Finding best counterstrategies for generalized Iterated Prisoner's Dilemma games

2000 J. Phys. A: Math. Gen. 33 L453

(http://iopscience.iop.org/0305-4470/33/47/101)

J. Phys. A: Math. Gen. **33** (2000) L453–L458. Printed in the UK PII: S0305-4470(00)15186-8

LETTER TO THE EDITOR

Finding best counterstrategies for generalized Iterated Prisoner's Dilemma games

Hendrik Moraal
Institute for Theoretical Physics, University of Cologne, D-50937 Cologne, Germany

Received 21 June 2000, in final form 23 October 2000

Abstract. A class of games with two players, who base their actions on the results of the previous round, is considered. These games are generalizations of the Iterated Prisoner's Dilemma. The best counterstrategy for player 1, given a strategy for player 2, is found by treating these games as Markov processes. Several transitions between such best strategies are found and shown to be akin to first-order phase transitions. The roles of a number of special strategies are elucidated. It is shown that there is a strategy for player 2 that never loses, even if this player does not know what kind of game is being played.

The Prisoner's Dilemma [1] in iterated form has become the leading paradigm for the explanation of the evolution of cooperation among selfish agents [2–4]. In the non-iterated form of this game, two prisoners must decide independently whether to confess to a crime (that they really committed together), although there is not enough evidence available to secure a conviction. If both decide to keep silent (i.e. they *cooperate*), they must be released; if one of them confesses (*defects*), he will be released and live a pleasant life as a state witness, while the other will go to prison for a long term; if both confess, they will also both be convicted, but their sentences will be reduced. Standard game theory [5] says that a rational, selfish agent should always defect, but the situation changes drastically if the game is repeated many times (Iterated Prisoner's Dilemma, IPD). This was shown conclusively in an experimental way by the computer tournaments organized by Axelrod [6–8], in which the 'tit-for-tat' strategy of Rapaport and Chammah [1] (do exactly what the other player did in the previous round) proved to be the most successful one. This may, however, no longer be true if spatial or ecological considerations also play a rôle [9, 10]; see also the discussion of feature (F) below.

For the purposes of this Letter, it is enough to define the iterated playing of a game as follows.

(i) A *game* with n players consists of a sequence of elementary *moves* by each of the players in turn.
(ii) A *round* is finished after all players have made a move. These are to be selected from a finite number of possibilities.
(iii) The *rules* of the game determine the possible moves of a player; these may depend on the moves already made in the present round as well as of the results of the previous rounds.
(iv) After each round, each player i obtains a *payoff* $F(i)$, which is completely determined by the moves of all players in the current round. If these payoffs are such that their sum over all players is zero, the game is a *zero-sum* one. A nonzero-sum game can be considered as a zero-sum one on introducing an extra player, who pays the balance of the n payoffs of the original players.

0305-4470/00/470453+06$30.00

(v) A *strategy* for a player is a 'recipe' telling this player what to do. It can be of the form 'if the history of the game is H (this is the set of the results of all previous rounds and of the earlier moves of the present one), then do move $s(H)$'; in this case the strategy is called *pure*. A *mixed* strategy is of the form 'if the history is H, then use a random device to select one of the possible rules s with probability $p_H(s)$'. A pure strategy is the limit of a mixed one for which $p_H(s) = 1$ for exactly one value of s.

These definitions can still be generalized; for the case where the game consists of one round only, a more general definition which includes the above can be found in the classic work of von Neumann and Morgenstern [5]. In the following, the number of players is $n = 2$ and the history H is restricted to the results of the previous round only.

In the rest of this Letter, results on generalizations of the IPD are obtained analytically by considering such games as Markov processes. In each round of such a generalized game, the two players independently select one of two possible actions, which will be denoted by 'cooperation' and 'defection' as in the IPD and are assigned values 1 and 0, respectively. The players have the result of the previous round available to base their choices on; therefore, the probability that the outcome is given by the pair (σ_1, σ_2) if the result of the previous round was (τ_1, τ_2) has the form of a product:

$$p(\sigma_1, \sigma_2; \tau_1, \tau_2) = p(\sigma_1; \tau_1, \tau_2) q(\sigma_2; \tau_1, \tau_2). \tag{1}$$

After each round, player 1 receives an amount given by a fixed *payoff matrix* h_{σ_1,σ_2}, whereas player 2 gets h_{σ_2,σ_1}. A *strategy* for player 1 consists of a set of four probabilities (p_0, p_1, p_2, p_3) so that

$$p(1; \tau_1, \tau_2) = p_{2\tau_1+\tau_2} \qquad p(0; \tau_1, \tau_2) = 1 - p(1; \tau_1, \tau_2) \tag{2}$$

holds. Since player 2 sees the game from the opposite perspective, her strategy (q_0, q_1, q_2, q_3) translates to

$$q(1; \tau_1, \tau_2) = q_{2\tau_2+\tau_1} \qquad q(0; \tau_1, \tau_2) = 1 - q(1; \tau_1, \tau_2). \tag{3}$$

Such a game corresponds to the IPD if one has $h_{0,1} > h_{1,1} > h_{0,0} > h_{1,0}$; see the discussion about the possible rewards and punishments of the two prisoners described above. For more technical reasons, one also has to choose $h_{0,1} + h_{1,0} < 2h_{1,1}$ [11]. Other variants are known under several suggestive names [11]: for *Deadlock* one has $h_{0,1} > h_{0,0} > h_{1,1} > h_{1,0}$, for *Chicken* $h_{0,1} > h_{1,1} > h_{1,0} > h_{0,0}$ and for *Stag Hunt* $h_{1,1} > h_{0,1} > h_{0,0} > h_{1,0}$. It will turn out that a comparison of the relative efficiencies of two strategies depends only on the sign of the difference $h_{0,1} - h_{1,0}$. In all of the examples above, this is positive; such games will be called *IPD-type* games since they all imply a dilemma. For games with $h_{0,1} - h_{1,0} < 0$, this is not the case; such games are called *COOP-type* games. These are related to associated IPD-type games by a symmetry, as will be shown below.

As noted before, the probabilities of equation (1) only depend on the results of the previous round, so that the iterated playing of games of this type can be described by a Markov process. Let $P_{\sigma_1,\sigma_2}(g_1, g_2, k)$ be the probability that the kth round has the result (σ_1, σ_2) and that the players have accumulated g_1 and g_2 as total payoffs. The master equation for these probabilities reads

$$P_{\sigma_1,\sigma_2}(g_1, g_2, k) = \sum_{\tau_1,\tau_2} p(\sigma_1, \sigma_2; \tau_1, \tau_2) P_{\tau_1,\tau_2}(g_1 - h_{\sigma_1,\sigma_2}, g_2 - h_{\sigma_2,\sigma_1}, k - 1). \tag{4}$$

The matrix of probabilities has the following explicit form in terms of the strategies of the two players; see equations (1)–(3):

$$\begin{pmatrix} (1-p_0)(1-q_0) & (1-p_1)(1-q_2) & (1-p_2)(1-q_1) & (1-p_3)(1-q_3) \\ (1-p_0)q_0 & (1-p_1)q_2 & (1-p_2)q_1 & (1-p_3)q_3 \\ p_0(1-q_0) & p_1(1-q_2) & p_2(1-q_1) & p_3(1-q_3) \\ p_0q_0 & p_1q_2 & p_2q_1 & p_3q_3 \end{pmatrix}. \tag{5}$$

Summing equation (4) over all gains g_1 and g_2 gives

$$Q_{\sigma_1,\sigma_2}(k) \equiv \sum_{g_1,g_2} P_{\sigma_1,\sigma_2}(g_1, g_2, k) = \sum_{\tau_1,\tau_2} p(\sigma_1, \sigma_2; \tau_1, \tau_2) Q_{\tau_1,\tau_2}(k-1). \tag{6}$$

The average gains $G_1(k)$ and $G_2(k)$ for the two players are obtained by multiplying equation (4) by g_1 and g_2, respectively, and summing over all variables:

$$\begin{aligned} G_1(k) &= G_1(k-1) + \sum_{\sigma_1,\sigma_2} h_{\sigma_1,\sigma_2} Q_{\sigma_1,\sigma_2}(k) \\ G_2(k) &= G_2(k-1) + \sum_{\sigma_1,\sigma_2} h_{\sigma_2,\sigma_1} Q_{\sigma_1,\sigma_2}(k). \end{aligned} \tag{7}$$

In the limit of $k \to \infty$, it is expected (by ergodicity) that the $Q_{\sigma_1,\sigma_2}(k)$ can be replaced by an eigenvector of the matrix of equation (5) with eigenvalue equal to 1. It will be assumed that this eigenvector, denoted by Q_{σ_1,σ_2}, is nondegenerate. The quantity w defined by

$$w = \lim_{k\to\infty} [G_1(k) - G_2(k)]/[k(h_{0,1} - h_{1,0})] = Q_{0,1} - Q_{1,0} \tag{8}$$

is then a measure of the difference between the efficiencies of the strategies of the two players. The larger it is (its possible values are $-1 \leqslant w \leqslant 1$), the more efficient is player 1's strategy as compared to that of player 2. This is the case if $h_{0,1} > h_{1,0}$ holds, i.e. in games of the IPD type (see the discussion above). For games of the COOP type, the terms on the right-hand side of equation (8) have to be exchanged. A general symmetry follows from the structure of the matrix of equation (5) and from (8).

(i) The COOP-type game obtained from an IPD-type game by exchanging $h_{1,0}$ and $h_{0,1}$ is called its associate game.

(ii) Let a game of IPD-type and strategies (p_0, p_1, p_2, p_3) and (q_0, q_1, q_2, q_3) lead to a specific value of w. Then the associate COOP-type game with strategies $(1-p_3, 1-p_2, 1-p_1, 1-p_0)$ and $(1-q_3, 1-q_2, 1-q_1, 1-q_0)$ leads to the same value of (a properly defined) w. This is due to the fact that this symmetry transforms the matrix $M(i, j)$, $i, j \in \{0, 1, 2, 3\}$, of equation (5) into $M(3-i, 3-j)$, so that its transformed eigenvector with eigenvalue 1 has $Q_{0,1}$ and $Q_{1,0}$ exchanged.

Due to this symmetry, only IPD-type games, for which equation (8) is correct as it stands, have to be considered.

The possible values of a strategy fill a four-dimensional hypercube. The best counterstrategies for player 1, i.e. those for which w is maximal given a fixed strategy for player 2, have been obtained analytically from the eigenvector Q_{σ_1,σ_2} of the matrix of equation (5) for the eight diagonals of the hypercube of the strategies of player 2. The results are listed in table 1. An explicit expression for the quantity $Q_{0,1} - Q_{1,0}$ has been obtained by computer algebra; this has been used to write a program to obtain the optimal strategies for player 1 if the strategy of player 2 does not lie on one of these diagonals.

A number of salient features of the results listed in table 1 or obtained numerically elsewhere in the hypercube will be discussed in the following; these features are listed in the last column of this table.

(A) The first striking feature of table 1 is the predominance of the strategy $(0, 0, 0, 0)$ for player 1. This means that this player always defects, independent of the outcome of the previous round. It is known as the ALL-D strategy. Indeed, if a strategy for player 2 is picked at random inside the hypercube, a numerical check for the best strategy of player 1 turns out to be ALL-D in about 95% of all cases. This is somewhat surprising at first sight, since it seems to corroborate the single-game analysis of the Prisoner's Dilemma. The special roles of other strategies will become clearer in the following.

Table 1. Best counterstrategies for player 1 if player 2 plays a strategy on one of the eight diagonals of the hypercube of probabilities.

Diagonal	Values	Best for player 1	w	Type
(r, r, r, r)	$r = 0$	$(0, p_1, p_2 < 1, p_3)$	0	(B)
	$0 < r < 1$	$(0, 0, 0, 0)$	r	(A)
	$r = 1$	$(p_0, 0, p_2, p_3 < 1)$	1	(C)
$(1 - r, r, r, r)$	$0 \leqslant r \leqslant \frac{1}{4}(\sqrt{17} - 1)$	$(0, 0, 0, 0)$	$\frac{1}{2}$	(A)
	$\frac{1}{4}(\sqrt{17} - 1) < r < 1$	$(1, 0, 0, 0)$	$r^2/(2 - r)$	(E)
	$r = 1$	$(p_0, 0, p_2, p_3 < 1)$	1	(C)
$(r, 1 - r, r, r)$	$0 \leqslant r < 1$	$(0, 0, 0, 0)$	r	(A)
	$r = 1$	$(p_0, 0, p_2, p_3 < 1)$	1	(C)
$(r, r, 1 - r, r)$	$r = 0$	$(0, p_1, p_2 < 1, p_3)$	0	(B)
	$0 < r \leqslant 1$	$(0, 0, 0, 0)$	$\frac{1}{2}$	(A)
$(r, r, r, 1 - r)$	$r = 0$	$(0, p_1, p_2 < 1, p_3)$	0	(B)
	$0 < r \leqslant 1$	$(0, 0, 0, 0)$	r	(A)
$(1 - r, 1 - r, r, r)$	$0 \leqslant r \leqslant 1$	$(0, 0, 0, 0)$	$\frac{1}{2}$	(A)
$(1 - r, r, 1 - r, r)$	$0 \leqslant r < 1$	$(0, 0, 0, 0)$	$1 - r$	(A)
	$r = 1$	(p_0, p_1, p_2, p_3)	0	(D)
$(1 - r, r, r, 1 - r)$	$0 \leqslant r \leqslant \frac{1}{4}(\sqrt{5} + 1)$	$(0, 0, 0, 0)$	$\frac{1}{2}$	(A)
	$\frac{1}{4}(\sqrt{5} + 1) < r \leqslant 1$	$(1, 0, 0, 1)$	$r(2r - 1)$	(F)

(B) There are three pure strategies for player 2:

(1) (0, 0, 0, 0) (ALL-D again, but now for player 2),
(2) (0, 0, 1, 0) (player 2 cooperates only if player 1 defected and she herself cooperated in the previous round, see equation (3)),
(3) (0, 0, 0, 1) (player 2 cooperates only if both players cooperated in the previous round),

where w is given by

$$w = w_1 = -p_0/(1 - p_2 + p_0). \tag{9}$$

Here the strategy to be adopted by player 1 does not depend on p_1 and p_3, but $p_0 = 0$ is necessary in order not to lose. This limit should, however, not be performed with p_2 set to 1, since then $w_1 = -1$ follows instead of $w_1 = 0$.

(C) Similarly, at the pure strategies:

(1) (1, 1, 1, 1) (this is the 'always cooperate' strategy known as ALL-C for player 2),
(2) (0, 1, 1, 1) (player 2 defects only if both defected in the previous round),
(3) (1, 0, 1, 1) (player 2 defects only if player 1 cooperated and she herself defected in the previous round),

where w is given as

$$w = w_2 = (1 - p_3)/(1 + p_1 - p_3) \tag{10}$$

and the best strategy for player 1 has $w_2 = 1$ for $p_1 = 0$, where this limit has to taken with $p_3 < 1$, else $w_2 = 0$ results.

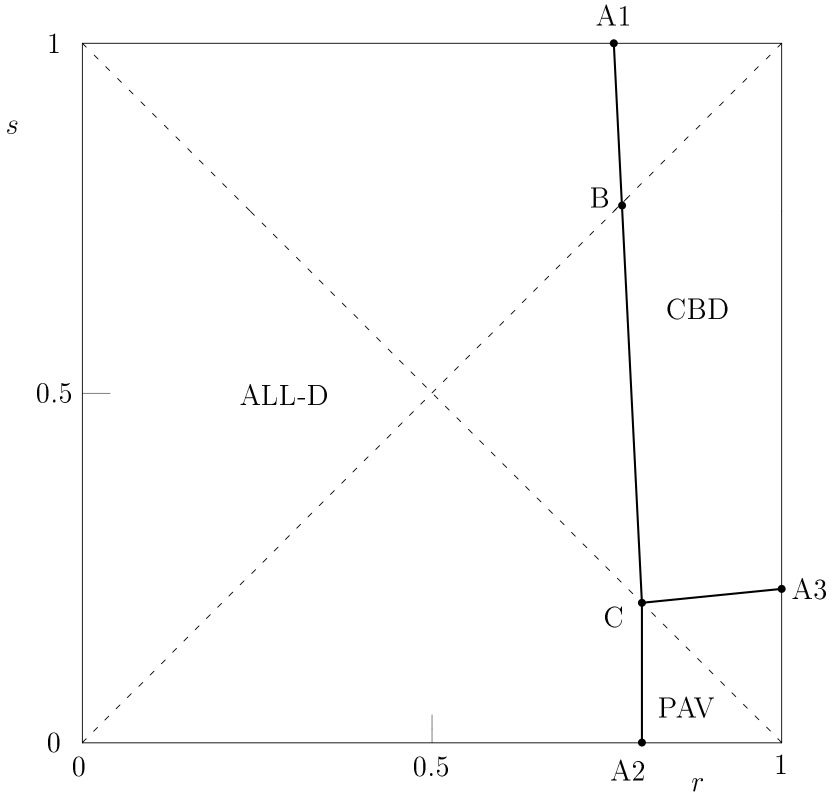

Figure 1. The best counterstrategies for the strategies of player 2 of the form $(1-r, r, r, s)$. The thin broken lines are the diagonals of features (E) and (F) above with the transition points B and C, respectively. The phase boundaries are the following: A1C, $s = (2-3r^2)/(1-r)$; A2C, $s = (r^2 - 7r + 5)/(1-r)$; A3C, $s = \frac{1}{6}[3r + 2 - \sqrt{33r^2 - 48r + 28}]$. These are indistinguishable from straight lines in this figure.

(D) If player 2 plays the stategy (0, 1, 0, 1), which corresponds to 'tit-for-tat' (TFT) or 'do exactly what player 1 did in the previous round', the strategy of player 1 is immaterial: w is always equal to 0 in this case.

Away from the pure strategies, there are two pockets where a strategy different from ALL-D is the best for player 1. In both cases there are transitions from ALL-D to these other strategies on the diagonals:

(E) On the diagonal $(1-r, r, r, r)$, there is a range of values for which the strategy (1, 0, 0, 0) (only cooperate if both players defected in the previous round; this will be called CBD from now on) is superior to ALL-D for player 1, although player 2 uses a mixed strategy.

(F) Similarly, there is a range of values on the diagonal $(1-r, r, r, 1-r)$ where the strategy (1, 0, 0, 1) (cooperate only if the players made the same choice in the previous round), known as Pavlov or PAV, is the best one for player 1. Interestingly, this is the strategy that outperforms TFT in a situation where mutation and natural selection play a rôle [10].

In both of the cases (E) and (F), the derivative of w with respect to r is discontinuous at the transition from ALL-D, so these transitions are akin to first-order phase transitions.

The strategies CBD and PAV extend from the diagonals to encompass finite volumes of the hypercube. Therefore, this consists of about 95% of ALL-D, the remaining 5% being occupied by CBD and PAV. Other strategies only occur as best ones at the corners of the hypercube: these are the cases (B), (C) and (D) above. There is also a transition (again of first-order type) between regions where CBD or PAV is the best counterstrategy. In figure 1, this is shown in the plane $(1 - r, r, r, s)$, which contains both of the diagonals of cases (E) and (F).

Up to now, the games have always been considered from the point of view of player 1. How about player 2? What strategy should she use in order not to lose to player 1, who always picks the best counterstrategy? From table 1, the strategies with $w = 0$ are the sought-for ones: (0, 0, 0, 0) (or ALL-D), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1) and (0, 1, 0, 1) (or TFT) will all guarantee that player 2 does not lose. If the game is not of the IPD but of the COOP type, the symmetry found above gives (1, 1, 1, 1) (or ALL-C), (1, 1, 0, 1), (1, 0, 1, 1), (0, 1, 1, 1) and (0, 1, 0, 1) (or TFT again) as solutions. Therefore, even if player 2 does not know what kind of game is being played or if player 1 changes the rules (i.e. the payoff matrix) as he goes along, there simply is no way to lose for player 2 if she sticks to the TFT strategy. This result shows the unique position of this strategy very clearly.

In conclusion, it has been shown that the best strategy for player 1 of a class of games which generalize the Iterated Prisoner's Dilemma is either ALL-D (always defect), CBD (cooperate only if both players defected in the previous round) or PAV (cooperate only if both players chose the same action in the previous round), except at special corners of the hypercube of strategies of player 2. This player, on the other hand, cannot lose if she plays the TFT strategy (always do what the other player did in the previous round), even if the payoff matrix of the game is unknown or changes with time. The iterated playing of this type of game is a dynamic process, so these results are not easily compared with those obtained for single game playing, where the best pair of strategies is such that none of the players can improve his or her payoff (this is called a Nash equilibrium [12]). Actually, the usefulness of this concept in economics is hampered by the difficulty of describing the dynamics necessary to reach such an equilibrium.

I wish to thank R J Moraal for a number of useful discussions.

References

[1] Rapaport A and Chammah A M 1965 *Prisoner's Dilemma* (Ann Arbor, MI: University of Michigan Press)
[2] Smith J M 1982 *Evolution and the Theory of Games* (Cambridge: Cambridge University Press)
[3] Dawkins R 1988 *The Selfish Gene* (Oxford: Oxford University Press)
[4] Sigmund K 1993 *Games of Life* (Oxford: Oxford University Press)
[5] Von Neumann J and Morgenstern O 1944 *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press)
[6] Axelrod R 1984 *The Evolution of Cooperation* (New York: Basic Books)
[7] Axelrod R and Hamilton W D 1981 *Science* **211** 1390
[8] Axelrod R and Dion D 1988 *Science* **242** 1385
[9] Nowak M and May R M 1992 *Nature* **359** 826
[10] Nowak M and Sigmund K 1993 *Nature* **364** 56
[11] Flake G W 1998 *The Computational Beauty of Nature* (Cambridge, MA: MIT Press)
[12] Nash J F 1951 *Ann. Math.* **54** 286