one agent at $i = p = \{0, 1/16, 2/16, ..., 15/16\} \times q = \{0, 1/16, 2/16, ..., 15/16\}$. The agents are randomly placed on a 2 dimensional world grid.

The interaction processes are described in (Cohen et al., 1999) and are summarized here:

- 2DK – 2 Dimensional grid (16×16 torus) with agents keeping the same position each time step and playing the same four neighbors (North-South-East-West) each time step.
- FRNE – Fixed random network of symmetrical neighbors. A random network of four paired neighbors is selected at time step 0 and kept for the entire run.
- FRN – Fixed random network of asymmetrical neighbors. Each agent picks four other agents at random to play at time step 0 and plays the same neighbors for the entire run.
- 2DS – 2 Dimensional grid (16×16 torus) with the agent's position shuffled on each time step and playing the four neighbors (North-South-East-West) each time step.
- RWR – Each agent picks four other agents at random (with replacement) to play at each time step.

Adaptive processes determine how the agents evolve over time. Two of the adaptive processes (1FGA and BMGAS) are described in (Cohen et al., 1999) and summarized here, while three others (MBMGAS, R1FGA and RBMGAS) are described here:

- 1FGA – For each agent, select another agent at random, if it played better, copy its strategy (with errors). There are two sources of errors, the comparison of the randomly chosen agent with the current agent and possible mutation when copying.
- BMGAS – For each agent, if the best agent it played was better, copy its strategy (with errors). There are two sources of errors, the comparison of the best agent with the current agent and possible mutation when copying.
- MBMGAS – Same as BMGAS except with errors while selecting the best agent played. That is, there are three sources of errors, selecting the best agent played, the comparison of the (so called) best agent with the current agent and possible mutation when copying.
- R1FGA – For each agent, select another agent at random, then select randomly between the two agents. This is a control for 1FGA.
- RBMGAS – For each agent, select randomly amoung the agent and the agents it played. This is a control for BMGAS and MBMGAS.

Each agent plays the IPD at each time step and each game consists of four moves. Which opponents are chosen and the total number of games played by each agent depends on the interaction process for that experiment. However, each agent will play at least four games each time step. The agent's final score is the average of its score across all of the games it played that time step. After all of the agents have played their games, the agent's strategies are synchronously updated based on the adaptive process for that experiment.

The non-control adaptive processes all involve choosing the best agent based on score and then copying that agent's strategy. When choosing the best agent, errors are made based on the selection error rate; when copying the agent's strategy, errors are made based on the mutation rate and the mutation spread. The selection error rate is the probability of choosing the wrong agent at each comparison; the mutation rate is the probability of independently mutating each of the genes; and the mutation spread is the standard deviation of the normal random number (with mean 0) which is added to mutate a gene. If after mutation a gene is greater than 1.0, it is set to 1.0 and if it is less than 0.0, it is set to 0.0.

For the control adaptive processes, instead of selecting the best agent based on score, the "best" agent is selected at random. Copying the selected agent's strategy is still subject to the same errors as in the non-control cases.

A history consists of 3000 time steps; and a complete run of an experiment consists of 1000 histories. The primary figure of merit, the average score for the population at each time step, gives a measure of the amount of cooperation in the population. To avoid biases relating to startup, data is only collected for the last 1500 time steps of each history. To obtain two numbers to describe the results of an experiment, the average and standard deviation of the population average score is computed for the last 1500 time steps for each history. Then the averages and standard deviations are averaged across the histories. However, we shall see that these two numbers do not tell the whole story.

The parameters varied in these experiments are the selection error rate, the mutation rate and the mutation spread.

Interaction Process	Adaptive Process	Mean Score (sd) (Cohen et al., 1999)	Mean Score (sd)	Mean Score Diff. (%)
2DK	1FGA	2.025 (0.069)	2.032 (0.064)	0.3
	BMGAS	2.554 (0.009)	2.555 (0.007)	0.0
	MBMGAS		2.502 (0.008)	
	R1FGA		2.247 (0.020)	
	RBMGAS		2.224 (0.019)	
FRNE	1FGA	2.035 (0.089)	2.032 (0.065)	−0.1
	BMGAS	2.572 (0.007)	2.572 (0.007)	0.0
	MBMGAS		2.526 (0.008)	
	R1FGA		2.247 (0.020)	
	RBMGAS		2.230 (0.020)	
FRN	1FGA	1.884 (0.120)	1.906 (0.093)	1.1
	BMGAS	2.476 (0.026)	2.420 (0.015)	−2.3
	MBMGAS		2.383 (0.016)	
	R1FGA		2.247 (0.020)	
	RBGMAS		2.239 (0.024)	
2DS	1FGA	1.484 (0.086)	1.495 (0.072)	0.7
	BMGAS	1.089 (0.003)	1.089 (0.004)	0.0
	MBMGAS		1.095 (0.004)	
	R1FGA		2.244 (0.020)	
	RBMGAS		2.245 (0.021)	
RWR	1FGA	1.502 (0.109)	1.487 (0.090)	-1.0
	BMGAS	1.098 (0.036)	1.103 (0.020)	0.5
	MBMGAS		1.110 (0.016)	
	R1FGA		2.246 (0.020)	
	RBGMAS		2.244 (0.022)	

Table 2: Comparison of results.

Results

As shown in Table 2, this study verifies the results in (Cohen et al., 1999) for the ten experiments that match. The selection error rate was 0.1, the mutation rate was 0.1 and the mutation spread was 0.4 for all of the experiments listed in the table.

Figure 1 is a histogram of average population score values for a single combination of interaction process and adaptive process (2DK and 1FGA). Each line shows the histogram with a selection error between 0.00 and 0.25 and with a mutation rate of 0.10 and a mutation spread of 0.20. The histogram for 2DK with R1FGA is also shown for reference.

We see that there is a relatively smooth transition from a cooperating population when the selection error is 0.00 to a non-cooperating population when the selection error is 0.25. When the selection error is near the middle of the range (0.10 and 0.15) the population alternates between cooperative and highly non-cooperative states, while spending relatively little time in between those two states.

From an evolutionary point of view, there are two stable states in the fitness landscape at about average score 1.25 and 2.3, and there is sufficient "energy" in the selection error rate, mutation rate and mutation spread to allow the popula-

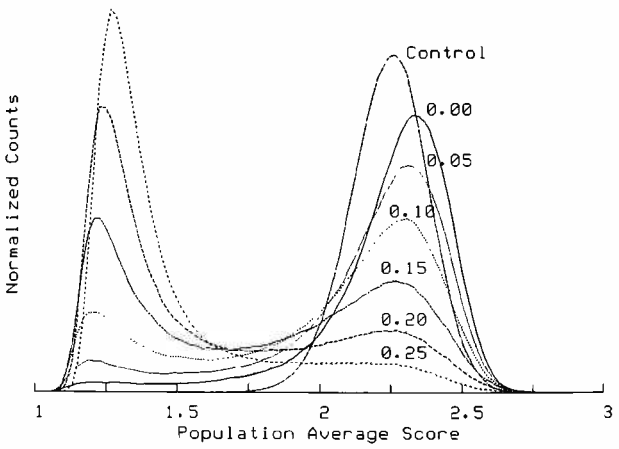

Figure 1: Score histograms for 2DK/1FGA with selection error 0.00-0.25, mutation rate 0.10 and mutation spread 0.20 and the control score histogram 2DK/R1FGA.

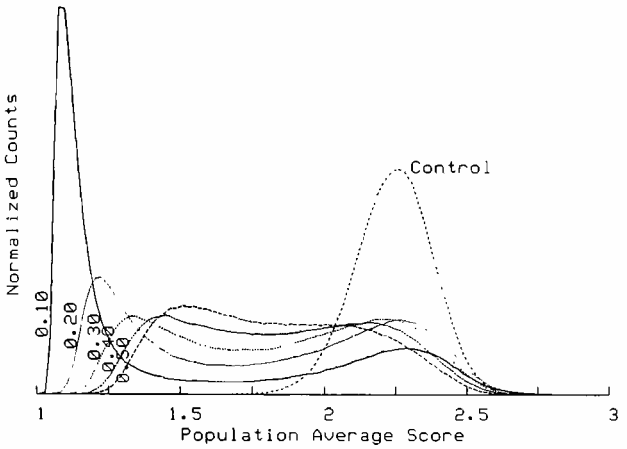

Figure 2: Score histograms for 2DK/1FGA with selection error 0.15, mutation rate 0.10 and mutation spread 0.10-0.50 and the control score histogram 2DK/R1FGA.

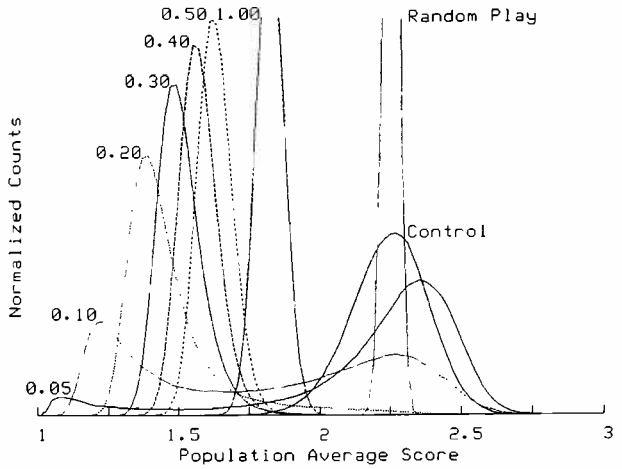

Figure 3: Score histograms for 2DK/1FGA with selection error 0.15, mutation rate 0.05-1.0 and mutation spread 0.20, the control score histogram 2DK/R1FGA and the score histogram for random play.

tion to jump between these two states. In addition, in Figure 2, we see that for larger values of the mutation spread, there is so much "energy" that all states between an average score of 1.5 and 2.5 become equally probable.

Figure 3 is a set of histograms of average population scores for the 2DK/1FGA experiment with varying values of mutation rate. For comparison, the histogram for random play is included.

The adaptive process used in the random play experiment sets each "gene" of each player's strategy to a random number uniformly distributed between 0 and 1, inclusive, at each time step.

For this set of experiments, adding "energy" to the system by increasing the mutation rate changes the fitness landscape from having two quasi-stable states to a single highly stable state. The stability of the state is inferred from the steepness of the distribution.

The mutation algorithm used in these experiments tends to scatter the values of the "genes" in the interval [0,1] with a bias to the boundary values. In this case, ALLD performs better than the other strategies. So, the selection pressure of the adaptive process will tend to favor the ALLD strategy which is what we see from the position of the histogram peaks in the figure.

Figure 4 shows histograms corresponding to Figure 1 for all of the experiments. That is, the average score histograms for all of the combinations of interaction processes and adaptive processes with a selection error rate between 0.00 and 0.25, a mutation rate of 0.10 and a mutation spread of 0.20. Each column is a different interactive process and each row is a different adaptive process.

Quasi-stable states only appear for the adaptive process 1FGA, that is, only for the global adaptive process. The two states are well defined for the interaction processes that have fixed local interactions, but the cooperate state is practically non-existent for the two interaction processes with a high degree of mixing. None of the BMGAS or MBMGAS cases have quasi-stable states. Thus, it appears that it is the global nature of 1FGA that gives rise to the two quasi-stable states.

Conclusion

One way to look at this data is to consider the surface $S: \mathbf{R}^3 \rightarrow \mathbf{R}^2$, where the domain is the three control parameters: selection error, mutation rate and mutation spread and the range is the population average score and the normalized population count. For these experiments, the range and domain were restricted to $[0,0.25] \times [0,1] \times [0,0.5] \rightarrow [1,3] \times [0,1]$. Figures 1–3 are orthogonal slices of this surface.

From the figures, we can see that the quasi-stable states are weakly dependent on the selection error but strongly dependent on the mutation rate and mutation spread. In fact, the quasi-stable states only occur in the region bounded by $[0.05,0.20] \times [0.05,0.20] \times [0.1,0.3]$ which is only about 4% of the entire domain.

These experiments show that a stable cooperative population can suddenly and rapidly become a stable non-cooperative population (and vice versa). In addition, this happens when the interactions between the agents is localized and fixed and when the agents have access to global information about the best strategy, and then use that information erroneously. That description bears an eerie resemblance to modern society, where most people interact mostly with nearby people, yet are informed of the actions of others via mass media.

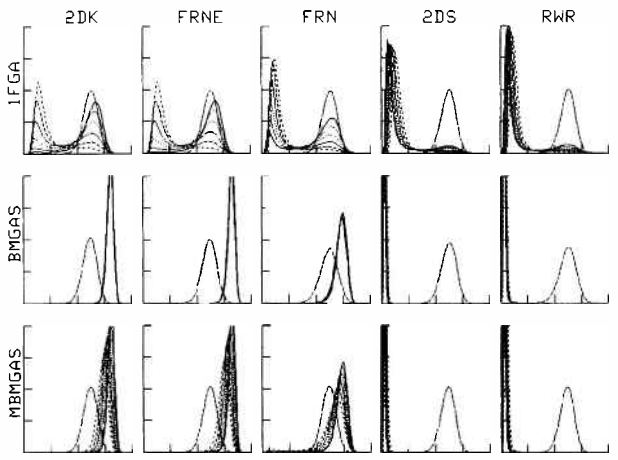

Figure 4: Score histograms for all experiments with selection error 0.00-0.25, mutation rate 0.10 and mutation spread 0.20 and their control score histograms.

Further Research

The experiments in this paper leave many questions unanswered. For example, what causes the population to transition from one quasi-stable state to the other? One approach to answering this question is to look at the distribution of agents in (p,q) space for different population average scores. That is, to look at the three dimensional histogram of population average score, p and q.

Other questions include: How does the population size affect the fitness landscape? Will smaller populations increase the stability of the two states reducing the time spent between them? Will larger populations take longer to transition from one state to the other thus blurring the distinctions between the states?

What about other interaction processes such as tag-mediated selection (Riolo, 1996; Cohen et al., 1999). Under what conditions, if any, does that system exhibit quasi-stable states?

Another area of interest is to investigate the effects of "teaching" the agents by biasing them toward a particular strategy. For example, biasing the agents toward ALLC by adding a small positive random offset to i, p and q after the mutation step of the adaptive process. The question is how do various biases affect the fitness landscape?

Acknowledgements

I would like to thank Chris Platt for the many conversations on this research and Mary Mueller for helping track down some of the references and for her encouragement. I would also like to thank all of the reviewers for their comments.

References

Angeline, P. J. (1994). An alternate interpretation of the iterated prisoner's dilemma and the evolution of non-mutual cooperation. *Proceedings 4th Artificial Life Conference*, pages 353–358.

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

Axelrod, R. (1997). *The Complexity of Cooperation. Agent-Based Models of Competition and Collaboration.* Princeton University Press, Princeton, New Jersey.

Bendor, J. (1993). Uncertainty and the evolution of cooperation. *Journal of Conflict Resolution*, 37(4):709–734.

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

Cohen, M. D., Riolo, R. L., and Axelrod, R. (1999). The emergence of social organization in the prisoner's dilemma: How context-preservation and other factors promote cooperation. Sante Fe Institute Working Paper 99-01-002, Sante Fe Institute, Sante Fe, NM.

Delahaye, J.-P. and Mathieu, P. (1996). Random strategies in a two levels iterated prisoner's dilemma : How to avoid conflicts. *Proceedings of the ECAI 96 Workshop: Modelling Conflicts in AI*, pages 68–72.

Hoffman, R. and Waring, N. (1996). The localisation of interaction and learning in the repeated prisoner's dilemma. Sante Fe Institute Working Paper 96-08-064, Sante Fe Institute, Sante Fe, NM.

Molander, P. (1985). The optimal level of generosity in a selfish, uncertain environment. *Journal of Conflict Resolution*, 29(4):611–618.

Nowak, M. and Sigmund, K. (1989). Oscillations in the evolution of reciprocity. *Journal of Theoretical Biology*, (137):21–26.

Nowak, M. and Sigmund, K. (1992). Tit for tat in heterogeneous populations. *Nature*, 355:250–253.

Nowak, M. and Sigmund, K. (1993). Chaos and the evolution of cooperation. *Proceedings of National Academy of Science, USA*, 90:5091–5094.

Nowak, M. and Sigmund, K. (1995). Invasion dynamics of the finitely repeated prisoner's dilemma. *Games and Economic Behavoir*, pages 364–390.

O'Riordan, C. (2001). Iterated prisoner's dilemma: A review. Technical Report NUIG-IT-260601, National University of Ireland, Galway, Ireland.

Poundstone, W. (1992). *Prisoner's Dilemma.* Doubleday, New York.

Riolo, R. L. (1996). The effects of tag-mediated selection of partners in evolving populations playing the iterated prisoner's dilemma. Sante Fe Institute Working Paper 97-02-016, Sante Fe Institute, Sante Fe, NM.

Wu, J. and Axelrod, R. (1995). How to cope with noise in the iterated prisoner's dilemma. *Journal of Conflict Resolution*, 39(1):183–189.

Yoshida, S., Inuzuka, N., Naing, T. T., Seki, H., and Itoh, H. (1998). A game-theoretic solution of conflicts among competitive agents. *Lecture Notes in Computer Science*, 1441:193–205.