It pays to be fair

Volker Schatz, June 2008
http://www.volkerschatz.com/science/nonpapers/

This work compares two strategies entered into the 2005 Iterated Prisoner's Dilemma Competition. Both strategies select their moves based on past experience of their opponent's. They differ in their objective: One of them aims at maximum gain. The other tries to mirror its opponent's move as closely as possible. The latter strategy is fair in that it forgoes taking advantage of the opponent's cooperation. Taking into account the Competition's scoring system, the gain from mutual defection does not outweigh the loss resulting from mutual cooperation, compared to choosing the opposite move of the opponent's. Nonetheless, the latter strategy performed significantly better than the former. This confirms that altruistic strategies tend to do better than strictly selfish ones in a one-to-one comparison with other things being equal.

Keywords: **Iterated prisoner's dilemma, fairness, machine learning, game theory**

I. Introduction

The Prisoner's Dilemma is a two-player non-zero sum game which has been extensively researched in philosophy, the social sciences and game theory.[Axe84, Pou92] It is about the situation in which two accomplices to a crime have been caught and are held separately. Either can reduce their own likely prison sentence by incriminating the other accomplice. However, if both try to inculpate each other, this reduction will be minimal; and if both accomplices stay silent, their combined sentences are shorter than in any other case.

The Prisoner's Dilemma game has been formalised by assigning a numerical value to the comparative benefit a player attains, which is called the payoff. When a player incriminates his accomplice, this is commonly called the DEFECT move, a player's supporting the other COOPERATE.

A single-round game of Prisoner's Dilemma is trivial in that DEFECT is a strictly dominant strategy. Playing a considerable number of rounds in sequence, however, allows players to punish their opponent's DEFECT in previous rounds. This is called the Iterated Prisoner's Dilemma game. This is a simple but useful model for human behaviour in a social environment and has been used to investigate topics as diverse as international arms races,[Kir98, SSS00] law enforcement[Wil88, BGBS07] and doping.[Bre87, BW97]

Quite some years after the original framing of the Prisoner's Dilemma game, interest has turned to the relative success of different strategies in Iterated Prisoner's Dilemma. Robert Axelrod organised two computer tournaments of Iterated Prisoner's Dilemma .[Axe80a, Axe80b] Strategies programmed in FORTRAN or BASIC by both academics and non-academics were pitted against themselves, every different strategy and the RANDOM strategy. The tournaments yielded the surprising result that it was not the most selfish strategies which did well, but that the readiness to cooperate was a prerequisite for success. The winner of both tournaments was the TIT FOR TAT strategy, which plays its opponent's move of the previous round.

In 2004 and 2005, the tournament of Axelrod has been repeated, with slight modifications, by Kendall, Darwen and Yao in the Iterated Prisoner's Dilemma Competition (IPDC).[KDY, KYC07] The opportunity of the 2005 competition was taken to compare two strategies which use the same method towards different aims. One tries to behave fairly, while the other is completely selfish. The following section describes the strategies in detail. The third section describes the competition and the results achieved by the two strategies in question. The last two sections discuss these results and draw the conclusions.

II. Strategies

A. General

Two strategies were entered into the Iterated Prisoner's Dilemma Competition. Both learn about their opponent's behaviour over the course of several rounds. They differ in how they use this information. The MIRROR strategy tries to play the same move as its opponent. This is a fair way of behaving: It rewards a cooperative opponent while denying a treacherous one its prize. The other strategy, AMBITIOUS, seeks to maximise its own payoff.

The two strategies share a common structure. Both build a record of past rounds in the form of an "experience table". Entries in the experience table reflect how the strategy should have played, with hindsight, to achieve its objective in past rounds. The experience values are dependent on moves played in previous rounds, and are inferred from similar series of previous moves if the situation at hand has not been encountered before. The creation and updating of the experience table is presented in detail in the following section.

Given the experience table, both strategies act as follows. The history of moves both by the opponent and the strategy itself in the last H rounds is recorded. Before deciding which move to play in the current round, the experience table is updated depending on the opponent's move of the previous round, which is now known. Then the element corresponding to the current move history is read from the experience table. A uniformly distributed random number is generated and compared with the experience table entry. Depending on the comparison, either COOPERATE or DEFECT is played, so that the move which has (or would have) proved better in the past has the higher probability.

This approach is designed to predict the moves of an opponent which acts deterministically depending on moves in previous rounds. The random element in the decision serves to avoid getting stuck with a bad strategy when facing an opponent which changes its behaviour or acts (partly) randomly. One modification to this algorithm was made: Both strategies play COOPERATE until their opponent has played DEFECT once. This allows the strategies to cooperate well with themselves, each other or any strategy which cooperates until cheated. Experience is already collected during this initial cooperation phase.

B. The experience table

The experience table utilised by both strategies is a look-up table mapping the move history $\boldsymbol{h} \in \{0,1\}^{2H}$ to a value $e(\boldsymbol{h}) \in [0,1]$ indicating the suitability of a current move. The move history $\boldsymbol{h}$ contains both players' moves in the previous H rounds. The DEFECT move is represented by 0, COOPERATE by 1. The current move is decided by comparing a uniformly distributed random number $r \in [0,1]$ with $e(\boldsymbol{h})$. If $r < e(\boldsymbol{h})$, COOPERATE is played, otherwise DEFECT.

The different aims of the two strategies are implemented by different definitions of e_{new}, the new experience to be entered into the table. For the MIRROR strategy, e_{new} is the opponent's last move coded as 0 for DEFECT or 1 for COOPERATE. For the AMBITIOUS strategy, the new experience value depends on the payoff achieved and on which move the strategy itself played in the last move:

$$e_{\text{new, AMB}} = \begin{cases} \tilde{P} & \text{if COOPERATE played} \\ 1 - \tilde{P} & \text{if DEFECT played,} \end{cases} \tag{1}$$

$$\text{where } \tilde{P} = \frac{P - P_{\text{min}}}{P_{\text{max}} - P_{\text{min}}}$$

The payoff P results from the IPDC's payoff matrix presented in Table 1. For this payoff table, $P_{\text{min}} = 0$ and $P_{\text{max}} = 5$.

e_{new} is entered into the experience table before the following move is taken. The table entry for which it is most relevant is that corresponding to the history $\boldsymbol{h}_{\text{prev}}$ of the previous move, which contains both players' moves in the last but one to the last but H rounds. However, since all 2^{2H} configurations in the table are unlikely to be learned from experience in the course of one game, the new experience is also made to affect those entries with a similar move history. A parameter $\tau \in [0,1]$ determines how much this influence is reduced for every differing move in the history — the contribution to a move history differing in n moves is weighted by τ^n. In addition, previous experience is not discarded, but averaged with the new experience value with weight $\varphi \in [0,1]$. Formally the update of the experience table can be written as follows:

$$e(\boldsymbol{h}) \leftarrow w(\boldsymbol{h})\, e_{\text{new}} + (1 - w(\boldsymbol{h}))\, e(\boldsymbol{h}), \tag{2}$$

$$\text{where } w(\boldsymbol{h}) = (1-\varphi)\, \tau^{\|\boldsymbol{h}-\boldsymbol{h}_{\text{prev}}\|_1}$$

$$\text{and } \|\boldsymbol{v}\|_1 = \sum_i |v_i| \quad \text{(the Manhattan length)}$$

This formula was applied to all $\boldsymbol{h} \in \{0,1\}^{2H}$ which had at least one vector component in common with $\boldsymbol{h}_{\text{prev}}$.

At the start of each game, all experience table entries are set to 0.5. This allows the strategies to start by experimenting without prejudice (apart from the policy of not being the first to defect, see previous section). The current move history is initialised to a history of continuous mutual cooperation.

Own move	Opponent's move	
	COOPERATE	DEFECT
COOPERATE	3	0
DEFECT	5	1

Table 1. The payoff matrix used in the Iterated Prisoner's Dilemma Competition. The table entries are the payoffs of the player whose move is given in the first column.

C. Choice of parameters

The algorithm for collecting experience described above has a number of parameter which should be tuned to achieve optimal results: H, the number of past rounds taken into account, the "transfer parameter" $\tau \in [0,1]$ which determines how much experience spills over to similar move histories, and the "fading parameter" $\varphi \in [0,1]$, the weight for retaining previous experience.

The number of past rounds entering into the history is subject to a tradeoff: Ideally, it should not be smaller than any deterministic opponent's memory. On the other hand, too long a memory slows down learning (τ and φ being equal), as the cumulative weight $W(H)$ of e_{new} will be smaller in relation to the total number of table entries 2^{2H}:

$$\begin{aligned} W(H) \,/\, 2^{2H} &= 2^{-2H}\,(1-\varphi) \sum_{n=0}^{H-1} \binom{H}{n} \tau^n \\ &= 2^{-2H}\,(1-\varphi) \left[(1+\tau)^H - \tau^H\right] \qquad (3) \\ &< (1-\varphi)\, 2^{-H} \quad \text{because } \tau \le 1 \end{aligned}$$

So an overly long memory has the two drawbacks that the experience gained in the course of one game may not suffice to fill the experience table with significant entries, and that the risk of lock-in is increased.

The parameter values used for the competition were not determined by a formal method. Rather, a number of test matches were played which pitted the strategies MIRROR and AMBITIOUS against standard strategies such as ALWAYS COOPERATE, ALWAYS DEFECT, RANDOM and TIT FOR TAT, as well as against themselves and each other. The history length H was also required to be small compared to the number of rounds to be played in each match in the tournament, which was given as averaging 200 .[KDY] On this basis, the parameters were chosen as $H = 5$, $\tau = 0.5$ and $\varphi = 0.4$.

III. COMPETITION RESULTS

A. The competitions

The Iterated Prisoner's Dilemma Competition comprised four separate competitions with slightly different rules. The first competition was a straightforward game of iterated prisoner's dilemma as explained in the Introduction. In the second competition, noise was introduced into a player's choice of move, simulating a misinterpretation of the

Strategy	Average score per game			Normalised Rank		
	Comp. 1	Comp. 2	Average	Comp. 1	Comp. 2	Average
MIRROR	418.6	369.3	393.9	0.339	0.491	0.415
AMBITIOUS	405.1	338.2	371.7	0.375	0.939	0.657

Table 2. Results of the MIRROR and AMBITIOUS strategies in the Iterated Prisoner's Dilemma Competitions 1 and 2. The scores and ranks were averaged over all five runs of both competitions, the number of won minus lost games was summed up. The normalised rank was obtained by dividing the strategies' rank according to its score by the number of strategies taking part. (As usual, a lower rank is better.) The MIRROR strategy consistently outperforms AMBITIOUS.

move decision. The third competition was a multi-player competition. In competition 4, every participant could submit only one strategy in order to prevent collusion between strategies.

The strategies MIRROR and AMBITIOUS were entered into competitions 1 and 2. The payoff matrix for both of them was the one given in Table 1. Each competition consisted of five runs with exactly 200 rounds each.[1] In each round, every participating strategy was pitted against every other, itself and the RANDOM strategy, which was added by the competition organisers.

B. Results

Table 2 shows the results achieved by our two strategies. The balance of won minus lost games was computed by adding up these entries in the competitions' results table which can be downloaded from the IPDC's web site .[KDY] The score per game was obtained by dividing the total scores by the number of games played ($5 \cdot 192$ in the first competition, $5 \cdot 165$ in the second). While the theoretical range of this quantity is from 0 to 1000, a more reasonable scale on which to evaluate strategies is from 200 to 600, which correspond to continued mutual defection and cooperation, respectively .[Axe80a] The score per game actually achieved in the first competition ranged from 221.5 to 525.0, in the second competition from 285.3 to 436.4. Finally, the participating strategies were ranked according to their total score in each competition. The ranks of MIRROR and AMBITIOUS were divided by the number of participants to make them comparable across competitions and averaged to give their overall normalised rank.

As can be seen from Table 2, MIRROR and AMBITIOUS did neither especially badly nor especially well. In the first competition, both were in the better half of the field, both by score and by rank. In the second competition, AMBITIOUS was ten places from the bottom by rank, but still just inside the middle third by score. In both competitions, MIRROR achieved better results than AMBITIOUS. This was also true for each run of the competitions individually.

IV. Conclusions and Discussion

As can be seen from the data presented in the previous section, the MIRROR strategy performed consistently better than AMBITIOUS despite being less intent on its own advantage. This confirms Axelrod's observation that forgiving strategies perform better.[Axe80a, Axe80b] By Axelrod's nomenclature, both strategies are nice, but only MIRROR is forgiving, as it will try to cooperate when its opponent does, even after having been cheated. Unlike previous work, this investigation compares two strategies which share most of their algorithm. Therefore advantages or drawbacks of the learning algorithm do not influence the comparison.

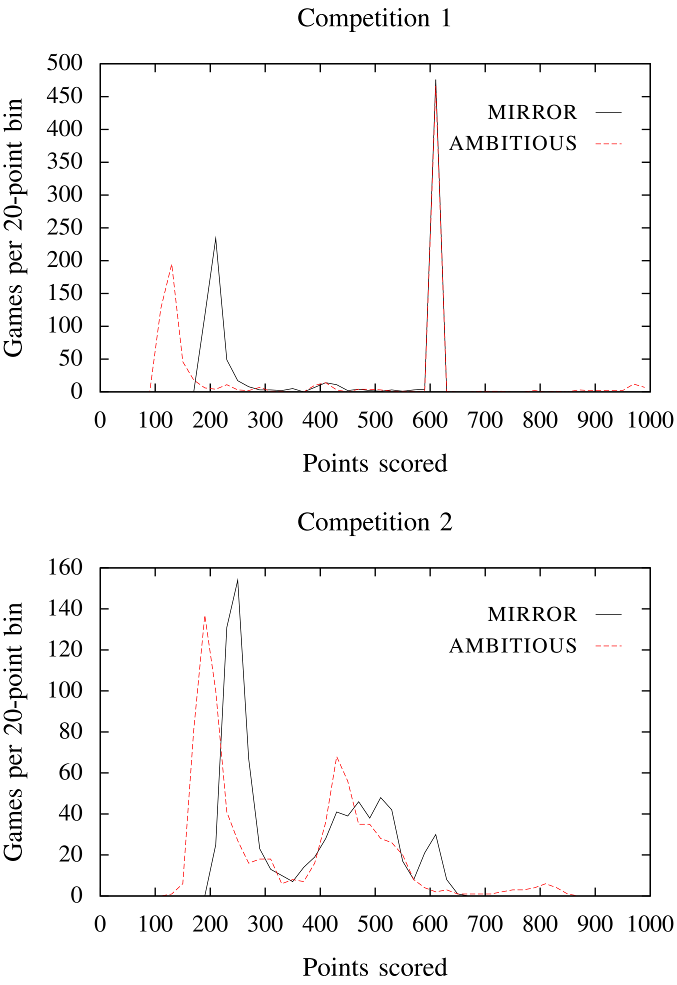

Figure 1. Histograms of both strategies' scores in both competitions.

It is interesting to see how the two strategies' different score came about. Figure 1 shows histograms of their scores in each competition. In the first competition, both share a high peak at 600 points, mutual cooperation. This reflects their "niceness", the inability to defect before their opponent does. The difference between the strategies lies in the region of scores which imply that their opponent must have been

[1] This was in contradiction to the Competition's announcement, which stated the number of rounds would vary.

defecting for most of the game.

With MIRROR behaving less egotistically than AMBITIOUS, any comparative advantage it enjoyed can have resulted only indirectly, from the way competitors' behaviour was influenced by its own. Figure 1 shows clearly that the field of competitors was sophisticated enough to recognise AMBITIOUS as selfish, stop cooperating with it and in the process successfully cheating it.

The second competition differed from the first in that a player's choice of move was not always accepted as is, but there was a small chance that it was "misinterpreted" and the opposite move was taken. As the second part of Figure 1 shows, this significantly changes the results.

MIRROR still scored better than AMBITIOUS in games where their opponent mostly defected. Though the difference was smaller than in the first competition, this may be due to the fact that not all strategies from the first competition also took part in the second. But the most obvious difference compared to the first competition is that the mutual cooperation peak has all but vanished, due to a run of mutual cooperation being interrupted by an unintended defection. The MIRROR strategy retains a small peak at the score 600, corresponding to mutual cooperation, which suggests it can recover from one or a few defections when playing a similarly forgiving opponent. When it cannot, its score is reduced to between 400 and 550 points. AMBITIOUS's score, by contrast, peaks near the low end of that range and shows no residual at the mutual cooperation value.

V. Summary

Two strategies submitted to the 2005 Iterated Prisoner's Dilemma Competition were presented. They share a common learning algorithm using an experience table, but differ in their aims: one tries to behave fairly, the other selfishly. The close similarities in implementation between the strategies allow a direct comparison with other things being equal. The fair strategy did consistently better than the selfish one. This is in line with well-known observations about the properties of winning strategies.

References

∴ prior work • reference ↭ related ◁ background

[Axe80a] ∴ R. Axelrod: *Effective Choice in the Prisoner's Dilemma*. Journal of Conflict Resolution **24** (1) (1980) 3–25.

[Axe80b] ∴ R. Axelrod: *More Effective Choice in the Prisoner's Dilemma*. Journal of Conflict Resolution **24** (3) (1980) 379–403.

[Axe84] • R. Axelrod: The Evolution of Cooperation. Basic Books. 1984. ISBN 0-465-02121-2.

[BGBS07] ↭ O. Bar-Gill & O. Ben-Shahar: *The prisoners' (plea bargain) dilemma*. 2007. Online. Discussion paper.

[Bre87] ↭ G. Breivik: *The doping dilemma: some game theoretical and philosophical considerations*. Sportwissenschaft **17** (1987) 83–94.

[BW97] ↭ E. Bird & G. Wagner: *Sport as a common property resource: a solution to the dilemmas of doping*. Journal of Conflict Resolution **41** (1997) 749–766.

[KDY] ∴ G. Kendall, P. Darwen & X. Yao: *The Iterated Prisoner's Dilemma Competition*. Website. Online.

[Kir98] ↭ D. P. Kirby: *The Strategic Defense Initiative and the Prisoner's Dilemma*. Tech. rep., Defense Technical Information Center. 1998.

[KYC07] ◁ G. Kendall, X. Yao & S. Y. Chong: The Iterated Prisoners' Dilemma: 20 Years on. World Scientific. 2007. ISBN 978-9812706973.

[Pou92] • W. Poundstone: Prisoner's Dilemma. Doubleday. Feb 1992. ISBN 0-385-41567-2.

[SSS00] ↭ R. Smith, M. Sola & F. Spagnolo: *The Prisoner's Dilemma and Regime-Switching in the Greek-Turkish Arms Race*. Journal of Peace Research **37** (6) (2000) 737–750. doi:10.1177/0022343300037006005.

[Wil88] ↭ J. Wiley: *Reciprocal Altruism as a Felony: Antitrust and the Prisoner's Dilemma*. Michigan Law Review **86** (1988) 1906–1928.