



Extending the Iterated Prisoner's Dilemma without Synchrony

CH. HAUERT AND H. G. SCHUSTER

*Department of Theoretical Physics, Christian-Albrechts-University of Kiel,
Leibnizstrasse 15/242, D-24098 Kiel, Germany*

(Received on 1 May 1997, Accepted in revised form on 11 November 1997)

In biology altruistic behaviour of selfish individuals often can be modelled by the iterated Prisoner's Dilemma game (PD). If the opponents are caught repeatedly in this dilemma reciprocity may lead to cooperative strategies adopted by the individuals. In this article we present results from numerical simulations of the infinitely iterated stochastic alternating PD. First we investigate influences of the memory size on strategies in the alternating two player PD. We show that *Firm but Fair* is a strong strategy hardly affected by memory size and different values of the temptation to defect. Second we discuss successful strategies in the alternating N player N step memory PD. In this situation we focus on the stability of cooperative strategies and compare the results to experiments of predator inspection by sticklebacks carried out by Milinski and co-workers.

© 1998 Academic Press Limited

1. Introduction

Since the early days of the Prisoner's Dilemma (PD) a very large number of articles have been published on the subject. Many researchers from various disciplines were attracted by this simple model showing such a rich dynamical behaviour and having applications in very different fields (see e.g. Dawes, 1980; Axelrod, 1984; Milinski, 1987; Nowak & Sigmund, 1993). Regardless of its simplicity, it is a powerful model for social interactions and to explain altruistic behaviour of selfish individuals. In the PD, any rational player will defect when played only one round but cooperation can be rewarded if repeatedly played with high probability. The game was originally formulated by Dresher & Flood as early as 1950 (Flood, 1958). For his extensive computer tournaments Axelrod (1984) defined the simultaneous PD as it is commonly used today. Nevertheless, in biology, political science and sociology often an alternating formulation, where players do not decide simultaneously but rather take their decisions in turns, seems to be more adequate. Examples from biology where reciprocal altruism based on alternating moves of the individuals has been observed include baboons

(Packer, 1977; Trivers, 1971) and vampire bats (Wilkinson, 1984). In the former case two young male baboons alternate the roles of one distracting the attention of a dominant male while the other has the opportunity of mounting an oestrous female. In the latter situation well fed vampire bats regurgitate some blood to feed their hungry fellow bats. Other examples from biology include predator inspection, alarm calls, grooming, etc. (see e.g. Axelrod & Hamilton, 1981; Milinski, 1987). Theoretically the alternating PD has been studied for example by Frean (1994), Nowak & Sigmund (1994), Boerlijst *et al.* (1997) and Leimar (1997).

2. The Game

In the Prisoner's Dilemma (PD) the players have two options, they must choose between cooperation (C) and defection (D). Depending on all players' decision every one obtains a certain number of points. A very important point is the timing of the decisions which has a fundamental effect on the strategies. In the simultaneous PD all players take their decisions at once whereas in the alternating game the players

take their decisions in turns. In order to keep strategies as simple as possible and to allow comparisons to results of other researchers we follow the approach chosen by Freat (1994) as well as Nowak & Sigmund (1994). They used slightly different methods to determine the pay-off for each player. Since we were not only interested in the generalization to N players with M steps memory but also in the influences of the temptation to defect, we follow the more general method described by Freat and will embed Nowak & Sigmund's arguments as a special case. Before a biological interpretation of the strategies can be given we must explain the calculation of the pay-off for each player.

Unlike in the simultaneous game, a round in the alternating PD consists of a single players decision. N rounds will be called one unit, where N refers to the number of players in the game. In one such unit every player makes one decision. The player deciding first in each unit will be referred to as the leader. After the last player in one unit made its decision the pay-offs for this unit are determined according to the following rule:

- (1) a cooperating player receives a reward of R points from every cooperating opponent but only the sucker's pay-off S from defecting ones;
- (2) a defecting player gets the temptation of T points from every cooperating opponent but only the punishment P from defecting ones.

In the PD the values of R , S , T and P must satisfy the following conditions:

$$(a) T > R > P > S \quad (b) R > (T + S)/2 \quad (1)$$

Due to condition (1a) in a single unit a player is better off choosing D instead of C . Condition (1b) states that a unit of mutual cooperation yields the highest average pay-off. This generalization of the pay-off values is identical to the one used in the generalization

of the simultaneous variant of the game (Hauert & Schuster, 1997). Thus, we have a general formulation of the game covering both, the simultaneous as well as the alternating variant of the N player M step memory PD.

Concerning the calculation of the pay-offs, we did not take into account that the players take their decisions in turns. Distributing the pay-offs as explained above would result in a heavily biased game. Considering for example the iterated game with N players having N steps memory (covering one unit of rounds), we see that the player deciding last in each unit of rounds has the most information available to determine the pay-off in this unit. All other players must find their decisions depending at least on one move of an opponent in the previous unit. Since we study the strictly alternating PD with a fixed sequence of the players, it follows that the leader has a serious handicap. Our approach leaves two possibilities to balance the game and give equal chances to every player. The first one follows Freat's (1994) work, where we have to repeat the game for each player being once in every position of deciding (see below). The second possibility implements Nowak & Sigmund's (1994) approach. Here, the player making a decision determines the pay-off for all players regardless of the opponents decisions. This results in a modified rule to determine the player's pay-off:

- (1) a cooperating leader obtains a points and all others b ;
- (2) a defecting leader gets c points leaving the others with d .

In analogy to eqn (1) the following conditions must hold for a , b , c , d :

$$(a) c > a \quad (b) c - a < b - d \quad (2)$$

TABLE 1
Pay-off values achieved by each player in the alternating three player PD after one unit of rounds

History	Alternating, symmetric payoff			Alternating, individual payoff		
	1	2	3	1	2	3
$C^1C^2C^3$	$a + b + b$	$b + a + b$	$b + b + a$	$R + R$	$R + R$	$R + R$
$C^1C^2D^3$	$a + b + d$	$b + a + d$	$b + b + c$	$R + S$	$R + S$	$T + T$
$C^1D^2C^3$	$a + d + b$	$b + c + b$	$b + d + a$	$S + R$	$T + T$	$R + S$
$C^1D^2D^3$	$a + d + d$	$b + c + d$	$b + d + c$	$S + S$	$T + P$	$T + P$
$D^1C^2C^3$	$c + b + b$	$d + a + b$	$d + b + a$	$T + T$	$S + R$	$S + R$
$D^1C^2D^3$	$c + b + d$	$d + a + d$	$d + b + c$	$T + P$	$S + S$	$P + T$
$D^1D^2C^3$	$c + d + b$	$d + c + b$	$d + d + a$	$P + T$	$P + T$	$S + S$
$D^1D^2D^3$	$c + d + d$	$d + c + d$	$d + d + c$	$P + P$	$P + P$	$P + P$

The pay-off values in the middle column are calculated according to Nowak & Sigmund (1994) and those in the right column follow Freat (1994).

Again, condition (2a) guarantees that in a single round the leader is better off choosing D instead of C . Condition (2b) then states that the cost occurring to the leader altruistically choosing C are less than the benefits to the other players. From the two conditions follows implicitly that $b > d$.

The above rule for the pay-off values by Nowak & Sigmund can be easily embedded in the generalized pay-off rule stated before. In Table 1 we have listed all possible outcomes of one unit of rounds in the three player three step memory PD as well as the respective pay-off values for each player according to the two pay-off rules. Comparing the pay-off values of the two approaches leads to the following restrictions for R, S, T, P :

$$\left. \begin{aligned} a + 2b &= 2R \\ a + b + d &= R + S \\ a + 2d &= 2S \\ c + 2b &= 2T \\ c + b + d &= T + P \\ c + 2d &= 2P \end{aligned} \right\} T + S = R + P \quad (3)$$

Equation (3) holds for an arbitrary number of players and results in the same restriction for R, S, T and P as found by Nowak & Sigmund (1994) in the alternating two player PD. Thus, if R, S, T, P satisfy eqn (3) the two approaches are identical. Later we will see that eqn (3) has further consequences concerning the dimensions of the strategy space.

Whether a player opts for C or D in the next round depends on its memory length. In this article we discuss two limiting cases: (a) two players with M steps memory and (b) N players with N steps memory which corresponds to one unit of rounds. The strategy of a player is encoded in a vector of conditional probabilities. Each element specifies the probability to

cooperate in the next round for a given history. This array consists of 2^M entries where M specifies the memory length. The different probabilities are indexed by the respective history. Most recent moves are noted to the right and those in the past to the left. The players own move is noted first in each unit of rounds. In Fig. 1 a sample sequence of moves is shown for two players with three steps memory and for three players with three steps memory as well as the probability to cooperate for the corresponding histories. For example, p_{cdc} specifies in the two players three steps memory game the probability to cooperate after experiencing a T in the previous unit and with the opponent cooperating before. In the three player three step memory PD p_{cdc} specifies the probability to cooperate given that the player itself opted for C followed by D and C of the other players. The probability to cooperate given a specified but arbitrary history x will be referred to as p_x .

We now want to calculate the mean pay-off for each player in the infinitely iterated alternating PD. Since we consider probabilistic strategies and infinitely iterated games the initial moves of the players are of no importance. Nevertheless it is important to keep in mind that the success of a strategy also depends on the player's position of deciding in each unit—an effect which must be balanced as mentioned before. The transition from one round to the next can be formulated as a Markov process. First, we look at the alternating PD for two players with M steps memory ($M \geq 1$). For simplicity, we show only an example for two players with three steps memory remembering the last unit of rounds plus the opponents move in the previous unit. If it is player 1's turn to decide, the transition matrix $A_{2,3}^1$ is given in (4). The lower index of $A_{N,M}^i$ specifies the number of players N and the memory length M . The upper index indicates that it is player number i 's turn to decide.

$$\begin{array}{c|cccccccc} & C^1C^2C^1 & C^1C^2D^1 & C^1D^2C^1 & C^1D^2D^1 & D^1C^2C^1 & D^1C^2D^1 & D^1D^2C^1 & D^1D^2D^1 \\ \hline C^2C^1C^2 & p_{ccc}^1 & 1 - p_{ccc}^1 & 0 & 0 & 0 & 0 & 0 & 0 \\ C^2C^1D^2 & 0 & 0 & p_{ccd}^1 & 1 - p_{ccd}^1 & 0 & 0 & 0 & 0 \\ C^2D^1C^2 & 0 & 0 & 0 & 0 & p_{cdc}^1 & 1 - p_{cdc}^1 & 0 & 0 \\ C^2D^1D^2 & 0 & 0 & 0 & 0 & 0 & 0 & p_{cdd}^1 & 1 - p_{cdd}^1 \\ D^2C^1C^2 & p_{dcc}^1 & 1 - p_{dcc}^1 & 0 & 0 & 0 & 0 & 0 & 0 \\ D^2C^1D^2 & 0 & 0 & p_{dcd}^1 & 1 - p_{dcd}^1 & 0 & 0 & 0 & 0 \\ D^2D^1C^2 & 0 & 0 & 0 & 0 & p_{ddc}^1 & 1 - p_{ddc}^1 & 0 & 0 \\ D^2D^1D^2 & 0 & 0 & 0 & 0 & 0 & 0 & p_{ddd}^1 & 1 - p_{ddd}^1 \end{array} \quad (4)$$

The transition matrix for the other player is given by a simple permutation of the indices. The joint behaviour of both players is then given by the matrix $A_{2,3}^1 \cdot A_{2,3}^2$. Since this matrix is irreducible, there exists a unique left eigenvector to the eigenvalue 1 (Frobenius' theorem). The elements of the eigenvector specify the frequencies of all possible patterns in history. To determine the pay-off of each player we now simply have to multiply the frequency of each pattern with the corresponding pay-off. Accumulating these values yields the mean pay-off of each player. So far, we assumed that player 1 is the leader in every unit of rounds, neglecting that this might be a serious handicap. For this reason this calculation has to be done for $A_{2,3}^1 \cdot A_{2,3}^2$ as well as $A_{2,3}^2 \cdot A_{2,3}^1$. In the former Player 1 is the leader and in the latter player 2. Only the average of both configurations leads to the mean pay-off for each player.

Second, we turn to the alternating PD for N players ($N > 2$) with N steps memory. The procedure to calculate the pay-off of the two players is very similar to the one described above. In order to keep things simple, we show again only an example of the transition matrix for three players with three steps memory. For player 1 the transition matrix $A_{3,3}^1$ is given in (5):

	$C^2C^3C^1$	$C^2C^3D^1$	$C^2D^3C^1$	$C^2D^3D^1$	$D^2C^3C^1$	$D^2C^3D^1$	$D^2D^3C^1$	$D^2D^3D^1$	
$C^1C^2C^3$	p_{ccc}^1	$1 - p_{ccc}^1$	0	0	0	0	0	0	(5)
$C^1C^2D^3$	0	0	p_{ccd}^1	$1 - p_{ccd}^1$	0	0	0	0	
$C^1D^2C^3$	0	0	0	0	p_{cdc}^1	$1 - p_{cdc}^1$	0	0	
$C^1D^2D^3$	0	0	0	0	0	0	p_{cdd}^1	$1 - p_{cdd}^1$	
$D^1C^2C^3$	p_{dcc}^1	$1 - p_{dcc}^1$	0	0	0	0	0	0	
$D^1C^2D^3$	0	0	p_{dcd}^1	$1 - p_{dcd}^1$	0	0	0	0	
$D^1D^2C^3$	0	0	0	0	p_{ddc}^1	$1 - p_{ddc}^1$	0	0	
$D^1D^2D^3$	0	0	0	0	0	0	p_{ddd}^1	$1 - p_{ddd}^1$	

Note that the two matrices are identical except for the interpretation of the states and the probabilities. As above, a simple permutation of the indices leads to the transition matrices for the other players. The joint behaviour of N players is then given by the matrix $A_{N,N}^1 \cdot A_{N,N}^2 \cdot \dots \cdot A_{N,N}^N$ with $M=N$. As above, this matrix is irreducible and the unique left eigenvector to the eigenvalue 1 specifies the frequency of the different patterns in history leading to each player's pay-off. Again, this implies that player 1 is the leader. Thus, we have to repeat the above calculation for every

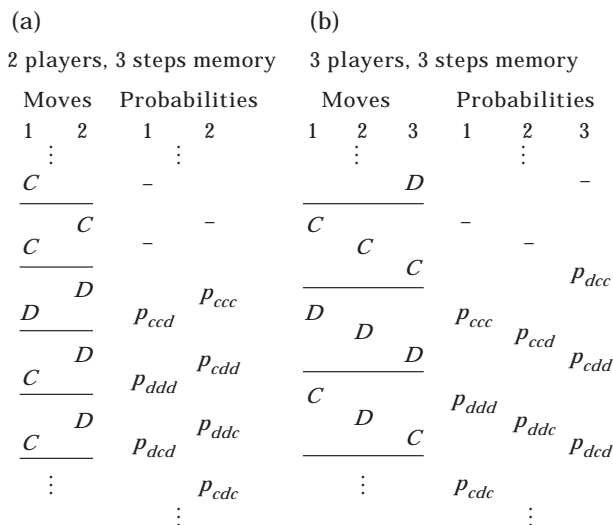


FIG. 1. Column (a) shows a sample sequence of moves in the alternating PD for two players with three steps memory and (b) for three players with three steps memory. The horizontal lines indicate borders between successive units of rounds which are used to calculate the pay-off (see text). Besides, the corresponding conditional probabilities p_x are shown, indicating the probability that the next player will cooperate for a given history x .

player being once in every position of deciding. The average of all these configurations yields the mean pay-off for each player.

Repeating the game with all players deciding once in every position and taking the average pay-off was useful to explain the above calculations but seems a rather artificial scenario to model biological situations. In the infinitely iterated PD an alternate but equivalent and biologically relevant interpretation of this averaged pay-off exists: after *each* players' decision *all* N players obtain their pay-off based on the last N moves.

Let us now have a closer look at the special case discussed by Nowak & Sigmund (1994). Choosing R ,

S , T and P according to eqn (3) simply means that it makes no difference which player is leading in a unit of rounds. In other words, the handicap of being the leader vanishes for certain pay-off values. It follows that only a single Markov chain has to be solved. For example $A_{2,3}^1 \cdot A_{2,3}^2$ and $A_{2,3}^2 \cdot A_{2,3}^1$ lead to the same mean pay-offs. Furthermore, this implies for N players with N steps memory ($N > 2$) that only the *number* of cooperating or defecting opponents count. Thus, for example in the three player PD the following conditional probabilities must be identical: $p_{ccd} \equiv p_{cdc}$. This reduces the strategy space dramatically from 2^N to $2N$ dimensions.

As Frean (1994) points out, if eqn (3) holds we model biological situations such as a well-fed vampire bat deciding whether to share its food, whereas hungry fellow bats are unable to influence the pay-off of either party. However, if eqn (3) does not hold, we model situations where the pay-off depends on the behaviour of all co-players such as body positioning in a shoal of fish on predator inspection.

In evolutionary game dynamics the pay-off of a player determines its reproductive success or similarly its frequency in the population (Maynard Smith, 1982). The frequencies f_i of I different strategies in the population at time t will change to f_i^{t+1} at time $t + 1$ according to the following equations:

$$f_i^{t+1} = f_i^t \cdot \frac{a_i}{a}$$

where

$$a_i = \sum_{n_1, n_2, \dots, n_{N-1} = 1}^I f_{n_1}^t f_{n_2}^t \dots f_{n_{N-1}}^t a_{i, n_1, n_2, \dots, n_{N-1}}$$

mean pay-off of player i , (6)

$$a = \sum_{i=1}^I f_i^t a_i \quad \text{population pay-off, (7)}$$

and

$$\sum_{i=1}^I f_i^t = 1 \quad \forall t \quad \text{normalization.}$$

The $a_{i, n_1, n_2, \dots, n_{N-1}}$ stand for the mean pay-off obtained by strategy i playing against strategies $n_1 \dots n_{N-1}$ as described above. A strategy i obtaining a higher total pay-off than the population pay-off ($a_i > a$) will spread.

At $t = 0$ we start with a homogenous population playing a completely random strategy with $p_x = 0.5$. Every 100 generations a mutation occurs on average and a new strategy is introduced in the population. In order to explore the corners of the search space the conditional probabilities to cooperate p_x of the new strategy are drawn according to the U-shaped density distribution $U(z) = (\pi z(1-z))^{-1/2}$ in the interval

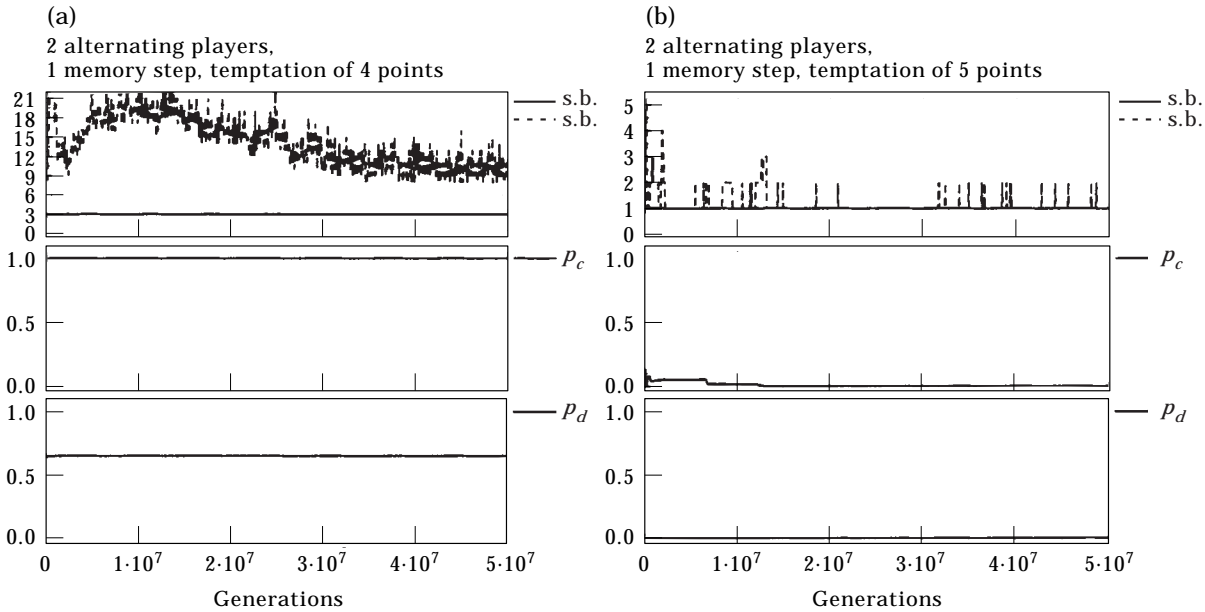


FIG. 2. Reactive strategies in the alternating PD crucially depend on the pay-off values. In (a) the cooperative strategy *generous TFT* wins in perfect agreement with theoretical calculations (Nowak & Sigmund, 1994). The increased temptation in (b) leaves cooperative strategies without hope. Soon *allD* establishes leading to a poor mean population pay-off close to 1 instead of 3 points. Legend of top graphs: — population pay-off; --- number of strategies.

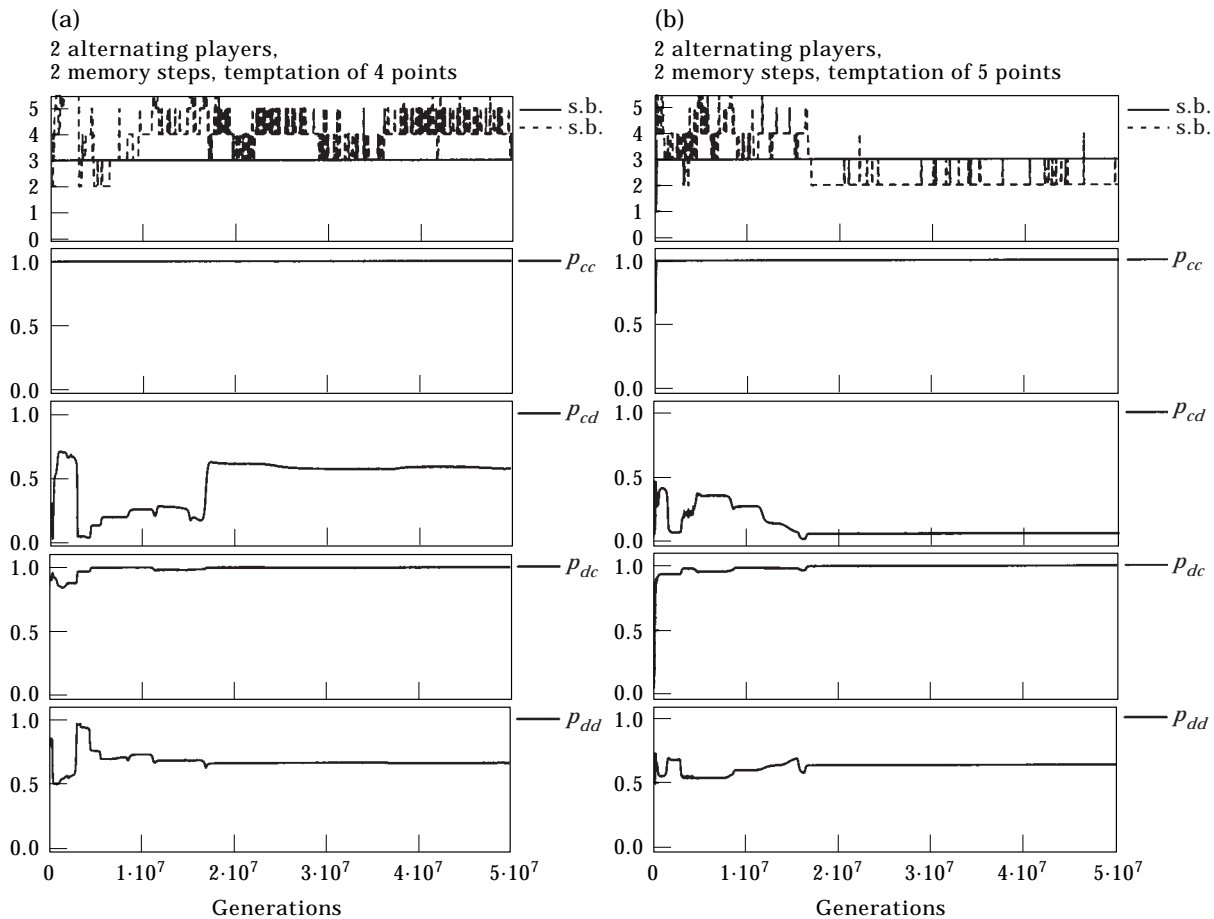


FIG. 3. In (a) we used the same configuration as Nowak & Sigmund (1994). The winning strategy is *generous TFT* as in Fig. 2(a). Reasons for differences to the cited article are given in the text; (b) shows simulations using an increased temptation of $T = 5$ as in Freat (1994). The winning strategy follows the rule of *Firm but Fair* and is in perfect agreement to Freat's results. Legend of top graphs: — population pay-off; --- number of strategies.

[0.001, 0.999]. The reduced interval specifies a minimal probability to misimplement a decision or to misinterpret the opponents moves. A new strategy is introduced with an initial frequency of 0.0011. Whenever the frequency of any strategy drops below a fixed noise level of 0.001 it is removed.

3. Results

Due to the complexity and nonlinearity of the system our results are based on extensive numerical simulations. Since the dimension of the strategy space increases exponentially with the memory length of the players, we were forced to limit our simulations to $N, M \leq 4$. In order to obtain reliable results we made several simulation runs for every configuration. Nevertheless, in most cases the strategy space is far too big to perform any statistics. For this reason all strategies present at the end of all simulation runs

were merged into a new population. In this population the frequencies of the strategies were set proportional to the population pay-off weighted by their frequencies at the end of the respective simulation run. The evolution of the strategy achieving the highest pay-off in this new population is shown in the figures.

Simulations were made for three different values of the temptation T in order to study its influences on the strategies adopted by the players. In figures labeled (a) the pay-off values $R = 3, S = 0, T = 4$ and $P = 1$ were used and in those labeled (b) the temptation was increased to $T = 5$. Results for $T = 3.5$ are discussed in the text. Only $T = 4$ satisfies the symmetry condition $R + P = T + S$. $T = 5$ corresponds to the traditional values introduced by Axelrod (1984) and simulations for $T = 3.5$ were made to investigate whether differences in the strategies result from the broken symmetry condition or the value of the temptation.

3.1. 2 PLAYERS, M STEPS MEMORY

First we consider the alternating iterated PD for two players having different memory lengths ($M = 1, \dots, 4$). Figures 2–5 display the time evolution of the most successful strategies out of 10 simulation runs. The top graph shows the mean population pay-off as well as the number of strategies in the population. The graphs below show the mean conditional probability to cooperate grouped according to the possible outcomes of the last unit of rounds.

Successful strategies adopted by the players differ considerably depending on their memory length. Players without memory in the iterated PD are in an analogous situation as those playing the PD only one round. Thus, defecting strategies will always win and may intrude into any population. Without memory, there is no possibility to reward cooperation without giving defectors a chance to exploit.

Players with one step memory adopt so-called reactive strategies considering only the opponents last

move (Fig. 2). Nowak & Sigmund (1994) showed that Tit for Tat (*TFT*)—do whatever the opponent did in the previous round ($p_c \approx 1, p_d \approx 0$)—may establish cooperation if eqn (3) holds. *TFT* will then be outperformed by a more tolerant strategy called *generous TFT* ($p_c \approx 1, p_d = (R - P)/(T - P) = 2/3$). Amazingly this corresponds exactly to the results for the simultaneous PD. Increasing the temptation to $T = 5$ destroys this analogy and again *allD* wins ($p_x \approx 0$). For $T \neq 4$ the symmetry between the players decisions is broken and the success of a player's strategy also depends on the position of making a decision in each unit of rounds. Surprisingly for $T = 5$ the leaders handicap can not be outweighed by averaging the achieved points over both positions. Since for $T = 3.5$ strategies as in Fig. 2(a) evolve we see that the symmetry condition has no effect on the success of cooperative strategies. The increased temptation is solely responsible for the breakdown of cooperation. For $T = 5$ no cooperative strategy is

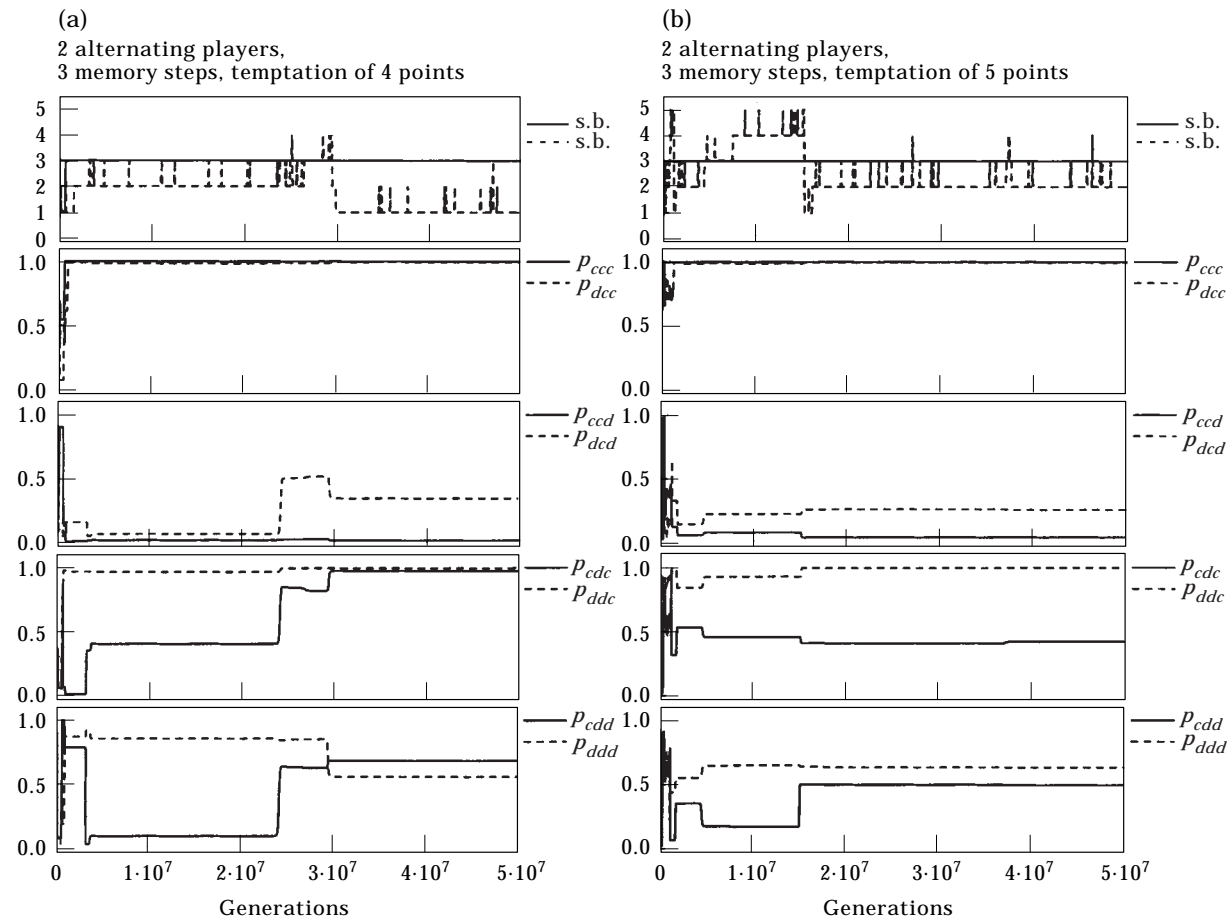


FIG. 4. Almost identical strategies establish for two players with three steps memory regardless of the temptation T . Unaffected by the longer memory both strategies follow the concept of *Firm but Fair*. The only, though important, difference concerns p_{cdc} which is much smaller for $T = 5$. This means that such strategies try to exploit dumb cooperators—a direct consequence of the increased temptation. Legend of top graphs: — population pay-off; --- number of strategies.

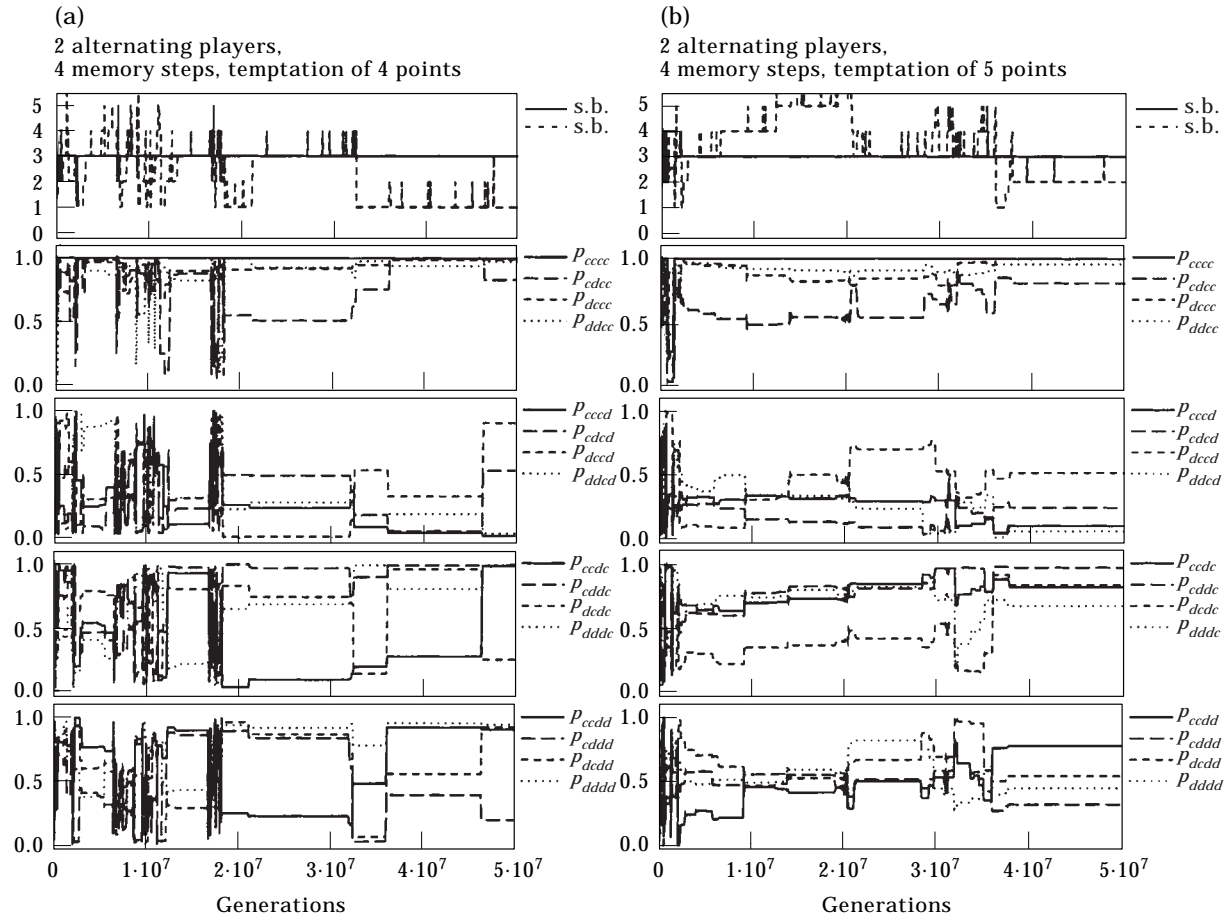


FIG. 5. For two players with four steps memory the winning strategy for $T = 4$ in (a) shows interesting reactions (see text) whereas for $T = 5$ in (b) the strategy does not take advantage of the longer memory still following the concept of *Firm but Fair*. Legend of top graphs: — population pay-off; --- number of strategies.

immune to invasion. For example the following cycle of strategies dominating the population is possible: *allD* followed by TFT, which is in turn vulnerable to the strategy $p_c \approx 0, p_d \approx 1$ and then back to *allD* [see e.g. Nowak *et al.* (1995) for an extensive discussion of such cycles in the simultaneous PD]. The simulations clearly show that *allD* is the strongest attractor. For $T \leq 4$ TFT is no more vulnerable to $p_c \approx 0, p_d \approx 1$ and thus prevents cycles as described above. For such values of T a population dominated by *generous TFT* can be invaded only by strategies with a higher p_d but these will be eliminated by defecting mutants. The five simulation runs for $T = 3.5$ ended with 31 nice strategies ($p_c > 0.99$), the 10 runs for $T = 4$ with 110 nice strategies and for $T = 5$ none of the 11 strategies from 10 runs was nice.

In Fig. 3 we show simulations of configurations discussed (a) by Nowak & Sigmund (1994) for $T = 4$ and (b) by Frean (1994) for $T = 5$. Both simulations are in good agreement with the work cited above except that for $T = 4$ the mean probability to

cooperate $p_{cd} \approx 0.6$ rather than $p_{cd} \approx 0$. Note that, nevertheless, the winning strategy is still in good agreement: $p_{cc} = 0.999, p_{cd} = 0.009, p_{dc} = 0.994$ and $p_{dd} = 0.807$. The reason for this difference is the small frequency of the winning strategy ($\sim 7\%$) in the population. Such strategies with $p_{cc} \approx 1, p_{cd} \approx 0, p_{dc} \approx 1$ and $p_{dd} = x$ are called *Firm but Fair*. They cooperate if the opponent opted for (C) in the last round, defect when played for a sucker and try to re-establish cooperation after a unit of mutual defection with the probability $p_{dd} = x$. For $T = 3.5$ strategies follow the same concept with $p_{dd} \approx 0.8$. Interestingly the success of *Firm but Fair* remains unaffected by the value of the temptation. The five simulation runs for $T = 3.5$ ended with 15 nice strategies ($p_{cc} > 0.99$), the 10 runs for $T = 4, 5$ ended with 27 nice strategies each.

Increasing the memory to three steps (Fig. 4) leads to strategies close to those with two steps memory following the concept of *Firm but Fair* [see Fig. 3(b)] regardless of the temptation ($T = 3.5, 4, 5$). It follows

that the increased memory has only minor influences on the strategies adopted by the players. Thus, the increased information available to the players does not seem to be crucial for next decision. The five simulation runs for $T = 3.5$ ended with nine nice strategies ($p_{ccc} > 0.99$), the 10 runs for $T = 4, 5$ ended with 25 nice strategies each.

In Fig. 5 the memory length is further increased to four memory steps. The resulting strategies are nice ($p_{ccc} > 0.99$) but strategies have a hard time to dominate the population. Rather short metastable states are observed and we may assume that continuing the simulations will lead to further similar transitions. Since the strategies remain cooperative regardless of these transitions the population payoff remains essentially unaffected. In general all strategies still fit well in the concept of *Firm but Fair* regardless of the temptation. The strategy in (a) for $T = 4$ shows interesting deviations from this concept: it returns to

business after punishing or being punished for defection ($p_{dcd}, p_{cdc} \gg 0$), tries to exploit dumb cooperators ($p_{dcd} \ll 1$) and returns to cooperation after *DD* except when played for a sucker in the previous round. Results for $T = 5$ (b) are surprisingly close to *Firm but Fair* with the only exception that as for $T = 4$ often punishment for defecting is tolerated ($p_{dcd} \approx 0.5$). For $T = 3.5$ (not shown) the following differences from *Firm but Fair* are observed: an opponent's defection may be punished one round later ($p_{dcd} \ll 1$) and dumb cooperators are exploited ($p_{dcd} \ll 1$) as for $T = 4$. The five simulation runs for $T = 3.5$ ended with eight nice strategies ($p_{ccc} > 0.99$), the 10 runs for $T = 4, 5$ ended with 22 nice ones out of 24 and 25 nice ones out of 31, respectively.

3.2. N PLAYERS, N STEPS MEMORY

Second we turn to results of simulations of the alternating iterated PD for N players remembering

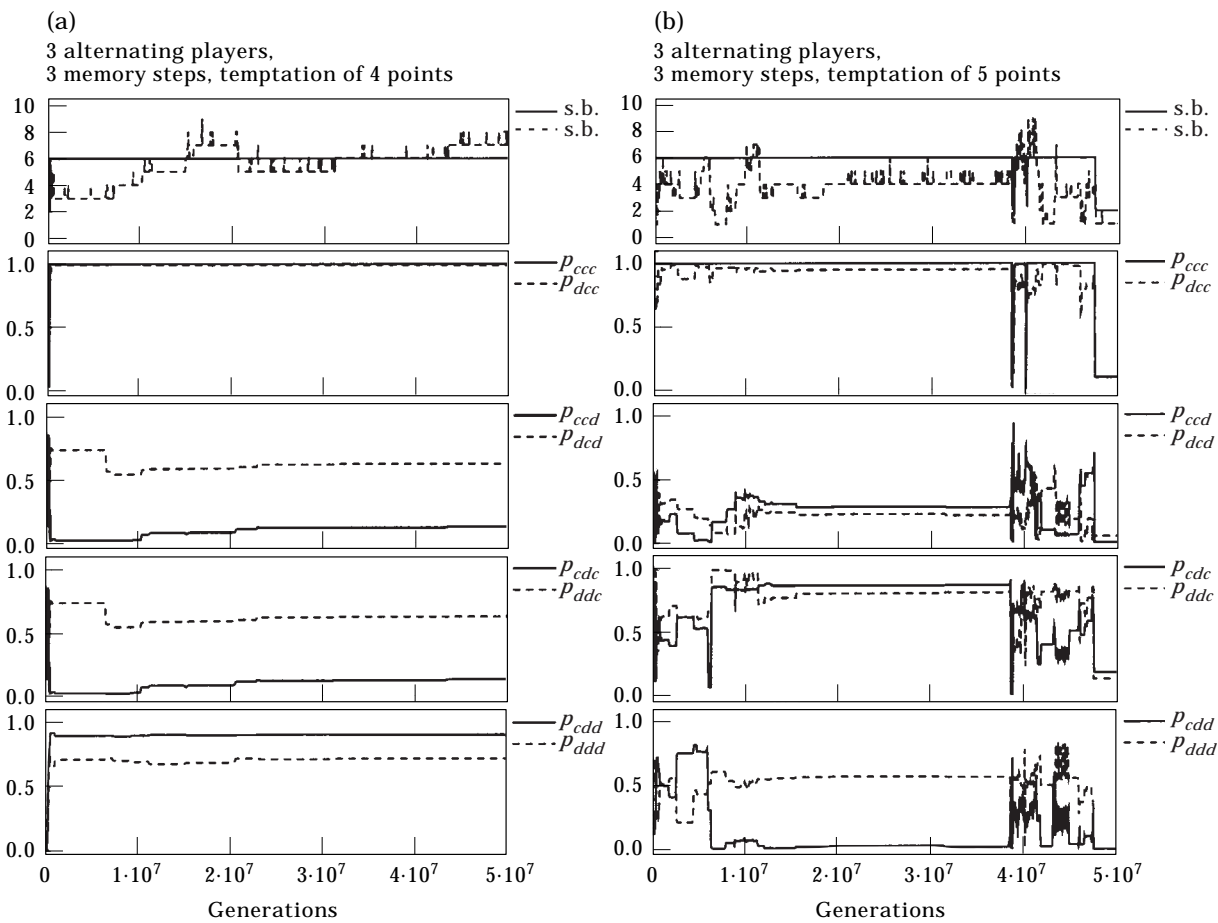


FIG. 6. Amazingly fast a successful and cooperative strategy is found in (a) for three players with three steps memory and $T = 4$. The steadily increasing number of strategies in the population indicates a lowered selection pressure. In (b) for $T = 5$ cooperative strategies dominate the population for a very long time but every now and then cooperation breaks down and *allD* establishes such as around $4.75 \cdot 10^7$ generations. These breakdowns are accompanied by a decrease in the population pay-off. Legend of tops graphs: — population pay-off; --- number of strategies.

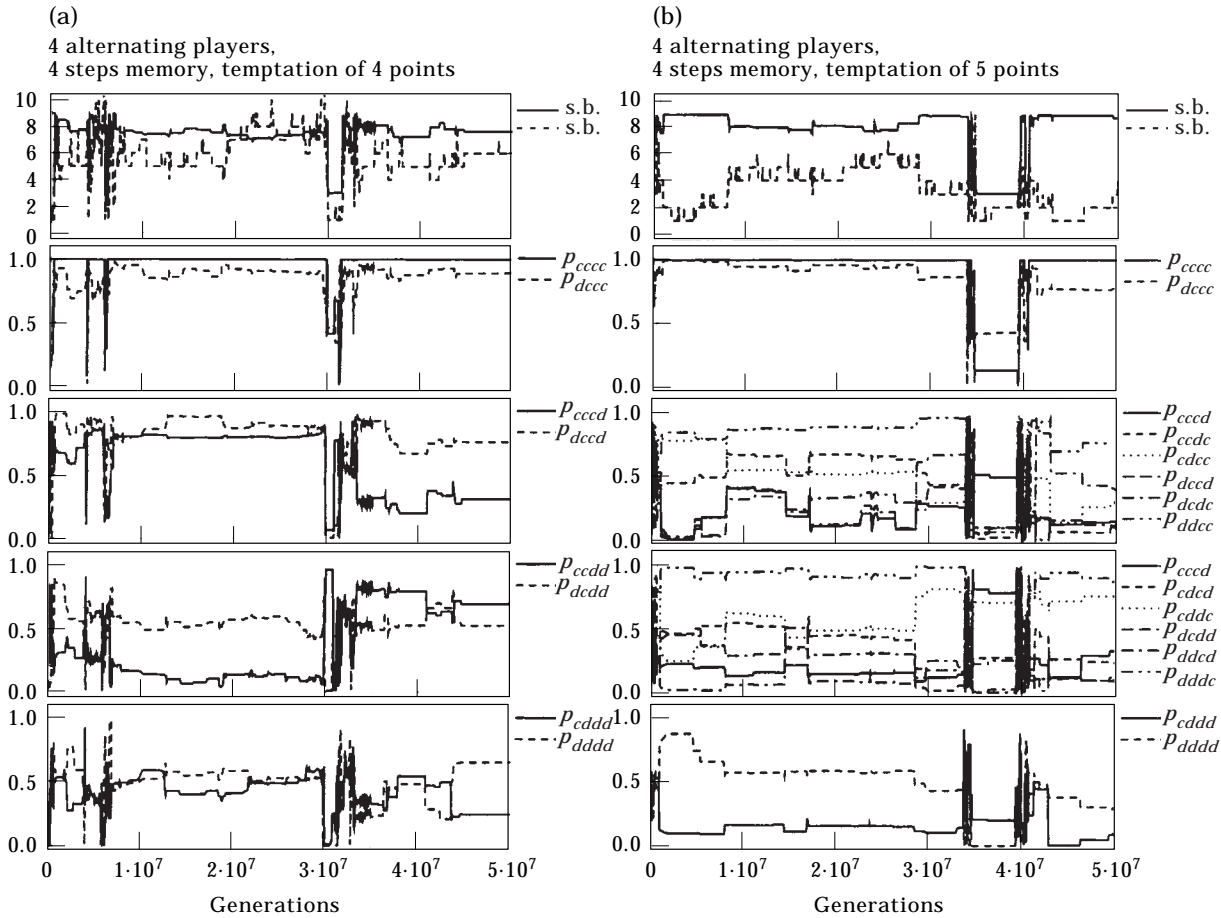


FIG. 7. Simulations of the four players four steps memory PD for $T = 4, 5$ are dominated by nice strategies for long periods but strategies seem to have a hard time to establish. Due to frequent transitions between different strategies neither for $T = 4$ in (a) nor for $T = 5$ in (b) the maximum population pay-off is achieved. Eventually cooperation breaks down entirely leading to a very poor population pay-off. Legend of top graphs: — population pay-off; --- number of strategies.

the last N moves. Figures 6 and 7 show the time evolution of the winning strategy for $N = 3, 4$. The top graph again shows the mean population pay-off as well as the number of strategies in the population. The graphs below show the mean conditional probabilities to cooperate grouped according to the number of defecting opponents given that the player itself opted for C or D .

Simulations for $N = 2$ are discussed in the previous section (see Fig. 3). In Fig. 6 we show simulations of the alternating PD for three players with three steps memory. In (a) with $T = 4$ the winning strategy was nice ($p_{ccc} > 0.99$) whereas in (b) for $T = 5$ nice strategies dominated the population for a very long period but cooperation breaks down towards the end of the simulation. Nevertheless *allD* does not seem to be the end of the story but rather a stage in the progression of evolution. Around generation $3.8 \cdot 10^7$ a similar breakdown is observed but cooperation is

re-established quickly. Note that due to the symmetries discussed in Section 2 for $T = 4$ the following identities hold: $p_{ccd} \equiv p_{cdd}$ and $p_{dcd} \equiv p_{ddc}$. For $T = 3.5$ very cooperative strategies dominate the population ($p_x \gg 0$ except $p_{cdd}, p_{ddd} \ll 1$; not shown). No breakdown of cooperation is observed. It follows that the stability of strategies is sensitive to changes in the temptation but it hardly matters whether symmetry condition given in eqn (3) holds. The five simulation runs for $T = 3.5$ ended with 23 nice strategies ($p_{ccc} > 0.99$), the 10 runs for $T = 4$, with 51 nice strategies but only 17 nice ones out of 30 in 10 runs for $T = 5$.

The time evolution of successful strategies in the alternating four player PD is shown in Fig. 7. Due to the enormous CPU time requirements of these simulations we limited our simulations to five simulation runs for temptations $T = 4, 5$ and three control runs for $T = 3.5$. For $T = 4, 5$ once in a while

cooperation breaks down and defecting strategies establish [see e.g. (b) from around generation $3.4 \cdot 10^7$ till $4 \cdot 10^7$. Simulations for $T=3.5$, however, do not display such breakdowns of cooperation (not shown). Nevertheless strategies are close to those for $T=5$ except that they are generally a bit more cooperative. Winning strategies may be characterized as follows: in all configurations the winning strategy was nice ($p_{cccc} > 0.99$) but none was able to achieve the maximum population pay-off of nine points. Often the change from $C \rightarrow D$ of a single opponent is punished (third graph from the top). The probability of punishing such a change increases with T as intuitively expected. The reverse change from $D \rightarrow C$ is rewarded for $T=3.5$ (not shown) and 5 [fourth graph from the top in Fig. 7(b); $p_{edcc}, p_{dddc} \gg 0$, all other $p_s \ll 1$] whereas for $T=4$ the strategies are rather undecided because it remains unclear which player switched from D to C . This follows from the symmetry condition discussed above leading to the following identities: $p_{cccd} \equiv p_{ccdc} \equiv p_{edcc}, p_{ccdd} \equiv p_{cedd} \equiv p_{cedc}$ and in analogy when the player itself opted for D . The five simulation runs ended with 25 nice strategies out of 26 for $T=4$ and with 15 nice strategies out of 16 for $T=5$.

4. Conclusions

In the alternating two player PD we see that cooperative solutions do exist regardless of the memory size M and the temptation T . The only exception is the configuration with one step memory and $T=5$ where *allD* is the clear winner. In this situation the increased temptation and not the broken symmetry between the players decisions hinders cooperation from establishing. *Generous TFT* turns out to be evolutionary stable only for $T \leq 4$. All strategies found for memory sizes $M \geq 2$ fit well in the concept of *Firm but Fair* outlined by Frean (1994) and Nowak & Sigmund (1994). The longer memory allows refined actions but generally strategies remain close to *Firm but Fair*. It is somewhat surprising to see that the players do not seem to take particular advantage of the longer memory. This implies that the crucial information is contained in the knowledge about the outcome of the last unit of rounds. Increasing the temptation from $T=3.5, 4$ to 5 decreases the readiness to restore cooperation. This obviously helps to prevent strategies from being played for a sucker. Such characteristics become the more important the higher the temptation T . For $T=3.5$ only nice strategies were found ($p_{c\dots c} > 0.99$) regardless of the memory size.

In the alternating N player PD ($N=3, 4$) cooperative solutions to the dilemma still exist but they do not seem to be evolutionary stable any more. For $T=3.5$ and $T=5$ cooperative strategies follow a pattern where a change from $D \rightarrow C$ is rewarded, the reverse is punished and after mutual defection an effort is made with a certain probability to re-establish cooperation. This pattern is less obvious for $T=4$ because these strategies decide only on the number of defecting opponents. Note that this pattern corresponds again to the concept of *Firm but Fair* for $N=2$.

From our simulation results we may assume that in the alternating N player N step memory PD ($N > 2$) at least for higher values of T no stable strategy exists. For $T=4, 5$ we observe every now and then a temporary breakdown of cooperation. This is in good agreement with experiments done by Milinski *et al.* (1990) studying the behaviour of a shoal of four sticklebacks on predator inspection. They observed mainly pairwise interactions between the fish during inspection as well as flight. Even when all four fish approached the predator they could make out two pairs. Our results support their conclusion that larger inspecting group sizes are possibly made up of several independent reciprocal pairs. Only in the alternating two player PD evolutionary stable cooperative strategies seem to exist.

We thank Olaf Stenull for extensive, stimulating discussions and for helpful comments on the manuscript.

REFERENCES

- AXELROD, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- AXELROD, R. & HAMILTON, W. D. (1981). The evolution of cooperation. *Science* **211**, 1390–1396.
- BOERLIJST, M. C., NOWAK, M. A. & SIGMUND, K. (1997). The logic of contrition. *J. theor. Biol.* **185**, 281–293.
- DAWES, M. R. (1980). Social dilemmas. *Ann. Rev. Psychol.* **31**, 169–193.
- FLOOD, M. (1958). Some experimental games. *Mgmt Sci.* **5**, 5–26.
- FREAN, M. R. (1994). The prisoner's dilemma without synchrony. *Proc. R. Soc. Lond. B* **257**, 75–79.
- HAUERT, C. & SCHUSTER, H. G. (1997). Effects of increasing the number of players and memory size in the iterated Prisoner's Dilemma: a numerical approach. *Proc. R. Soc. Lond. B* **264**, 513–519.
- LEIMAR, O. (1997) Repeated games: a state space approach. *J. theor. Biol.* **184**, 471–498.
- MAYNARD SMITH J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
- MILINSKI, M. (1987). Tit for Tat in sticklebacks and the evolution of cooperation. *Nature, Lond.* **325**, 434–435.
- MILINSKI, M., PFLUGER, D., KÜLLING, D. & KETTLER, R. (1990). Do sticklebacks cooperate repeatedly in reciprocal pairs? *Behav. Ecol. Sociobiol.* **27**, 17–21.

- NOWAK, M. A. & SIGMUND, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature, Lond.* **364**, 56–58.
- NOWAK, M. A. & SIGMUND, K. (1994). The alternating prisoner's dilemma. *J. theor Biol.* **168**, 219–226.
- NOWAK, M. A., SIGMUND, K. AND EL-SEDY, E. (1995). Automata, repeated games and noise. *J. Math. Biol.* **33**, 703–722.
- PACKER, C. (1977). Reciprocal altruism in *Papio anubis*. *Nature, Lond.* **265**, 441–443.
- TRIVERS, R. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57.
- WILKINSON, G. S. (1984). Reciprocal food-sharing in the vampire bat. *Nature, Lond.* **308**, 181–184.