How Important Is Your Reputation in a Multi-Agent Environment

Xin Yao
School of Computer Science
The University of Birmingham
Edgbaston, Birmingham B15 2TT
United Kingdom
x.yao@cs.bham.ac.uk

Paul J. Darwen
Dept. of Computer Science and Electrical Engineering
The University of Queensland
Brisbane, Queensland 4072
Australia
darwen@csee.uq.edu.au

Abstract

Most work on the evolutionary approach to the iterated prisoner's dilemma (IPD) game uses a binary model where the choice of each player can only be cooperation or defection. However, we rarely commit ourselves to complete cooperation or defection in the real world. This paper examines the continuous IPD game and similarities and differences between the discrete and continuous games. The paper also studies the issue of reputation of a player, following Nowak and Sigmund's recent work, and how it affects the evolution of cooperation. This study differs from Nowak and Sigmund's in that players in a population can have more than two levels of cooperation (or even continuous). The players are also changing all the time under the influence of selection, crossover and mutation. We think that this is a more realistic model of the evolution of society in the real world.

1 Introduction

Co-evolutionary learning can discover solutions to problems, without knowledge from human experts. The method has worked well on simple games such as Iterated Prisoner's Dilemma (IPD) [2] [4] [7] [11], and on non-game tasks, such as creating a sorting algorithm [9].

This paper looks at two recently-studied additions to the original IPD game, which have never been studied together.

Firstly, IPD usually allows players only have two choices of action: full cooperation or full defection [2]. Recent papers have considered what happens more choices than those two extremes [8] [15].

Secondly, IPD games usually last a long time, to allow future retaliation (the "shadow of the future" [1, page 13]) and thus make mutual cooperation more likely. Short games tend to encourage defection.

A recent paper [13] looked at "reputation" in IPD — information about players' past actions are available to future opponents. In the real world, most short-term interactions are nonetheless cooperative, motivated by records of previous interactions: credit ratings, criminal records, etc. In IPD, reputation would allow the accumulation of knowledge from previous (short) games, which should have the same encouragement on cooperation as long games [13].

That recent paper [13] only considered IPD with two choices, full cooperation or full defection. This paper will consider IPD with reputation, with more choices than the two extremes.

This paper brings together IPD with a wide choice of cooperation levels, and reputation caused by earlier games.

Section 2 briefly reviews recent work on IPD. Section 3 describes how the co-evolution was implemented. Results are in Section 4: Section 4.1 considers the case of shorter-length interactions in IPD, and Section 4.3 considers IPD games with fine-grained choices of action.

Section 5 discusses a relevant real-world situation: anti-ballistic missile (ABM) systems, the kind that would defend Japan from North Korean missiles. Should ABMs be outlawed, as they were in the 1970's, because they supposedly promote nuclear war? The paper concludes in Section 6.

2 Background

2.1 Co-Evolutionary Learning

Evolutionary computation maintains a population of trial solutions to some problem. An *evaluation function* measures the quality of each trial solution. Usually, a human expert programs that evaluation function, knowing what a good solution should be. Unfortunately, for many game-like applications, experts are often unavailable.

Co-evolution is where each trial solution in the evolving population is evaluated by its peers in the same population (or perhaps by another population evolving in parallel). It optimizes to a moving target. As evolution proceeds, population members become better judges of each other.

Co-evolution's most popular application is to learn a game. Each individual in the population is a strategy, and plays the game against every other individual in the current population. A strategy's fitness is the average score of those games. These are usually 2-player (i.e., pairwise) interactions, but they can also be n-player interactions [17].

Starting from a random initial population, co-evolution requires no *a priori* knowledge of how to play the game well. As the population evolves, the individual strategies improve at playing the game. This means the co-evolutionary evaluation function escalates its level of difficulty, causing an "arms race" of ability.

Numerous studies have followed this general approach on a variety of games such as Backgammon [5] [14] and game-like tasks [16] [10]. Some studies use two populations, where a member of one population is evaluated by

members of the *other* population, and vica-versa [9].

2.2 Classic Iterated Prisoner's Dilemma

Iterated Prisoner's Dilemma is an abstract mathematical game which is widely studied in political science, biology, and artificial intelligence [1]. In its basic form, the Prisoner's Dilemma is a two-player game where each player has two choices, and the payoff matrix for those two choices (Figure 1) satisfies the following conditions:

- $T > R$ and $P > S$ (Defection always pays more)
- $R > P$ (Mutual cooperation beats mutual defection)
- $R > (S + T)/2$ (Alternating doesn't pay)

	Cooperate	Defect
Cooperate	R, R	T, S
Defect	S, T	P, P

Figure 1: The payoff matrix for the 2-player prisoner's dilemma game. The values S, P, R, T must satisfy $T > R > P > S$ and $R > (S + T)/2$.

There are many choices of parameters to satisfy the conditions of an IPD shown in Figure 1. This paper uses the values $R = 4$, $T = 5$, $S = 0$, and $P = 1$.

In *iterated* prisoner's dilemma, the game in Figure 1 is played not just once, but many times, with the memory of previous iterations. This allows the possibility of retaliation and mutual cooperation.

The iterated game is widely studied because it contains the basics of many real-world situations: examples include the U. S. Watergate scandal of 1972-75 [12], the Cold War of 1945-1990 [3], and life in the trenches during the First World War of 1914-1918 [1, pages 73-87].

As in the real world, co-evolutionary learning applied to IPD demonstrates that sometimes mutual cooperation can dominate, even without some central authority to enforce cooperation [2]. This "evolution of cooperation" is occasionally punctuated by sudden mass extinctions [6, page 287] [8, page 141] [11].

2.3 Prisoner's Dilemma with More Choices

In the original version of 2-player IPD, each player only has two choices: full cooperation or full defection. In reality, however, cooperation is rarely all-or-nothing [15, page 175] — instead, players can choose among varying degrees of cooperation. Recent papers add continuous levels of cooperation [8] [15] to the game of IPD.

	-1	$-\frac{1}{3}$	$+\frac{1}{3}$	$+1$
-1	1	$2\frac{1}{3}$	$3\frac{2}{3}$	5
$-\frac{1}{3}$	$\frac{2}{3}$	2	$3\frac{1}{3}$	$4\frac{2}{3}$
$+\frac{1}{3}$	$\frac{1}{3}$	$1\frac{2}{3}$	3	$4\frac{1}{3}$
$+1$	0	$1\frac{1}{3}$	$2\frac{2}{3}$	4

Figure 2: The payoff matrix for the 2-player prisoner's dilemma game, with four choices of cooperation.

In this paper, more choices for cooperation is a straightforward generalization of the 2-choice case. For example, if each player has four choices (instead of the usual two), the payoff matrix we use is in Figure 2, which shows the payoff to the player on the left. Note that the four corners of Figure 2 are the same payoffs as for the two-choice game — it is merely interpolation.

In Figure 2, the payoff to player A is given by:

$$p_A = 2.5 - 0.5c_A + 2c_B, (-1 \leq c_A, c_B \leq 1) \quad (1)$$

where c_A and c_B are the cooperation levels of the two players, which are discretized into two, four, or however many fine-grained choices of cooperation are allowed.

2.4 Prisoner's Dilemma with Reputation

A recent paper added "reputation" (a.k.a. indirect reciprocity) to IPD [13]. They found that, making available the results of previous games to future players, would increase the likelihood of cooperation. Unfortunately, these results were only considered when players have only two choices — full cooperation or defection. This paper considers reputation when players have a wider range of choices.

3 Setup And Implementation

Following Axelrod [2], a genetic algorithm (GA) maintains a population of trial strategies for IPD. Every generation of the GA, each strategy plays against other members of the current population. A player's fitness is its average score over all these games. Opponents are chosen in random order, to avoid predictability for reputation (see Section 3.4).

In each game of IPD between two players, each player can see the opponent's recent actions, and their own recent actions — in the runs shown here, this history is limited to only the most recent iteration, i.e., what just happened.

To start a game, each player has in its genotype what it presumes are the pre-game "history", that affects its first move.

3.1 The Shadow of the Future

To end a game, the "shadow of the future" [1, page 13] is an issue. Briefly, if players could count, and if the number of iterations in a game of IPD were fixed, then players would have no incentive to cooperate on the very last move — but

given that, they would have no incentive to cooperate on the *second* last move, and so on back to the start of the game.

One way to keep game length uncertain is to have a probability of ending the game on this move. By reducing this probability, games can be shorter, and still of uncertain length (to encourage cooperation).

However, in this paper, individual strategies are represented in such a way (feedforward neural networks) that they cannot count. This allows us to use a fixed-length game, with no chance of the players figuring out that games are fixed-length – to them, it is the same situation as using a probability of ending the game.

3.2 Representation of Individual Strategies

Each member of the GA population is a fixed-length string of floating-point numbers. These represent the weights and biases of a feed-forward neural network with 20 hidden nodes, and one output node. For each remembered iteration of IPD, there are four inputs (five with reputation, see Section 3.4):

1. One for one's own earlier level of cooperation.
2. One for the opponent's earlier level of cooperation.
3. An input which is 1 if the opponent exploited the player, and zero otherwise.
4. An input which is 1 if the opponent was exploited by the player, and zero otherwise.

For the last two inputs, "exploited" means that the player's level of cooperation is different. Even if payoffs are different by a tiny amount, the corresponding input will still be 1. These last two inputs avoid learning to subtract the other two inputs — with this implementation, players know immediately if they are being exploited, even by a small amount.

3.3 Discretized Cooperation and Payoff

Each individual is a neural network with one output node. Like all the hidden nodes, that output node is sigmoided, to put its value in the range $[-1, +1]$.

After sigmoiding, the output is still continuous. This output is discretized in such a way that all the values are equidistant from each other. For example, if there were 4 choices of cooperation, the output node's value in the range $[-1, +1]$ would be adjusted to one of the values $(-1, -\frac{1}{3}, +\frac{1}{3}, +1)$. This discretized output is the player's choice of cooperation for the current move.

For two players A and B, with discretized levels of cooperation of a player c_A and its opponent c_B, the payoff to player A is given by Equation 1:

$$p_A = 2.5 - 0.5c_A + 2c_B, (-1 \leq c_A, c_B \leq 1)$$

3.4 Reputation

In addition to the four inputs described in Section 3.2, there is one additional input. For the 2-player game, a player sees the "reputation" of its opponent (not its own).

Each generation, the new population's reputations are initialized to zero. For the first five games, a player is given the benefit of the doubt by having zero reputation. After that, a player's reputation is the sign (+1 or -1) of the average difference in payoff of its previous games this generation — using the sign is to make it more obvious to neural networks. For example, if a particular player is being greedy and exploitative, then the payoff difference between it and its unlucky opponents will be positive. After the first 5 games, opponents will see +1 in their input for opponents' reputation.

As Nowak and Sigmund [13] point out, reputation should counteract the tendency of IPD games of short duration to cause defection. With reputation, the results of previous games (even short ones) are available to later players, which should have a similar effect as a long game of IPD.

4 Results

4.1 Shorter Games Discourage Cooperation

In most studies of IPD, each game between two players lasts for around 150 iterations, following Axelrod [2]. This is so long, that players have a huge incentive to establish a cooperative relationship.

Shorter-length interactions offer less incentive to cooperate. Figure 3 shows ten runs of IPD with game length 10 iterations (instead of 150), for a co-evolving population of 100, where players can choose from 2 levels of cooperation.

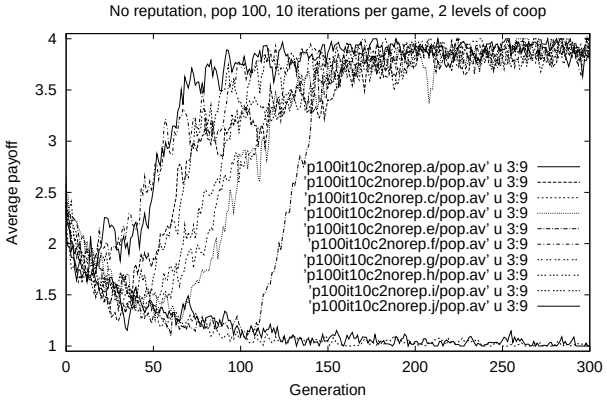

Figure 3: Game length is only 10 iterations, so there is less incentive to cooperate than for games of 150 iterations.

In Figure 3, 10 iterations is still long enough to encourage cooperation in most co-evolutionary runs — only 2 out of 10 runs has defection dominant. Shortening that to only 5 iterations in Figure 4, shows that cooperation is rare — only 1 out of 10 runs rises above defection. Any shorter than that, and cooperation becomes extremely unlikely.

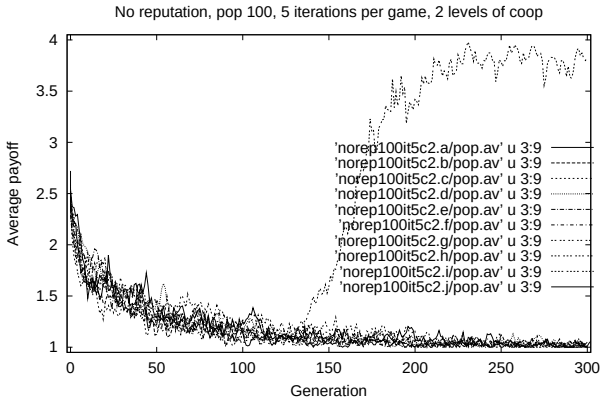

Figure 4: Game length is only 5 iterations, so there is less incentive to cooperate.

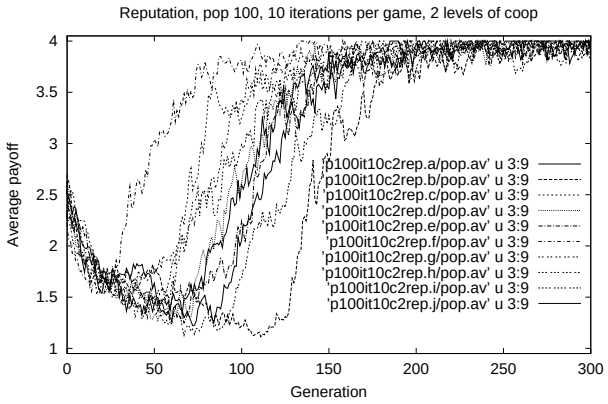

Figure 5: Game length is only 10 iterations, so there is less incentive to cooperate, but reputation makes cooperation more likely.

Reducing game length to only 5 iterations again, reputation still makes cooperation attractive — 9 out of 10 runs show mutual cooperation dominating.

Figure 6 reduces game length completely, to only 1 iteration — non-iterated one-shot prisoner's dilemma, so players decide their single move based only on the opponent's reputation. Figure 6 shows at least some level of mutual cooperation in 8 out of 10 runs. There is, however, much more volatility in the runs with some cooperation.

Figures 5 through 6 validate some of the results of Nowak and Sigmund [13]. Reputation can encourage mutual cooperation in short-length interactions, by making information about previous interactions available for the future.

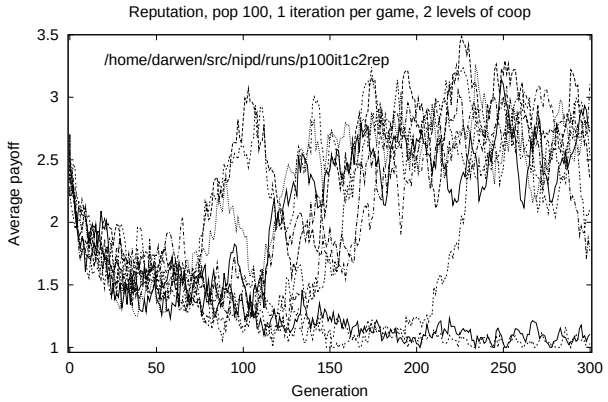

Figure 6: Game length is only 1 iteration, so there is less incentive to cooperate. But reputation keeps it a somewhat attractive option.

4.2 Reputation Reduces Tendency to Defection

Just as for Figure 3, Figure 5 also has a population of 100, player IPD games of length 10 iterations, with players choosing from only 2 levels of cooperation. The difference is, Figure 5 has reputation — this makes cooperation more popular than in the no-reputation case of Figure 3.

4.3 More Choices of Cooperation Discourages Cooperation

Cooperation was popular in Figure 3, which only offered two choices of cooperation. What happens if players have a more fine-grained choice? In Figure 7, players have 8 choices instead of 2. To give cooperation a chance, game length is increased to 20 iterations, and a larger population of 150 promotes searching the more complicated game. The extra choices completely demolish mutual cooperation — Figure 3 shows that defection always dominates.

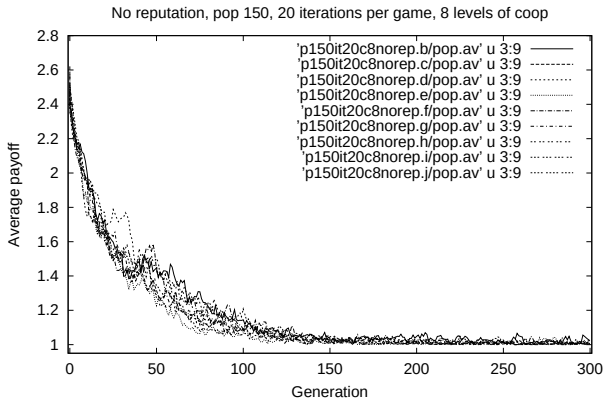

Figure 7: Game length is 20 iterations, with 8 choices of cooperation, and defection dominates.

However, the same situation with reputation makes cooperation more likely, as shown in Figure 8, with 9 out of 10 runs showing some degree of mutual cooperation. However, the extra choices do greatly complicate matters — it cannot be said that reputation has made cooperation as pop-

ular as it is in longer games [2] [6].

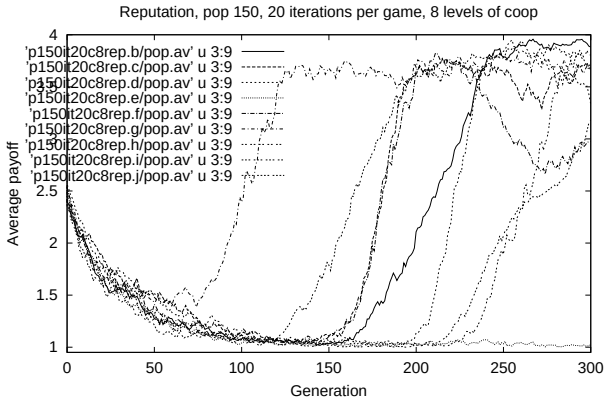

Figure 8: Game length is 20 iterations, with 8 choices of cooperation. With reputation, cooperation is much more likely.

4.4 Game Length Still Has Same Effect

The dependence on game length remains the same with the extra choice. In Figure 8 with its 8 choices of cooperation, game length was a generous 20 iterations. If we reduce this, it makes defection more likely, just as it did for only 2 choices in Section 4.1.

Reducing game length to only 10 iterations in Figure 9 still causes some degree of mutual cooperation, although it is not as overwhelming as it was for 20 iterations.

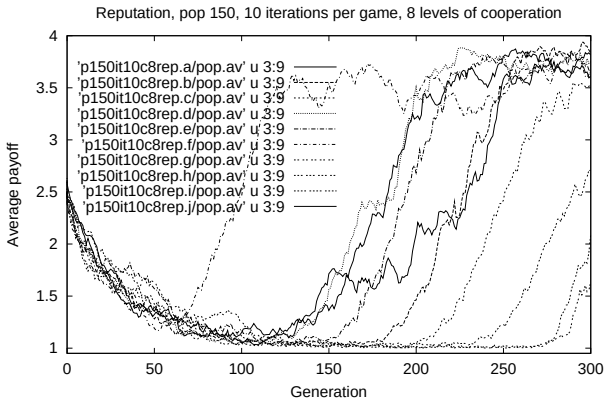

Figure 9: Game length is 10 iterations, with 8 choices of cooperation. With reputation, cooperation is still likely.

Reducing game length to only 5 iterations in Figure 10 means that fully half of the runs have defection dominating. This demonstrates that, for 2 or 8 choices, reputation can mitigate but not avoid the tendency to defect in short-length interactions.

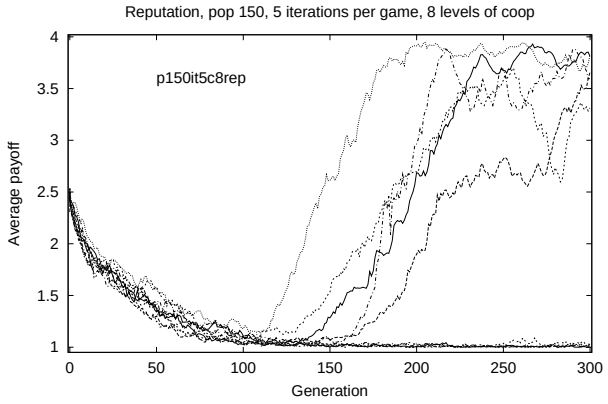

Figure 10: Game length is 5 iterations, with 8 choices of cooperation. With reputation, cooperation is still likely.

4.4.1 Number of Choices Doesn't Affect Reputation

This raises a question: just how good is reputation at promoting mutual cooperation, for the tricky situation of many choices of cooperation? If we keep the other parameters the same as for Figure 9 — population 150, games 10 iterations long — and vary the number of available levels of cooperation, does cooperation become more or less likely?

Surprisingly, no. Even though the number of choices has a devastating effect on the case without reputation, varying the number of choices has little effect when reputation is available. Figure 11, with a huge 64 choices available, doesn't look any different from the case with only 8 choices in Figure 9.

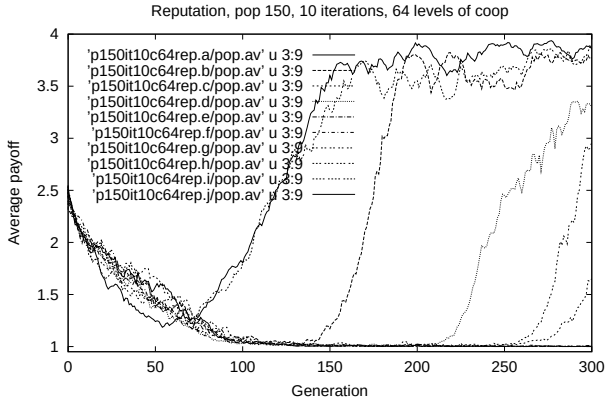

Figure 11: Game length is 10 iterations, with 64 choices of cooperation. With reputation, cooperation is still likely.

5 Discussion

Iterated Prisoner's Dilemma provides clues for real-world situations. One of these is nuclear warfare, the ultimate IPD game [3]. The above results give insights into a problem that currently faces Japan, Taiwan, and the United States —

anti-ballistic missiles (ABMs).

During the Cold War, ABM systems were outlawed by the Strategic Arms Limitation Treaty (SALT-2). The justification for this comes from IPD considerations. Today, ABM systems are back on the agenda.

From Figure 3, having only two stark choices (full cooperation or full defection) generally encourages cooperation, compared to Figure 7 (with 8 choices instead of 2) where mutual cooperation is very improbable — a chilling possibility if nuclear weapons are involved. This observation — fewer choices of cooperative levels promotes cooperation — led to the doctrine of Mutually Assured Destruction (MAD): if the only choices are peace or full-scale all-out nuclear war, then it makes the latter less likely. ABM systems, by offering more fine-grained choices in nuclear brinkmanship, went against the MAD doctrine — if ABM systems increase choice, then they could cause a war. So ABM systems were outlawed.

But classic studies of IPD do not consider reputation. This is a gross simplification of international diplomacy. Countries have reputations. The MAD doctrine is incorrect, according to Figure 11: with reputation, you can have many fine-grained choices in cooperation, without threatening mutual cooperation.

Although computer models are not the same as the real thing, we cannot run experiments on the real thing — we cannot run two countries as a laboratory — but Figure 11 suggests that ABM systems do not (by themselves) necessarily encourage nuclear war, due to the effects of reputation.

6 Conclusion

This paper has demonstrated two trends in games of IPD with reputation, and with a more fine-grained range of choices than merely full cooperation or full defection:

- Shorter game length in IPD does encourage mutual defection, and reputation mitigates this to a great degree, but not completely.
- More choices in cooperative level also encourages mutual defection, but again reputation mitigates this so that (for games of reasonable length) the degree of choice has no effect on the dominance of cooperation.

Both trends are important and relevant to the real world. More work is currently being done to explore various aspects of our simulation models.

ACKNOWLEDGMENTS — This work is partially supported by the Australia Research Council.

REFERENCES

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

[2] Robert M. Axelrod. The evolution of strategies in the iterated prisoner's dilemma. In *Genetic Algorithms and Simulated Annealing*, chapter 3, pages 32–41. Morgan Kaufmann, 1987.

[3] Steven J. Brams. *Superpower Games*. Yale University Press, 1985.

[4] David M. Chess. Simulating the evolution of behavior: the iterated prisoners' dilemma problem. *Complex Systems*, 2:663–670, 1988.

[5] Paul J. Darwen and Jordan B. Pollack. Co-evolutionary learning on noisy tasks. In *Congress on Evolutionary Computation*, Washington DC, July 1999. In press.

[6] Paul J. Darwen and Xin Yao. On evolving robust strategies for iterated prisoner's dilemma. In *Progress in Evolutionary Computation*, volume 956 of *Lecture Notes in Artificial Intelligence*, pages 276–292. Springer, 1995.

[7] David B. Fogel. Evolving behaviours in the iterated prisoner's dilemma. *Evolutionary Computation*, 1(1):77–97, 1993.

[8] Paul G. Harrald and David B. Fogel. Evolving continuous behaviors in the Iterated Prisoner's Dilemma. *Biosystems*, 37:135–145, 1996.

[9] W. Daniel Hillis. Co-evolving parasites improve simulated evolution as an optimization procedure. In *Artificial Life 2*, pages 313–323. Addison-Wesley, 1991.

[10] Hugues Juillé and Jordan B. Pollack. Co-evolving intertwined spirals. In *Fifth Annual Conference on Evolutionary Programming*, pages 461–468, 1996.

[11] Kristian Lindgren. Evolutionary phenomena in simple dynamics. In *Artificial Life 2*, pages 295–312. Addison-Wesley, 1991.

[12] Douglas Muzzio. *Watergate Games: strategies, choices, outcomes*. New York University Press, 1982.

[13] Martin A. Nowak and Karl Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393:573–577, 11 June 1998.

[14] Jordan B. Pollack and Alan D. Blair. Co-evolution in the successful learning of Backgammon strategy. *Machine Learning*, 32(3):225–240, 1998.

[15] Gilbert Roberts and Thomas N. Sherratt. Development of cooperative relationships through increasing investment. *Nature*, 394:175–179, 9 July 1998.

[16] Christopher D. Rosin and Richard K. Belew. Methods for competitive co-evolution: Finding opponents worth beating. In *Sixth International Conference on Genetic Algorithms*, pages 373–380. Morgan Kaufmann, 1995.

[17] Xin Yao and Paul J. Darwen. An experimental study of N-person iterated prisoner's dilemma games. *Informatica*, 18:435–450, 1994.