ta Applicandae Mathematicae 20: 247–265, 1990.



# The Evolution of Stochastic Strategies in the Prisoner's Dilemma

MARTIN NOWAK
*Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, England*

and

KARL SIGMUND
*Institut für Mathematik der Universität Wien, Strudlhofg. 4, A-1090 Vienna, Austria and IIASA, Laxenburg, Austria*

(Received: 21 December 1989; in final form; 13 June 1990)

**Abstract.** The evolution of reactive strategies for repeated 2 × 2-games occurring in biology is investigated by means of an adaptive dynamics.

**AMS subject classifications (1980).** 34C35, 90D15, 80D45, 92A12.

**Key words.** Prisoner's Dilemma, reactive strategies, iterated games, adaptive dynamics.

## 1. Introduction

Games get more interesting if they are repeated over and over again. Children know this, and so do game theorists. A prime example is the Prisoner's Dilemma. Since 'Defect' dominates 'Cooperate', there is, as has been said, not properly any dilemma in this game. But if it is repeated an unknown number of times, cooperation becomes a promising option.

In Axelrod's computer tournaments (Axelrod, 1984), Tit For Tat (cooperate in the first round, and then do whatever the other did last) scored extremely well. This success encouraged biologists to try and explain the evolution of cooperation in natural populations by reciprocal interactions based on repeated encounters (Axelrod and Hamilton, 1981; Axelrod, 1987; Wilkinson, 1984; May, 1987; Milinski, 1987). This motivates the study of *reactive strategies* (where the decision in each round depends on the opponent's behaviour in the previous round). Since biological interactions, in contrast to interactions between computer programs, teem with uncertainties, the reactions will not be clear cut, in general, but rather stochastic: an increase or decrease of the probability to cooperate.

In an error-prone world, Tit For Tat loses much of its lustre, since a single mistake between two Tit For Tat players leads to a sequence of mutual recriminations which can only be broken by a further mistake. A certain level of generosity (i.e. a tendency to cooperate even after a defection by the opponent) is much more appropriate. We

shall see that in the presence of 'noise', it is best to forget sometimes (not always!) a bad turn but never a good one.

There are other repeated biological games, besides the Prisoner's Dilemma. We mention 'Hawk-Dove' (called 'Chicken' by classical game theorists) where the options are to escalate the conflict or not. If the expected gain is less than the cost of losing an escalated contest, neither 'Hawk'- nor 'Dove'-strategy dominates and some mixture is evolutionarily stable. Again, it is plausible that the two opponents meet repeatedly and react to each other's behaviour in the last round.

Other plausible assumptions would be: (a) that a larger part of the past than just the last round should be taken into account; (b) that the history of one's own decisions should be included; (c) that the possibilities to increase and decrease the chance of future encounters, and to learn during the game, should be allowed (Axelrod, 1984; Feldman and Thomas, 1987). All these aspects will be omitted here: our players are pretty dumb.

This is acceptable in a 'game theory without rationality' (the expression is from Rapaport) which aims at modelling the evolution of biological populations. We are not interested in finding strategies which optimize the benefit of the group, or best replies against a given behaviour, but rather in the *directional change* favoured by selection. If an individual with a deviant strategy does better than the rest, it is likely that more and more individuals will adopt this strategy. This can occur in several ways, through imitation or learning or – if the strategy is a hereditary trait, and the payoff reproductive success – through natural selection. What we will model is an essentially homogeneous population sprinkled with a few 'mutants' using strategies which differ only slightly from the original one. The population then moves into the direction which is most promising. Of course, this is rather a caricature of an evolutionary process. In particular, the assumptions that the population remains homogeneous and that only small deviations are taken into account are debatable. We will return to this in the discussion, but mention now that computer simulations based on less restrictive evolutionary mechanisms agree quite well with these simplifications.

The (already) classical ESS-theory is also based on the assumption of a homogeneous population. In particular, a strategy is said to be *evolutionarily stable* (i.e. an ESS) if an infinite homogeneous population adopting it cannot be invaded by mutants under the action of natural selection (Maynard Smith, 1982). This explains how such a strategy is able to hold itself, but not how it can get established. We will see that this can be quite a problem. An ESS need not be attainable: every other homogeneous population can be proof against invasion by the mutant using this ESS. Such an ESS is then a *Garden of Eden* configuration, in the sense that it is not the outcome of a sequence of adaptations. The adaptive dynamics that we are proposing casts an interesting light on this phenomenon, which has first been noted – in another context – by Eshel and Motro (1981).

In Section 2, we deal with reactive strategies and their stationary outcomes, and in Section 3, with the payoff function. In Section 4, the important limiting case of

probability one for the next round is investigated. In this case, where the invadability of a homogeneous population is well understood, we introduce and analyse the adaptive dynamics which describes the optimal directional change. In Section 5, another special case is studied, where the expected gain for switching from one type of move to the other does not depend on the adversary's move. In Section 6, the adaptive dynamics for the general case is investigated. In particular, it is shown that in the two special cases, whenever it pays to increase gratitude (i.e. readiness to cooperate after a cooperation by the adversary), it also pays to increase forgiveness (the readiness to cooperate after a defection by the adversary), and vice versa. In the general case, one direction still holds, but not always its converse.

## 2. Reactive Strategies and *C*-Levels

We shall consider games with two strategies $C$ and $D$ and assume that there exists a constant probability $w \in ]0, 1]$ to repeat the game. The *reactive strategies* to which we limit our attention will be determined by triples $(y, p, q) \in [0, 1]^3$, where $y$ is the probability to play $C$ in the first round, while $p$ and $q$ are the conditional probabilities to play $C$ after an opponent's $C$ (resp. $D$) in the previous round (Nowak and Sigmund 1989a, b).

Strategies where $r := p - q$ vanishes, are said to be *unconditional*. They do not properly depend on the adversary's move. A strategy with $r = 1$ (i.e. $p = 1$, $q = 0$) is said to be *reciprocal*, while it is said to be *paradoxical* if $r = -1$ (i.e. $p = 0$, $q = 1$). *Tit For Tat* and *Suspicious Tit For Tat* are reciprocal (with $y = 1$, resp. $y = 0$). Such strategies are not stochastic. Neither are *AllD* ($y = p = q = 0$) and *AllC* ($y = p = q = 1$). Since we are mostly interested in noisy interactions (caused, for example, by uncertainties in the perception of the opponent's move or identity, or by the lacking control over one's own actions), we shall usually consider only strategies with $|r| < 1$. The special cases of reciprocal or paradoxical strategies are always easy to deal with separately (see Aumann, 1981).

The *C-level* $c_n$ is the probability to play $C$ in the $n$th round. For a $(p, q)$-strategist, it is given by the response function

$$\alpha(x) = px + q(1 - x) = q + rx, \tag{1}$$

where $x$ is the $C$-level of the opponent in the previous round.

If a player with strategy $E = (y, p, q)$ is matched against an opponent using strategy $E' = (y', p', q')$, then his $C$-level $c_n$ determines his opponents $C$-level $c'_{n+1}$, which, in turn, determines $c_{n+2}$. This 'echo-effect' is described by the recursive relations $c'_{n+1} = \alpha'(c_n)$ (with $\alpha'(x) = q' + r'x$) and $c_{n+2} = \alpha(c'_{n+1})$, so that

$$c_{n+2} = \alpha\alpha'(c_n) \quad \text{and} \quad c'_{n+2} = \alpha'\alpha(c'_n). \tag{2}$$

This yields directly

THEOREM. *The C-levels $c_n$ and $c'_n$ converge at the geometric rate $|rr'|^{1/2}$ to the stationary values $c$ and $c'$ satisfying*

$$(1 - rr')c = \alpha(q') = \alpha\alpha'(0),$$
$$(1 - rr')c' = \alpha'(q) = \alpha'\alpha(0), \tag{3}$$
$$\alpha(c') = c, \qquad \alpha'(c) = c'.$$

For unconditional strategies the convergence is in one step. If $|rr'| = 1$, the $C$-levels oscillate periodically.

The stationary $C$-level of strategy $E$ against itself is the solution of $x = \alpha(x)$ and, hence, given by

$$s = \frac{q}{1 - r}. \tag{4}$$

In the $(p, q)$-space, all strategies on the straight line through $E$ and the reciprocal strategy $(1, 0)$ have the same stationary $C$-level against themselves, and consequently against each other. Indeed, if two of the stationary $C$-levels $s$ ($E$ against itself), $s'$ ($E'$ against itself), $c$ ($E$ against $E'$) and $c'$ ($E'$ against $E$) are equal, then so are all four. More precisely, the differences $c - c'$, $c - s'$ and $s - s'$ have always the same sign.

We close this section with a remark on more general strategies. Depending on the simultaneous moves of the two players, the 'state' of the game in the $n$th round is $CC$, $CD$, $DC$ or $DD$. If we assume that the first player uses $C$ in the next round with probabilities $p_1, p_2, p_3$ or $p_4$, respectively, and the second player similarly with $p'_i$, then the transition probabilities to the state in the $n + 1$th round are given by the matrix

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

For properly stochastic strategies, $P$ is mixing and the elements $p_{ij}^{(n)}$ of $P^n$ converge (independently of $i$) to the components $\pi_j > 0$ of a stochastic vector $\pi$. In this paper, we consider only *opponent-determined* strategies, i.e.

$$p_1 = p_3 = p, \qquad p_2 = p_4 = q \quad \text{and} \quad p'_1 = p'_2 = p', \qquad p'_3 = p'_4 = q'.$$

In this case $\pi_1\pi_4 = \pi_2\pi_3$, i.e. the events that the players use $C$ in the stationary state are independent. The same independence holds if the strategies are *self-determined*, i.e.

$$p_1 = p_2 = p, \qquad p_3 = p_4 = q \quad \text{and} \quad p'_1 = p'_3 = p', \qquad p'_2 = p'_4 = q'.$$

In fact the 'linkage' $x_1x_4 - x_2x_3$ converges in both cases to 0 at the geometric rate $|rr'|$, since $\mathbf{z} = \mathbf{x}P$ implies $z_1z_4 - z_2z_3 = rr'(x_1x_4 - x_2x_3)$. In the general case described by $P$, however, this does not hold: one has

$$z_1z_4 - z_2z_3 = \sum_{i<j} x_ix_j(p_i - p_j)(p'_i - p'_j)$$

and the stationary solution $\pi = \pi P$ need not satisfy the 'linkage equilibrium' condition $\pi_1\pi_4 = \pi_2\pi_3$, even if the initial probability did.

## 3. The Payoff Function

For each round of the game, the payoff is given by the matrix

$$\begin{array}{c} \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} R & S \\ T & P \end{pmatrix}. \end{array} \tag{5}$$

The letters are meant to suggest the Prisoner's Dilemma, where $C$ stands for 'cooperate' and $D$ for 'defect'. If both players opt for $C$ they get a 'reward' $R$ which is higher than the 'punishment' $P$ obtained if both choose $D$. But if one player chooses $D$ and the other $C$, the defector gets away with a payoff $T$ (for 'temptation') which is higher than the reward, and the cooperator gets the 'sucker's payoff' $S$ which is even smaller than $P$. We assume furthermore that the joint payoff for two cooperators, $2R$, is larger than $T+S$, the joint payoff for one cooperator and one defector. Thus the Prisoner's Dilemma is characterized by

$$T > R > P > S \quad \text{and} \quad 2R > T + S. \tag{6}$$

Since $D$ dominates $C$ (no matter what the other does, it is best to defect), both players will defect and end up with $P$ as payoff. The pure strategy $D$ is a Nash equilibrium – no player can improve his lot as long as his adversary sticks to it.

But (5) serves to describe all $2 \times 2$-games and not just the Prisoner's Dilemma. In the *Chicken*-game, for example, $D$ means 'escalate' the confrontation and $C$ means 'don't' (i.e. 'chicken out'). The payoff $P$ is both parties escalate is smaller than the payoff if they don't. The payoff $T$ for one-sided escalation is highest again. But – and this is the difference with the Prisoner's Dilemma – the payoff $S$ for a player who does not escalate, but is faced with an escalating adversary, lies between $R$ and $P$. Hence

$$T > R > S > P \tag{7}$$

and the mixed strategy with a ratio $S - P : T - R$ between $C$ and $D$ is the unique Nash equilibrium.

We shall use as our examples only Chicken and the Prisoner's Dilemma, in view of their biological significance; but if not specified otherwise, the results will hold for all $2 \times 2$-games.

Player $I$ obtains as payoff in the $n$th round

$$A_n = Rc_nc'_n + Sc_n(1 - c'_n) + T(1 - c_n)c'_n + P(1 - c_n)(1 - c'_n).$$

His payoff for the iterated game is given by

$$A(E, E') = \sum_{n=0}^{\infty} A_n w^n \tag{8}$$

if $w < 1$ (since each round is 'discounted' by the probability $w$ to stop the interaction). For $w = 1$, this series diverges. In this case we use as payoff

$$A(E, E') = \lim_n A_n. \tag{9}$$

In both cases, it is useful to write the payoff as sum

$$A(E, E') = G_1\Gamma_1 + G_2\Gamma_2 + G_3\Gamma_3 + G_4\Gamma_4, \tag{10}$$

with

$$G_1 = (R - T) + (P - S), \qquad G_2 = S - P, \qquad G_3 = T - P \quad \text{and} \quad G_4 = P.$$

THEOREM (Nowak and Sigmund, 1989b): *The payoff for the iterated game is given by* (10) *with*

$$\Gamma_1 = cc', \qquad \Gamma_2 = c, \qquad \Gamma_3 = c', \qquad \Gamma_4 = 1 \tag{11}$$

*if* $w = 1$, *and, if* $w < 1$, *with*

$$\begin{aligned}
\Gamma_4 &= \frac{1}{1 - w}, \\
\Gamma_3 &= \frac{1}{1 - w}\frac{1}{1 - uw^2}(e' + wr'e), \\
\Gamma_2 &= \frac{1}{1 - w}\frac{1}{1 - uw^2}(e + wre'), \\
\Gamma_1 &= \frac{1}{1 - u^2w^2} \\
&\quad \left(yy' + wzz' + \frac{w^2}{1 - w}\left(vv' + \frac{u}{1 - uw^2}[v'(e + wre') + v(e' + wr'e)]\right)\right),
\end{aligned} \tag{12}$$

*where*

$$\begin{aligned}
&e = (1 - w)y + wq, \qquad e' = (1 - w)y' + wq', \\
&u = rr', \qquad z = \alpha(y'), \qquad z' = \alpha'(y), \qquad v = \alpha(q') \qquad v' = \alpha'(q).
\end{aligned} \tag{13}$$

The proof is a straightforward computation. We note that for $w = 1$, the initial probabilities $y$ and $y'$ play no role. Also, if one denotes by $A_w$ and $A_1$ the payoffs given by (12) and (11), then $\lim(1 - w)A_w = A_1$ for $w \uparrow 1$, as has to be expected.

## 4. No Discount of the Future

For $w = 1$, we can neglect the initial probability $y$ and speak of $(p, q)$-strategies. This case is well understood (Nowak, 1990) and can be used to illustrate the proposed adaptive dynamics.

The group-selectionist approach would be to look for a policy optimizing the total

payoff. If all members of a population use the same strategy $E = (p, q)$, which choice would be best? This means to maximize $A(E, E)$, i.e. the expression $G_1 s^2 + (G_2 + G_3)s$, with $s = (1 - r)^{-1} q \in [0, 1]$. The solution $s_0$ yields a straight line in the $(p, q)$-space through the reciprocal strategy $(1, 0)$. In most cases it reduces to $s_0 = 0$ or $s_0 = 1$ (or both), corresponding to $q = 0$ or $p = 1$.

But this approach is not in the Darwinian spirit, where group benefit is of secondary importance. For Axelrod's values $T = 5$, $R = 3$, $P = 1$ and $S = 0$, for example, the maximal payoff is attained whenever $p = 1$. But a population with $p = 1$ and $q = 1$ would be invaded by less cooperative strategists. It is the *individual* success that counts.

Thus, we will look for the best strategy of an individual mutant. If the rest of the population is homogeneous, this reduces to finding the best reply $E = (p, q)$ against a known strategy $E' = (p', q')$. Maximizing $A(E, E')$, i.e. the expression $G_1 cc' + G_2 c + G_3 c' + G_4$ means in view of $c' = \alpha'(c)$ to maximize $c^2 G_1 r' + c(G_1 q' + G_2 + G_3 r')$ as a function of $c$. In most cases (for example if $G_1 r' > 0$) this maximum is attained for $c = 0$ or $c = 1$, i.e. *AllD* or *AllC*. If the expression is maximized for a $c_0 \in (0, 1)$, the corresponding solutions in the $(p, q)$-space are the points on the straight line $p(q' + c_0 r') + q(1 - q' - c_0 r') = c_0$. (For Axelrod's values, if $p' = 1$ and $q' = \frac{1}{2}$, all strategies with $3p + q = 2$ are optimal replies.)

In biological applications, of course, it is not likely that a mutant will jump right away to his optimal strategy. It seems more plausible to assume that small individual deviations will explore the strategy space and that the population will evolve under selection into the direction which seems most promising.

If the homogeneous population is in the state $E' = (p', q')$, this direction is given by the gradient of the payoff $A(E, E')$ of a mutant strategy $E = (p, q)$, with components $\partial A/\partial p$ and $\partial A/\partial q$ evaluated at $E = E'$.

This defines a vectorfield in the strategy space pointing into the direction which optimizes a mutants increase in payoff. The corresponding *adaptive dynamics* is given by

$$\dot{p} = \frac{\partial A}{\partial p}(E, E'), \qquad \dot{q} = \frac{\partial A}{\partial q}(E, E') \tag{15}$$

where the derivatives are evaluated at $E = E'$.

We emphasize that (15) is not the gradient of $A(E, E)$, which would, in our situation, correspond to a group-selectionist approach. The vector field (15) points into the direction which is most advantageous for the single mutant. Under selection, the population as a whole moves into this direction; this alters the fitness landscape in such a way that the optimal direction changes gradually. We note that in more elaborate models involving genetic or developmental constraints, the vector field has to be multiplied by some covariance matrix. This more general model of frequency dependent selection is studied in Hofbauer and Sigmund (1990), where it is shown that such adaptations can lead to cyclic or even chaotic dynamics. But nothing of the sort happens for the present model.

THEOREM. *The adaptive dynamics is given by*

$$\dot{p} = \frac{q}{(1-r)^3}\left[G_1 q + (G_2 + G_3 r)\frac{1-r}{1+r}\right],$$

$$\dot{q} = \frac{1-p}{(1-r)^3}\left[G_1 q + (G_2 + G_3 r)\frac{1-r}{1+r}\right]. \tag{16}$$

This follows by a straightforward computation.

We note that *both components have the same sign.* The region in the strategy space where this sign is positive will be called the *C-region*, and the other the *D-region.*

If $G_1 = 0$, the $C$-region is bounded by the straight line $G_2 + G_3 r = 0$. If $G_1 \neq 0$, it is bounded by the graph of

$$q = f(r) := \frac{G_2 + G_3 r}{-G_1}\,\frac{1-r}{1+r}.$$

We are only interested in values $r \in (-1, 1)$. In this strip $f$ is convex if $G_1^{-1}(G_2 - G_3)$ is negative, and concave if it is positive. It converges to 0 for $r \uparrow 1$ and to infinity (with the sign of $-G_1^{-1}(G_2 - G_3)$) for $r \downarrow -1$. If $|G_2| < |G_3|$ it has a zero for $r = -G_3^{-1}G_2$ in $(-1, 1)$.

The vector $(\dot{p}, \dot{q})$ at the point $E = (p, q)$ is orthogonal to the line from $E$ to the reciprocal strategy $(1, 0)$. In a homogeneous population, therefore, there is no

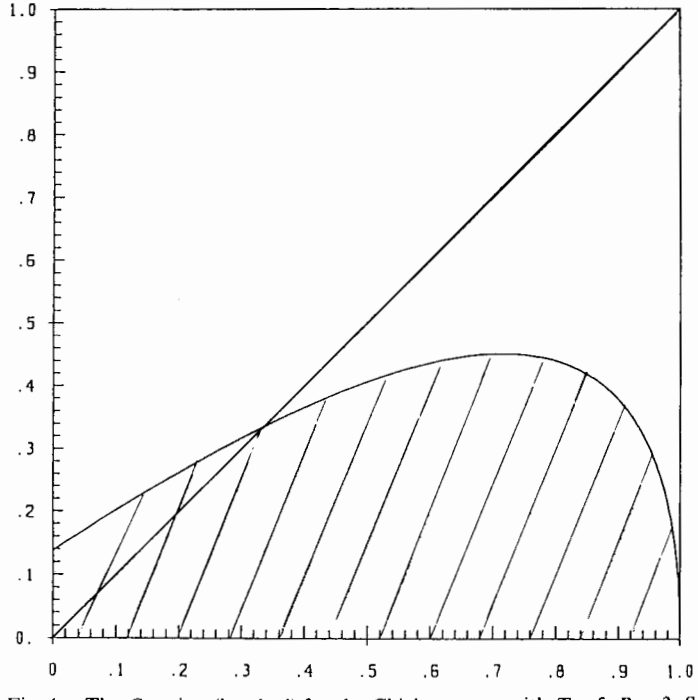


Fig. 1. The $C$-region (hatched) for the Chicken game with $T = 5$, $R = 3$, $S = 1$, $P = 0$.

evolutionary tendency towards this strategy. Reciprocity is the pivot rather than the aim of this evolution. We also note that the magnitude of (16) is (roughly) proportional to the distance from the reciprocal strategy.

Let us consider Chicken as an illustration. In this case $G_1 < 0$ and $G_3 > G_2 > 0$. Thus $f$ is concave and has a root in $(-1, 0)$. The $C$-region is sketched in Figure 1 (we used that $f'(1) \in (0, 1)$ and $f(0) = -(G_2/G_1) \in (0, 1)$). This region contains all strategies with $q = 0$, but no strategy with $q = 1$ or $p = 1$. The equilibria of (16) are on the curve $q = f(r)$. Among the strategies which are unconditional, the only equilibrium is given by

$$p = q = \frac{S - P}{S - P + T - R}$$

which is the unique Nash solution for the nonrepeated game.

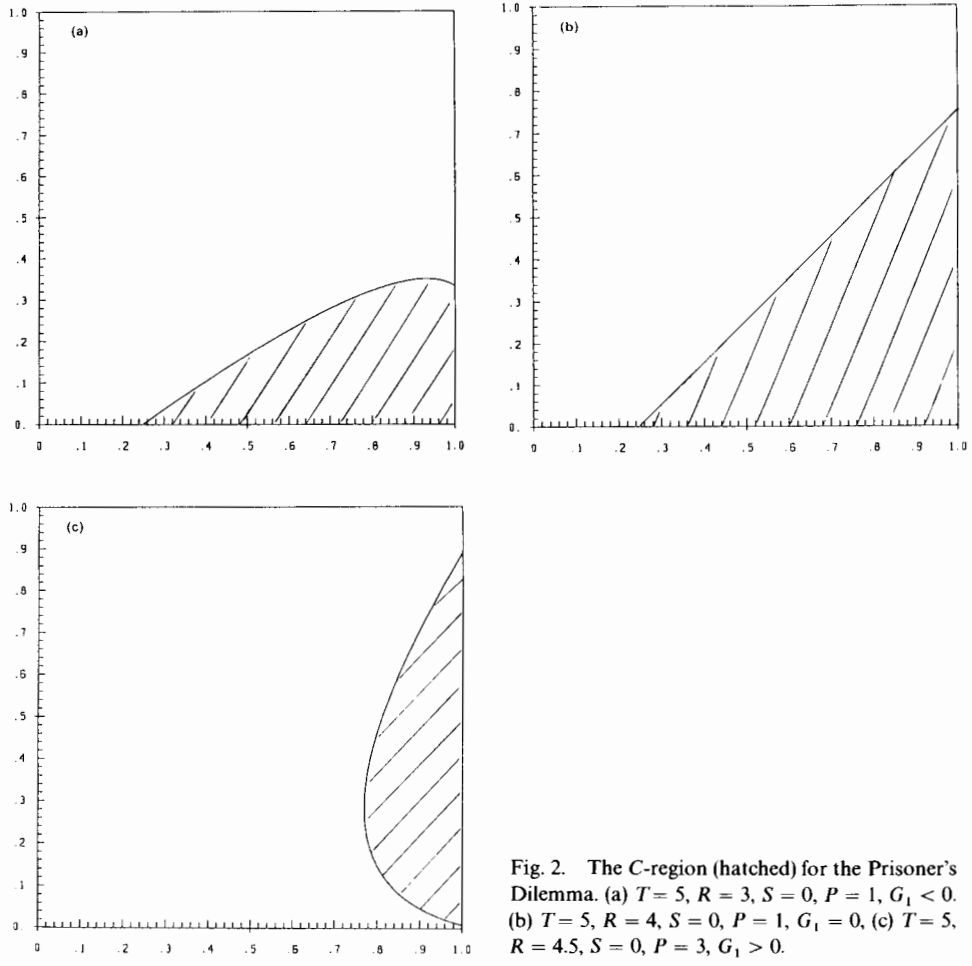


Fig. 2. The $C$-region (hatched) for the Prisoner's Dilemma. (a) $T = 5$, $R = 3$, $S = 0$, $P = 1$, $G_1 < 0$. (b) $T = 5$, $R = 4$, $S = 0$, $P = 1$, $G_1 = 0$, (c) $T = 5$, $R = 4.5$, $S = 0$, $P = 3$, $G_1 > 0$.

For the Prisoner's Dilemma, we have $G_3 > 0 > G_2$, while $G_1$ can have any sign. The function $f(r)$ has a zero at $(P - S)/(T - P)$ in $(0, 1)$ iff $2P < S + T$. In this case (which always holds if $G_1 < 0$) the strategies $(p, 0)$ with $(P - S)/(T - P) < p < 1$ belong to the $C$-region. So do (in every case) the strategies $(1, q)$ with $0 < q < 1 - (T - R)/(R - S)$. The unconditional strategies do not belong to the $C$-region (see Figure 2).

We briefly sketch how to determine the set of all strategies $E'$ which can invade a population of $E$-strategists in the sense that $A(E', E) > A(E, E)$ (Nowak, 1990). With $c' = c(E', E)$ and $s = c(E, E)$ one obtains

$$A(E', E) - A(E, E)$$
$$= G_1(rc'^2 + qc' - s^2) + G_2(c' - s) + G_3(rc' + q - s)$$

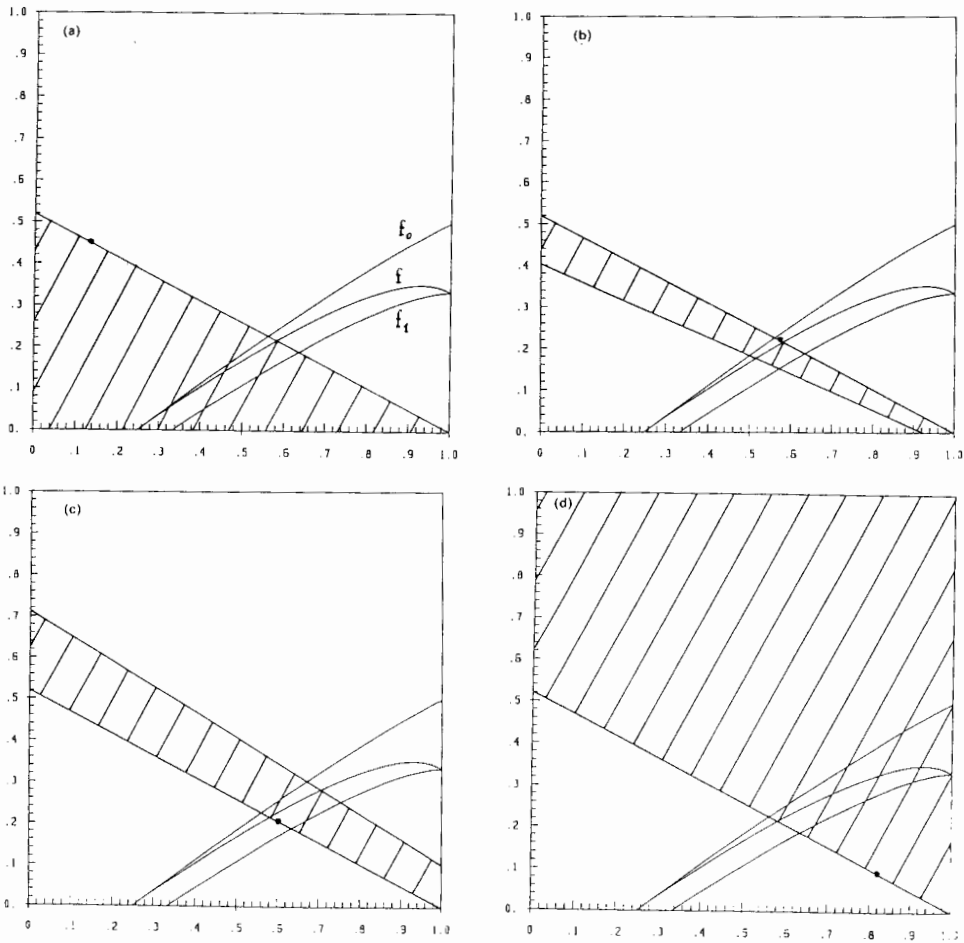


Fig. 3. The Prisoner's Dilemma with $G_1 < 0$ (the numerical values are the same as for (a) in Fig. 2). The hatched region corresponds to the strategies which can invade $E$ (represented by the dot). In (a), (b), (c) and (d) four different positions of $E$ are shown. In the four cases $E$ lies on the same line through $(1, 0)$, and the $C$-level $s$ of $E$ against itself (see Eq. (4)) is therefore the same.

which vanishes for $c' = s$ and (if $G_1 r \neq 0$) for

$$c' = \bar{s} = -\frac{q}{r(1-r)} - \frac{G_2 + G_3 r}{G_1 r}. \tag{17}$$

One has $s = \bar{s}$ iff $q = f(r)$, $\bar{s} = 0$ iff

$$q = f_0(r) := (1 + r) f(r) \tag{18}$$

and $\bar{s} = 1$ iff

$$q = f_1(r) := f_0(r) - r(1 - r).$$

For the Prisoner's Dilemma, the curves given by $f_0$ and $f$ intersect in $q = 0$, $p = (P - S)/(T - P)$ (provided $2P < S + T$) and those given by $f_1$ and $f$ in $p = 1$, $q = 1 - (T - R)/(R - S)$. The points in the strategy space satisfying $q = f_0(r)$ are in the $C$-region if $G_1 > 0$ and in the $D$-region if $G_1 < 0$. For $q = f_1(r)$ the opposite holds. If $r \neq 0$, the strategies $E'$ with $c'$ between $s$ and $\bar{s}$ can invade if $G_1 < 0$ (see Figure 3) those in the complement can invade if $G_1 > 0$.

Let us consider $G_1 < 0$ and choose a value $\hat{q} > 0$ such that there is a unique equilibrium $(\hat{p}, \hat{q})$ in the interior of the state space. Let us assume that $\hat{q}$ is evolutionarily fixed in some way or other and that variation can only occur in $p$. Then $A(p, \hat{p}) < A(\hat{p}, \hat{p})$ for all $p \neq \hat{p}$, so that $\hat{p}$ is a strict Nash equilibrium and, in particular, an evolutionarily stable strategy: no 'mutant' $p$-value can invade. However, it is difficult to see how $\hat{p}$ can become established in the first place. Indeed, as can easily be seen from Figure 3, we have for any $p \neq \hat{p}$ that $A(\hat{p}, p) < A(p, p)$ so that the $\hat{p}$ values cannot invade a homogeneous $p$-population. The adaptive dynamics – which now reduces to the $p$-component of (16) – always points away from the ESS-value $\hat{p}$. In this sense *an evolutionarily stable strategy can be inaccessible.*

We note that the strategies $E$ with $A(AllD, E) = A(E, E)$ are those with $q = f_0(r)$ (and *AllD* itself). The strategies with $A(AllC, E) = A(E, E)$ are those with $q = f_1(r)$ (and *AllC* itself).

If we assume some minimal *noise-level*, so that the strategy space reduces to $\varepsilon < p$, $q < 1 - \varepsilon$, then there is a tendency, within the $C$-region, to approach $p = 1 - \varepsilon$, $q = f(1 - \varepsilon - q)$, i.e.

$$p \sim 1, \qquad q \sim 1 - \frac{T - R}{R - S} \tag{20}$$

in the sense that every small deviation decreasing neither $p$ nor $q$ will succeed. However, if $G_1 > 0$, a population evolving towards the limit given by (20) will leave the south-east corner of the strategy space which is bounded by $q = f_1(r)$, and then *AllD* can invade. By looking only at the dynamics given by (16) one does not notice the important role of the sign of $G_1$.

In a population belonging to the $C$-region, *AllD* can only appear through a large mutational jump. The gradient approach is *myopic* in the sense that it evaluates only those fluctuations which are close to the established strategy.

In an interesting paper, Molander (1985) has shown that for a small noise level $\varepsilon$, the strategy with

$$p \sim 1, \qquad q \sim \min\left[1 - \frac{T-R}{R-S}, \frac{R-P}{T-P}\right] \tag{21}$$

maximizes the payoff of the population, subject to being invasion-proof against defectors using strategies with $q = 0$. The sign of $G_1$ determines for which value the minimum in (21) is attained. Our approach takes into account a larger class than the $(p, 0)$-strategies.

THEOREM. *Let us assume that $\varepsilon < p, q < 1 - \varepsilon$. Then the strategy with*

$$p = 1 - \varepsilon, \qquad q = \min[f(1 - \varepsilon - q), f_0(1 - \varepsilon - q)] \tag{22}$$

*maximizes the payoff of the population, subject to being invasion-proof under selection against defectors using any strategy with a smaller p- or q-value.*

Of course (22) converges for $\varepsilon \to 0$ to the limiting value (21).

## 5. The Special Case of Equal Gains from Switching

We now turn to the case $w < 1$, where the initial probability $y$ plays a role. For the sake of simplicity, we first consider the special case where the gain $T - R$ obtained by switching from $C$ to $D$ against a $C$-opponent is the same as the gain $P - S$ against a $D$-opponent. This special case $G_1 = 0$ holds for example for the Prisoner's Dilemma with $T = 4$, $R = 3$, $P = 2$, $S = 1$ as considered by Smale (1980), or for the numerical simulations in Müller (1987) and in Boyd (1988). It also holds for zero sum games, where $R = P = 0$ and $S = -T$.

The dynamics (15) is now supplemented by the equation

$$\dot{y} = \frac{\partial A}{\partial y}(E, E'), \tag{23}$$

where the right-hand side is evaluated at $E = E'$.

THEOREM. *The adaptive dynamics for $w < 1$ and $G_1 = 0$ is given by*

$$\dot{y} = d, \qquad \dot{p} = d\,\frac{w}{1-w}\,\frac{e}{1-wr}, \qquad \dot{q} = d\,\frac{w}{1-w}\left(1 - \frac{e}{1-wr}\right), \tag{24}$$

*where*

$$d = d(r) = \frac{G_2 + G_3 wr}{1 - w^2 r^2}, \qquad e = (1 - w)y + wq. \tag{25}$$

Since $0 < e < 1 - wr$, the components of the dynamics all have the same sign, given by $G_2 + G_3 wr$. Again we can speak of a $C$-region (where all the parameters $y, p, q$ tend to increase) and a $D$-region. They are separated by a plane parallel to the plane $p = q$.

We note that

$$\frac{w}{1-w}\dot{y} - \dot{p} - \dot{q} = 0 \tag{26}$$

and

$$y\dot{y} + \left(p - \frac{1}{w}\right)\dot{p} + q\dot{q} = 0 \tag{27}$$

so that the vector defined by (24) at $E = (y, p, q)$ lies in the plane orthogonal to $(1 - w, -w, -w)$ and is tangent to the sphere with center $(0, 1/w, 0)$ through $E$.

For fixed $E$, the set of strategies having a given payoff $A(E', E)$ lies on a plane. In particular, the set of strategies $E'$ with $A(E', E) = A(E, E)$ is the intersection of the strategy space with the plane through $E$ which is orthogonal to $(\dot{y}, \dot{p}, \dot{q})$. We know that this plane contains $(0, 1/w, 0)$ and is parallel to $(1 - w, -w, -w)$. Hence

THEOREM. *The strategies $E'$ which can invade $E = (y, p, q)$ in the sense that $A(E', E) > A(E, E)$ are those in the open half-space into which* (24) *is pointing.*

One can verify that

$$A(AllD, E) - A(E, E) = -\frac{d}{1-w}(1 + wr)e$$

and

$$A(AllC, E) - A(E, E) = \frac{d}{1-w}(1 + wr)(1 - wr - e).$$

This implies the following theorem.

THEOREM. *For the vector field given by* (24), *one has*

$$\dot{y} = \frac{1-w}{1-w^2r^2}[A(AllC, E) - A(AllD, E)],$$

$$\dot{p} = \frac{w}{1-w^2r^2}[A(E, E) - A(AllD, E)], \tag{28}$$

$$\dot{q} = \frac{w}{1-w^2r^2}[A(AllC, E) - A(E, E)].$$

In the case of the Prisoner's Dilemma with equal gains from switching, the $C$-region is nonempty iff $w > (P - S)/(T - P)$. (For this to be possible we need $2P < S + T$). In this case the $C$-region is a prism, one edge of which consists of the reciprocal strategies $(y, 1, 0)$.

If there exists a minimal noise-level $\varepsilon$, the same reasoning as in the last section yields the following theorem.

THEOREM. *If we assume that* $\varepsilon < p, q < 1 - \varepsilon$ *then the strategy given by*

$$p = y = 1 - \varepsilon, \qquad q = 1 - \varepsilon - \frac{1}{w}\frac{P - S}{T - P} \tag{29}$$

*maximizes the payoff of the population, subject to being uninvadable by any defectors using smaller p-, q- or y-values. (If the value q in* (29) *is negative, there exists no strategy uninvadable by defectors).*

## 6. The General Case

Due to the complicated form of $\Gamma_1$, the general case $G_1 \neq 0$ leads to some rather tedious computations. One gets for the adaptive dynamics

$$\begin{aligned}
&\dot{y} = d - g[y - wr(y - q)],\\
&\frac{1 - w}{w}\dot{p} = d\frac{e}{1 - wr} - \frac{gF}{(1 - wr)(1 - wr^2)},\\
&\frac{1 - w}{w}\dot{q} = d\left[1 - \frac{e}{1 - wr}\right] + \frac{gF}{(1 - wr)(1 - wr^2)} - g[\alpha(y) - wr(y - q)]
\end{aligned} \tag{31}$$

with $d$ as in (25), $e$ as in (13),

$$g = \frac{-G_1}{(1 - wr)(1 - wr^2)}$$

and

$$\begin{aligned}
F = {} & y^2 r(1 - w)(1 - wr)^2 + yq(1 - w)(1 + wr^2 - 2w^2r^3) + \\
& + wq^2(1 + r - wr^2 - w^2r^3).
\end{aligned}$$

For $G_1 = 0$ we are back to (25). We see directly that (31) leads to

$$\frac{w}{1 - w}\dot{y} - \dot{p} - \dot{q} = g[\alpha(y) - y] \tag{32}$$

which has the sign of $G_1(y - s)$, where $s$ is the stationary $C$-level in the population given by (3).

We will consider first the evolution of the initial probability $y$ only, for fixed $p$ and $q$. From (31) we obtain

$$\dot{y} = g\left[\frac{d}{g} - y(1 - wr) - wrq\right],$$

an expression which vanishes for

$$y = (1 - wr)^{-1}\left(\frac{d}{g} - wrq\right). \tag{33}$$

Let us denote by $\bar{y} = \bar{y}(r, q)$ a truncated form of this, given by

$$\bar{y} = (1 - wr)^{-1}\left(\frac{d}{g} - wrq\right) \quad \text{if this value is in } (0, 1),$$

$$= 1 \quad \text{if it is larger than 1,}$$

$$= 0 \quad \text{if it is smaller than 0.}$$

Since the payoff $A(y, y')$ as given by (12) is linear in $y$ and $y'$, we can write it – up to a constant – in the form of an Euclidean inner product $\mathbf{y} \cdot A\mathbf{y}'$ with $\mathbf{y} = (1 - y, y)$ $\mathbf{y}' = (1 - y', y')$ and

$$A = \begin{pmatrix} A(0, 0) & A(0, 1) \\ A(1, 0) & A(1, 1) \end{pmatrix}.$$

A direct computation shows that $A(1, 0) - A(0, 0) = d - gwrq$ and $A(0, 1) - A(1, 1) = g(1 + wrq - wr) - d$. These expressions sum up to $g(1 - wr)$. It follows that

$$\frac{A(1, 0) - A(0, 0)}{[A(1, 0) - A(0, 0)] + [A(0, 1) - A(1, 1)]} = \frac{1}{(1 - wr)}\left(\frac{d}{g} - wrq\right).$$

If $\bar{y} \in (0, 1)$, this implies that it corresponds to an equilibrium: $A(0, \bar{y}) = A(1, \bar{y})$ $(= A(y, \bar{y})$ for all $y \in [0, 1])$. Hence

THEOREM. *For fixed $p$ and $q$, $\bar{y}$ corresponds to a strict Nash equilibrium if $G_1 < 0$. If $G_1 > 0$, the boundary points of $[0, 1]$ which are distinct from $\bar{y}$ are strict Nash equilibria.*

In the $(p, q)$-square, the region with $\bar{y} = 0$ is bounded by the curve $q = (gwr)^{-1}d$. This curve has the point $\mathbf{P}_1 (p = (P - S)[w(T - P)]^{-1}, q = 0)$ in common with the curve $q = \bar{y}$. Similarly the region $\bar{y} = 1$ is bounded by the curve $q = (wr)^{-1}((d/g) - 1 + wr)$, which has the point $\mathbf{P}_2 (p = 1, q = 1 - (T - R)[w(R - S)]^{-1})$ in common with the curve $\bar{y} = p$. Finally, the curve $\bar{y} = s$ also contains the point $\mathbf{P}_1$ and $\mathbf{P}_2$ (always assuming that these points belong to the strategy space).

We have seen that for $w = 1$, the signs of $\dot{p}$ and $\dot{q}$ agree; similarly, if $w < 1$ but $G_1 = 0$, the signs of $\dot{y}$, $\dot{p}$ and $\dot{q}$ agree. In general, this is not the case however. The surfaces $\dot{y} = 0$, $\dot{p} = 0$ and $\dot{q} = 0$ do not coincide. On the other hand, they are not 'in general position' (i.e. transversal) either.

THEOREM. *The equilibria of* (31) *in the $(y, p, q)$-space form a line which is given by the intersection of the surfaces $y = s$ and $y = \bar{y}$.*

Indeed, if $\dot{y} = \dot{p} = \dot{q} = 0$ then (trivially) $y = \bar{y}$ and (by (32)) $y = \alpha(y)$ which is equivalent to $y = s$. Conversely, if $y = s$, then $\dot{y}$, $\dot{p}$ and $\dot{q}$ are positive multiples of $d - gs(1 - wr^2)$, which implies the theorem.

We note in passing that for $E \in \{y = s\}$ one has

$$A(E, E) = \frac{s}{1 - w}(G_1 s + G_2 + G_3)$$

and that the equality $A(AllD, E) = A(E, E)$ holds if and only if $s = -G_1^{-1}(G_2 + G_3 wr)$, i.e.

$$q = (1 - r)\frac{d}{g}\frac{(1 + wr)}{(1 - wr^2)} \tag{34}$$

which means that $E$ is an equilibrium. In the $(p, q)$-space, both the curve (34) and the curve $q = \bar{y}$ converge for $w \to 1$ to $q = f_0(r)$ (where $f_0$ is given by (18)). In this sense, $q = \bar{y}$ bounds the region where *AllD* can invade a stationary strategy. Similarly, $p = \bar{y}$ converges to $q = f_1(r)$ and bounds the region where *AllC* can invade a stationary strategy.

One can also check that for $w \uparrow 1$, the surfaces $\dot{p} = 0$ and $\dot{q} = 0$ both converge to the vertical face defined by $q = f(r)$ and $y \in [0, 1]$, while $\dot{y} = 0$ converges to $y = (1 - r)^{-1}[(1 + r)f(r) - qr]$.

THEOREM. *For the Prisoner's Dilemma, the faces $\dot{p} = 0$ and $\dot{q} = 0$ do not intersect transversally.*

It follows that the faces only touch tangentially.

Indeed, it is easy to see that $\dot{p} = 0$ or $\dot{q} = 0$ can hold only if $r > 0$, as we shall now assume. One then writes $\dot{p} = G_1 U$ and $\dot{q} = G_1(V - U)$ with $U = Ay^2 + Byq + Cq^2 + Dy + Eq$ and $V = Fy + Gq + H$, the coefficients $A$ to $H$ being given by (31):

$$A = r(1 - w)(1 - wr)^2, \qquad B = (1 - w)(1 + wr^2 - 2w^2r^3),$$

$$C = w(1 + r - wr^2 - w^2r^3), \qquad D = (1 - w)(1 - wr^2)\left(-\frac{d}{q}\right),$$

$$E = w(1 - wr^2)\left(-\frac{d}{q}\right), \qquad F = r(1 - w)(1 - wr)(1 - wr^2),$$

$$G = (1 - wr)(1 + wr)(1 - wr^2), \qquad H = (1 - wr)(1 - wr^2)\left(-\frac{d}{g}\right).$$

One notes that $A, B, C, F, G > 0$ while $\dot{p} = 0$ or $\dot{q} = 0$ implies $d/g < 0$, i.e. $D, E, H < 0$. For fixed $y$, $U$ is quadratic in $q$ and convex, while $V$ is affine linear and increasing. If $U(q) = 0$ then $V(q) \leqslant 0$. Indeed, $U(q) = 0$ means

$$q = (2C)^{-1}\{-(By + E) \pm [(By + E)^2 - 4Cy(Ay + D)]^{1/2}\}.$$

One has to show that $V(q) \leqslant 0$, i.e. that

$$G[(By + E)^2 - 4Cy(Ay + D)]^{1/2} \leqslant G(By + E) - 2C(Fy + H). \tag{35}$$

We check first that the term on the right-hand side is positive, since the coefficient of $y$ is

$$GB - 2CF = (1 - w)(1 - wr)^2(1 - wr^2)^2$$

and the remaining term is

$$\frac{d}{g} w(1 - wr)^2(1 - wr^2)(1 + 2r + wr^2) \geqslant 0.$$

Next, the expression

$$[G(By + E) - 2C(Fy + H)]^2 - G^2[(By + E)^2 - 4Cy(Ay + D)]$$

is just

$$wr(1 - w)(1 - wr)^2(1 - wr^2)2\left[y(1 - wr^2) - \frac{d}{g}\right]^2$$

which is nonnegative (and 0 iff $y = d/(g(1 - wr^2))$, which is just $\bar{y}(p, q)$ for this value of $q$). Thus (35) holds.

Hence $\dot{p} = 0$ implies $G_1\dot{q} \leqslant 0$, which shows that $\dot{p} = 0$ and $\dot{q} = 0$ cannot intersect transversally.

We conjecture that more is true: for all reactive games

(a) if $G_1 r < 0$ then $\dot{p} > 0$ implies $\dot{q} > 0$,
(b) if $G_1 r > 0$ then $\dot{q} > 0$ implies $\dot{p} > 0$.

The condition $G_1 < 0$ means that by a 'reform' (i.e. switching from $D$ to $C$) one gains more against a $C$-opponent than against a $D$-opponent, and the condition $r > 0$ means that such a reform increases the frequency of ulterior $C$'s. The implication in statement (a) means that if it pays to be more grateful, it pays to be also more forgiving.

It is easy to check that with $r = 0$, i.e. for unconditional strategies,

$$\frac{1 - w}{w}\dot{p} = e(G_2 + G_1 q) \quad \text{and} \quad \frac{1 - w}{w}\dot{q} = (1 - e)(G_2 + G_1 q)$$

have always the same sign, so that $\dot{p} = 0$ iff $\dot{q} = 0$ iff $p = q = -(G_2/G_1)$. This value is in $(0, 1)$ for the Chicken game, but for the Prisoner's Dilemma, it is always negative.

## 7. Discussion

In this paper, we have dealt mostly with the evolution of *homogeneous* populations. In Nowak and Sigmund (1989a, b) it is shown that heterogeneous populations consisting of three or four subtypes can already exhibit a remarkable variety of selection dynamics, including limit cycles and heteroclinic cycles. It seems difficult to decide how general these examples are, and to go beyond numerical simulations. Computer experiments tend to suggest that heterogeneous populations end up in more or less stationary distributions smeared out along arcs of the line of equilibria of (29), or clustering near *AllD*.

The fact that homogeneous populations do not evolve towards *TFT* is not surprising. Tit for Tat fares never better than its opponent (and sometimes a bit

worse). Its success in Axelrod's tournaments is due to the composite structure of the 'population' of contestants, which frequently harmed each other while settling into a cooperative mode with *TFT*. The fact that in a heterogeneous population of reactive $(y, p, q)$-strategies, *TFT* does not emerge as winner, could possibly mean that these reactive strategies form an ensemble which too narrow to be representative.

Apart from computer simulations, there are few treatments of the heterogenous case. The importance of composite populations emerges from Boyd and Lorberbaum (1987) where it is shown that every pure strategy can be invaded by the joint effect of two deviating strategies, if $w$ is sufficiently large (see also May, 1987). This does not imply, as Boyd and Lorberbaum claim, that pure strategies are not evolutionary stable, which means proof against invasion by any *one* deviant strategy.

There are several papers dealing with the iterated Prisoner's Dilemma in a noisy environment. Axelrod (1986) has shown that with an error rate of 1%, *TFT* still finishes first in the tournament, while Donninger (1986) showed that with 10%, it finishes sixth. Actually noise hurts *TFT* mostly in conflicts against itself, so that the variegated composition of the tournament does not really display the most obvious weakness of *TFT* which is its echoing effect. Axelrod has suggested that a strategy which is sometimes generous leads to a better performance. The results in Molander (1985) and in our paper confirm this.

An essential notion for games with uncertainties is the concept of perfect equilibrium (Selten, 1975), Boyd (1989) has shown that if the probability of a mistake is always positive, then a pure strategy which is a strong perfect equilibrium against itself is evolutionarily stable. Boyd has shown that *AllD* is a strong perfect equilibrium against itself, and claims the same for *contrite Tit for Tat*, where players cooperate if they are not in good standing or if the opponent is in good standing, and otherwise defect. (One starts in good standing, remains in good standing as long as one follows this strategy, and returns to good standing by cooperating in one round.) This allows to apologize for a mistake. However, if one of two contrite *TFT*-players mistakenly believes to be in good standing, this leads again to endless recriminations. Hence, Boyd's argument is valid if there are no errors in perception but only in implementation, i.e. if the occasional mistake is due to a 'trembling hand', but not if it is due to a 'fuzzy mind'.

Contrite Tit for Tat has to take into account more than just the last move by the adversary and hence is not reactive in our sense. Another class of strategies with a longer memory are the *CC*-strategies (cooperate conditionally) studied by Müller (1987). Such strategies are determined by two parameters, the probability to retaliate after an opponents defection, and the 'relaxation time' of the retaliation. An extreme example is *GRIM*, a strategy which always defects after the first defection of the adversary and never cooperates again. Müller has shown that in a noisy environment, this is the *CC*-strategy best suited to invade an *AllD* population. After such an invasion, the level of forgiveness can be raised (it would be interesting to know by how much, in term of the two parameters). But if *AllC* becomes too frequent, *AllD* can take over again.

## Acknowledgements

Part of this paper was written while one of the authors (M.N.) participated in IIASA's Young Scientist's Summer Programme. We gratefully acknowledge support from the Austrian Forschungsförderungs Fond project P6866.

## References

Aumann, R. J. (1981), Survey of repeated games, in R. J. Aumann *et al.* (eds.), *Essays in Game Theory and Mathematical Economics in Honor of Oscar Morgenstern.*

Axelrod, R. (1984), *The Evolution of Cooperation*, Basic Books, New York.

Axelrod, R. and Hamilton, W. D. (1981), The evolution of cooperation, *Science* **211**, 1390–1396.

Axelrod, R. and Dion, D. (1988), The further evolution of cooperation, *Science* **242**, 1385–1390.

Axelrod, R. (1987), The evolution of strategies in the iterated prisoner's dilemma, in Davis, D. (ed.), *Genetic Algorithms and Simulated Annealing*, Pitman, London.

Boyd, R. and Lorberbaum, J. P. (1987), No pure strategy is evolutionarily stable in the Repeated Prisoner's Dilemma, *Nature* **327**, 58–59.

Boyd, R. (1989), Mistakes allow evolutionary stability in the repeated prisoner's dilemma game, *J. Theoret. Biol.* **136**, 47–56.

Donninger, C. (1986), Is it always efficient to be nice? in A. Dieckmann and P. Mitter (eds.) *Paradoxical Effects of Social Behaviour: Essays in Honor of Anatol Rapoport*, Physica, Heidelberg, pp. 123–134.

Eshel, I. and Motro, A. (1981), Kin selection and strong evolutionary stability of mutual help, *Theoret. Population Biol* **19**, 420–433.

Feldman, M. and Thomas, E. (1987), Behavior-dependent contexts for repeated plays of the prisoner's dilemma II: Dynamical aspects of the evolution of cooperation, *J. Theoret. Biol.* **128**, 297–315.

Hofbauer, J. and Sigmund, K. (1988), *Dynamical Systems and the Theory of Evolution*, Cambridge University Press.

Hofbauer, J. and Sigmund, K. (1990), Adaptive dynamics and evolutionary stability, to appear in *Letters Appl. Math.*

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

Maynard Smith, J. (1982), *Evolution and the Theory of Games*, Cambridge University Press.

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

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

Müller, U. (1987), Optimal Retaliation for Optimal Cooperation, *J. Conflict Resolution* **31**, 692–724.

Nowak, M. and Sigmund, K. (1989a), Oscillations in the Evolution of Reciprocity, *J. Theor. Biol.*, **137**, 21–26.

Nowak, M. and Sigmund, K. (1989b), Game dynamical aspects of the Prisoner's Dilemma, *J. Appl. Math. Comp.* **30**, 191–213.

Nowak, M. (1990), An evolutionarily stable strategy may be inaccessible, *Theoret Population Biol.* **142**, 237–241.

Selten, R. and Hammerstein, P. (1984), Gaps in Harley's argument on evolutionarily stable learning rules and in the logic of *TFT, Behavioural and Brain Sci.* 7, 115–116.

Selten, R. (1975), Reexamination of the perfectness concept for equilibrium points in extensive games, *Internat. J. Game Theory* **4**, 25–55.

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

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