J. theor. Biol. (1997) **185**, 281–293

The Logic of Contrition

MAARTEN C. BOERLIJST*‡, MARTIN A. NOWAK* AND KARL SIGMUND†

Department of Zoology, University of Oxford, Oxford OX1 3PS, U.K. and the † Institut für Mathematik, Universität Wien Strudlhofgasse 4, A-1090 Vienna, Austria

(*Received on* 11 *January* 1996, *Accepted on* 10 *June* 1996)

A highly successful strategy for the Repeated Prisoner's Dilemma is Contrite Tit For Tat, which bases its decisions on the "standings" of the two players. This strategy is as good as Tit For Tat at invading populations of defectors, and much better at overcoming errors in implementation against players who are also using it. However, it is vulnerable to errors in perception. In this paper, we discuss the merits of Contrite Tit For Tat and compare it with other strategies, like Pavlov and the newly-introduced Remorse. We embed these strategies into an eight-dimensional space of stochastic strategies which we investigate by analytical means and numerical simulations. Finally, we show that if one replaces the conventions concerning the "standing" by other, even simpler conventions, one obtains an evolutionarily stable strategy (called Prudent Pavlov) which is immune against both mis-perception and mis-implementation.

1. Introduction

Tit For Tat has an Achilles' heel: it is vulnerable to errors [see Axelrod & Hamilton (1981), Axelrod (1984), Molander (1985), Müller (1987), Axelrod & Dion (1988), Bendor *et al*. (1991), Bendor (1993), Kollock (1993), Nowak & Sigmund (1993b), Nowak *et al*. (1995a)]. If a Tit for Tat (*TFT*) player erroneously plays *Defect* against another *TFT*-player, this leads to a long vendetta. There are several ways to overcome this problem. One can, for instance, play *Generous Tit For Tat* (*GTFT*): always cooperate if the other player cooperated in the previous round, but defect only with a certain probability if he defected [see Molander (1985) and Nowak & Sigmund (1992)]. Alternatively, one could use the strategy *PAVLOV*: cooperate if and only if you and your opponent used the same move in the previous round [see Kraines & Kraines (1988), Fudenberg & Maskin (1990) or Nowak & Sigmund (1993b)]. Both strategies are error-proof: a mistaken defection is quickly corrected, and mutual cooperation resumed.

Another error-correcting strategy has been proposed by Sugden (1986) in his seminal book on '*The Evolution of Rights, Co-operation and Welfare*'. This is *Contrite Tit For Tat*, or *cTFT* [see also Boyd (1989) and Wu & Axelfod (1995)]. Like *GTFT* and *PAVLOV*, this is a memory one-strategy: it decides according to the outcome of the previous round. However, in contrast to its two rivals, this outcome does not only depend on the moves of the two players (which can be *C* or *D*, cooperate or defect), but also on their *standing*, which can be *g* ("good") or *b* ("bad"). A player is in good standing if he has cooperated in the previous round, or if he has defected while provoked (i.e. while he was in good standing and the other player was not). In every other case defection leads to a bad standing. The strategy *cTFT* begins with a cooperative move, and cooperates except if provoked.

If two *cTFT*-players engage in a repeated Prisoner's Dilemma, and if the first player defects by mistake, he loses his good standing. In the next round, he will cooperate, whereas the other *cTFT*-player will defect without losing his good standing. Then both players will be in good standing and resume their mutual cooperation in the following round. This strategy is

‡ Author to whom correspondence should be addressed.

0022–5193/97/070281 + 13 $25.00/0/jt9600326

related to Dawkins' (1989) Remorseful Prober, who defects once in a while but accepts retaliation in the following round without complaint.

As Sugden has shown, *cTFT* is evolutionarily stable. Moreover, it is as good as *TFT* in invading a population of defectors. In contrast, *PAVLOV* and *GTFT* fare both very poorly in such an environment, and need a "catalyser" to create the type of cooperative environment in which they can thrive.

On the other hand, the additional complexity of the *cTFT* strategy has its drawbacks. In particular, while *cTFT* is immune to errors in the implementation of a move, it is not immune to errors in the perception of a move. If, in a match between two *cTFT* players, one player mistakenly believes that the other is in bad standing, this leads to a sequence of mutual backbiting, just as with *TFT*. [Errors in perception—rather than implementation—have been studied in Miller (1996), Kollock (1993), Nowak *et al*. (1995b).]

In this paper, we discuss the relative merits of all (stochastic or deterministic) memory one strategies with or without standing. *cTFT* is not the only evolutionarily stable rule which is Pareto-optimal (and hence yields the maximal pay-off if the whole population adopts it). Depending on the exact pay-off values, either *PAVLOV* or another strategy called *REMORSE* has the same qualities. A player using the *REMORSE* strategy cooperates if he was in bad standing in the previous round, or if both players cooperated. This strategy, again, is error-correcting. Indeed, suppose that both players use *REMORSE*. If the second player defects by mistake, he cooperates in the next round, whereas the first player defects and remains in good standing. In the following round, both players defect and obtain a bad standing; from then onward, both resume cooperation.

We discuss *cTFT*, *PAVLOV* and *REMORSE* with analytical methods and numerical simulations, embedding them in a large class of stochastic strategies. Finally, we show that by replacing the conventions concerning the "standing" by another set (which is even easier to implement, and only depends on an "internal variable") one is led to a *PRUDENT-PAVLOV* strategy which is an ESS and immune against errors both in implementing and in perceiving moves.

2. Preliminaries on the Repeated Prisoner's Dilemma

The Prisoner's Dilemma (or PD) is a game between two players each having two options, namely to cooperate (play C) or to defect (play D). If both cooperate, they get a reward R higher than the punishment P which they receive if both defect. If one player defects and the other cooperates, the defector get the pay-off T (for temptation) and the cooperator the sucker's pay-off S. We shall always assume

$$T > R > P > S \quad (1)$$

so that the option D dominates C (it is better no matter what the other player chooses). But if both players use D, they fail to get the reward.

In the iterated PD, the game is played for several rounds. We shall assume that there is a constant probability w for another round. The length of the game is a stochastic variable with mean value $1/(1-w)$. A strategy for the iterated PD is a program telling the player in each round whether to chose C or D (this can be a randomised decision: cooperate with such and such a probability). If A_n is the pay-off for one player in the n-th round, his expected pay-off is $\Sigma A_n w^n$ (note that w^n is the probability that an n-th round occurs). We shall mostly be interested in large w (close to 1). Frequently, the limiting case $w = 1$ is considered (the infinitely repeated PD). In this case, the pay-off is the limit of the mean

$(1/n)(A_1 + \cdots + A_n)$, for $n \to \infty$ (if it exists).

We shall assume

$$2R > T + S \quad (2)$$

so that it is better for the two players to cooperate jointly rather than to alternately defect.

Let us now assume that in every round, each player is provided with a *standing*, which can be g (good) or b (bad). In the following round, the player acts (i.e. opts for C or D) and obtains a new standing which depends on his action and on the previous standing of both players. As mentioned in the introduction, the rules for updating the standing are the following: if the other player has been in good standing, or if we both have been in bad standing, I receive a good standing if I cooperate, and a bad standing otherwise. If I have been in good standing and the other player in bad standing, I receive a good standing no matter what I am doing.

Thus, if I cooperate in a given round, I will always obtain a good standing: but if I defect, I will be in good standing only if, in the previous round, I have been in good standing and my opponent has been in bad standing.

In a given round, a player can be in three possible states: Cg, Dg and Db: the first means that he has cooperated (which automatically entails good standing), the second that he has defected with good reason, the third that he has wantonly defected. The *state* of the game in a given round is made up of the states of the first and the second player. There are nine such combinations: (Cg, Cg), (Cg, Dg),

(Cg, Db), (Dg, Cg), (Dg, Db), (Db, Cg), (Db, Dg), (Db, Db) and (Dg, Dg). It is easy to check that this last state can never be reached: we therefore omit it, and number the remaining eight states in this order.

cTFT is the strategy which cooperates except if it is in good standing and the other player is not, whereas *REMORSE* is the strategy which cooperates only if it is in bad standing, or if both players had cooperated in the previous round.

3. In Search of Stability

A strategy $\hat{S}$ is said to be an evolutionarily stable strategy, or ESS, if in a population where all members adopt it, no other strategy can invade under the effect of selection. More precisely, if $A(S, S')$ is the expected pay-off for an S-player in a population of S'-players, then $\hat{S}$ is an ESS if for all strategies S different from $\hat{S}$ one has $A(S, \hat{S}) \leqslant A(\hat{S}, \hat{S})$ and, if equality holds, $A(S, S) < A(\hat{S}, S)$ [see Maynard Smith (1982)].

It is easy to see that for the infinitely repeated Prisoner's Dilemma, i.e. for $w = 1$, there exists no ESS. This is simply due to the fact that two strategies differing only in their first—say—three hundred moves will have exactly the same pay-off.

But as shown in Sugden (1986), for $w < 1$ the strategy *cTFT* is evolutionarily stable in a very important sense: if there is a small, but non-vanishing probability of mis-implementing a move, every strategy that deviates, against a *cTFT*-player, from what the *cTFT*-rule would prescribe, fares less well than it would have by following this rule. Note that if there is such an error probability, every finite sequence of moves will have a positive probability. [See Selten (1975), Selten & Hammerstein (1984), and Boyd (1989) where the connection with Selten's concept of a perfect equilibrium is discussed.]

The basic idea of Sugden's proof allows to decide for *every* deterministic rule $\hat{S}$ based on finitely many states whether it is evolutionarily stable in the sense defined above. Because of the error probability, every state can be reached with positive probability. Let us start in any of the possible states, assuming for the moment that no error will occur in the following rounds, and let us follow the fate of a player invading a $\hat{S}$ population. Since the next move of his adversary is always specified, there are only two possible states that can be reached in the next round, depending on whether the invader uses C or D. From each of these states, two states can be reached in turn, etc. Since there are only finitely many states, each branch of the game-tree must eventually return to a state it had visited before. Therefore, it is possible to compute the pay-off along every branch, discounting by the factor w at every step. One of the branches issued from each state describes what happens if I use $\hat{S}$ myself. If this always yields the highest pay-off, and no alternative does, then $\hat{S}$ is evolutionarily stable, provided the probability for mistakes in implementation is sufficiently small.

In Fig. 1 we check this for *PAVLOV*. Two arrows issue forth from each state, depending on whether the invader plays C or D against his $\hat{S}$-adversary. The vertices of the graphs describe the invader's state in the first (or upper) position, and the state of his opponent in the second (or lower) position. The arrows describe the possible transitions, which only depend on the invaders choice, since the opponent's moves are specified by $\hat{S}$. The solid arrow indicates the move the invader would choose if he were also a $\hat{S}$-strategist. We see in Fig. 1 that *PAVLOV* is an ESS if and only if $T + wP < R + wR$, as has been shown by Harrington & Axelrod (unpublished data). The critical decision occurs when we are in (D, D) or (C, C) and I have to decide whether to get two R's in succession, or a T followed by a P. On the other hand, Fig. 2 shows that *TFT* is never an ESS if w is close to 1: in (C, D), my best move leads to (C, C).

In Fig. 3, we see that *cTFT* is always an ESS if w is close to 1, and in Fig. 4 that *REMORSE* is an ESS if and only if $T + wP > R + wR$ (the opposite as with *PAVLOV*). The critical case, here, comes when in state (Dg, Db) or (Cg, Db). Defecting twice (as *REMORSE* specifies) will get me $T + wP$. Cooperating twice yields $R + wR$. We note that *REMORSE* can handle *AllD* very well and is threatened by more cooperative

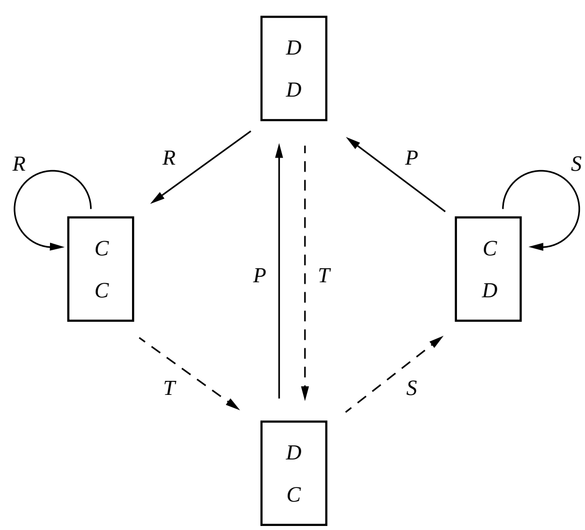

FIG. 1. *PAVLOV* is an ESS if $T + wP < R + wR$. Solid lines indicate the moves specified by the *PAVLOV* strategy; dotted lines indicate the alternative moves. See text for further explanation.

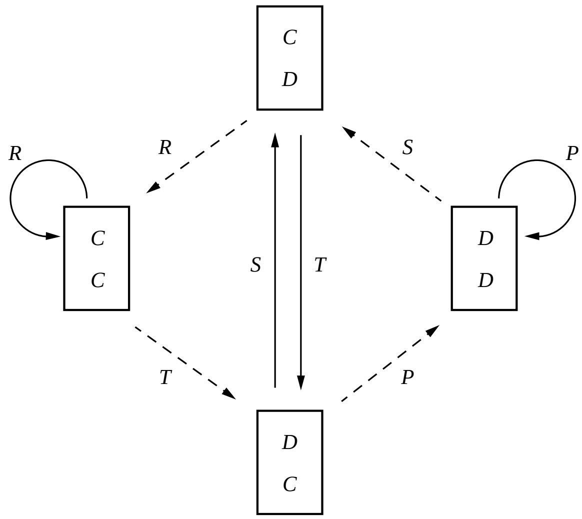

FIG. 2. *TFT* is no ESS for large *w*.

strategies; *PAVLOV* exploits *AllC* to the hilt, but is endangered by *AllD*.

One can use the same method to verify, for instance, that *AllD* and *GRIM* are evolutionary stable rules (*GRIM* cooperates only if both players cooperated in the previous round. If one defects against a *GRIM*-player, that player will never revert to cooperation.) For certain pay-off values, the strategy *WEAKLING* is also an ESS: it cooperates if and only if it is in bad standing. However, these strategies are far from optimal. If a population is stuck with such a strategy, it does very poorly (the average pay-off is P for *AllD* and *GRIM*, and $(R + P)/2$ for *WEAKLING*). In contrast, if a whole population adopts *PAVLOV*, *GTFT*, *cTFT* or *REMORSE*, it will on average obtain the pay-off R per round.

So far, we looked at errors in implementing a move. But there also exist, as we know from everyday life, errors in understanding which can threaten cooperation. *cTFT* is not immune to misperception of the other's move, as can be seen from the following table, where the first row is the sequence of my states, as I perceive them; the second the sequence of the opponent's states, as I perceive them (my error occurs in the second round, indicated by the asterisk) whereas the third and fourth row are the sequences of my (resp. my opponent's) true moves.

Cg	*Cg*	*Dg*	*Cg*	*Dg*	...
Cg	*Db**	*Cg*	*Db*	*Cg*	...
Cg	*Cg*	*Db*	*Cg*	*Db*	...
Cg	*Cg*	*Cg*	*Dg*	*Cg*	...

The average pay-off, after the mistake, is $(T + S)/2$, which is less than R.

Similarly, *REMORSE* is not immune to misperception of the other's move:

Cg	*Cg*	*Dg*	*Db*	*Cg*	*Dg*	*Dg*	...
Cg	*Db**	*Cg*	*Db*	*Db*	*Db*	*Cg*	...
Cg	*Cg*	*Db*	*Db*	*Cg*	*Db*	*Db*	...
Cg	*Cg*	*Cg*	*Dg*	*Dg*	*Db*	*Cg*	...

The average pay-off, after the mistake, is $(T + S + 2P)/4$, which is less than R.

In contrast to this, *PAVLOV* is immune to misperception of the other's move (or the own, for that matter):

C	*C*	*D*	*D*	*C*	...
C	*D**	*C*	*D*	*C*	...
C	*C*	*D*	*D*	*C*	...
C	*C*	*C*	*D*	*C*	...

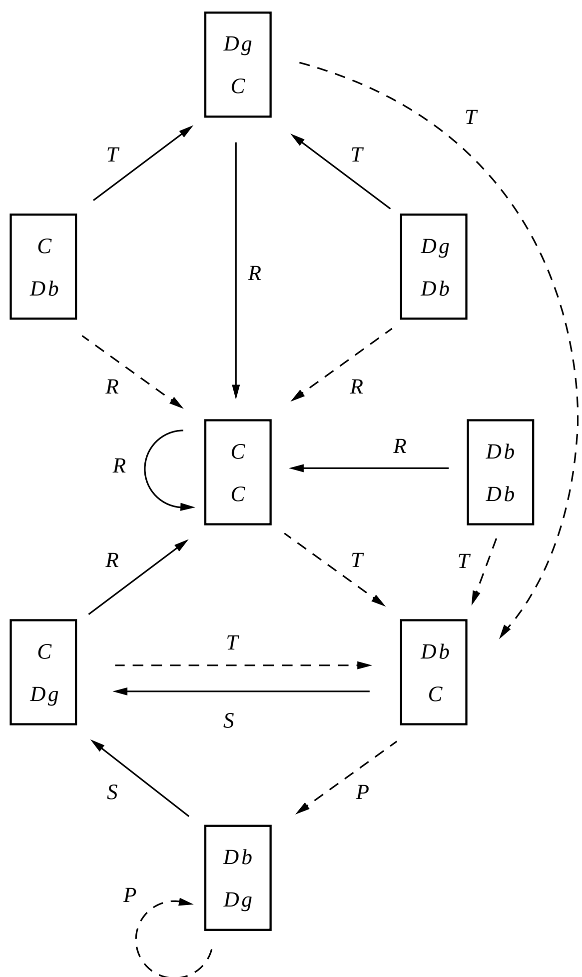

FIG. 3. *cTFT* is an ESS for large *w*.

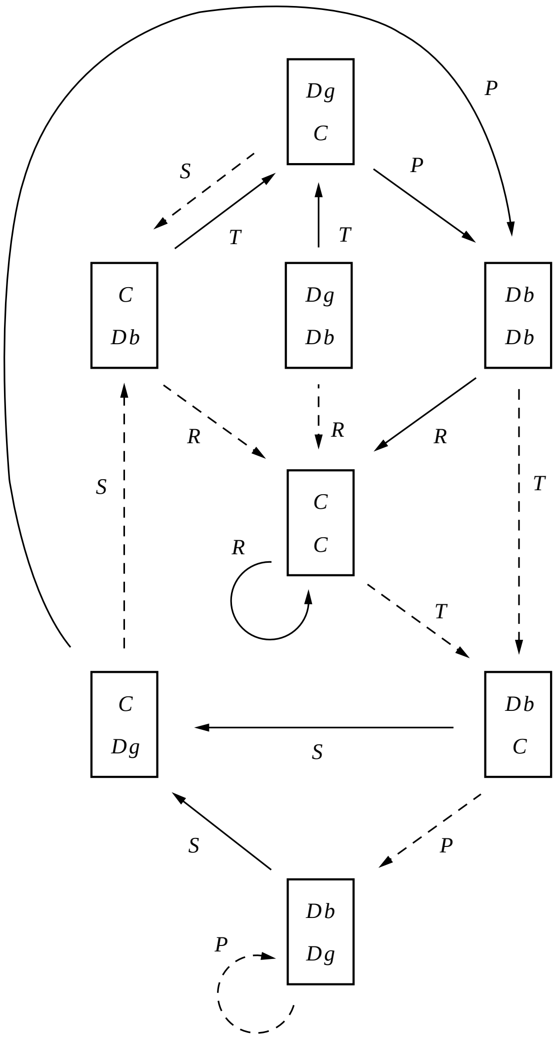

FIG. 4. *REMORSE* is an ESS if $T + wP > R + wR$.

The error is quickly corrected and the average pay-off remains R. [For a precise computation of the effect of the errors in perception, we refer to Nowak *et al.* (1995b).]

4. Stochastic Strategies with Standing

If we assume that each move can be mis-implemented with a certain probability, we are encountering stochastic strategies. As the example of *Generous Tit For Tat* (*GTFT*) shows, such strategies can be important in their own right, not just as imperfect realisations of deterministic strategies [see e.g. May (1987) and Sigmund (1995)].

Within the huge class of strategies for the iterated PD, we shall concentrate on the memory one strategies, where the decision, for each move, is uniquely based on the outcome of the previous move. Let us first omit the "standing". The outcome in every round, then, can be completely characterised by the pay-off for the first player, which is R, S, T or P. We shall number these outcomes by 1 to 4 (in this order) and consider strategies given by $\mathbf{p} = (p_1, \ldots, p_4)$ where p_i is the probability to cooperate after outcome i. For instance, *AllD*, the strategy that always defects, is given by (0, 0, 0, 0) and *TFT* by (1, 0, 1, 0). These are so-called *reactive* strategies, where the decision depends only on the *other* player's previous move, not on the own, i.e. where $p_1 = p_3$ and $p_2 = p_4$ [see Nowak (1990) and Nowak & Sigmund (1990)]. Examples of non-reactive strategies are *GRIM* (1, 0, 0, 0) and *PAVLOV* (1, 0, 0, 1). These are deterministic strategies, where the p_i are 0 or 1. If we assume that errors occur, we obtain stochastic versions, for instance $(1-\epsilon, \epsilon, 1-\epsilon, \epsilon)$ as an approximation to *TFT* [cf. Nowak & Sigmund (1993a) and (1995)].

If the rule $\mathbf{p}$ is matched against a rule $\mathbf{p}' = (p'_1, p'_2, p'_3, p'_4)$, this yields a Markov process where the transitions between the four possible states R, S, T and P are given by the matrix

$$\mathbf{T} =$$

$$\begin{bmatrix} p_1p'_1 & p_1(1-p'_1) & (1-p_1)p'_1 & (1-p_1)(1-p'_1) \\ p_2p'_3 & p_2(1-p'_3) & (1-p_2)p'_3 & (1-p_2)(1-p'_3) \\ p_3p'_2 & p_3(1-p'_2) & (1-p_3)p'_2 & (1-p_3)(1-p'_2) \\ p_4p'_4 & p_4(1-p'_4) & (1-p_4)p'_4 & (1-p_4)(1-p'_4) \end{bmatrix} \quad (3)$$

(note that p_2 is matched with p'_3 and vice versa; one player's S is the other player's T). If $\mathbf{p}$ and $\mathbf{p}'$ are in the interior of the strategy cube, then all entries of this stochastic matrix are strictly positive, and hence there exists a unique stationary distribution $\mathbf{s} = (s_1, s_2, s_3, s_4)$ such that $p_i^{(n)}$, the probability to be in state i in the n-th round, converges to s_i for $n \to \infty$ $(i = 1, 2, 3, 4)$. The components s_i are strictly positive and sum up to 1. They denote the asymptotic frequencies of R, S, T and P. The stochastic vector $\mathbf{s}$ is a left eigenvector of $\mathbf{T}$ for the eigenvalue 1, i.e. satisfies $\mathbf{s} = \mathbf{sT}$.

It follows that for $w = 1$, the pay-off for a player using $\mathbf{p}$ against an opponent using $\mathbf{p}'$ is given by

$$A(\mathbf{p}, \mathbf{p}') = Rs_1 + Ss_2 + Ts_3 + Ps_4. \quad (4)$$

If, for instance, a TFT player is matched against another TFT player, and if errors occur, the pay-off is reduced to $(R + S + T + P)/4$, which is less than R. On the other hand, two *PAVLOV*-players receive R (up to an ϵ-term) because their errors are quickly corrected. We note that the s_i and hence also the pay-off in (4) are independent of the initial condition, i.e. of the moves of the players in the first round. For $w < 1$, the pay-off has a more complicated expression and depends on the initial move, see Nowak & Sigmund (1995).

Let us now take the "standing" into account. A stochastic strategy based on the outcome of the previous round is now given by a vector $\mathbf{q} = (q_1, \ldots, q_8)$ where q_i is the probability to play C if the state in the previous round was i (we keep the ordering as described at the end of Section 2). There are $2^8 = 256$ *deterministic* strategies (where all q_i are 1 or 0).

The strategies $\mathbf{p} = (p_1, \ldots, p_4)$ considered previously do not depend on the standings, but only on the actions of the two players in the previous round. Such a $\mathbf{p}$-strategy can be viewed as a $\mathbf{q}$-strategy, with

$$\mathbf{q} = (p_1, p_2, p_2, p_3, p_4, p_3, p_4, p_4).$$

Tit For Tat, for instance, is (1, 0, 0, 1, 0, 1, 0, 0) and Pavlov is (1, 0, 0, 0, 1, 0, 1, 1). The strategy *cTFT* is given by (1, 1, 0, 1, 0, 1, 1, 1) and *REMORSE* by (1, 0, 0, 0, 0, 1, 1, 1).

If the first player is a $\mathbf{q}$-strategist and the second a $\mathbf{q}'$-strategist, the transition probabilities from one state of the game to the next are given by the following matrix $\mathbf{T}$:

$$\begin{bmatrix} q_1q_1' & 0 & q_1(1-q_1') & 0 & 0 & (1-q_1)q_1' & 0 & (1-q_1)(1-q_1') \\ q_2q_4' & 0 & q_2(1-q_4') & 0 & 0 & (1-q_2)q_4' & 0 & (1-q_2)(1-q_4') \\ q_3q_6' & 0 & q_3(1-q_6') & (1-q_3)q_6' & (1-q_3)(1-q_6') & 0 & 0 & 0 \\ q_4q_2' & 0 & q_4(1-q_2') & 0 & 0 & (1-q_4)q_2' & 0 & (1-q_4)(1-q_2') \\ q_5q_7' & 0 & q_5(1-q_7') & (1-q_5)q_7' & (1-q_5)(1-q_7') & 0 & 0 & 0 \\ q_6q_3' & q_6(1-q_3') & 0 & 0 & 0 & (1-q_6)q_3' & (1-q_6)(1-q_3') & 0 \\ q_7q_5' & q_7(1-q_5') & 0 & 0 & 0 & (1-q_7)q_5' & (1-q_7)(1-q_5') & 0 \\ q_8q_8' & 0 & q_8(1-q_8') & 0 & 0 & (1-q_8)q_8' & 0 & (1-q_8)(1-q_8') \end{bmatrix}. \tag{5}$$

Note that, due to the rules about standing, there are four vanishing entries in each row of this 8×8 matrix. In spite of these zeros, $\mathbf{T}$ is irreducible, and even mixing, provided all q_i are distinct from 0 and 1; indeed, the entries of $\mathbf{T}^n$ are all strictly positive for $n > 2$. It follows that there exists a uniquely defined strictly stochastic vector $\mathbf{s}$ such that $\mathbf{sT} = \mathbf{s}$, yielding the stationary probabilities of the eight states. The pay-off obtained by the $\mathbf{q}$-player against the $\mathbf{q}'$-player is

$$Rs_1 + S(s_2 + s_3) + T(s_4 + s_6) + P(s_5 + s_7 + s_8). \tag{6}$$

Let us compute this, for example, if a *REMORSE*-player (whose strategy, if the error probability is ϵ, is given by $(1-\epsilon, \epsilon, \epsilon, \epsilon, \epsilon, 1-\epsilon, 1-\epsilon, 1-\epsilon)$) confronts a *cTFT*-player with strategy $(1-\epsilon, 1-\epsilon, \epsilon, 1-\epsilon, \epsilon, 1-\epsilon, 1-\epsilon, 1-\epsilon)$. The transition matrix $\mathbf{T}$ is given by

$$\begin{bmatrix} (1-\epsilon)^2 & 0 & (1-\epsilon)\epsilon & 0 & 0 & \epsilon(1-\epsilon) & 0 & \epsilon^2 \\ \epsilon(1-\epsilon) & 0 & \epsilon^2 & 0 & 0 & (1-\epsilon)^2 & 0 & (1-\epsilon)\epsilon \\ \epsilon(1-\epsilon) & 0 & \epsilon^2 & (1-\epsilon)^2 & (1-\epsilon)\epsilon & 0 & 0 & 0 \\ \epsilon(1-\epsilon) & 0 & \epsilon^2 & 0 & 0 & (1-\epsilon)^2 & 0 & (1-\epsilon)\epsilon \\ \epsilon(1-\epsilon) & 0 & \epsilon^2 & (1-\epsilon)^2 & (1-\epsilon)\epsilon & 0 & 0 & 0 \\ (1-\epsilon)\epsilon & (1-\epsilon)^2 & 0 & 0 & 0 & \epsilon^2 & \epsilon(1-\epsilon) & 0 \\ (1-\epsilon)\epsilon & (1-\epsilon)^2 & 0 & 0 & 0 & \epsilon^2 & \epsilon(1-\epsilon) & 0 \\ (1-\epsilon)^2 & 0 & \epsilon(1-\epsilon) & 0 & 0 & \epsilon(1-\epsilon) & 0 & \epsilon^2 \end{bmatrix}. \tag{7}$$

We write $\mathbf{T} = \mathbf{P} + \epsilon\mathbf{Q}_1 + \epsilon^2\mathbf{Q}_2$ and $\mathbf{s} = \mathbf{x} + \epsilon\mathbf{y} + \epsilon^2\mathbf{z}$, where $\mathbf{x}$ is a stochastic vector, so that the components of $\mathbf{y}$ and $\mathbf{z}$ both sum up to 0. Developing $\mathbf{sT} = \mathbf{s}$ in powers of ϵ we obtain $\mathbf{xP} = \mathbf{x}$, $\mathbf{xQ}_1 + \mathbf{yP} = \mathbf{y}$ and $\mathbf{zP} + \mathbf{yQ}_1 + \mathbf{Q}_2 = \mathbf{z}$. The first equation yields $\mathbf{x} = (1-2a, a, 0, 0, 0, a, 0, 0)$ for unknown a. Hence $\mathbf{xQ}_1 = (-2 + 6a, -2a, 1-2a, 0, 0, 1-4a, a, a)$ so that the second equation yields $a = \frac{2}{7}$. Hence $\mathbf{x} = (\frac{3}{7}, \frac{2}{7}, 0, 0, 0, \frac{2}{7}, 0, 0)$. It follows that the pay-off for *REMORSE* against *cTFT* is given, up to ϵ, by

$$\tfrac{3}{7}R + \tfrac{2}{7}(S + T), \tag{8}$$

which is the same as the pay-off for *cTFT* against *REMORSE*. Since both *cTFT* and *REMORSE* are error-correcting, and therefore obtain pay-off R against their like, the competition between these two strategies leads to a bi-stable situation which is symmetric: both basins of attraction are equally large. If it had been otherwise, this would have suggested that one strategy is stronger than the other.

A similar situation holds between *cTFT* and *PAVLOV*, i.e. $(1-\epsilon, \epsilon, \epsilon, \epsilon, 1-\epsilon, \epsilon, 1-\epsilon, 1-\epsilon)$. The stationary distribution (up to ϵ) is now $(\frac{3}{9}, 0, \frac{2}{9}, \frac{2}{9}, \frac{2}{9}, 0, 0, 0)$ so that the pay-off for *PAVLOV* against *cTFT* is now

$$\tfrac{1}{3}R + \tfrac{2}{9}(S + T + P). \tag{9}$$

We can easily compute the perturbation term for the pay-off: in the above case, for instance, it is $\epsilon(-6R - 13S + 23T - 4P)/81$.

If a *PAVLOV*-player plays against *REMORSE*, the pay-off is R (up to ϵ). Indeed, this interaction is error correcting. The reason is that the two strategies (which both are error-correcting against their own) obey quite similar rules: as long as both players are in good standing, they follow the same program. (However, *REMORSE* does not exploit suckers, i.e. *AllC*-players, whereas *PAVLOV* does.)

We mention in passing that there exist equalizers within the class of **q**-strategies. More precisely, every pay-off between P and R can be written as $P + \pi$. Against a strategy of the form

$$\mathbf{q} = \Big(1 + \pi a - a(R-P), 1 + \pi a - a(T-P), 1 + \pi b - b(T-P), 1 + \pi a + a(P-S) - \frac{a}{b}, \pi b, \pi a + a(P-S), \pi a, \pi a\Big)$$

(where a and b are real parameters such that all q_i lie between 0 and 1) every strategy obtains the same pay-off, namely $P + \pi$. This can be shown by a computation similar to that in Boerlijst *et al*. (1997), but considerably more tedious. For $a = b$ we obtain the $(p_1, \ldots, p_4)$-strategies described in Boerlijst *et al*. (1997).

5. Numerical Simulations

In this section we present results of random mutation experiments in order to enhance the understanding of the dynamics and attainability of the different ESS's. In these experiments a population of strategies is simulated for 1 million time steps (and more, if no steady state is reached). Pay-off values between strategies are computed on the assumption that $w = 1$. The next fraction of a strategy X_i is computed by:

$$X_i(t+1) = X_i(t)\frac{\sum_j X_j(t)A(i,j)}{\sum_k X_k(t)\sum_j X_j(t)A(k,j)} \tag{10}$$

where $A(i,j)$ is the pay-off that strategy i gets when playing against strategy j and $X_i(t)$ is the frequency of strategy i at time t. In eqn (10) the change of a fraction is determined by the average score of the strategy divided by the average score of the population [comparable to replicator dynamics, see Hofbauer & Sigmund (1988)]. Whenever a fraction drops below 0.001, it is regarded as extinct and set to zero. Therefore, the total number of different strategies can never exceed 1000. Mutant strategies are introduced at a fraction of 0.0011. The chance of the appearance of a mutant is 0.01 per time step. After mutation and extinction events the population is rescaled to 1. Strategies are given by a vector $\mathbf{q} = (q_1, \ldots, q_8)$. There is a background noise $\epsilon = 0.001$. Mutants have a random set of q-values, with a bias towards pure strategies. q-values are set to ϵ or $(1-\epsilon)$, each with probability 1/3, or to the U-shaped distribution $(1 + \cos(\pi\rho))/2$ (with random variable ρ uniform between 0 and 1), if necessary rounded to ϵ or $(1-\epsilon)$. In this way the chance of obtaining a particular pure strategy is $(1/3 + \cos^{-1}(1-2\epsilon)/\pi)^8$, and hence the chance that a particular pure strategy appears within a simulation exceeds 99%.

We simulate for two different sets of pay-off values, which differ in dynamics. The first set of ($S = 0$, $P = 1$, $R = 3$, $T = 5.5$) is at high temptation to defect, whereas the second set of ($S = 0$, $P = 1$, $R = 3$, $T = 3.5$) is at low temptation. The two sets differ on whether $2R > T + P$ or not.

HIGH TEMPTATION ($T = 5.5$)

At high temptation we find the ESS's: *ALLD*, *GRIM*, *cTFT* and *REMORSE*. Simulations starting with just one of these strategies show that populations of *ALLD* and *GRIM* do not persist for a long time, whereas populations consisting of *cTFT* and *REMORSE* do persist. This still holds if w is slightly smaller than 1. The apparent contradiction that an ESS population can be invaded by mutants can be explained by the fact that in our model the score of a newly introduced mutant is (marginally) influenced by the mutant playing against itself. We argue that ESS's that are not stable against such small perturbations are structurally unstable: biologically, we assume that mutant strategies invade in small clusters, or clones.

Simulations starting from *ALLD* sooner or later end up in populations of either *cTFT*(-like) or *REMORSE*(-like) strategies. Figure 5 shows two typical runs: Figure 5(a) settling in *cTFT*, and Fig. 5(b) settling in *REMORSE*. The average population score very quickly approaches 3, indicating cooperation. Before the population reaches the steady state, periods of relative stasis alternate with periods of rapid change, comparable to e.g. Lindgren (1991). In Fig. 5(b) the population initially shows alternations between *PAVLOV*-like, and *REMORSE*-like dominance. In fact, these two types of strategies behave similarly in most cases.

Some *cTFT*-like and *REMORSE*-like strategies play almost neutral against pure (up to ϵ) *cTFT* and pure *REMORSE*. Often the final state is composed of a mixture of either these *cTFT*-like or *REMORSE*-like strategies. Figure 6 shows a simulation that ends in a *cTFT*-like population. The scores within such a mixture are all alike, so that the dynamics are governed by the score against "background mutants". This explains the drift and the accumulation of neutral mutants. Note that pure *cTFT* is also present in Fig. 6, but it fails to dominate the population.

To explore the basins of attraction of the ESS's we ran 100 simulations starting from *ALLD*: 68 ended in *cTFT*-like mixtures, 11 ended in pure *cTFT*, 15 ended in *REMORSE*-like mixtures, and six ended in pure *REMORSE*. It seems that competition is decided on the base of which strategy first exceeds a certain threshold. The fact that there are more neutral mutants around *cTFT* than around *REMORSE* explains the bias towards the former strategy. Simulations starting from 100 random mutants show similar statistics.

LOW TEMPTATION ($T = 3.5$)

Known ESS's at low temptation are *ALLD*, *GRIM*, *cTFT*, *PAVLOV* and *WEAKLING*. Again, *ALLD* and *GRIM* are easily invaded, whereas the other strategies persist. Starting 100 simulations from *ALLD* we get 63 *cTFT*-like mixtures, eight pure *cTFT*, 17 *PAVLOV*-like mixtures, three pure *PAVLOV*, six *WEAKLING*-like mixtures, and three times pure *WEAKLING*. The dynamics resembles that described for high temptation. Figure 7 shows a simulation that ends in pure *WEAKLING*. It can be seen that the appearance of *WEAKLING*-like strategies causes a drop in the score. Pure *WEAKLING* will slowly outcompete the other e *WEAKLING*-like strategies, and the population stays fixed in a sequence of alternating mutual cooperation and defection, giving a score of $(R+P)/2$. Only nine out of 100 simulations end in this non-cooperative mode, *PAVLOV*(-like) and *cTFT*(-like) populations both reach a score close to R.

OTHER PAY-OFF VALUES

Results for other pay-off values resemble the results of either of the above described situations. At the bifurcation point $T = 5$ the main attractor of simulations is again pure *cTFT* or *cTFT*-like mixtures. At this value also stable *REMORSE* or *REMORSE*-like mixtures, and *PAVLOV*-like mixtures are observed. Pure *PAVLOV* is no longer an ESS for this T-value. Another bifurcation point is at $T = 4$. Above this T-value *WEAKLING* is no longer an ESS (more generally, the condition is $T + S < R + P$).

To conclude, we see that the addition of a standing in the Prisoner's Dilemma facilitates the evolution of cooperation. Populations with random mutations in most cases quickly adapt to a cooperative mode, and only rarely the population gets trapped in the *WEAKLING* strategy. Surprisingly, this suboptimal trapping is only observed in situations with low temptation to defect.

6. The Alternating PD

One can also investigate *cTFT* in the context of the alternating Prisoner's Dilemma [see Boyd (1988), Nowak & Sigmund (1994) and Frean (1995)]. In the strictly alternating case, the two players take turns in deciding which move to chose: either to offer or to withhold assistance (C or D). As shown in Nowak & Sigmund (1994) the pay-off values must then satisfy $T - R = P - S$. In the alternating game, not only the state (Dg, Dg) but also the states (Db, Db) and (Cg, Dg) are unreachable. [The state (Cg, Db) for instance means: the first player has cooperated—he is by definition in good standing—and then, in the following round, the second player has defected, but nevertheless is in good standing, clearly an impossibility. We shall only consider the states where the first player's move has been answered by a move of the second player.] We denote the remaining states (Cg, Cg), (Cg, Db), (Dg, Cg), (Dg, Db), (Db, Cg) and (Db, Dg) by 1 to 6 (in this order), and consider stochastic strategies of the form $\mathbf{q} = (q_1, \ldots, q_6)$. If a

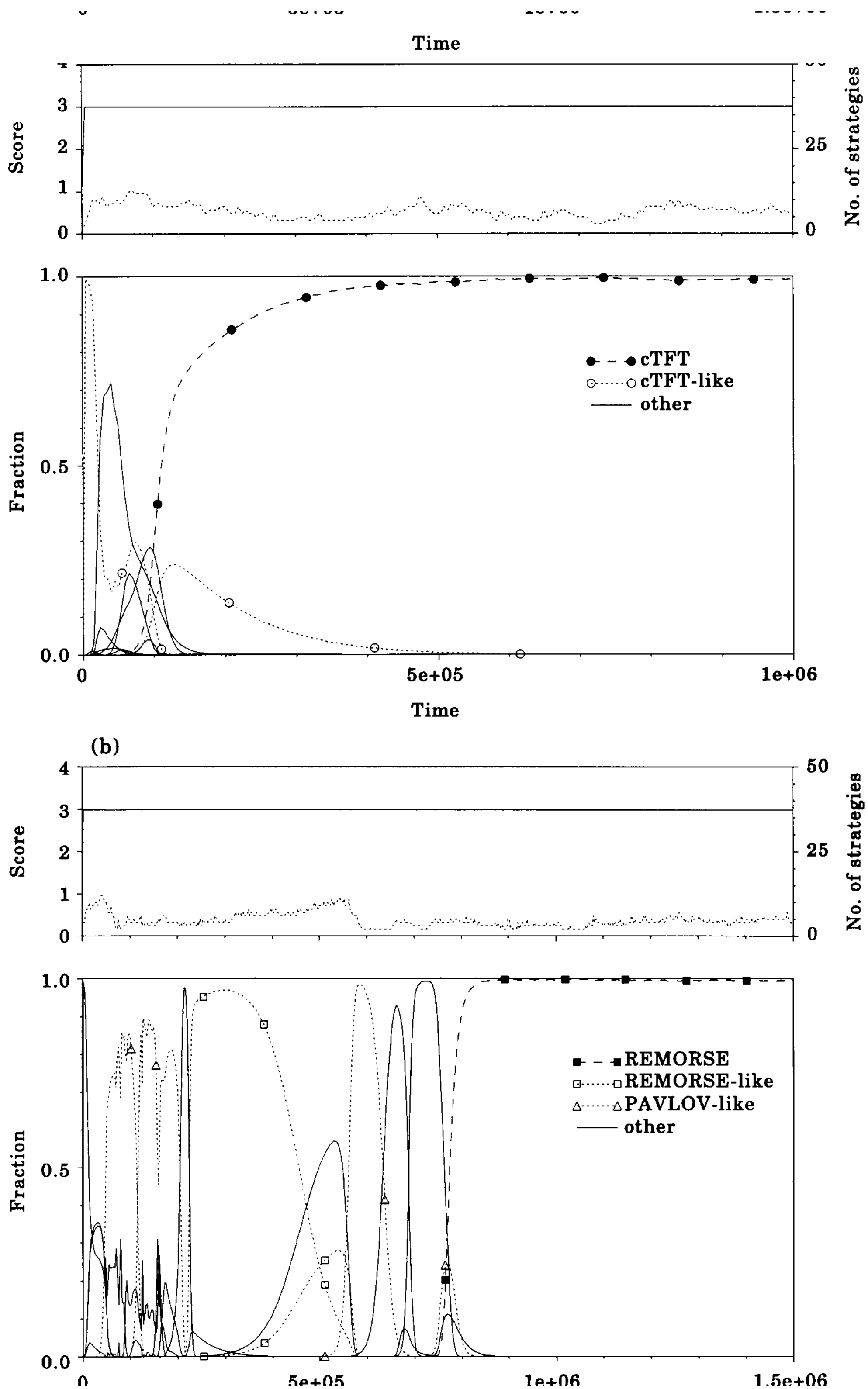

FIG. 5. Evolution of a population of strategies starting from pure *ALLD* with high temptation to defect ($T = 5.5$). In the upper panel the solid line indicates the average population score whereas the dotted line indicates the number of diferent strategies. (a) Settling in *cTFT*. (b) Settling in *REMORSE*. After Lindgren (1991).

q-player meets a **q**′-player, the transition matrix is given by

$$\mathbf{T} = \begin{bmatrix} q_1q_1' & q_1(1-q_1') & 0 & 0 & (1-q_1)q_2' & (1-q_1)(1-q_2') \\ q_2q_5' & q_2(1-q_5') & (1-q_2)q_6' & (1-q_2)(1-q_6') & 0 & 0 \\ q_3q_1' & q_3(1-q_1') & 0 & 0 & (1-q_3)q_2' & (1-q_3)(1-q_2') \\ q_4q_5' & q_4(1-q_5') & (1-q_4)q_6 & (1-q_4)(1-q_6) & 0 & 0 \\ q_5q_1' & q_5(1-q_1') & 0 & 0 & (1-q_5)(1-q_2') & (1-q_5)(1-q_2') \\ q_6q_3' & q_6(1-q_3') & 0 & 0 & (1-q_6)q_4' & (1-q_6)(1-q_4') \end{bmatrix}. \quad (11)$$

If **s**, again, denotes the stationary vector, then the pay-off for the **q**-player is

$$s_1R + s_2S + (s_3 + s_5)T + (s_4 + s_6)P. \quad (12)$$

We note that again, *cTFT* is evolutionarily stable. Numerical simulations (as described in the previous chapter) show that *ALLD* populations do not persist. All simulations settle in *cTFT*(-like) mixtures, making the alternating Prisoner's Dilemma a favourite playground for *cTFT*.

7. Discussion

All strategies considered in this paper can be implemented by finite automata. For the extensive theory in this field, we refer to Binmore & Samuelson (1992). One might ask whether the *cTFT*-strategy can be implemented by a strategy uniquely based on a finite (but possibly very long) memory of the moves of the two players, and not using the notion of standing. This, however, is not the case. If, for instance, a sequence of alternating defections occurs,

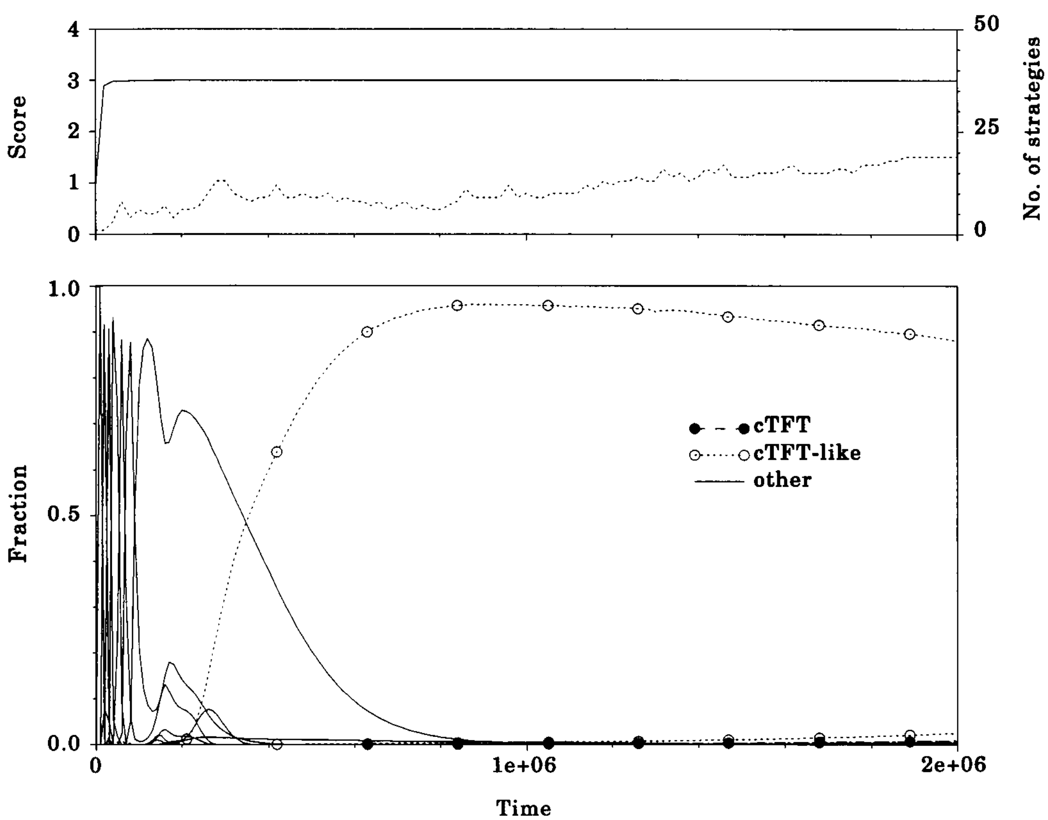

FIG. 6. Simulation settling in a mixture of *cTFT*-like strategies.

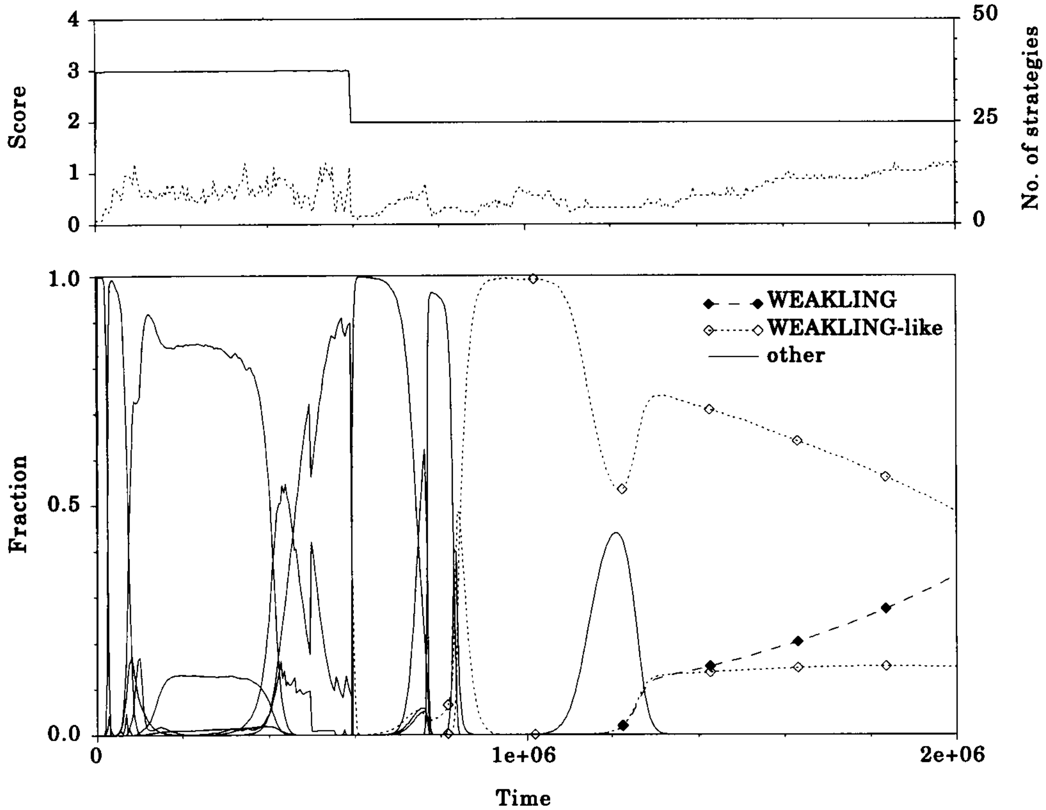

FIG. 7. Evolution of a population of strategies starting from pure *ALLD* with low temptation to defect ($T = 3.5$).

only the player that started to defect will have a bad standing. The next move is not specified by a finite memory of previous moves in case the initial defection happened prior to the memorised moves.

The concept of a "standing" introduces an interesting new twist to the theory of iterated games played by finite automata. The most immediate step, there, is certainly to study decision rules based on the outcome of the previous round, and the most immediate extension is to consider rules based on two, three or more previous rounds. Both Axelrod (1987) and Lindgren (1991) have studied by means of genetic algorithms the evolution of strategies with memory two or three. In particular, Lindgren has pointed out the very robust success of a class of memory-two strategies which usually cooperate with each other and where a unilateral defection (due to a mistake in implementation) entails *two* rounds of mutual defections (a kind of domestic row) before bilateral cooperation is resumed. Such strategies are similar to *PAVLOV*, but use the outcome of the last two rounds.

cTFT and *REMORSE* are of a different nature. They only depend on the outcome of the previous round, but this outcome, now, is more complex: it does not consist only on the actions *C* or *D* of the two players, but on the standing—good or bad—after a defection. The rules for determining this standing seem quite natural: we can identify with a player who feels bad after having committed erroneously a defection, or who feels provoked by the unilateral defection of the co-player after a string of mutual cooperation. The rules embody a certain notion of "fairness" which seems to be rather common. If it should indeed turn out that this notion is a human universal, we would have to explain how it emerged.

In principle, one could apply other rules of "standing". To start with, we should replace this term by a more neutral one, in order not to get trapped by its connotations, and think only of an arbitrary "tagging" of the states without specifying which is "good" or "bad". A strategy is now specified by the probability to cooperate and/or change the standing in the next round, depending on the current state (including the current standing) of both opponents. It is plausible that we can obtain some evolutionarily stable strategies for many such codes.

Here is, as an intriguing example, the strategy *Prudent-PAVLOV* (*pPAVLOV*). This strategy follows in most cases the *PAVLOV*-strategy, as the name suggests. However, after any defection it will only resume cooperation after two rounds of mutual defection. This is achieved by normally playing defections with standing D_1, and only playing D_0 after a mutual defection or an erroneous defection. The strategy only cooperates after mutual cooperation or after mutual defection with standing D_0. Suppose that two *pPAVLOV*s are engaged in a match. They usually both cooperate. If one defects by mistake, the state is (C, D_0). In the next round, the state is (D_1, D_1); in the second-next round, it is (D_0, D_0), and hereafter mutual cooperation is resumed. This strategy, which depends only on the previous round, acts to all purposes like Lindgren's (1991) memory two strategy. An erroneous defection against its like entails two rounds of mutual defection, and then leads back to mutual cooperation. An *AllC*-opponent will be exploited ruthlessly; but against an *AllD* opponent, *pPAVLOV* will be suckered every third round. It is easy to see (cf. Fig. 8) that this strategy is an ESS whenever

$$R + wR + w^2R > T + wP + w^2P \qquad (13)$$

and numerical simulations show that it attracts very well.

Moreover, *pPAVLOV* has the big advantage to be immune to errors in perception, as can be seen from the following table, which shows the evolution first from my (erroneous) point of view (first row: my moves, including my standing; second row: my opponents moves) and then from my co-player's point of view (third row: my moves; fourth row: his moves, including his standing). The mistake occurs in the second round (indicated by the asterisk).

C	C	D_1	D_1	D_0	C	...
C	D*	C	D	D	C	...
C	C	D	D	D	C	...
C	C	C	D_1	D_0	C	...

This is what happens if one of the *pPAVLOV*-players mis-interprets the other player's C for a D. Something similar happens if he mis-interprets his own C for a D (a less likely, but not completely impossible occurrence).

Altogether, we can interpret *pPAVLOV* as a sophisticated offspring of *PAVLOV*.

An interesting point about this strategy is that it distinguishes between D_0 and D_1 only for the own defections, but not for the other player's defection. We can view the "tagging" by 0 or 1 as an internal action. The *pPAVLOV* strategy does not monitor the standing of the adversary. This seems simpler than strategies like *cTFT* or *REMORSE*, which also keep track of the other fellows standing.

It seems highly plausible that there exists a wide variety of workable "taggings" which yield interesting ESS's. The question is whether an evolution based on mutation and selection would tend to lead to one form of "tagging" rather than another. This could ultimately shed light on why humans developed a sense of fairness, feelings of guilt, and highly effective social norms [see also Sugden (1986) and Young (1993) on the evolution of conventions]. The sheer combinatorial complexity of encompassing all conceivable codes, or taggings, is enormous, and the costs (in fitness) for reckoning with these "tags" seem difficult to evaluate. But it is a tempting problem.

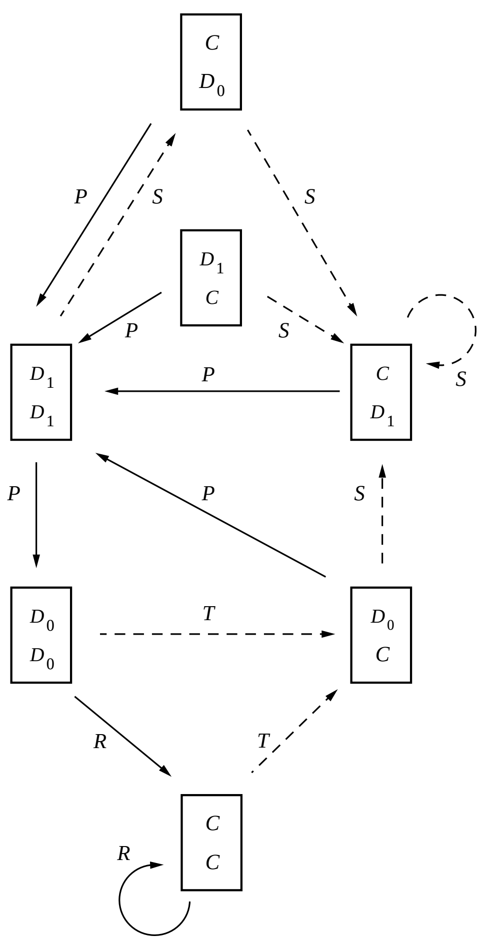

FIG. 8. *pPAVLOV* is an ESS if $T + wP + w^2P < R + wR + w^2R$.

Great progress on automata playing the repeated PD has recently been obtained by Olof Leimar in a paper that is due to appear in JTB soon. M.C.B. is funded by E.C. grant number ERBCHBICT941834, M.A.N. is funded by the Wellcome Trust.

REFERENCES

AXELROD, R. (1984). *The Evolution of Cooperation*. New York: Basic Books (reprinted 1989 in Penguin, Harmondsworth).

AXELROD, R. & HAMILTON, W. D. (1981). The evolution of cooperation. *Science* **211,** 1390–1396.

AXELROD, R. (1987). The evolution of strategies in the iterated Prisoner's Dilemma. In: *Genetic Algorithms and Simulated Annealing* (Davis, D., ed.). London: Pitman.

AXELROD, R. & DION, D. (1988). The further evolution of cooperation. *Science* **242,** 1385–1390.

BENDOR, J. (1993). Uncertainty and the evolution of cooperation. *J. Conflict Resolution* **37,** 709–734.

BENDOR, J., KRAMER, R. M. & STOUT, S. (1991). When in doubt: Cooperation in a noisy Prisoner's Dilemma. *J. Conflict Resolution* **35,** 691–719.

BINMORE, K. G. & SAMUELSON, L. (1992), Evolutionary stability in repeated games played by finite automata. *J. Economic Theory* **57,** 278–305.

BOERLIJST, M. C., NOWAK, M. A. & SIGMUND, K. (1997). Equal Pay for all Prisoners, *American Mathematical Monthly* (in press).

BOYD, R. (1988). Is the repeated Prisoner's Dilemma a good model of reciprocal altruism? *Ethnology and Sociobiology* **9,** 278–305.

BOYD, R. (1989). Mistakes Allow Evolutionary Stability in the Repeated Prisoner's Dilemma Game. *J. theor. Biol*. **136,** 47–56.

DAWKINS, R. (1989). *The Selfish Gene*. Oxford: Oxford University Press (New Edition).

FREAN, M. (1994). The Prisoner's Dilemma without synchrony. *Proc. R. Soc. Lond. B* **257,** 75–79.

FUDENBERG, D. & MASKIN, E. (1990). Evolution and cooperation in noisy repeated games. *American Economic Review* **80,** 274–279.

HOFBAUER, J. & SIGMUND, K. (1988). *The Theory of Evolution and Dynamical Systems*. Cambridge: Cambridge University Press.

KOLLOCK, P. (1993). An eye for an eye leaves everyone blind: Cooperation and accounting systems. *American Sociological Review* **58,** 768–786.

KRAINES, D. & KRAINES, V. (1988). Pavlov and the Prisoner's Dilemma. *Theory and Decision* **26,** 47–79.

LINDGREN, K. (1991). Evolutionary phenomena in simple dynamics. In: *Artificial life II* (Langton, C. G. *et al*., eds), Santa Fe Institute for Studies in the Sciences of Complexity, Vol. X, 295–312.

MAY, R. M. (1987). More evolution of cooperation. *Nature* **327,** 15–17.

MAYNARD SMITH, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

MILLER, J. H. (1996). The coevolution of automata in the Repeated Prisoner's Dilemma, *J. Econ. Behav. Organ*. **29,** 87–103.

MOLANDER, P. (1985). The optimal level of generosity in a selfish, uncertain environment. *J. Conflict Resolution* **29,** 611–618.

MÜLLER, U. (1987). Optimal retaliation for optimal cooperation. *J. Conflict Resolution* **31,** 692–724.

NOWAK, M. A. (1990). Stochastic strategies in the Prisoner's Dilemma. *Theor. Pop. Biol*. **38,** 93–112.

NOWAK, M. A., MAY, R. M. & SIGMUND, K. (1995a). The arithmetics of mutual help. *Sci. Am*. **272**(6), 76–81.

NOWAK, M. A., SIGMUND, K. & EL-SEDY, E. (1995b). Automata, repeated games, and noise. *J. Math. Biol*. **33,** 703–732.

NOWAK, M. A. & SIGMUND, K. (1990). The evolution of reactive strategies in iterated games. *Acta Applicandae Math*. **20,** 247–265.

NOWAK, M. A. & SIGMUND, K. (1992). Tit for tat in heterogeneous populations. *Nature* **355,** 250–252.

NOWAK, M. A. & SIGMUND, K. (1993a). Chaos and the evolution of cooperation. *Proc. Natl. Acad. Science U.S.A*. **90,** 5091–5094.

NOWAK, M. A. & SIGMUND, K. (1993b). Win-stay, lose-shift outperforms tit-for-tat. *Nature* **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. (1995). Invasion dynamics for the finitely repeated Prisoner's Dilemma. *Games and Economic Behaviour*, **11,** 364–390.

SELTEN, R. (1975). Re-examination of the perfectness concept for equilibrium points in extensive games. *Int. J. Game Theory* **4,** 25–55.

SELTEN, R. & HAMMERSTEIN, P. (1984). Gaps in Harley's argument on evolutionarily stable learning rules and in the logic of tit for tat. *Behav. Brain Sci*. **7,** 115–116.

SIGMUND, K. (1995). *Games of Life*. Harmondsworth: Penguin.

SUGDEN, R. (1986). *The Economics of Rights, Co-operation and Welfare*. Oxford: Blackwell.

WU, J. & AXELROD, R. (1995). How to cope with noise in the iterated Prisoner's Dilemma. *J. Conflict Resolution* **39,** 183–189.

YOUNG, H. P. (1993). The evolution of conventions. *Econometrica* **61,** 57–84.