Evolution of Iterated Prisoner's Dilemma Strategies with Different History Lengths in Static and Cultural Environments

Richard Brunauer, Andreas Löcker, Helmut A. Mayer
Gerhard Mitterlechner, Hannes Payer
Department of Computer Sciences
University of Salzburg
Jakob–Haringer–Straße 2
A-5020 Salzburg, AUSTRIA
{rbrunau, aloecker, helmut, gmitter, hpayer}@cosy.sbg.ac.at

ABSTRACT

We investigate evolutionary approaches to generate well–performing strategies for the iterated prisoner's dilemma (IPD) with different history lengths in static and cultural environments. The length of the history determines the number of the most recent moves of both players taken into account for the current move decision. The static environment constituting the opponents of the evolved players is made up of ten standard strategies known from the literature. The cultural environment starts with the standard strategies and gradually increases by addition of the best evolved players representing a culture. The performance of the various evolved strategies is compared in specific tournaments. Also, the behavior of an evolved player is analyzed in more detail by looking at the specific game sequences (and corresponding decisions), which out of all possible sequences are actually utilized in a tournament.

Keywords

Iterated Prisoner's Dilemma, Evolutionary Computation, Cultural Algorithms

1. INTRODUCTION

The *Prisoner's Dilemma* (PD) is a *non–zero sum* game formulated by the mathematician Tucker building on the ideas of Flood and Dresher in 1950 [2]. Since then, it has been discussed extensively by game theorists, economists, mathematicians, political scientists, biologists, philosophers, ethicists, sociologists, and computer scientists. Many variations of the Prisoner's Dilemma have been devised, one of them being the *Iterated Prisoner's Dilemma* (IPD), which is at the center of attention in this paper.

The original PD is the *Two-Person PD*, which can be formulated as follows:

Anthony and Bertram are arrested because they are suspected of committing a bank robbery. The two suspects are charged with illegal possession of weapons. Each of the suspects is kept in an isolation cell in order to be questioned independently. As the investigator wants Anthony and Bertram to confess the crime, he imposes the following conditions on both of them:

"If you confess and your accomplice denies the crime, you will go free for your testimony, but your accomplice will get five years prison for bank robbery. If your accomplice confesses but you deny, it will be the other way round. If both of you confess, you will get only three years prison for the robbery, but if both of you deny, you will get one year prison for illegal possession of a firearm."

The goal of each prisoner is to minimize his or her time in prison. At first sight, from the view of the individual prisoner, it seems to be evident that he should confess the crime resulting in a maximum of three years prison. But when taking a closer look the dilemma gets revealed by the fact that denial of both prisoners results only in one year prison. Not only is confession of both worse than the denial of both, mutual confession even yields the worst outcome possible, when considering the total number of years in prison.

In general, the PD is a model for scenarios in which individual interests lead to worse individual results for each participant than following collective interests. In other words, in a PD local optimization leads to the worst possible outcome globally. As the gains of a prisoner are not balanced by the losses of the other prisoner, the PD is a non–zero sum game.

Commonly, the terms denial and confession are replaced by *Cooperation* (C) and *Defection* (D), respectively, as the prisoners are players receiving (positive) points according to a *Payoff Matrix* (Table 1) instead of (negative) years in prison.

SAC'07 March 11-15, 2007, Seoul, Korea

Table 1: The prisoner's dilemma payoff matrix.

	Cooperate	Defect
	3	5
Cooperate	3	0
	0	1
Defect	5	1

The four different game outcomes and their corresponding payoff are often labeled as *Sucker S* (CD, 0 points), *Punishment P* (DD, 1 point), *Reward R* (CC, 3 points), and *Temptation T* (DC, 5 points), with points awarded to the first player. The game constitutes a dilemma, if $T > R > P > S$.

In the original one–shot PD the players only make a single decision resulting in the corresponding payoff (playing one move). Playing a number of moves and summing up all payoffs constitutes the iterative version of the PD. Here, the goal of a player is to collect as many points as possible. In order to force the players to favor cooperation (R) rather than exploiting the opponent (T) and being exploited by the opponent (S), the following must hold: $2R > (T + S)$.

In order to collect a maximal number of points, players can follow a strategy based on the previous decisions of their opponent (e.g., echoing the opponent's last move is called *Tit–For–Tat* (TFT)). Of course, this implies that the player has access to the opponent's decision in the previous round. The player that finishes the game with more points is said to be the winner. We use the term *game* for playing a number of moves against a fixed opponent.

If there are more than two players, the overall winner can be determined in a tournament played in a round robin format yielding $n(n-1)/2$ games per round with n players (excluding games of a player against itself). The player with the highest total payoff collected in all games is the overall winner. One could think that the (globally) optimal solution of an IPD is a strategy beating each competitor. However, this is incorrect, as a strategy A winning all games in a tournament may accumulate less payoff than a strategy B having lost against A.

For theoretical investigations of the IPD (and games in general) some definitions have been established. A strategy is said to be *pure*, if the decisions of a player depend on past decisions of the opponent, e.g., relentlessly playing D in an IPD is not a pure strategy. A strategy X is an *Evolutionarily Stable Strategy* (ESS), if there is no strategy Y that can invade it [10]. This definition comes from the realm of population genetics, and refers to invasion as transferring a new strategy Y into a population of strategy X players causing the latter on average to receive less payoff than Y. It can be shown that no pure strategy is evolutionarily stable in the IPD [5], which in effect rules out the possibility to discover an unbeatable strategy generating larger payoff than "Always Defect", which is an ESS. However, it has been pointed out that the exact formal definitions of an ESS vary, and the number of invading strategies plays an important role [4].

Inevitably, the success of a strategy always depends on its opponents, which may explain why no formal criterion for the identification of an optimal strategy exists. Moreover, the dependence on the environment leads to profound doubts, if there exists a single globally optimal strategy at all. Naturally, these conditions makes the IPD a promising candidate for evolutionary approaches to generate well–performing strategies, but as we will observe in the following, again, the evolved strategies depend on their environment.

Axelrod (1984) turned the attraction of computer scientists to the IPD when generating strategies by means of *Evolutionary Computation* [2]. Since then, despite or because of its simplicity, the TFT strategy has been proven to be very efficient.

Golbeck (2002) analyzed the behavior of evolved populations with a constant history of three (the three last moves of both players are used to decide on the current move). She found that two important traits always emerge in the evolved populations, namely, the ability to defend against defectors, and to respect cooperators [6]. In this work we are not concerned with collective traits of the population, but in optimizing the performance of single strategies with different history lengths.

Hingston and Kendall (2004) compared evolved strategies to strategies acquired during play by learning. They observed that the learners were able to compete at the same level with evolved players [7].

2. EVOLUTION OF STRATEGIES

Our general concept of evolution of IPD strategies is based on pure strategies whose fitness is assessed by playing games against a given set of standard strategies (augmented by fixed evolved strategies in a specific experiment). As pure strategies use previous decisions of both players for the rules generating the actual move, a player has to have access to the *History* of moves.

2.1 Deterministic Strategies

We call a strategy *deterministic*, if identical move sequences always generate the same move. A history of one yields $2^2 = 4$ different game situations, as given by the binary decisions (C or D) of the players. This can be easily generalized to $2^{2n} = 4^n$ distinct game sequences for a history of n.

2.1.1 Genotype Encoding

As we want to assign a move to each game situation, we simply interpret the locus (index) of each base (decision) of the binary chromosome as its corresponding game situation. E.g., for a strategy with a history of six there are $4^6 = 2^{12} =$ 4096 different sequences. Hence, the chromosome length is 4096, and the 12–bit history (read as a positive integer) refers to the locus, while the corresponding base encodes the decision ($0 = C, 1 = D$). Note that with a history of six evolution may choose out of 2^{4096} strategies.

With a history of four, the history `00100111` means: for the last move both, player X, and player Y played C; for the last but one move X chose D, and Y said C; for the last but second move X said C, and Y chose D, and four moves ago both opted for D.

Converting the binary history into a decimal index yields 36, which is the position of the bit indicating the decision of the player. If the number of moves played is smaller than the length of the history, the decision is generated by assuming all C in the missing history. Though, this assumption only affects the first moves of a game, the missing history could be evolved additionally. E.g., in a history of three with only one move played the history may be `01`, which in our approach is extended to `010000`.

If we replace the deterministic decision (single base) at each locus of the genotype by a probability (to play C), we can easily evolve non–deterministic strategies. The length of the chromosome increases to $l = k * 4^n$, where k is the number of bits used to encode a real–valued number in the unit interval. We did not investigate these strategies in this work, however, preliminary results have been encouraging.

2.2 Evolution in a Static Environment

In this scenario we evolve deterministic strategies playing against a fixed number of known strategies described in the following section. We run a set of experiments evolving strategies with histories ranging from one to six. For each history length a number of evolutionary runs is made resulting in different winning strategies emerging in each run. The best of these strategies is selected to be the final result generated by evolution, which in total gives six evolved players.

2.2.1 Standard Strategies

For the evolution experiments we used the following standard strategies from the IPDLX package (Section 3.1.2), and the additional strategies *Gradual* [3] and *Forgiving* [8].

- ALLC Unconditionally plays C with every move of a game.
- ALLD Analogous to ALLC playing D.
- RAND Plays randomly C or D with a probability of 0.5.
- GRIM *Grim Trigger* starts by playing ALLC, but after the first D of the opponent it switches to ALLD.
- TFT *Tit–For–Tat* always plays the opponent's last move. In the first round it plays C.
- STFT *Suspicious–TFT* is a TFT variant playing D as the opening move.
- TFTT *Tit–For–Two–Tats* unlike TFT forgives a single defection of the opponent, but plays D, if the the opponent has played two Ds in a row.
- Pavlov Repeats its move, if the payoff was T or R. Otherwise (P and S payoff) it plays the opposite move.
- Gradual Answers a C always with C. Gradual responds to Ds of opponent with a fixed sequence: after the first D Gradual plays DCC, the second D (overall) is answered by DDCC, hence, in general the n–th overall opponent D evokes n Ds and two Cs.
- Forgiving Plays always C, if opponent did so. If opponent plays D, then it plays one more C (two consecutive Cs) in the next move (it forgives). If opponent continues with D, Forgiving also plays D. After five consecutive mutual Ds, Forgiving makes a new try and plays C, again.

2.3 Evolution in a Cultural Environment

The second approach taken in the evolution of IPD strategies evolves players not only against the standard strategies, but also evolved strategies found in previous runs. With this technique, which may be viewed as a form of *Coevolution*, a higher degree of generalization could be achieved in the evolved players, as they do not always face the same set of strategies. As the additional evolved strategies resemble a *Culture* accumulating the knowledge of previous evolution runs (or ancestors), this approach, inspired by the concept of *Cultural Algorithms* [9], may be also termed *Cultural Coevolution*.

In more detail, this set of experiments starts using the standard strategies augmented by the best history 1 strategy generated in the static environment (Section 2.2). Then, the best history 2 strategy is added to the pool for evolution of history 3 strategies. This scheme is repeated until a history length of 6 is reached, hence, the result of cultural coevolution are five strategies (with histories from 2 to 6).

3. EXPERIMENTAL SETUP

In this section we describe details of the software implementation and specific parameter values used in the evolution experiments.

3.1 Software Pieces

The technical base for our application evolving and testing strategies are the two Java frameworks *JEvolution* supporting artificial evolution, and the *eXtended IPD Library* (IPDLX) [1] managing IPD tournaments.

3.1.1 JEvolution

JEvolution is a lean and compact Java framework for *Evolutionary Algorithms* supporting standard EA components, e.g., different genotype encodings, common mutation and recombination operators, and an interface for problem–specific code, i.e., the fitness evaluation.

3.1.2 IPDLX

This framework provides the infrastructure to conduct IPD tournaments. It supports standard strategies, a single player environment, a multi–player environment, and a tournament environment. Also, an interface for the implementation of new strategies is offered. Note that each player has access to its opponent's decision of the previous move, hence it is able to collect the full history of moves.

3.1.3 The IPD Application

Basically, IPDLX enables the fitness evaluation of the individuals (strategies) generated in JEvolution. Each phenotype plays in a tournament and is assigned a fitness identical to the total payoff earned in the game.

To avoid superfluent matches in the evolution process, we added a specific tournament mode. It only conducts games of the evolved strategy against all available opponents, as all other games do not contribute to the fitness. Evidently, when finally assessing the performance of the best evolved strategy after a run, we have to play a (single) full tournament. Eliminating the redundant games a single tournament is of linear instead of quadratic complexity speeding up evolution considerably.

3.2 Evolution Parameters

The JEvolution framework has been parameterized with the following values for the evolution of IPD strategies:

The chromosome length $l = 4^h + 1$ with $h >= 1$, where h is the history length and the extra–bit encodes the first round move. The bit mutation rate $p_m = \frac{1}{l}$ with $h \leq 4$ and $\frac{3}{l}$ with $h > 4$. The number of generations $g = 500$ ($h = 1$), $2,000$ ($h = 2$), $3,500$ ($h = 3$), $5,000$ ($h = 4$), $7,500$

[1] http://www.prisoners-dilemma.com/java/ipdlx/ipdlx_javadocs

($h = 5$), and $10,000$ ($h = 6$). The population size $n = 50$. 2–point crossover with a crossover rate $p_c = 0.6$ is used for recombination. Selection is realized by binary tournament selection without replacement. The fitness of an individual (strategy) is the average score per move in a tournament.

A tournament consists of 20 games against each player (Section 3.1.3) with 200 moves per game with a payoff matrix according to Table 1. We repeated each evolution run 20 times to obtain more reliable results for the analysis of evolved strategies.

4. EXPERIMENTAL RESULTS

In this section we present the results evolving strategies with different histories in the static (Section 2.2) and the cultural environment (Section 2.3) along with outcomes of specific tournaments and a deeper analysis of evolved strategies.

4.1 Static Environment

In Figure 1 the mean of the individuals with best fitness per generation is shown for various history lengths.

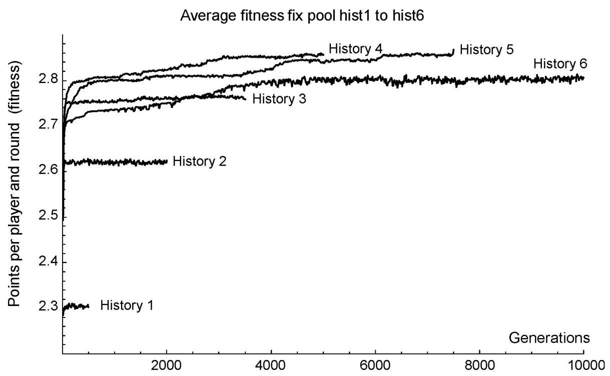

Figure 1: Evolution of best fitness of strategies with histories from one to six (averaged on 20 runs).

As can be seen the best strategies with history 5 collect the most points, followed by history 4 and history 3. We expected history 6 to end up in first place, as each strategy with history h may emulate the strategy of a $h-1$–player with the benefit of the additional information on another move. We could speculate that the number of generations ($10,000$) for history 6 evolution is too small considering the huge search space. Looking at Figure 1 it seems that history 6 evolution is stagnating, however, with a fitness jump similar to that at the end of history 5 evolution things could have changed.

The average score per move averaged on the 20 best individuals (one of each run) from history 1 to history 6 (Figure 2) is 2.6893, 2.8304, 2.8986, 2.9470, 2.9522, and 2.8728. Thus, the scores are slightly below 3.0, the score for permanent mutual cooperation.

The single best individuals (best of 20 runs for each history) scored 2.6950, 2.8380, 2.9480, 3.0530, 3.0660, and 3.0610. Again, history 5 is slightly above history 6, which makes the former the overall winner amongst evolved strategies in the static environment. The evolution of these best individuals is depicted in Figure 3.

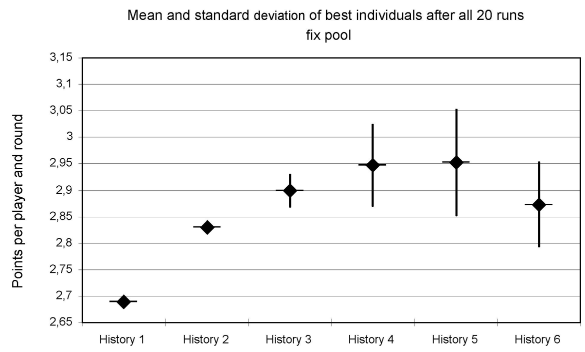

Figure 2: Mean score and standard deviation of the best individuals per history (averaged on 20 runs).

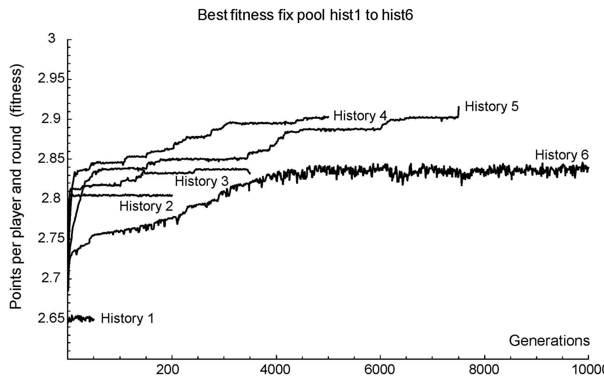

Figure 3: Evolution of the best individual (out of 20 runs) for each history.

4.2 Cultural Environment

In Figure 4 the mean of the best fitness during cultural evolution of strategies with histories from 2 to 6 is shown. Recall that with cultural evolution, the best strategy with history h found in all evolution runs is added to the opponents for evolution of $h+1$–players. Hence, we start with history 2, as the best player with history 1 can be taken from the static environment results. Consequently, the opponents during evolution of an $H6$–player (history 6) are the standard strategies (Section 2.2.1) and the five best evolved players with lower histories.

The mean fitness of the best individuals (Figure 5) from history 2 to history 6 are 2.8462, 2.9953, 2.9823, 3.0822, 3.1199, respectively.

For each history the fitness is above the values generated in the static environment. However, the fitness values cannot be directly compared as they were obtained in different environments. Also, the fact that histories 5 and 6 exceed the cooperation level of 3 points could mostly stem from the more complex evolved opponents, which allows evolution of more sophisticated strategies. The latter could be the reason for the constant improvement of the best $H6$–players in Figure 4, as they might elicit and exploit specific behaviors of the other evolved players, which they do not (have to) show against standard strategies. With a small exception (history 3 to 4) longer histories gain more pay–off.

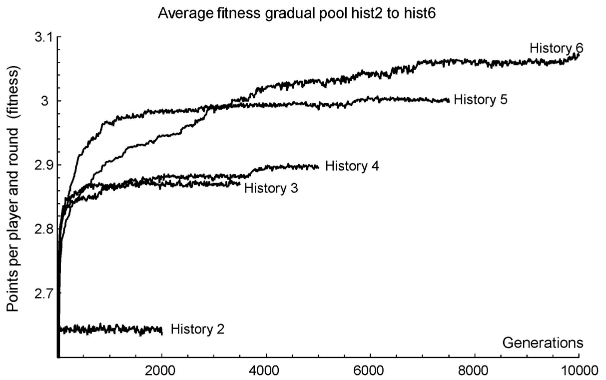

Figure 4: Evolution of best fitness of strategies with histories from two to six (averaged on 20 runs).

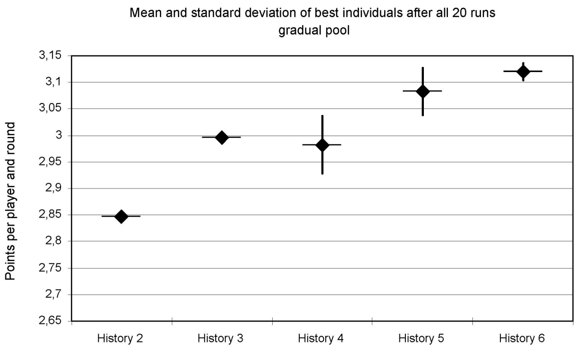

Figure 5: Mean score and standard deviation of the best individuals per history (averaged on 20 runs).

When looking at the evolution of the best individuals per history length (Figure 6) we observe fitness values of 2.8527, 3.0008, 3.1012, 3.2300, and 3.1423. Though, the best $H6$–players exhibit on average the highest fitness, the overall best individual is an $H5$–player, as is the case in the static environment.

4.3 Tournament with Best Static Strategies

In order to test the generalization capabilities of the evolved strategies in the static environment, we conducted round robin tournaments with the standard strategies and each best individual of histories from 1–6 (Section 4.1). For reliable results the average point scores given in Table 2 (and all following tournaments) are based on 1,000 rounds with 200 moves per game.

The clear winners in this tournament are Gradual and TFT. Note that the evolved strategy StaticH1 is identical to TFT, which could be expected, as TFT is known to be a high–performing strategy. We may conclude that the evolved strategies, which have been exposed to the standard strategies only during evolution, are taking away from each other at the benefit of the winning standard strategies. However, at this point we do not even know, if single evolved strategies are able to beat the standard strategies, as we "only" maximized the average score of the evolved strategy without playing a complete tournament (Section 3.1.3). Thus, we conducted tournaments with each single best static

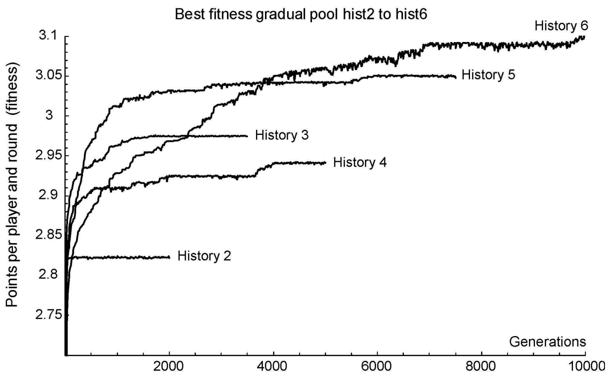

Figure 6: Evolution of the best individual (out of 20 runs) for each history.

Table 2: Tournament with standard strategies and best static strategies with different histories.

Rank	Score	Strategy
1	2.81	Gradual
2	2.74	TFT
3	2.74	StaticH1 (=TFT)
4	2.55	STFT
5	2.475	StaticH5
6	2.47	StaticH6
7	2.405	StaticH3
8	2.345	TFTT
9	2.335	StaticH4
10	2.33	StaticH2
11	2.21	Forgiving
12	2.205	GRIM
13	1.985	Pavlov
14	1.86	ALLC
15	1.68	RAND
16	1.67	ALLD

strategy. All evolved strategies with histories larger than 1 win their respective tournaments convincingly. Roughly, larger histories gain more payoff with the exception of history 6. The latter may be due to the huge search space, but could be also a consequence of the specific opponents. Table 3 shows the tournament with StaticH4 collecting the highest payoff of all static players.

4.4 Tournament with Best Culture Strategies

Next, all the best evolved culture strategies (histories 2–6) play a tournament with all standard strategies, whose outcome is presented in Table 4.

Expectedly, the culture strategies exhibit improved performance with increasing history length with the top strategies CultureH5 and CultureH6 exceeding the three–point limit. Though, H5 evolved to a better fitness than H6 (Section 2.3), it is not surprising that H6 won the tournament, as it has faced H5 during evolution (but not vice versa). CultureH4 and CultureH3 are beaten by Gradual and TFT, because H5 and H6 have learned to exploit the former. Without the latter two strategies H4 and H3 win the tournament easily (Table 5).

This might lead to the conclusion that the culture strategies are unstable and rely on each other for good perfor-

Table 3: Tournament with standard strategies and the best static strategy with history 4.

Rank	Score	Strategy
1	3.01	StaticH4
2	2.735	Gradual
3	2.62	TFT
4	2.46	TFTT
5	2.44	Forgiving
6	2.395	GRIM
7	2.38	STFT
8	2.24	Pavlov
9	2.095	ALLC
10	2.025	RAND
11	1.995	ALLD

Table 4: Tournament with standard strategies and best culture strategies with different histories.

Rank	Score	Strategy
1	3.095	CultureH6
2	3.045	CultureH5
3	2.81	Gradual
4	2.73	TFT
5	2.645	CultureH4
6	2.63	CultureH3
7	2.555	STFT
8	2.525	CultureH2
9	2.395	TFTT
10	2.36	Forgiving
11	2.29	GRIM
12	2.245	ALLC
13	2.115	Pavlov
14	1.72	ALLD
15	1.64	RAND

mance against the standard strategies. However, in tournaments, where each evolved culture strategy is left on its own, all of them outperform the standard strategies with rather similar payoffs around 2.8. As an example the results of the tournament with the best performing CultureH3 are shown in Table 6.

4.5 All Star Tournament

In Table 7 the results of a tournament with all standard and the best evolved strategies (StaticH1–H6 and CultureH2–H5) are shown.

The standard strategies Gradual and TFT are at the top of the field. Next come the culture strategies fairly ordered with respect to decreasing history length reflecting the cultural development. The weakest culture strategy CultureH2 beats all of the static strategies except StaticH2 (and StaticH1 being TFT). More surprisingly, the static strategies with greater history perform worse than those with shorter history lengths. This may likely be attributed to the large variety of strategies available with greater history. Hence, facing a culture strategy with possibly complex play the static strategies experience game sequences (with their corresponding decisions), which have never been activated during evolution.

The evolved strategy CultureH5 with the best overall fitness (Section 2.3) came up at sixth place beaten by CultureH6 and StaticH1 (=TFT). Obviously, the evolved strategies take away from each other, which enables Gradual and TFT to win the tournament, although, with the exception of StaticH1 (=TFT) *each* evolved strategy playing in a tournament on its own beats the standard strategies.

Table 5: Tournament with standard strategies and best culture strategies with histories 2–4.

Rank	Score	Strategy
1	2.975	CultureH4
2	2.815	CultureH3
3	2.695	Gradual
4	2.68	TFT
5	2.575	CultureH2
6	2.485	STFT
7	2.48	Forgiving
8	2.42	TFTT
9	2.335	GRIM
10	2.265	Pavlov
11	2.245	ALLC
12	1.835	ALLD
13	1.79	RAND

Table 6: Tournament with standard strategies and best culture strategy with history 3.

Rank	Score	Strategy
1	2.85	CultureH3
2	2.735	Gradual
3	2.62	TFT
4	2.495	Forgiving
5	2.485	Pavlov
6	2.46	TFTT
7	2.4	GRIM
8	2.38	STFT
9	2.25	ALLC
10	2.025	RAND
11	1.995	ALLD

4.6 Coverage Analysis

In order to get a deeper insight into evolved strategies, we looked at the number and distribution of the specific game sequences (and according moves) activated during a tournament, i.e., the genes actually used (or covered). In the exemplary tournament with our StaticH6 shown in Table 8 (1 round, 200 moves per game) we found that approximately four percent (161 bases) of all bases have been covered.

50 percent of these game sequences are played more than once. The large number of sequences occurring only a few times is the trace of the Random player. Less than seven percent of the covered sequences are played more than ten times. The distribution among these sequences is depicted in Figure 7.

Naturally, the largest portion is occupied by the sequence with 0s (cooperate) only, as mutual cooperation emerges frequently(e.g., against ALLC or TFT). The other sequences reflect strategies of StaticH6 against individual players (recall that the most recent decision of the evolved player is the left–most bit). E.g., the sixth and seventh history (both with 5.37%) fully exploit the TTFT player, which always forgives a single defect, hence, StaticH6 plays C and D in

Table 7: Tournament with standard, best static, and best culture strategies.

Rank	Score	Strategy
1	2.82	Gradual
2	2.805	TFT
3	2.805	StaticH1 (=TFT)
4	2.69	CultureH6
5	2.66	STFT
6	2.545	CultureH5
7	2.465	CultureH3
8	2.305	CultureH4
9	2.29	StaticH2
10	2.285	TFTT
11	2.275	CultureH2
12	2.23	StaticH4
13	2.175	Forgiving
14	2.125	StaticH5
15	2.11	GRIM
16	2.06	StaticH3
17	2.055	StaticH6
18	1.925	ALLC
19	1.87	Pavlov
20	1.51	ALLD
21	1.43	RAND

Table 8: Tournament results and coverage of StaticH6.

Rank	Score	Strategy
1	2.9975	StaticH6
2	2.574	TFT
3	2.4925	Gradual
4	2.394	TFTT
5	2.3785	STFT
6	2.2635	ALLC
7	2.2215	Pavlov
8	2.2095	Forgiving
9	2.189	GRIM
10	2.06	RAND
11	2.006	ALLD

Coverage for StaticH6: 161 bits

strict alternation.

In the following we take a look at games of StaticH6 against the standard strategies ALLC, ALLD, Gradual, and TFT. Table 9 gives the details of a game against ALLC.

Recall that a specific base in the genotype encodes the first move of the evolved player. Our StaticH6 always plays D as its first move enabling a maximal score of five points per move against ALLC, as it answers all of the six covered sequences with D.

In Table 10 the evolved player plays against ALLD.

It can be seen that StaticH6 also adheres to a strict ALLD strategy with the exception of the second move, where the evolved player offers a C, hereby losing a point.

A game against Gradual is analyzed in Table 11.

Here, within two moves a mutual cooperation is established giving the edge to StaticH6 due to the first move D.

Finally, the evolved player faces TFT in Table 12.

Against TFT, the evolved player again benefits from the opening move D, but receives the sucker payoff when in transition to mutual cooperation offers a C after two mutual Ds. Consequently, the game ends in even point scores. As can be seen from these examples, the evolved player is rapidly adjusting to different standard strategies, and yields (nearly) maximal payoff. Based on an interesting technique presented in [1] we hope to be able to categorize the overall behavior of the evolved players in terms of similarity to standard strategies in future work.

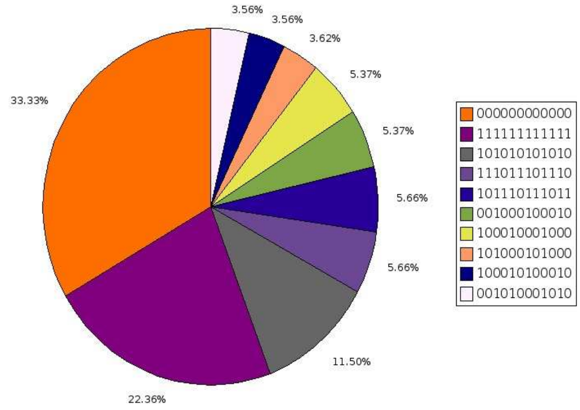

Figure 7: Distribution of game sequences played by StaticH6 (Table 8).

Table 9: Coverage of StaticH6 vs. ALLC.

Rank	Score	Strategy
1	5.0	StaticH6
2	0.0	ALLC

Coverage for StaticH6: 6 bits

```
H: 100000000000 Decision: 1 Moves: 1
H: 101000000000 Decision: 1 Moves: 1
H: 101010000000 Decision: 1 Moves: 1
H: 101010100000 Decision: 1 Moves: 1
H: 101010101000 Decision: 1 Moves: 1
H: 101010101010 Decision: 1 Moves: 193
```

5. SUMMARY AND CONCLUSIONS

We have presented evolutionary approaches to generate well–performing strategies for the *Iterated Prisoner's Dilemma* (IPD) considering previous decisions of both, the opponent, and the evolved player, up to a history length of six. We also compared evolution in a static environment, where evolved strategies with a specific history always play tournaments against the same standard strategies, to evolution in a cultural environment. In the latter, when evolving a strategy with history n, the single best evolved players from history 1 to $n-1$ are additional opponents in the tournament. Though, the single best static players (with the exception of StaticH1 (=TFT)) win each tournament with the standard strategies, the picture changes when all static players are in the tournament. Then, the standard strategies Gradual, TFT, and STFT are in front of the static players, which is the result of specialization of the evolved static players to the standard strategies. As soon, as they face their "colleagues", whom they did not face during evolution, they take away from each other to the benefit of the standard

Table 10: Coverage of StaticH6 vs. ALLD.

```
Rank    Score   Strategy
1       1.02    ALLD
2       0.995   StaticH6
Coverage for StaticH6: 8 bits

H: 110000000000 Decision: 0 Moves: 1
H: 011100000000 Decision: 1 Moves: 1
H: 110111000000 Decision: 1 Moves: 1
H: 111101110000 Decision: 1 Moves: 1
H: 111111011100 Decision: 1 Moves: 1
H: 111111110111 Decision: 1 Moves: 1
H: 111111111101 Decision: 1 Moves: 1
H: 111111111111 Decision: 1 Moves: 191
```

Table 11: Coverage of StaticH6 vs. Gradual.

```
Rank    Score   Strategy
1       3.01    StaticH6
2       2.96    Gradual
Coverage for StaticH6: 9 bits

H: 100000000000 Decision: 1 Moves: 1
H: 111000000000 Decision: 1 Moves: 1
H: 101110000000 Decision: 0 Moves: 1
H: 001011100000 Decision: 0 Moves: 1
H: 000010111000 Decision: 0 Moves: 1
H: 000000101110 Decision: 0 Moves: 1
H: 000000001011 Decision: 0 Moves: 1
H: 000000000010 Decision: 0 Moves: 1
H: 000000000000 Decision: 0 Moves: 190
```

strategies.

This specialization is avoided when evolving in the cultural environment, where strategies are evolved also in the presence of previously evolved players. Not surprisingly, two evolved players (CultureH6 and CultureH5) win the tournament with all evolved cultural and the standard strategies. However, all single best culture players also win the tournament without their "colleagues" against the standard strategies, which is a strong indication for the flexibility of these strategies.

With the static players there is a clear trend that a greater history length produces higher payoff, when single evolved players are in the tournament. This is not the case for the culture players, and in tournaments with all best evolved culture players the fact that the longest histories win may be mainly attributed to the evolution procedure (the longest history facing all others during evolution).

We have also demonstrated that the problem of the large search space induced by longer histories is alleviated by the small coverage of game sequences. In a specific tournament with StaticH6 only about four percent of all bases (each representing a game sequence) have been activated. However, the "unused" bases may be activated in games against players never seen during evolution, most likely leading to random behavior of the evolved player in this situation. Consequently, longer histories should be evolved with a large number of opponents so as to sample the search space more thoroughly.

In a tournament with all evolved and standard strategies (Table 7) won by Gradual and TFT, the culture players clearly outperform the static players, which puts cultural evolution in the focus of continuing work.

Table 12: Coverage of StaticH6 vs. TFT.

```
Rank    Score   Strategy
1       2.975   StaticH6
2       2.975   TFT
Coverage for StaticH6: 10 bits

H: 100000000000 Decision: 1 Moves: 1
H: 111000000000 Decision: 1 Moves: 1
H: 111110000000 Decision: 0 Moves: 1
H: 011111100000 Decision: 0 Moves: 1
H: 000111111000 Decision: 0 Moves: 1
H: 000001111110 Decision: 0 Moves: 1
H: 000000011111 Decision: 0 Moves: 1
H: 000000000111 Decision: 0 Moves: 1
H: 000000000001 Decision: 0 Moves: 1
H: 000000000000 Decision: 0 Moves: 189
```

6. REFERENCES

[1] D. Ashlock and E.-Y. Kim. Techniques for Analysis of Evolved Prisoner's Dilemma Strategies with Fingerprints. In *Proceedings of the Congress on Evolutionary Computation 2005*, September 2005.

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

[3] B. Beaufils, J.-P. Delahaye, and P. Mathieu. Our Meeting With Gradual: A Good Strategy for the Classical Iterated Prisoner's Dilemma. In *Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems 1996*, Cambridge, MA, May 1996. MIT Press.

[4] J. Bendor and P. Swistak. The Evolutionary Stability of Cooperations. *American Political Science Review*, 91(2):290–307, June 1997.

[5] R. Boyd and J. Lorberbaum. No Pure Strategy is Evolutionarily Stable in the repeated Prisoner's Dilemma Game. *Nature*, 327:58–59, May 1987.

[6] J. Golbeck. Evolving Strategies for the Prisoner's Dilemma. In *Advances in Intelligent Systems, Fuzzy Systems, and Evolutionary Computation 2002*, pages 299–306, February 2002.

[7] P. Hingston and G. Kendall. Learning versus Evolution in Iterated Prisoner's Dilemma. In *Proceedings of the Congress on Evolutionary Computation 2004*, pages 364–372, June 2004.

[8] C. O'Riordan. A forgiving strategy for the Iterated Prisoners Dilemma. *Journal of Artificial Societies and Social Simulation*, 3(4), October 2000. (`http://jasss.soc.surrey.ac.uk/3/4/3.html`).

[9] R. G. Reynolds and W. Sverdlik. Problem Solving Using Cultural Algorithms. In *Proceedings of the First IEEE Conference on Evolutionary Computation*, pages 645–650. IEEE, 1994.

[10] M. J. Smith. *Evolution and the Theory of Games*. Cambridge University Press, UK, 1982.