*J. theor. Biol.* (1994) **168**, 219–226

# The Alternating Prisoner's Dilemma

MARTIN A. NOWAK† AND KARL SIGMUND‡

† *Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, U.K.*
*and* ‡ *Institut für Mathematik, Universität Wien, Strudlhofgasse 4, A-1090 Wien, Austria*

(*Received on 2 October, 1993, Accepted on 13 December, 1993*)

Reciprocal altruism can often be modelled by a variant of the iterated Prisoner's Dilemma where players alternate in the roles of donor and recipient, rather than acting simultaneously. We consider strategies realised by simple transition rules based on the previous encounter, and show that the evolutionary outcome for the alternating Prisoner's Dilemma can be quite different from the simultaneous case. In particular, the winner of a simultaneous Prisoner's Dilemma is frequently a "win-stay, lose-shift" strategy based on the payoff experienced in the last round, whereas in the alternating Prisoner's Dilemma, the trend leads towards a "Generous Tit For Tat" strategy. If one allows only for reactive strategies based on the other player's last move, the overall payoff is the same for the alternating or the simultaneous version, although the sequence of moves can be different. In the alternating game "win-stay, lose-shift" strategies can only be successful if there is a longer memory of past encounters. The alternating and simultaneous Prisoner's Dilemma are two very different situations, and the whole existing literature is based on the simultaneous game.

## 1. Introduction

In many instances of reciprocal altruism, the partners alternate in their roles of donor and recipient. Trivers (Trivers, 1985: 393) summarizes that "reciprocal altruism is expected to evolve when two individuals associate long enough *to exchange roles frequently* as potential altruist and recipient" (our italics). Nevertheless, in the prevalent model for reciprocal altruism, the iterated Prisoner's Dilemma (PD), the player's roles are symmetric in each round: both have the same two options, namely to co-operate (i.e. to play $C$) or to defect (play $D$). It is usually assumed that this symmetry is of small effect. Axelrod states that "it makes little difference whether the choices are treated as simultaneous or as sequential" (see Axelrod, 1991: 85). In this paper we show that in some situations (as, for instance, when players make occasionally mistakes), important differences can arise.

One of the best known examples of reciprocal altruism is offered by South American vampire bats, who in their nightly excursions fasten on cattle and feast on their blood. Bats who have found a good meal are liable to help their hungry fellow-bats with some of their surplus (Wilkinson, 1984). Such blood donations are of critical, occasionally even of life-saving importance to the recipient. They seem to offer an instance of reciprocal altruism—which is, in Trivers' definition, "the trading of altruistic acts in which benefit is larger than cost, so that over a period of time both enjoy a net gain" (Trivers, 1985: 361). According to a ringing sequence by Richard Dawkins, "vampires could form the vanguard of a comfortable new myth of sharing, mutualistic cooperation" (Dawkins, 1989).

Another well-documented example concerns young male baboons (cf. Packer, 1977; and Trivers, 1985): one of them picks a fight with a dominant male, while the other profits from the diversion by mounting an oestrous female. On the next occasion, the roles of the youngsters may be reversed. Other real-life examples for reciprocal altruism include helping in fights, allogrooming, alarm calls, etc (see Axelrod & Hamilton, 1981). In most of these examples, it makes no sense to co-operate simultaneously; the partners have to take turns. To quote Trivers again: "reciprocal

0022–5193/94/100219 + 08 $08.00/0

altruism can also be viewed as a symbiosis, each partner helping the other while he helps himself. The symbiosis has a time lag, however: one partner helps the other and must then wait a period of time before he is helped in turn" (Trivers, 1971).

To be sure, there are examples of mutual help where the actions of the two players are simultaneous. The most clear-cut of these is described in a well-known investigation of predator inspection by a pair of stickleback (Milinski, 1987). In Milinski's experimental set-up, a single stickleback is confronted with a dummy predator and deluded by a mirror into believing that another stickleback is in the water tank. Inspection of the predator demands a careful approach by a sequence of short, halting moves. Depending on the inclination of the mirror, the image stickleback either keeps abreast, thus seemingly sharing the risk, or else lags consistently behind. The moves of the two sticklebacks could not be more simultaneous: one is the mirror-image of the other. (Milinksi found that the stickleback moved significantly closer up to the predator if the mirror image keeps abreast.)

Both the simultaneous and the alternating PD are therefore relevant for reciprocal altruism. In an alternating Prisoner's Dilemma, to co-operate must mean to act altruistically by playing *C when it is one's turn to do so*. This slight modification can affect strategies and payoffs to a considerable extent. For instance, if two Tit For Tat (*TFT*) players engage in an iterated PD of the usual, *simultaneous* kind, and if one of them defects by mistake, both players will subsequently play *C* and *D* in turns. On the other hand, if two Tit For Tat players engage in an *alternating* PD, and if again a defection occurs by mistake, then the result will be a sequence of mutual defections (Fig. 1).

In this paper, we shall discuss two varieties of the alternating PD. First, we assume that the roles are exchanged in every encounter: this is the *strictly alternating* PD. Next, we deal with the case where the roles alternate in a haphazard way. This second case is usually more realistic. It could happen, for instance, that the same bat is lucky two nights in succession, or that the same young baboon needs the help of his comrade on two or three consecutive occasions. But since this *randomly alternating* PD is slightly more complicated, we assume first that roles alternate strictly.

(i) The simultaneous game
Player 1 *CCCD'CDCD ... DD'DD ... DC'D ... DC CC ...*
Player 2 *CCCC DCDC ... CD DD ... DD C ... CC'CC ...*

(ii) The alternating game
Player 1 *C C D' D D D ... D C' C C C ...*
Player 2 *C C D D D D ... D C C C C ...*

FIG. 1. The effect of occasional errors on the game between two Tit For Tat players for the simultaneous PD and for the strictly alternating PD. There are three possible kinds of run for the simultaneous game, and two for the alternating game; but the average payoff is the same. The primes indicate mistakes.

## 2. The Description of the Game

Let us start by describing the simultaneous PD (see Axelrod, 1991). In each round, the two players have the options to play *C* (i.e. to co-operate) or play *D* (to defect). If both co-operate, both earn as payoff a "reward" *R* which is larger than the payoff *P*, the "punishment", which they receive if they both defect. But if one players opts for *D* and the other for *C*, then the defector receives a payoff *T* (the "temptation") which is larger than *R*, while the co-operator's payoff *S* (for "sucker") is even smaller than *P*. We assume furthermore that the two players earn more if both co-operate than if they agree to choose different moves and then share the total payoff: this means that $R > \frac{1}{2}(T + S)$. The payoff, according to evolutionary game dynamics, is assumed to be an increment in fitness, or reproductive success (see Maynard Smith, 1982).

In the alternating PD, the symmetry between the two players is broken. In each single round, one of the players is the "leader" (to use a game-theoretic expression), i.e. able to decide what the outcome is going to be. The leader has to choose between two options, which we denote again by *C* and *D*. Option *C* means that the leader receives payoff *a* and the other player payoff *b*. Option *D* means that the leader receives payoff *c* and the other player *d*. Such payoff values can be negative, as in the case of a vampire bat feeding another. But we always assume

$$\text{(i) } c > a \quad \text{(ii) } c - a < b - d. \tag{1}$$

Condition (i) means that in a single round, option *D* is better than *C* for the leader. Condition (ii) means that the cost occurring to the leader by altruistically playing *C* is less than the benefit to the other player [which, together with (i), implies $b > d$].

Conditions (i) and (ii) are surely satisfied if a well-fed bat shares part of its meal with a starving colleague. We shall not address the important question of how to signal one's need, and the attendant problem of how to assess the other's honesty—cf. the "Philip Sydney game" (Maynard Smith, 1991; and Grafen, 1990).

Let us now consider a "unit" of two consecutive rounds. Under the assumption of strict alternation, each of the players is leader once. If both partners

play $C$, both earn $a+b$ points in one unit. If both play $D$, both earn $c+d$. If one player plays $C$ and the other $D$, then the defector earns $c+b$ and the co-operator $a+d$. These payoff values for one unit are like those in one round of the simultaneous PD. We have only to set

$$R = a+b,\ P = c+d,\ T = c+b,\ S = a+d. \quad (2)$$

The inequality $c > a$ implies $T > R$ and $P > S$, whereas $c - a < b - d$ implies $R > P$ and $S + T < 2R$. Since $T > R > P > S$ and $S + T < 2R$ are the two conditions which define a simultaneous PD, two consecutive rounds of the alternating game are equivalent to one round of the simultaneous game. On the other hand, if $R$, $T$, $S$ and $P$ are given as in (2), then necessarily

$$T + S = P + R. \quad (3)$$

This is satisfied, for instance, if $T = 4$, $R = 3$, $P = 1$ and $S = 0$ (cf. Smale, 1980), the values which we shall use in our simulations, but it does not hold for $T = 5$, $R = 3$, $P = 1$ and $S = 0$, the values used by Axelrod in his computer tournaments (Axelrod, 1991). It follows that not every simultaneous PD corresponds to an alternating PD. But for every alternating PD, there is a corresponding simultaneous PD. The two games are quite different, however, because they admit different strategies (in the alternating PD, a strategy can depend on whether one starts by being the donor, for instance), and more importantly, because the same strategies can lead to different outcomes.

Condition (3) greatly simplifies the analysis of the simultaneous (iterated) PD (see Nowak & Sigmund, 1990). It means that the cost of switching from $D$ to $C$ is the same against a defector as against a cooperator. In the context of the alternating PD, it means that if both players alternate between $C$ and $D$, it does not matter whether we view this game as composed of units where $(C,C)$ follows $(D,D)$, or where $(D,C)$ follows $(C,D)$. We note that (2), as a linear equation in the four unknowns $a, b, c, d$, is of rank 3. The set of solutions is either empty or, if (3) is satisfied, a one-dimensional subspace. For our simulations, we shall take $a = 2$, $b = 1$, $c = 3$, $d = -2$ (hence $R = 3$, $S = 0$, $T = 4$, $P = 1$).

In the following, we shall consider the infinitely iterated PD only. This is, of course, an idealization, approximated by real-life interactions only if the probability of further encounters is very high. Furthermore, we shall restrict our attention to strategies which are conditional upon the last few rounds only, and probabilistic. This means that the memories of the players are short, and that their behavioural rules are not clear-cut, but rather fuzzy: a higher or lesser propensity to opt for $C$ or for $D$, given the outcome of the last interactions. In particular, this approach takes account of errors in implementing a rule, which are unavoidable in any biological context (May, 1987).

### 3. The Simultaneous PD

For expository reasons, we begin by briefly sketching the situation for the simultaneous PD, which is well understood if we assume that the players base their decisions on the previous round only, (see Nowak 1990; Nowak & Sigmund 1990, 1992, 1993; Lindgren 1991). The outcome in such a round is specified by the payoff, which can be $R$, $S$, $T$ or $P$. These outcomes will be numbered 1, 2, 3, 4 (in this order). Thus outcome 3 means "I got $T$", or more explicitly "I opted for $D$ in the previous round, and my co-player opted for $C$". A player's strategy will be given by a quadruple $\boldsymbol{p} = (p_1, p_2, p_3, p_4)$, where $p_i$ is a number between 0 and 1 and denotes the probability to play $C$, given that the outcome of the previous round was $i$ (with $i = 1, 2, 3$ or 4). The strategy Tit For Tat, for instance, is given by the rule $(1, 0, 1, 0)$; it simply repeats the previous move of the other player. If $\varepsilon$ is the probability of an error in implementing a move, then this becomes $(1-\varepsilon, \varepsilon, 1-\varepsilon, \varepsilon)$. If a player with strategy $\boldsymbol{p}$ meets a player with strategy $\boldsymbol{p}' = (p_1', p_2', p_3', p_4')$, then the transition from one round to the next is given by the Markov chain

$$\begin{pmatrix} p_1p_1' & p_1(1-p_1') & (1-p_1)p_1' & (1-p_1)(1-p_1') \\ p_2p_3' & p_2(1-p_3') & (1-p_2)p_3' & (1-p_2)(1-p_3') \\ p_3p_2' & p_3(1-p_2') & (1-p_3)p_2' & (1-p_3)(1-p_2') \\ p_4p_4' & p_4(1-p_4') & (1-p_4)p_4' & (1-p_4)(1-p_4') \end{pmatrix} \quad (4)$$

For instance, the transition from outcome 1 to outcome 3 means that the $\boldsymbol{p}$-player defects after experiencing an $R$ (which happens with probability $1-p_1$) and that the $\boldsymbol{p}'$-player co-operates (which happens with probability $p_1'$). We note that one player's $T$ is the other player's $S$.

If the $p_i$ and $p_i'$ are strictly positive (as is always the case, if we take errors into account), matrix (4) has a unique left eigenvector $\mathbf{S} = (s_1, s_2, s_3, s_4)$ to the eigenvalue 1 with the property that the $s_i$ are all strictly positive and sum up to 1. The asymptotic frequency of outcome $i$ is then given by $s_i$, so that the payoff for the $\boldsymbol{p}$-player in the iterated PD, i.e. the limit in the mean of the payoff per round, is simply given by

$$s_1R + s_2S + s_3T + s_4P. \quad (5)$$

This is independent of the first moves of the players. The initial condition has only a transitory effect, as it is blurred by mistakes.

In Nowak & Sigmund (1992, 1993), we have analysed by computer experiments the evolution of large, well-mixed populations under the effect of selection and mutation. Each individual has some strategy $\boldsymbol{p}$ and interacts with all other individuals in the population. Since its average payoff (which depends on the strategies of the other players, and hence on the composition of the population) is a measure of its reproductive success, the frequency of its offspring (which inherits strategy $\boldsymbol{p}$) is proportional to this success. This yields the frequency of $\boldsymbol{p}$ in the following generation, and reflects the action of natural selection. Furthermore, we introduce occasionally a tiny minority using a new, randomly chosen strategy $\boldsymbol{q}$: depending on its fitness, it spread or vanished. This mutational process provides an endless source of variability.

The evolutionary process displayed in these computer chronicles is difficult to analyse in details, and exhibits a bewildering richness of dynamical features. But basically, most chronicles show a distinct trend towards an increasingly stable co-operation. Most co-operative populations, in the long run, adopt a strategy close to $p_1 = 1$, $p_2 = 0$, $p_3 = 0$, $p_4 = 1$. This strategy co-operates if and only if both players, in the previous round, opted for the same move. This is a "win-stay, lose-shift" strategy: players stick to their former move if it was rewarded with a $T$ or an $R$, but switch to the other option if they received only a $P$ or an $S$.

## 4. The Strictly Alternating PD

Now let us consider the infinitely iterated PD with strictly alternating rounds. Let us assume that the memory of each player covers the previous two rounds, i.e. one "unit" of the alternating game. Again, there are four possible outcomes, depending on the choices of the player and of the co-player (the latter was the leader in the previous round, and the former in the round before). These outcomes are experienced as $R$, $S$, $T$ or $P$, and will be numbered from 1 to 4, again. But we must note that within one unit, the memories of the two players are based upon different (though overlapping) past units. The leader in round $2n$ considers the outcome of round $(2n-2)$ and $(2n-1)$, the leader in round $(2n+1)$ considers rounds $(2n-1)$ and $2n$. If we denote by $p_i$, again, the propensity to play $C$ after outcome $i$ in the previous unit, and by $p_i'$ that of the other player, then we obtain as transition matrix, from one round to the next, the Markov chain

$$\begin{bmatrix} p_1p_1' & p_1(1-p_1') & (1-p_1)p_2' & (1-p_1)(1-p_2') \\ p_2p_3' & p_2(1-p_3') & (1-p_2)p_4' & (1-p_2)(1-p_4') \\ p_3p_1' & p_3(1-p_1') & (1-p_3)p_2' & (1-p_3)(1-p_2') \\ p_4p_3' & p_4(1-p_3') & (1-p_4)p_4' & (1-p_4)(1-p_4') \end{bmatrix} \tag{6}$$

which is distinct from (4), but has a related structure. The average payoff, again, is computed as in (5).

There is a considerable difference between the simultaneous and the strictly alternating PD. This can be seen, for instance, by watching two "win-stay, lose-shift" players matched against each other (Fig. 2). In the simultaneous PD, a mistaken defection by one of the players leads to one round where both players defect, and then to a return to co-operation. In this sense, "win-stay, lose-shift" is error-correcting: a brief burst of hostility, and then back to business. The average payoff, in a population adopting this strategy, is close to $R$. In the strictly alternating PD, this is quite different. A mistaken $D$ is answered by a $D$, which elicits a $C$, which is followed by a $D$ in turn. Thus each player, after a mistake, keeps playing two $D$'s and one $C$, periodically. With probability $\frac{2}{3}$, the next mistake does not affect this regime; only with probability $\frac{1}{3}$ will it redress the game to a run of mutual co-operation. The average payoff is $\frac{1}{2}(R+P)$, and hence lower than $R$.

Let us now assume a small noise level $\varepsilon$. Among the strategies which are "almost pure", in the sense that all $p_i$ differ from 0 or 1 by $\varepsilon$ only, the strategy closest to $(1, 0, 1, 1)$, i.e. $(1-\varepsilon, \varepsilon, 1-\varepsilon, 1-\varepsilon)$, emerges almost always as the winner of the evolutionary race, especially if we assume that $2R > T + P$ (which holds precisely if $c - a < \frac{1}{2}(b-d)$, i.e. if the cost to the donor is smaller than *half* the benefit to the recipient, and is satisfied for our numerical values). This condition is equivalent to the requirement that the strategy close to $(1, 0, 1, 1)$ cannot be invaded by the always defecting strategy which is close to $(0, 0, 0, 0)$, or, for that matter, by the strategies close

(i) The simultaneous game
Player 1 *CCCD'DCCC* ...
Player 2 *CCCC DCCC* ...

(ii) The alternating game
Player 1 *C C D' C D D C*...*D C D C' C C*...
Player 2 *C C D D C D D*...*D D C C C C*...

FIG. 2. Occasional errors in the game between two win-stay, lose-shift players. In the simultaneous game, the strategy is error correcting and leads back to mutual co-operation. In the strictly alternating PD, a mistake leads to a run where each player plays two $C$s and one $D$, periodically. A further mistake can redress the situation only if it changes a $D$ into a $C$, which happens only for one in three errors. The primes indicate mistakes.

to $(0,0,1,0)$, $(0,1,0,0)$, $(0,1,1,0)$, $(1,0,0,0)$, or $(1,1,0,0)$. Furthermore, the other strategies can never invade, with the exception of the strategy close to $(1,1,1,0)$ (which does exactly as well, and can therefore enter by random drift) and the Tit For Tat strategy, which is close to $(1,0,1,0)$: the outcome is a mixture where a minority (with frequency $\varepsilon 2(T-R)/R-P)$ plays Tit For Tat and the overwhelming majority adopts the strategy close to $(1,0,1,1)$. In the simultaneous PD, by contrast, the condition $2R > T+P$ leads to the non-invadability of the win-stay, lose-shift strategy close to $(1,0,0,1)$.

The strategy $(1,0,1,1)$ is a very tolerant strategy, which only defects if it has been played for a sucker, i.e. if it has experienced an $S$ during the last unit. In contrast to "win-stay, lose-shift", it does not defect after a $T$, which means that it does not try to exploit suckers. A strategy $(1,0,1,x)$, with $0<x<1$, is somewhat less tolerant than this, but more tolerant than Tit For Tat: with probability $x$ it forgives a $D$ of the co-player, if this was in response to one's own $D$ in the previous round. If $2R > T+P$, a population playing a strategy close to $(1,0,1,x)$ cannot be invaded by *All D* invaders, for any $x$.

Let us now look how these generous strategies do against each other. A $(1-\varepsilon,\varepsilon,1-\varepsilon,x)$-player in a population of $(1-\varepsilon,\varepsilon,1-\varepsilon,x')$ strategists has a payoff

$$R+\varepsilon\left[\frac{-R(2+x'-xx')+T(2x'-xx')+P(2-x')}{x+x'-xx'} - (R-S)\right]. \quad (7)$$

This can be shown by using the same straightforward perturbation method as in Nowak *et al.* (1994). For our numerical values, this yields

$$F(x,x') = 3+\varepsilon\left[-3+\frac{-4+4x'-xx'}{x+x'-xx'}\right].$$

It follows that $F(x,x')-F(x',x')$ is of the same sign as $(x-x')(3x'^2-8x'+4)$. This last term has a zero at $x'=\frac{2}{3}$. As long as $x'<\frac{2}{3}$, an $x$-player does better than average in an $x'$-population whenever $x>x'$; it follows that higher $x$-values can invade as long as $x'<\frac{2}{3}$, and conversely; among the strategies of the form $(1-\varepsilon,\varepsilon,1-\varepsilon,x)$, evolution will lead towards an $x$-value of $\frac{2}{3}$. Moreover, it can easily be shown that this value is *evolutionarily stable* (see Maynard Smith, 1982), in the sense that a homogeneous population of the form $(1-\varepsilon,\varepsilon,1-\varepsilon,\frac{2}{3})$ cannot be invaded by a $(1-\varepsilon,\varepsilon,1-\varepsilon,x)$ minority, for any value of $x$.

Let us now consider evolutionary chronicles, starting from a random strategy and introducing a small minority of a randomly generated mutant strategy every 100 generations. We can observe that natural

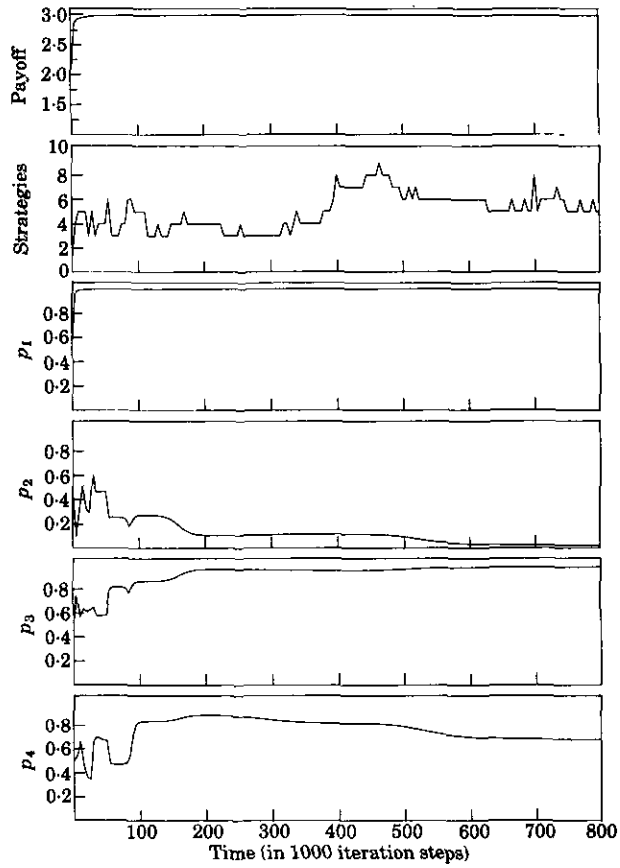


FIG. 3. Evolutionary simulation of the strictly alternating Prisoner's Dilemma with all stochastic strategies that remember the last two moves. The payoffs were evaluated according to eqns (4) and (5). We started with a homogeneous population of the random strategy (0·5, 0·5, 0·5, 0·5). Every 100 generations (on average) a new mutant was introduced. New mutants were chosen at random; the probabilities $p_i$ were taken from the U-shaped density distribution $[\pi x(1-x)]^{-1}$ to get more bias towards the boundaries of the four-dimensional strategy space (because the relevant strategies like *All D*, *TFT* or *GTFT* are close to the boundary). But we only admit $p_i$ values within 0·001 and 0·999 (to have some minimum level of noise). As population dynamics we use the difference equation $x_i' = x_i f_i/f$, where $x_i$ denotes the frequency of strategy $i$, $f_i$ its payoff in the population, and $f = \Sigma x_i f_i$ the average payoff of the population. Thus payoff is related to fitness, successful strategies leave more offspring. If the frequency of a strategy drops below 0·001 it is removed from the population (new strategies start with a frequency of 0·0011).

This simulation is representative for the overwhelming majority of simulations we performed for the strictly alternating Prisoner's Dilemma. Here a strategy close to $(1,0,1,2/3)$ dominates after some 600 000 iterations of the difference equation. We performed 40 such simulations (each for $10^7$ iterations yielding a total of about $10^5$ mutants per run) and strategies very close to this strategy won in 39 runs. The figure shows (from the top): the average population payoff (for one double move), the number of different strategies in the system and the population averages of the probabilities $p_1, p_2, p_3, p_4$. The payoff values are $a=2$ $b=1$ $c=3$ $d=-2$, corresponding to $R=3$ $S=0$ $T=4$ $P=1$.

selection leads in the overwhelming majority of the simulations to a co-operative society where most members adopt a strategy close to $(1, 0, 1, \frac{2}{3})$ (see Fig. 3). This variant of the Generous Tit For Tat (GTFT) forgives (roughly) two thirds of all those defections of the adversary which follow one's own $D$, but it does not forgive those defections which were unwarranted. If all play this strategy, a defection by mistake would entail a $D$ as reply, but a subsequent $D$ would only follow with a probability of $\frac{1}{3}$. A run of defections is therefore very short; evolution has found again a strategy which is error-correcting. A "win-stay, lose-shift" strategy, which does so well in the simultaneous PD (Nowak & Sigmund, 1993), does rather poorly in the strictly alternating case, because it is no longer error-correcting against its own.

## 5. Reactive Strategies: an Unexpected Equivalence

Let us now restrict our attention to the so-called *reactive strategies*, where each move depends only on the previous move of the *other* player. Tit For Tat is an example for such a reactive strategy. These strategies are characterised, both in the simultaneous and in the strictly alternating case, by $p_1 = p_3$ and by $p_2 = p_4$. We shall denote the first of these values by $p$, and the second by $q$. Thus $p$ is the probability to co-operate after a $C$ of the other player, and $q$ the probability to co-operate after a $D$. In the simultaneous PD, the average probability to play $C$ converges to

$$s = \frac{(p-q)q' + q}{1-(p-q)(p'-q')}, \tag{8}$$

(where $p'$ and $q'$ are the corresponding probabilities of the co-player). This was shown, by different methods, in Nowak (1990) and Nowak & Sigmund (1990). Exactly the same argument holds for the strictly alternating PD. It follows that the payoff, for a $(p, q)$ player meeting a $(p', q')$ co-player is given in the simultaneous game by

$$Rss' + Ss(1-s') + T(1-s)s' + P(1-s)(1-s') \tag{9}$$

and in the strictly alternating game by

$$as + c(1-s) + bs' + d(1-s') = P + s(a-c) + s'(b-d). \tag{10}$$

If the simultaneous PD corresponds to the alternating PD, i.e. if (3) holds, the two expressions given by (9) and (10) reduce to the same value. This is somewhat surprising, as the sequence of moves, in the two games, can be quite different. For instance, if two Tit For Tat players are matched against each other, and if occasional mistakes occur, then in the simultaneous case, all possible outcomes (i.e. $R, S, T$ or $P$) occur with the same frequency; indeed, if we start with a run of all-out co-operation, a mistake will cause a $D$ and thus a run of alternating co-operation with defection, before another mistake turns the game into either a co-operative run again or, with the same probability, into a run of mutual defections. In contrast, to this, two Tit For Tat players in a strictly alternating game experience units of $R$ or of $P$ with the same frequency; indeed, a mistake turns a run of all-out cooperation into a run of mutual defection, and vice versa (Fig. 1). In the former case, the payoff is $\frac{1}{4}(R+S+T+P)$, and in the latter case $\frac{1}{2}(R+P)$, but by (3) this reduces to the same value.

If we view each reactive strategy $(p, q)$ as a point in the unit square, we can easily describe its game dynamics, i.e. the evolution of a population of players meeting at random and spreading at a rate proportional to their payoff. The straight line $p - q = (T-R)/(T-P)$ divides the square into two parts. In the part which contains the lower right corner (1,0), i.e. the strategy Tit For Tat, there is a tendency towards cooperation: if all members of a population play a strategy $(p,q)$ which belongs to this set, invaders using a strategy with a higher $p$- or $q$-value, i.e. a higher propensity to play $C$, will spread and eventually take over. On the other hand, in a population whose members all stick to a $(p, q)$ strategy belonging to the other part of the unit square invaders with a tendency to defect more frequently will spread. The strategy

$$p = 1, \quad q = \frac{R-P}{T-P} \tag{11}$$

is optimal in the sense that a population where all members adopt it is (i) immune to defectors, and (ii) receives the highest payoff among all populations enjoying such an immunity (see also Molander, 1985). This strategy, which always responds to a $C$ with a $C$, but tolerates $D$ with a certain probability, has been called Generous Tit For Tat. As shown in Nowak & Sigmund (1992) by means of extensive computer simulations, in a population which is a mixture of many $(p, q)$ strategies, selection usually leads to one of two possible outcomes, namely either to a population consisting of all-out defectors, or to a population very close to Generous Tit For Tat. This last outcome occurs only if the initial population contains a small minority of stern Tit For Tat players (strict retaliators, where $p$ is very close to 1 and $q$ very close to 0), or if such a minority is introduced by a

random mutation. Tit For Tat is a catalyser for the establishment of Generous Tit For Tat.

This, then, is what happens within the range of reactive strategies, both for the simultaneous and for the strictly alternating PD.

### 6. The Randomly Alternating PD

Let us now turn to the *randomly* alternating PD. We assume here that in every round, each player has a 50% chance of being the leader. It no longer makes sense to divide the games into "units" of two rounds each. Instead, we assume that players base their propensity to play $C$ on the outcome of the last round only; this includes knowledge of who was the leader, of course. Again, each player experiences this outcome through the payoff received in the last round, which can be $a$ (the player was leader, and opted for $C$), or $b$ (the co-player was leader, and chose $C$), or similarly $c$ or $d$. Again, we enumerate the outcomes $a$, $b$, $c$, $d$ by 1, 2, 3, 4 (in this order), and denote by $p_i$ the probability that the player opts for $C$, given that the outcome of the previous round was $i$. (For instance, $p_3$ denotes the propensity of the player to co-operate after outcome 3, i.e. after having been leader in the previous round and having opted for $D$. The probability that the player can actually implement this decision, i.e. the chance of being leader in the next round, is $\frac{1}{2}$, independent of the player's decision). The transition rule, from one round to the next, is given by the Markov chain

$$\frac{1}{2}\begin{pmatrix} p_1 & p_2' & (1-p_1) & (1-p_2') \\ p_2 & p_1' & (1-p_2) & (1-p_1') \\ p_3 & p_4' & (1-p_3) & (1-p_4') \\ p_4 & p_3' & (1-p_4) & (1-p_3') \end{pmatrix} \tag{12}$$

The corresponding left eigenvector s is of the form $(s, s', \frac{1}{2}-s, \frac{1}{2}-s')$, with $0 < s, s' < \frac{1}{2}$. The values $s$ and $s'$, whose sum is the asymptotic probability for a move $C$, satisfy

$$\begin{aligned} s(4-2p_1+2p_3)+s'(2p_4-2p_2) &= p_3+p_4 \\ s'(4-p_1'+2p_3')+s(2p_4'-2p_2') &= p_3'+p_4' \end{aligned} \tag{13}$$

The limit in the mean payoff, for the $p$-player, is

$$\frac{P}{2}+s(R-T)-s'(T-P) = \frac{c+d}{2}+s(a-c)+s'(b-d). \tag{14}$$

In a homogeneous population (where all members adopt the same strategy $p$), we have $s = s'$. The maximal payoff is $\frac{R}{2}$, but unconditional co-operation (i.e. $s = s' = \frac{1}{2}$) could be subverted by exploiters.

The evolutionary races, in this case, lead less frequently to co-operation than in the strictly alternating game. Nevertheless, if $2R > T + P$, one observes in most evolutionary runs the establishment either of all-out defectors, or else of a co-operative strategy with $p_1 = p_2 = p_3 = 1$ and $p_4 = 2R - T - P/T - P$, which is $\frac{1}{3}$ for our numerical values (Fig. 4). In fact, one can directly check that among populations close to $(1, 1, 1, x)$, selection leads towards ever higher $x$-values; but among those close to $(1, 1, 1, x)$ which are immune to invasion by *All D* players, the maximal value of $x$ is $2R - T - P/T - P$. It is interesting to see that once co-operation becomes preponderant, $p_1$ and $p_2$ evolve much faster towards 1 than does $p_3$. A population discovers soon that it pays to

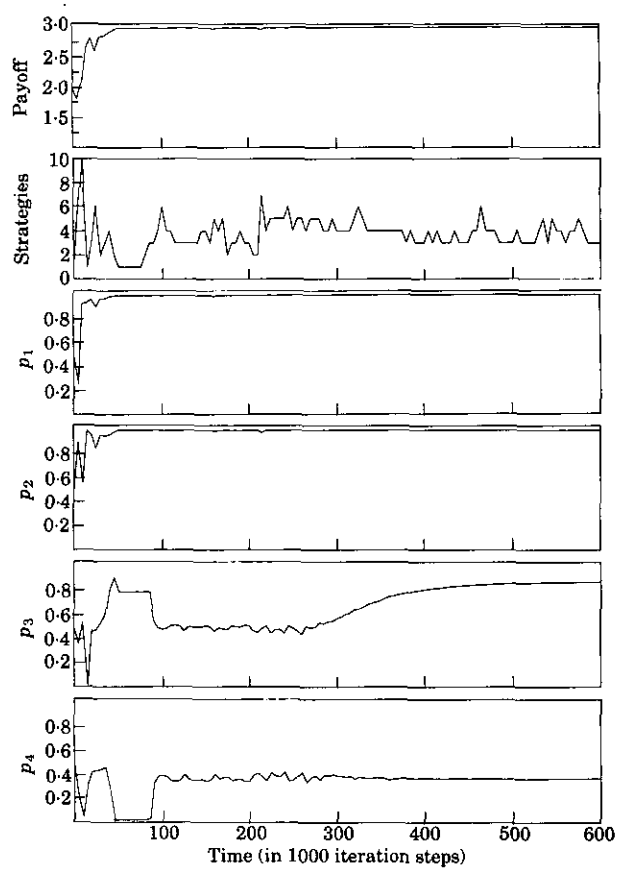


FIG. 4. An evolutionary simulation of the stochastically alternating Prisoner's Dilemma for all strategies that consider the last move. The simulation was performed exactly as outlined in the legend of Fig. 3. For the payoff evaluation we used eqn (13). Here it is more difficult to achieve cooperation, probably because the invasion barrier of *TFT* into *All D* is twice as high as for the simultaneous or strictly alternating game (the invasion barrier is the smallest frequency of *TFT* players capable of invading an *All D* population). Co-operation was reached in eight runs out of 40. Co-operative populations were always dominated by strategies very close to the *GTFT* variant (1, 1, 1, 1/3). Note that the probabilities $p_i$ have different meanings for the strictly and randomly alternating game (see text for details). Payoff values as Fig. 3.

co-operate after a $C$ (i.e. a co-operative move by oneself or by the other player). It takes much longer to realize that it pays also to co-operate after one's own $D$ (i.e. to extend an olive-branch by reverting to $C$, and thus to make up for one's defective move from the previous round). Altogether, this strategy can again be viewed as a variant by Generous Tit For Tat: it always repays a kind move, tolerates a defection by the co-player with a certain probability, and punishes at most by a single $D$.

We can again define reactive strategies, i.e. those for which only the co-player's last move is of relevance. Since these strategies have to repeat their previous move if they happened to have been leader, they are of the form $(1,p,0,q)$. Tit For Tat, for instance, is now realized by the transition rule $(1, 1, 0, 0)$. For reactive strategies, (13) yields $2s = q + (p-q)q'/1-(p-q)(p'-q')$, cf. (8), and the payoff, up to the obvious factor $\frac{1}{2}$, is the same as the payoff for the strictly alternating PD given in (10), and hence also the same as for the simultaneous PD. The evolutionary winner among reactive strategies is $(1, 1, 0, \frac{1}{3})$, again.

## 7. Conclusion

To conclude, we can say that evolutionary simulations show, both for the simultaneous and for the alternating PD, a strong tendency (but no necessity) to evolve towards co-operation. The co-operative strategies achieving this are quite different in the simultaneous and the alternating case. In the simultaneous case, they embody a "win-stay, lose-shift" principle which does not shrink from exploiting a sucker. In the alternating case, we find that some sort of Generous Tit For Tat emerges, (just as it did within purely reactive strategies in the simultaneous case). In the simultaneous PD, the emerging "win-stay, lose-shift" strategy has two important properties: (i) it is error-correcting in the sense that if a mistake occurs against players using the same strategy, then mutual co-operation is quickly re-established; and (ii) it exploits unconditional co-operators, thus preventing the spreading of suckers through random drift in the population. In this sense, "win-stay, lose-shift" is proof against both strategical and mutational noise. For the alternating PD, there is no strategy with memory two (i.e. taking account of the last two moves) which satisfies both (i) and (ii). It is only strategies with memory longer than three which combine both these properties. For the strictly alternating PD such as strategy is the following: remain co-operative (=play $C$ after $C$–$C$); reciprocate an unwarranted defection (=play $D$ after $C$–$D$); offer a handshake (=play $C$ after a $D$–$D$); recognise a handshake (=play $C$ after $C$–$D$–$D$–$C$); and finally, never give a sucker an even break (=play $D$ after $D$–$C$–$D$–$C$). This strategy has the same features as "win-stay, lose-shift" in the simultaneous game.

We emphasize the difference between (i) strategies like Tit For Tat and its variants, which more or less mimic the co-players moves, and (ii) "win-stay, lose-shift"-strategies which depend only on the payoff experienced by the player. A "win-stay, lose-shift" principle can work even for players which do not understand that they are engaged in a game, and which are unaware of encountering another decision-maker. There are many differences between the simultaneous and the alternating Prisoner's Dilemma. This paper does some first steps to explore a new world. The dilemma has no end.

This work was partly supported by the Austrian Forschungsförderungsfonds P8043, the British Council and the Wellcome Trust.

## REFERENCES

AXELROD, R. (1991). *The Evolution of Cooperation.* Harmondsworth: Penguin Books.

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

DAWKINS, R. (1989). *The Selfish Gene*, 2nd edn. Oxford: Oxford University Press.

GRAFEN, A. (1990). Biological signals as handicaps. *J. theor. Biol.* **144**, 517–546.

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

MAY, R. M. (1987) More evolution of co-operation. *Nature, Lond.* **327**, 15–17.

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

MAYNARD SMITH, J. (1991). Honest Signalling: the Philip Sydney game. *Anim. Behav.* **42**, 1034–1035.

MILINSKI, M. (1987). Tit for Tat in sticklebacks and the evolution of cooperation. *Nature, Lond.* **325**, 434–435.

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

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

NOWAK, M. A. & SIGMUND, K. (1990). The evolution of stochastic strategies in the Prisoner's Dilemma. *Acta Applic. Math.* **20**, 247–265.

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

NOWAK, M. A. & SIGMUND, K. (1993). Win-stay, lose-shift outperforms tit for tat, *Nature, Lond.* **364**, 56–58.

NOWAK, M. A., SIGMUND, K. B., EL-SEDY, E. (1994). Automata, respected games and noise. *J. math. Biol.*, in press.

PACKER, C. (1977). Reciprocal altruism in *Papio anubis*. *Nature, Lond.* **265**, 441–443.

SMALE, S. (1980). The Prisoner's Dilemma and dynamical systems associated to non-cooperative games. *Econometrica* **48**, 1617–1634.

TRIVERS, R. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57.

TRIVERS, R. (1985). *Social Evolution.* Menlo Park, CA: Benjamin Cummings.

WILKINSON, G. S. (1984). Reciprocal food-sharing in the vampire bat. *Nature, Lond.* **308**, 181–184.