GAME DYNAMICAL ASPECTS OF THE PRISONER'S DILEMMA

Martin Novak
Karl Sigmund

December 1988
WP-88-125

IIASA
International Institute
for Applied Systems Analysis

GAME DYNAMICAL ASPECTS OF THE PRISONER'S DILEMMA

Martin Novak
Karl Sigmund

December 1988
WP-88-125

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

Foreword

A game dynamical analysis of the Iterated Prisoner's Dilemma reveals its complexity and unpredictability. Even if one considers only those strategies where the probability for cooperation depends on the last move, one finds stable polymorphisms, multiple equilibria, periodic attractors and heteroclinic cycles.

Alexander Kurzhanski
Program Leader
System and Decision Sciences Program

Game Dynamical Aspects of the Prisoner's Dilemma

Martin Nowak
Institut für Theoretische Chemie der Universität Wien
Währingerstr. 17 , A-1090 Wien , Austria

Karl Sigmund
Institut für Mathematik der Universität Wien
Strudlhofg. 4 , A-1090 Wien , Austria
and IIASA, Laxenburg, Austria

1.Introduction

Ever since the publication of Axelrod's basic book (1984), the Iterated Prisoner's Dilemma (*IPD*) is generally viewed as the major game theoretical paradigm for the evolution of cooperation based on reciprocity. In repeated encounters, two players are faced with the choice to cooperate or to defect (C or D). If both cooperate, their payoff R (reward) is higher than the payoff P (punishment) obtained if both defect. But if one player defects while the other cooperates, then the defector's payoff T (temptation) is higher than R, while the cooperator's payoff S (sucker) is smaller than P. It is furthermore assumed that $R > \frac{1}{2}(S+T)$, so that joint cooperation is more profitable than alternating C and D.

If the game consists of a single encounter, the best option is to defect, no matter what the other player does. Since both players will resort to this solution, they end up with the punishment instead of the reward. A simple argument shows that the same holds if the game consists of a fixed number of encounters (known to both players): one has just to apply the previous reasoning to the last move and then to work backward. But if the length of the game is unknown, as for example if there is a fixed probability w for a further encounter, then the players may 'learn' that it is in their interest to cooperate.

In Axelrod's well known computer tournaments, the simplest strategy did best. This was Tit For Tat (TFT), submitted by Anatol Rapaport : it consists of starting with a cooperative move and then doing whatever the opponent did on his previous move. Most strategies among the 'runner's up' shared with TFT the properties of being nice (i.e. starting with C), provokable and forgiving.

The assessment in Axelrod's contests was established by round robin tournaments. For applications to evolution, Axelrod and Hamilton (1981) stressed the 'ecological approach' and hence the underlying dynamics of the game : each strategy participates to the next generation in proportion to its present success. Thus good strategies spread in the population at the expense of weaker ones, but what is good and what is weak depends on the composition of the population and hence varies in time : it may happen, for instance, that a strategy does well when rare but poorly when it meets itself too often, so that it chokes on its own success. This view of 'frequency dependent fitness values' is at the core of Maynard Smith's applications of game theoretical arguments to evolutionary models (1982), and in particular of his notions of uninvadable phenotype and evolutionarily stable strategy (ESS).

In spite of its success, TFT is not an ESS. For sufficiently high w, it cannot be invaded by All Defect ($ALLD$), as Axelrod has shown. But $ALLC$ for example, does as well as TFT in a population consisting only of itself and TFT, and hence can spread by genetic drift. Once its frequency is sufficiently high, $ALLD$ can take advantage and invade, since it has to fear less retaliation than against *TFT* alone. This argument is due to Selten and Hammerstein (1984), who also pointed out another weakness of *TFT*: if by mistake, one of two *TFT*-players makes a wrong move, this locks the two opponents into a hopeless sequence of alternating *D*'s and *C*'s.

Such a mistake is unlikely to occur in a computer tournament, but has to be expected in 'real life'. Actual biological situations are fraught with errors and uncertainties. The answer to the opponents last move (which may be misperceived in the first place) is only an increase or decrease in the readiness to cooperate. This emerges quite clearly from Milinski's (1987) experiments on sticklebacks or Lombardo's (1987) data on tree swallows.

As May (1987) points out, it is important to 'take more account of intrinsic stochasticities and of evolutionary stability against representative ensembles of mutant strategies'.

This suggests considering stochastic strategies given by three parameters (y, p, q), where y is the probability to cooperate in the first move, and p and q the conditional probabilities to cooperate, given that the adversary's last move was a C or a D. Thus a strategy is defined by a triple $(y, p, q) \in [0, 1]^3$. For example, $ALLC = (1, 1, 1)$ or $TFT = (1, 1, 0)$ are extremal representatives. A p-value of 0.95 can be interpreted as a mixed strategy, or as a decision to cooperate after C, subject to an error rate of 0.05 due to incomplete control over one's own action. Tit For Two Tats (*TFTT*, which defects only after two consecutive *D*'s from the opponent) is not a member of this class, and neither is a strategy taking also account of one's own previous move. Most of the programs submitted to Axelrod's tournaments were much more complex. But in spite of their limitations, strategies of type (y, p, q) already display a remarkable variety of interactions.

There are several candidates for an appropriate evolutionary dynamics, all leading more or less to the same outcome. We shall use here the Ansatz given by Taylor and Jonker (1979): the rate of increase of a strategy is the difference between its payoff and the average payoff in the population. This game dynamics, which relates well to the theory of evolutionary stability, has been studied extensively, e.g. by Zeeman (1980) or by Schuster and Sigmund

(1985). We refer to Hofbauer and Sigmund (1988) for a recent treatment.

If only two strategies are competing, one can find (i) dominance (e.g. $E_1 = ALLD$ always outcompetes $E_2 = ALLC$),(ii) bistability ($E_1 = ALLD$ and the stochastic TFT $E_3 = (y, 1, 0)$, with $0 < y < 1$, do never coexist, but which one wins depends on the initial frequencies) and (iii) stable polymorphism ($E_2 = ALLC$ and $E_3 = (y, 1, 0)$ settle down to a predetermined equilibrium). Roughly speaking, case (ii) occurs frequently if the two competing strategies differ only in p, and case (iii) if they differ only in q. If we consider a simulated evolutionary process consisting of alternating (a) periods of selection described by the game dynamics and (b) mutations introducing a small population which differs from the current population by a slight deviation , sometimes in p and sometimes in q, then the tendency is either towards $ALLD$ or towards a state with $p = 1$ and some well defined q-value, but usually not towards TFT. Thus one should sometimes forget a bad turn, but never a good one.

For three competing strategies, it may depend on the initial condition whether a polymorphic state gets established or not. An example is obtained by the three strategies E_1, E_2 and E_3 above : most initial conditions lead to an equilibrium of all three strategies which is asymptotically but not evolutionarily stable, other initial conditions lead to a monomorphic state consisting of E_1 only. For other choices of E_1, E_2 and E_3, a 'stone-scissors-paper' effect can occur :E_1 dominates E_2, which dominates E_3, which in turn dominates E_1 again. The presence of all three strategies in the population can lead to neutral oscillations or to a stable polymorphism. It can also lead to a heteroclinic cycle a la May and Leonard (1975): for a long time, one of the strategies will seem to have complete ascendancy, until it is replaced , quite suddenly, by the next strategy, which will seem to have the upper hand until it is superseded in its turn by the next one etc... in an endless cyclic alternation, proceeding in fits and starts, the time between the revolutions growing exponentially. In practice, this behaviour means that a random fluctuation (or a computer round-off) will wipe out one of the strategies during its 'weak' phase, and so lead to the fixation of the dominating of the two remaining strategies. But it is completely impossible to predict which one will turn out to be the ultimate winner. (This case can occur even if y, the initial readiness for cooperation, is the same one for all three strategies).

With four competing strategies, one can find oscillations which damp down to some equilibrium, or which 'explode' in the aforementioned way, or which settle down to some predetermined amplitude and period. Such limit cycles can be found quite frequently, in fact. Whether chaotic oscillations occur is probable but we have found none so far.

In section 2 , we introduce the explicit game dynamics and compute the payoff matrix for (y, p, q)-strategies. In section 3, we investigate the evolution if a single parameter is varied, and in section 4 we study examples of oscillating behaviour in low dimensions. In the discussion in section 5, we refer to other dynamical approaches to the *IPD* and suggest some further lines of investigation.

2. The Payoff-Matrix and the Game Dynamics

Each game consists of a sequence of rounds between two players having the options C and D each. The probability that the game is extended by another round will be denoted by $w \in [0,1]$. This parameter w can also be viewed as a discount factor for the future payoff.

We denote by a_n and a'_n the probability that the first (resp. second) player cooperates in the n-th round. If the first player uses strategy (y, p, q) and the second player (y', p', q') , then $(a_0, a'_0) = (y, y')$, $(a_1, a'_1) = (z, z')$ and

$$\begin{aligned} a_{n+2} &= u a_n + v \\ a'_{n+2} &= u a'_n + v' \end{aligned} \tag{1}$$

with

$$z = p y' + q(1 - y')$$

$$z' = p' y + q'(1 - y)$$

$$v = pq' + q(1-q')$$

$$v' = p'q + q'(1-q)$$

$$u = (p-q)(p'-q')$$

By A_n and A'_n we denote the expected payoff for the first and second player in the n-th round , and by $A = \sum A_n w^n$ and $A' = \sum A'_n w^n$ their total payoff. Clearly

$$A_n = a_n a'_n (R - S - T + P) + a_n (S - P) + a'_n (T - P) + P$$

and for A'_n the same with S and T exchanged.

For $w < 1$, (1) allows to compute the payoff by a simple geometric sum which yields

$$A = (R - S - T + P)\Gamma_1 + (S - P)\Gamma_2 + (T - P)\Gamma_3 + P\Gamma_4 \tag{2}$$

where

$$\Gamma_1 = \frac{1}{1-u^2w^2}[yy' + wzz' + \frac{w^2}{1-uw^2}[uv'(y+wz) + uv(y'+wz') + \frac{vv'(1+uw^2)}{1-w}]]$$

$$\Gamma_2 = \frac{1}{1-uw^2}[y + wz + \frac{w^2}{1-w}v]$$

$$\Gamma_3 = \frac{1}{1-uw^2}[y' + wz' + \frac{w^2}{1-w}v']$$

$$\Gamma_4 = \frac{1}{1-w}$$

if $|u| < 1$. For $|u| = 1$ we have deterministic strategies, i.e. $p, q \in \{0,1\}$. This yields four special cases for Γ_1, Γ_2 and Γ_3 (Γ_4 remains unchanged):

(i) $p = p' = 1, q = q' = 0$ (Tit For Tat against itself). The sequence a_n is $y, y', y, y', \ldots$ periodically.

$$\Gamma_1 = \frac{yy'}{1-w}$$

$$\Gamma_2 = \Gamma_3 = \frac{y' + wy}{1-w^2}$$

(ii) $p = p' = 0, q = q' = 1$ (the paradoxical strategy against itself). The sequence a_n is $y, 1-y', y, 1-y', \ldots$ periodically. We have

$$\Gamma_1 = \frac{1}{1-w^2}[yy' + w(1-y)(1-y')]$$

$$\Gamma_2 = \Gamma_3 = \frac{1}{1-w^2}[y + w(1-y')]$$

(iii) $p = q' = 1, p' = q = 0$ (Tit For Tat against the paradoxical strategy). The sequence a_n is now $y, y', 1-y, 1-y', \ldots$ with period 4. We have

$$\Gamma_1 = \frac{1}{1-w^4}[yy' + wy'(1-y) + w^2(1-y)(1-y') + w^3 y(1-y')]$$

$$\Gamma_2 = \frac{1}{1-w^4}[y' + wy + w^2(1-y') + w^3(1-y)]$$

$$\Gamma_3 = \frac{1}{1-w^4}[y + wy' + w^2(1-y) + w^3(1-y')]$$

(iv) $p = q' = 0, p' = q = 1$ is like (iii) with roles reversed.

In more general situations, the conditional strategy in each move may be determined by the outcome of the k previous moves of both players, for some fixed memory length k. This can be modelled as a Markov chain. In our case, for instance, the states in the n-th round are the pairs (C,C),(C,D),(D,C) and (D,D) of possible moves by the two players, and the transition to the state in the (n+1)-th round is given by the stochastic matrix

$$\Pi = \begin{pmatrix} pp' & p(1-p') & (1-p)p' & (1-p)(1-p') \\ qp' & q(1-p') & (1-q)p' & (1-q)(1-p') \\ pq' & p(1-q') & (1-p)q' & (1-p)(1-q') \\ qq' & q(1-q') & (1-q)q' & (1-q)(1-q') \end{pmatrix}$$

The initial probability distribution is $(yy', y(1-y'), (1-y)y', (1-y)(1-y'))$ and the stationary distribution (for irreducible Π) is given by

$$(1-u)^{-2}[vv', v(1-v'), (1-v)v', (1-v)(1-v')]$$

as can be checked easily. We shall not pursue this approach here, which is of special interst in the case $w = 1$ (no discount of the future), but refer to Nowak(1989).

We now turn to the game dynamics. In principle the strategy set is the three-cube $[0,1]^3 = Q$ and the state of the population is a probability distribution on Q . It is possible to write down some plausible dynamics for the evolution of this distribution in time , but rather difficult to analyse it. We shall therefore assume that only finitely many strategies are present in the population , denoting them by E_1 to E_n and their frequencies by x_1 to x_n. Thus the state of the population at time t is given by the vector $\underline{\mathrm{x}} = \underline{\mathrm{x}}(t)$ in the unit simplex S_n . Since we know the payoff a_{ij} for strategy E_i agaist E_j , i.e. the payoff matrix A , we can compute the average payoff $(A\underline{\mathrm{x}})_i = \sum a_{ij}x_j$ for strategy E_i in the population, and the mean payoff $\underline{\mathrm{x}}A\underline{\mathrm{x}} = \sum x_i(A\underline{\mathrm{x}})_i$ within the population. The game dynamical Ansatz by Taylor and Jonker consists in assuming that $\frac{\dot{x}_i}{x_i}$, the rate of growth of strategy E_i , is given by its relative success, i.e. by the difference $(A\underline{\mathrm{x}})_i - \underline{\mathrm{x}}A\underline{\mathrm{x}}$ between the payoffs for E_i and the mean payoff. This yields

$$\dot{x}_i = x_i((A\underline{\mathrm{x}})_i - \underline{\mathrm{x}}A\underline{\mathrm{x}}) \tag{3}$$

on the (invariant) state space S_n . This type of equations occurs in many biological contexts: we refer to Hofbauer and Sigmund (1988) for a recent survey. We shall use it to study the evolution of a small number $n = 2, 3$ or 4 of competing strategies of the IPD in order to get a feeling for the complexity involved in the full game with its continuum of strategies .

The faces $x_i = 0$ of the population simplex are invariant : if the stategy E_i is missing , it will not be introduced through the competition described by (3) (but possibly by other mechanisms , like mutation , migration ,etc.). On the other hand , it may be that $x_i(0) > 0$ but $\liminf_{t\to\infty} x_i(t) = 0$, which means elimination of E_i . A fixed point $\hat{\underline{\mathbf{x}}}$ in the interior of S_n (i.e. satisfying $\hat{x}_i > 0$ for all i) is a solution of the linear equations

$$(A\underline{\mathbf{x}})_1 = \ldots = (A\underline{\mathbf{x}})_n$$

Generically there is one or no such solution in $intS_n$, in exceptional cases we may have linear manifolds of fixed points . We obtain similarly the fixed points in the lower dimensional faces making up the boundary of S_n . In particular the corners of S_n , i.e. the unit vertices coresponding to the presence of a unique strategy E_i in the population are fixed points .

If (3) is permanent , in the sense that there exists a compact set in $intS_n$ where all orbits in the interior end eventually up , then all strategies present in the population will survive (their frequencies will be bounded away from 0) . In this case there always exists a unique polymorphic equilibrium $\hat{\underline{\mathbf{x}}} \in intS_n$, but it need not be stable. For $n \geq 4$ (but not for $n < 4$) the orbits can converge to a periodic or chaotic attractor. Their time averages, however, converge to $\hat{\underline{\mathbf{x}}}$:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T x_i(t)dt = \hat{x}_i \tag{4}$$

Several conditions for permanence are known (see Hofbauer and Sigmund (1988)). In particular the system cannot be permanent if there exists a Nash equilibrium on the boundary (i.e. a fixed point such that $(A\underline{\mathbf{x}})_i \leq \underline{\mathbf{x}}A\underline{\mathbf{x}}$ whenever $x_i = 0$; recall that for $x_i > 0$ we have $(A\underline{\mathbf{x}})_i = \underline{\mathbf{x}}A\underline{\mathbf{x}}$).

3. Variation of a Single Parameter

For a preliminary orientation we keep two of the three parameters (y, p, q) fixed and consider populations of competing strategies which differ only in the third parameter. For illustrations we shall use Axelrod's payoff values $T = 5, R = 3, P = 1$ and $S = 0$ if not otherwise stated. As discount factor, we shall use $w = 0.9$ for our numerical examples.

(A) Variation of y

This case is the one which is easiest to analyse. Indeed , the payoff given by (2) is affine linear in y and y' . This allows us to use the results from Sigmund (1987).

Let us consider first the pure strategist case : $n = 2, y_1 = 0, y_2 = 1$. The 2×2-matrix A is then easily computed. Let

$$\hat{y} = \frac{a_{12} - a_{22}}{(a_{12} - a_{22}) + (a_{21} - a_{11})} \tag{5}$$

If $a_{12} > a_{22}$ and $a_{21} < a_{11}$, then x_2 converges to 0, the uncooperative strategy wins. If $a_{12} < a_{22}$ and $a_{21} > a_{11}$, then x_1 vanishes. In both cases $\hat{y} \notin [0,1]$. If $a_{12} - a_{22}$ and $a_{21} - a_{11}$ both have the same sign , then $\hat{y} \in (0,1)$. In this case , both terms have to be positive since their sum ,i.e. the denominator of (5), is $(1 - uw)^{-1}$, which is larger than 0. In this case x_2 converges to $\hat{y}$ for all initial conditions $x_2 \in (0,1)$. For Axelrod's values we have $\hat{y} > 1$ for $p - q < 0.4$ (roughly) , $\hat{y} < 0$ for $p - q > 0.8$, and $\hat{y} \in (0,1)$ for values in between (see Fig.1).

Let us now turn to the general case of n strategies $y_1 < \ldots < y_n$ with frequencies x_1 to x_n. The expression

$$\bar{y} = \sum x_i y_i$$

is the average readiness for cooperating in the first move. In Sigmund (1987) it is shown that

$$V(\underline{x}) = \frac{1}{2(1 - uw)} (\bar{y} - \hat{y})^2$$

is a potential for (3), with $\hat{y}$ given as in the pure strategist case (5). More precisely there exists a Riemannian metric on S_n (the so-called Shashahani-metric) with respect to which (3) is a gradient . We may distinguish generically three cases :

(i) if $\hat{y} < y_1$ then $x_1 \to 1$. This means that the strategy least prepared to cooperate wins out. This happens e.g. for $p \leq q$ (for $p = q$ we have $\hat{y} = -1$, which is negative).

(ii) if $\hat{y} > y_n$, then $x_n \to 1$. This means that ultimately , there will be as much cooperation as possible within the population. This is the case in particular for Tit-For-Tat players when $\hat{y} \simeq 2.26 > 1$.

(iii) if $y_1 < \hat{y} < y_n$, there exists a linear manifold of fixed points in $intS_n$, given by $\bar{y} = \hat{y}$. All orbits approach this set (actually along invariants of motion). This means that the population converges to a polymorphic state , where all strategies have the same payoff. This happens, for example, for $(p, q) = (0.75, 0.25)$ where $\hat{y}$ is approximately 0.43. The evolution for $y_1 = 0.2, y_2 = 0.6$ and $y_3 = 0.8$ is sketched in Fig.2.

(B) Variation of q

Since the payoff given by (2) is not linear in q and q' but fractional quadratic we cannot use the same method as before. We are unable to give a full global analysis of the resulting system , and can only offer some arguments supported by numerical simulation.

Let us consider first the special case $R + P = S + T$ (which includes for example Smale's (1980) values: $T = 3, R = 2, P = 1, S = 0$). Then the contribution of Γ_1 in (2) vanishes and we are left with a payoff function A which is fractional linear in q:

$$A(q) = \frac{aq + b}{cq + d}$$

where the a, b, c, d are expressions in the parameter $y = y', p = p'$ and q' (cf. part D). Since these values are all in $[0, 1]$ the denominator is always well defined (we recall that $w < 1$). Thus $q \to A(q)$ is monotonically increasing or decreasing in [0,1] depending on whether $ad - bc$ is positive or negative. This in turn depends on q' and p, but interestingly not on y. More precisely, we shall show in part D that there are two possible cases, depending on the value of

$$\hat{q} = p - \frac{P - S}{T - P} \frac{1}{w}$$

(which is always < 1).

(i) If $\hat{q} < 0$ then $q \to A(q)$ is monotonically decreasing in $[0,1]$ for all values of q'. Thus if $q_1 < q_2$ we have $A(q_1,q_2) > A(q_2,q_2)$ and $A(q_1,q_1) > A(q_2,q_1)$ and so q_1 dominates q_2. Hence it always pays to defect if $\hat{q} < 0$ (i.e. for small values of p -little gratitude from the opponent- and small values of w -small risk of further encounters).

(ii) If $0 < \hat{q} < 1$ then $q \to A(q)$ is increasing for $q' < \hat{q}$ and decreasing for $q' > \hat{q}$. For two strategies $q_1 < q_2$, with frequencies x_1 and x_2, there are three possibilities:

(a) If both q_1 and q_2 are smaller than $\hat{q}$, then $A(q_1,q_1) < A(q_2,q_1)$ and $A(q_1,q_2) < A(q_2,q_2)$. Thus q_2 dominates q_1, and $x_1 \to 0$.

(b) If both q_1 and q_2 are larger than $\hat{q}$, then $A(q_1,q_1) > A(q_2,q_1)$ and $A(q_1,q_2) > A(q_2,q_2)$. Then q_1 dominates q_2, and $x_2 \to 0$.

(c) If $q_1 < \hat{q} < q_2$, then $A(q_1,q_1) < A(q_2,q_1)$ and $A(q_2,q_2) < A(q_1,q_2)$. In this case x_1 converges to the value

$$\frac{A(q_2,q_1) - A(q_1,q_1)}{A(q_2,q_1) - A(q_1,q_1) + A(q_1,q_2) - A(q_2,q_2)}.$$

Hence we obtain a stable polymorphism of the two strategies.

Thus for $\hat{q} > 0$ (a high probability for a return in cooperation), a small increase in cooperation ($q_2 = q + \varepsilon$) will succeed if the overall cooperation ($q_1 = q$) is smaller than $\hat{q}$, but not if it is larger than $\hat{q}$ .The value $\hat{q}$ can be viewed as a stable level of 'forgiveness'.

Let us consider now the case of several strategies $q_1 < \ldots < q_n$, with frequencies x_1 to x_n. Numerical simulations indicate that the following holds :

(i) if $q_n < \hat{q}$, then $x_n \to 1$;

(ii) if $q_1 > \hat{q}$, then $x_1 \to 1$;

(iii) if $q_k < \hat{q} < q_{k+1}$, then x_k and x_{k+1} converge to some strictly positive values summing up to 1, and all other strategies vanish. Thus a mixture of the two strategies 'closest' to the value $\hat{q}$ gets established.

The effect of a large number of mutations introducing new q-values into the population will eventually lead to a population which is almost homogeneous and consists only of strategies very close to $\hat{q}$.

In the general case $R - S - T + P \neq 0$, the term Γ_1 introduces complications which we cannot fully analyse. The overall effect is to blur the sharp transition, at $\hat{q}$, from defection to cooperation. For Axelrod's values, for example, this blurring effect is quite small, and the overall picture for most parameter values very similar to the special case : the population converges to a more or less homogeneous state with a q-value as close as possible to $\hat{q}$ (0 if $\hat{q} < 0$). For example , we can observe by computation that

$$\lim_{y=p\to 1} \hat{q} \simeq 0.26.$$

If nice ($y \simeq 1$) and grateful ($p \simeq 1$) strategies vary in their readiness to forgive, then the evolutionary tendency is towards $q \simeq 0.26$). It must be stressed, however, that for a small range of parameter values y and p, a more complex outcome is conceivable, especially for large $|R - S - T + P|$.

(C) Variation of p

The situation is closely related, but in some sense almost complementary to the previous one. Again, it is useful to consider first the special case $R+P = S+T$. The payoff function A is fractional linear in p, and hence $p \to A(p)$ is monotonically increasing or decreasing in $[0,1]$, depending on p' and q (but not on y). The crucial parameter is now

$$\hat{p} = q + \frac{P-S}{T-P}\frac{1}{w}$$

(which is always > 0).

(i) if $\hat{p} > 1$, then $p \to A(p)$ is monotonically decreasing in $[0,1]$ for all values of p'. Thus if $p_1 < p_2$, then p_1 dominates p_2. Hence it always pays to defect if $\hat{p} > 1$ (i.e. for large values of q -the readiness to forgive - and small values of w - the risk of further encounters).

(ii) if $0 < \hat{p} < 1$, then $p \to A(p)$ is decreasing for $p' < \hat{p}$ and increasing for $p' > \hat{p}$. For two strategies $p_1 < p_2$ there are three possibilities:

(a) if $p_2 < \hat{p}$, then p_1 dominates p_2;

(b) if $p_1 > \hat{p}$, then p_2 dominates p_1;

(c) if $p_1 < \hat{p} < p_2$, we obtain an unstable equilibrium. Depending on the initial frequency, p_1 or p_2 will outcompete its rival strategy.

Hence for $\hat{p} < 1$ (a small expectation to get away with a defection) a slight increase in cooperation ($p_2 = p + \varepsilon$) will succeed if the overall cooperation ($p_1 = p$) is larger than $\hat{p}$, but not if is smaller. The value $\hat{p}$ can be viewed as a reciprocity threshold: if the average tendency is to defect, then it pays to defect, while if it is to cooperate, then the more one cooperates the better.

In this case, the effect of an evolutionary process of mutation and selection drives the population to the fixation of a pure strategy $p = 0$ or $p = 1$. Which of these alternatives holds depends on the initial phase of the process.

Again the general case is similar if $R - S - T + P$ is relatively small (for example Axelrod's values). The population converges in most cases to a homogeneous state with $p = 0$ or (if the initial population is cooperative, and defection punished severely) with $p = 1$.

(D) More on the special case $R + P = S + T$

For the expression

$$A(q) = \frac{aq + b}{cq + d}$$

given in (B), we have

$$a = (S - P)[w(1-y) + f(1-q')] + (T-P)fr$$

$$b = (S-P)[y + wyp + fpq'] + (T-P)(y' + wyr + wq' + fq')$$

$$c = rw^2$$

$$d = 1 - prw^2$$

where $r = p - q'$ and $f = w^2(1-w)^{-1}$. A rather tedious computation shows that

$$ad - bc = (1-w)^{-1}w(1+rw)(qw + y(1-w))[(T-P)rw - (P-S)].$$

The first four factors on the right hand side are always positive, so that $q \longrightarrow A(q)$ is strictly increasing in $[0,1]$ if and only if the parameter q' satisfies $q' < \hat{q}$, with $\hat{q} = p - \frac{P-S}{T-P}\frac{1}{w}$ which is independent of y.

In the same manner, we obtain that $p \longrightarrow A(p)$ is strictly increasing if and only if the parameter p' satisfies $p' > \hat{p}$, with $\hat{p} = q + \frac{P-S}{T-P}\frac{1}{w}$.

It is interesting that the two conditons for $A(q)$ and $A(p)$ to increase are actually the same. Thus let us consider a population with strategy (p,q) and a small mutant population with a strategy slightly differing in either its p or its q value. If

$$p - q > \frac{P-S}{T-P}\frac{1}{w} \tag{6}$$

then the mutant can invade and take over iff its strategy is more cooperative (higher p or q); if the inverse inequality holds, the mutant can invade iff its strategy is less cooperative.

Since $R < \frac{1}{2}(T+S)$ and $R + P = S + T$ by assumption, we have $\frac{P-S}{T-P} < 1$. If $w < \frac{P-S}{T-P}$, there is no region in the (p,q)-space $[0,1]^2$ where more cooperative strategies succeed. The evolution tends to the fixation of $ALLD$. If $w > \frac{P-S}{T-P}$, then there exist a region (in the south-east corner of the parameter square $[0,1]^2$) where cooperativity is favoured, while in the remaining zone it is discriminated against (see Fig.3). An evolutionary alternation of selection and small mutations tends to $ALLD$ if it starts in this zone of defection and to $p = 1, q = 1 - \frac{P-S}{T-P}\frac{1}{w}$ if it starts in the zone of cooperation. (Larger fluctuations, however, can lead the evolutionary path frome one zone to the other and hence complicate the outcome.) There is no tendency to approach TFT. On the other hand, a result by Axelrod implies in the present case that for $w > \frac{P-S}{T-P}$, the strategy TFT cannot be invaded by $ALLD$. This agrees well with our result that for such w, no strategy near TFT (large p, small q) can be invaded by a less cooperative strategy (with lower p or q value).

4.Oscillating behaviour

(A) Three strategies

The most interesting phenomenon, in the case of three strategies, is that of cyclic competition: strategy E_1 dominates E_2 , E_2 dominates E_3 and E_3 in its turn dominates E_1 . This occurs if the modified payoff matrix $(a_{ij} - a_{ii})$, whose diagonal is zero, has the sign structure

$$A' = \begin{pmatrix} 0 & + & - \\ - & 0 & + \\ + & - & 0 \end{pmatrix}.$$

This happens for a fairly substantial set of strategies in the (y, p, q) space. As example we mention (a) $E_1 = (0.40, 0.75, 0.75), E_2 = (0.40, 0.75, 0.25), E_3 = (0.40, 0.95, 0.25)$

(b) $E_1 = (0.75, 0.75, 0.75), E_2 = (0.75, 0.75, 0.25), E_3 = (0.54, 0.95, 0.30)$.

(We remark that in the first case all y values are the same.) The results from section 3(D) suggest that this can only happen if one of the strategies has large p- and small q-value , i.e. is a neighbour of TFT. This cyclic 'stone-scissors-paper'-structure determines the behaviour at the boundary of the state space S_3 . Its corners are saddles, and its edges saddle-connections (i.e. orbits having one corner as α-limit and another as ω-limit). The cyclic arrangement of these saddle connections forms a so-called 'heteroclinic cycle'. With respect to generic perturbations of a dynamical system, such a cycle is not structurally stable : within the class of game dynamical equations of type (3) however, it is stable in general.

The behaviour on the boundary does not specify the behaviour in the interior of the state space. There exists a unique interior fixed point $\hat{\underline{\mathbf{x}}}$ (the unique Nash solution of the game), but two generic cases can occur.

(i) $\hat{\underline{\mathbf{x}}}$ is globally stable. All orbits in $intS_3$ converge in an oscillatory manner towards $\hat{\underline{\mathbf{x}}}$. The system is permanent. This occurs if $detA' > 0$, as for example (a). (See Fig.4)

(ii) $\hat{\underline{\mathbf{x}}}$ is unstable. In this case all orbits in $intS_3$ (with the exception of the fixed point itself) converge to the boundary. More precisely, their ω-limit is the whole heteroclinic cycle. Hence the orbits follow the boundary , remaining for exponentially increasing times near a corner and switching after such a period of 'near-rest' suddenly to the next corner. Such a behaviour has been described (in other contexts) by several authors (for a survey see Hofbauer and Sigmund (1988)) . It seems particularly interesting that the time averages (4) do not converge in this case. Their accumulation points form a triangle contained in $intS_3$ and containing $\hat{\underline{\mathbf{x}}}$. Numerically the round-off error will wipe out one species , and one strategy reaches eventually fixation, but it is impossible to predict which. This type of behaviour occurs if $detA' < 0$ and example (b) is the case in point (Fig.5).

Of course it can also happen that $detA' = 0$. The point $\hat{\underline{\mathbf{x}}}$ is a center surrounded by closed orbits filling up $intS_n$. This case of 'neutral oscillations' is highly degenerate, of course.

Another interesting case, mentioned in the introduction, is that of the three strategies $E_1 = ALLD$, $E_2 = ALLC$ and $E_3 = (y, 1, 0)$ with $0 < y < 1$. This last strategy can be viewed a a sort of suspicious TFT, which starts with a random move. The uncertainty of TFT in the first move seems to be realistic in a biological context, because in contrast to $ALLD$ and $ALLC$, TFT is able both to cooperate and to defect. There exists one three-species equilibrium which can be shown to be a sink, the two-species equilibria between $(y, 1, 0)$ and $ALLC$ resp. $(y, 1, 0)$ and $ALLD$, one of which is stable and the other unstable in the two-strategy subsystem, and the three one-species equilibria of which one, namely *ALLD*, is a sink and hence evolutionarily stable. It follows that the three-species equilibrium cannot be an *ESS*. In fact, we have here Zeeman's (1980) example of an attractor which is not evolutionarily stable. It is surprising that *ALLC* and *ALLD* can coexist if there is a sufficient amount of suspicious *TFT* in the population (Fig. 6).

In the case $w = 1$ (i.e. no discount of the future) E_1, E_2 and E_3 as above form a 'stone-scissors-paper' cycle and the state space S_n is filled with periodic orbits around the stable (but no longer asymptotically stable) polymorphic equilibrium. If $w < 1$ and $y = 1$, then most initial conditions yield a mixture between *ALLC* and *TFT*, with *ALLD* eliminated. (Fig. 7)

(B) Four strategies

For $n = 4$, the behaviour of (3) is not yet fully classified, but the permanence criteria are fairly well understood (Kirlinger (1986), Hofbauer (1987)). Our feeling is that whatever can happen with (3) can be realized by a suitable *IPD* game dynamics. Numerically, we have found limit cycles, e.g. for

$E_1(0.75, 0.75, 0.75), E_2(0.75, 0.75, 0.24), E_3(0.40, 1.00, 0.30), E_4(0.70, 1.00, 0.00)$

(see Fig. 8).

What one can show analytically in this case is that the system is permanent (no strategy will get eliminated), and that the unique interior fixed point $\hat{\underline{\mathbf{x}}}$ is unstable. Thus the ω-limit of all interior orbits is disjoint from the boundary, and for almost all interior orbits it does not reduce to $\hat{\underline{\mathbf{x}}}$.

5. Discussion

There are several other dynamical approaches to the *IPD* in the literature. We mention in particular Feldman and Thomas (1987), where it is shown , using a discrete version of game dynamics, that if the probability w of continuing the game is part of the strategy and depends on the previous move, then a polymorphism of TFT and $ALLD$ can get established. Another investigation, by Blad (1986), studies game dynamics for a modified version of TFT: the players are not allowed any memory of earlier outcomes, but can use a third ('adaptive') strategy A besides C and D, which 'splits' the game into two subgames, playing C on the first game and whatever the other did on the second. The encounters are repeated infinitely often ($w = 1$), and the dynamics is given by (3), with $E_1 = C, E_2 = D, E_3 = A$. A small perturbation yields a structurally stable dynamics. It is shown that A is the unique 'good' locally stable Nash solution ($x_2 = 0$) of the stabilized game, and D the unique 'bad' one ($x_2 = 1$). This is related to (but different from) a general theory of Smale (1980) on dynamical systems associated with non-cooperative games, where strategies have a bounded memory, where the evolution is based on the players average accumulated payoff and where 'good' strategies lead to 'good' solutions (i.e. equilibria of the dynamics) which are stable.

Of special interest is the approach of Axelrod (1987), which reports the effect of a genetic algorithm of Holland-type upon an ensemble of strategies where each move is determined by the history of the last three interactions. It is shown that an evolutionary dynamics leads frequently to the establishment of strategies which are quite different from TFT, and that algorithms mimicking 'sexual' recombination are much faster than 'asexual' algorithms in promoting strategies doing considerably better than TFT against eight 'representative strategies' cuddled from Axelrods previous round-robin tournaments.

Our approach emphasizes the dynamical complexity and unpredictability for small numbers of competing strategies using very simple stochastic decision rules. It would seem interesting to extend this by taking into accound more diversified ensembles of strategies with a longer memory. It could well be that this 'smoothens' the dynamics.

As a first step, one can approach this question by numerical experiments, starting with a random distribution of strategies and introducing from time to time a mutant close to the prevailing ensemble. If one considers only (y, p, q)- strategies differing in a single parameter, the situation is reasonably clear :(i) variation of y leads, depending on p and q, either to a pure state with $y = 0$ or $y = 1$ or to a rich mixture of strategies with a predetermined average y-value ; (ii) variation of p leads to an extremal value 0 or 1, depending on y, q and possibly the initial state of the population ; (iii) variation of q leads to a monomorphic population with a predetermined q-value. If one admits variation in two or three parameters, the result is less predictable. It depends obviously on the initial conditions and the history of mutational events. In view of the preceding results, it seems highly unlikely that TFT is the evolutionary outcome. A solid statistical analysis is required to settle this question. It seems difficult to rely on intuition in this field, and our few experiments to date are far from conclusive.

References

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. (1987), The evolution of strategies in the Iterated Prisoner's Dilemma, in Davis,D.(ed), Genetic Algorithms and Simulated Annealing, Pitman.

Blad,M.C. (1986), A dynamic analysis of the repeated Prisoner's Dilemma Game, Int.J.Game Theory **15**, 83-99.

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. Theor. Biol. **128**, 297-315.

Hofbauer,J. (1987), Heteroclinic cycles on the simplex, Proc.Int.Conf. Nonlinear Oscillations, Janos Bolyai Math.Soc. Budapest.

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

Kirlinger,G. (1986), Permanence in Lotka-Volterra equations: linked predator-prey systems. Math. Biosci. **82**.165-191.

Lombardo,M.P. (1985), Mutual Restraint in Tree Swallows: A Test of the TIT FOR TAT Model of Reciprocity, Science **227**, 1363-1365.

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

May, R.M. and Leonard, W. (1975),Nonlinear aspects of competition between three species, SIAM J. Appl. Math. **29**, 243-252.

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.

Nowak, M. (1989), Stochastic Strategies in the Prisoner's Dilemma, preprint.

Schuster,P. and Sigmund,K. (1985), Towards a dynamics of social behaviour: strategic and genetic models for the evolution of animal conflicts, J.Soc.Biol. Structures **8**, 255-277.

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

Sigmund,K. (1987) Game dynamics, mixed strategies and gradient systems. Theor.Pop.Biol. **32**, 114-126.

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

Taylor, P. and Jonker, L. (1979),Evolutionarily stable strategies and game dynamics, Math. Biosc. **40**, 145-156.

Zeeman,E.C. (1980), Population dynamics from game theory, in: Global Theory of Dynamical Systems, Springer Lecture Notes in Mathematics **819**.

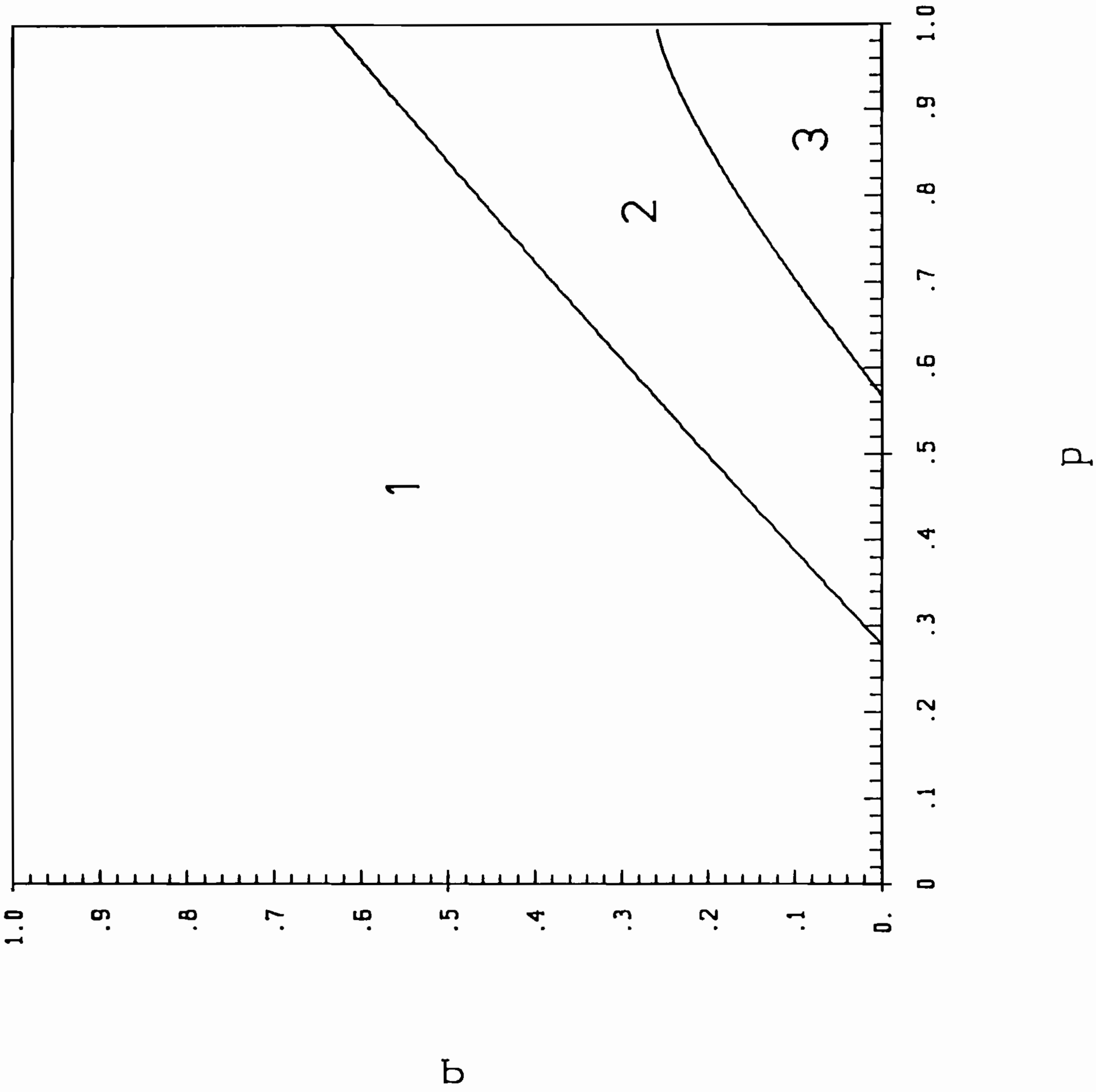

Fig.1: In region 1, $\hat{y} > 1$; in region 2, $\hat{y} \in (0,1)$ and in region 3, $\hat{y} < 0$.

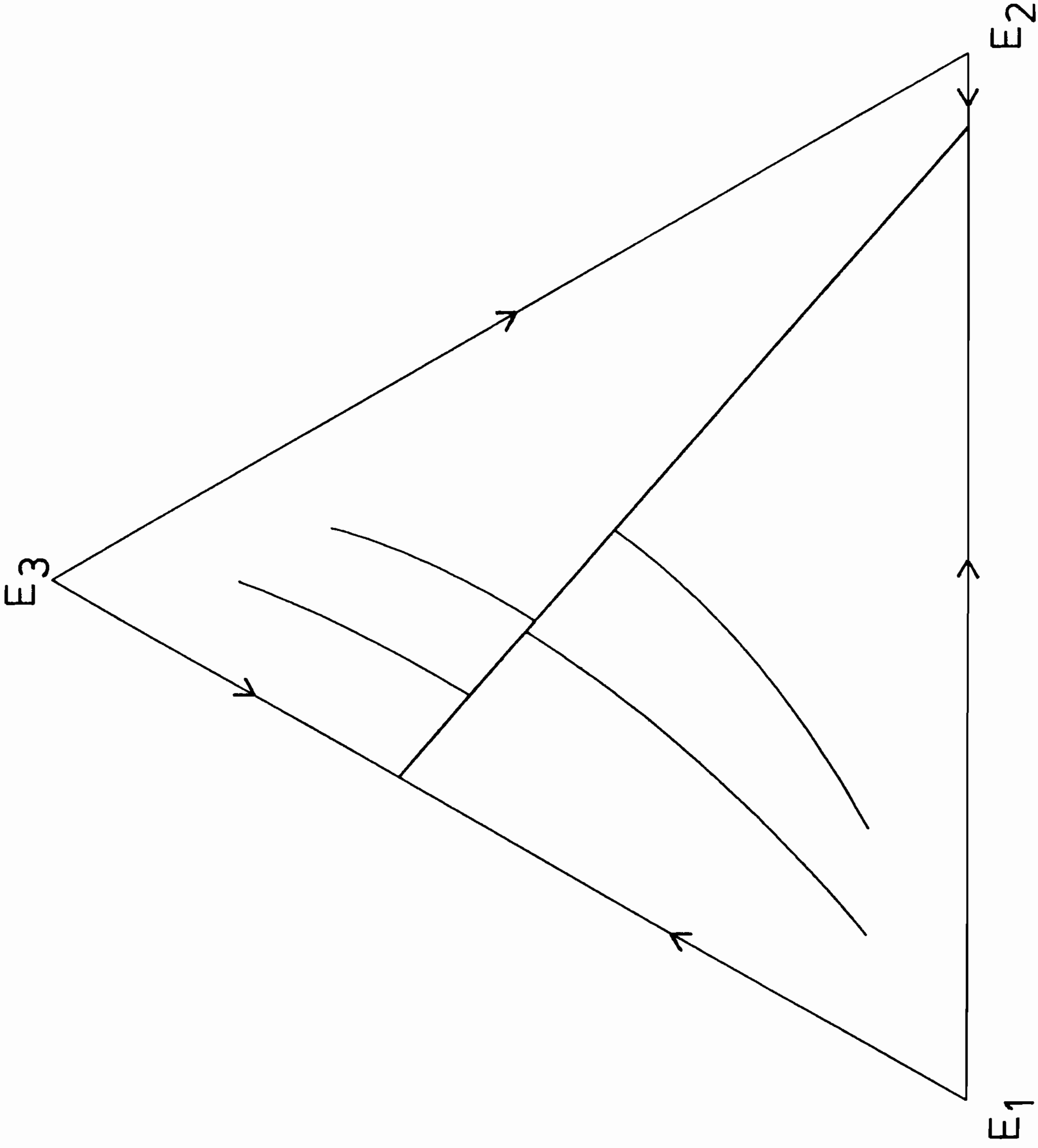

Fig.2: Phase portrait of (3) with $E_1 = (0.20, 0.75, 0.25)$, $E_2 = (0.60, 0.75, 0.25)$, $E_3 = (0.80, 0.75, 0.25)$. In this case $\hat{y} \simeq 0.43$.

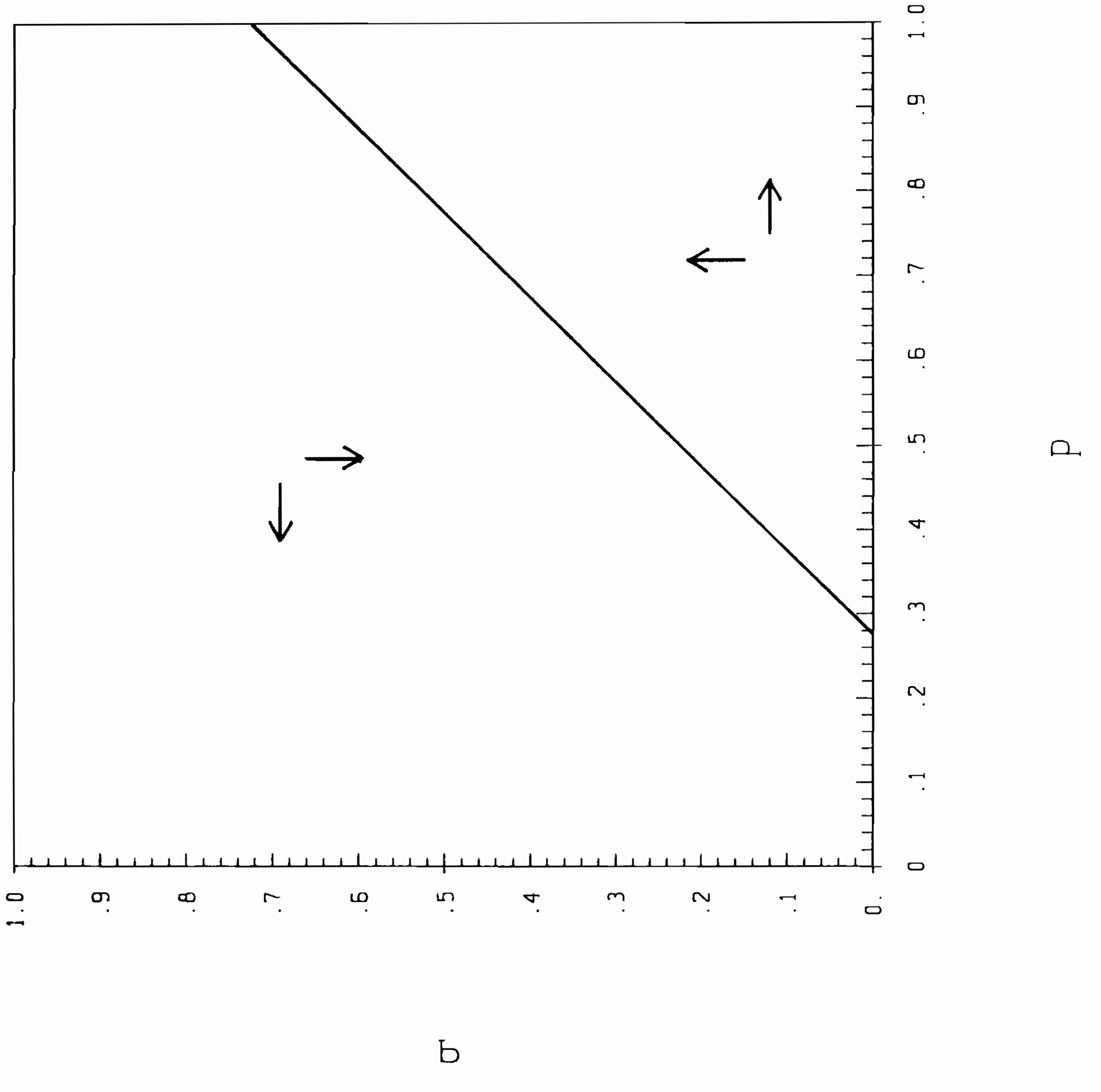

Fig.3: Cooperation increases in the corner defined by equation (6).

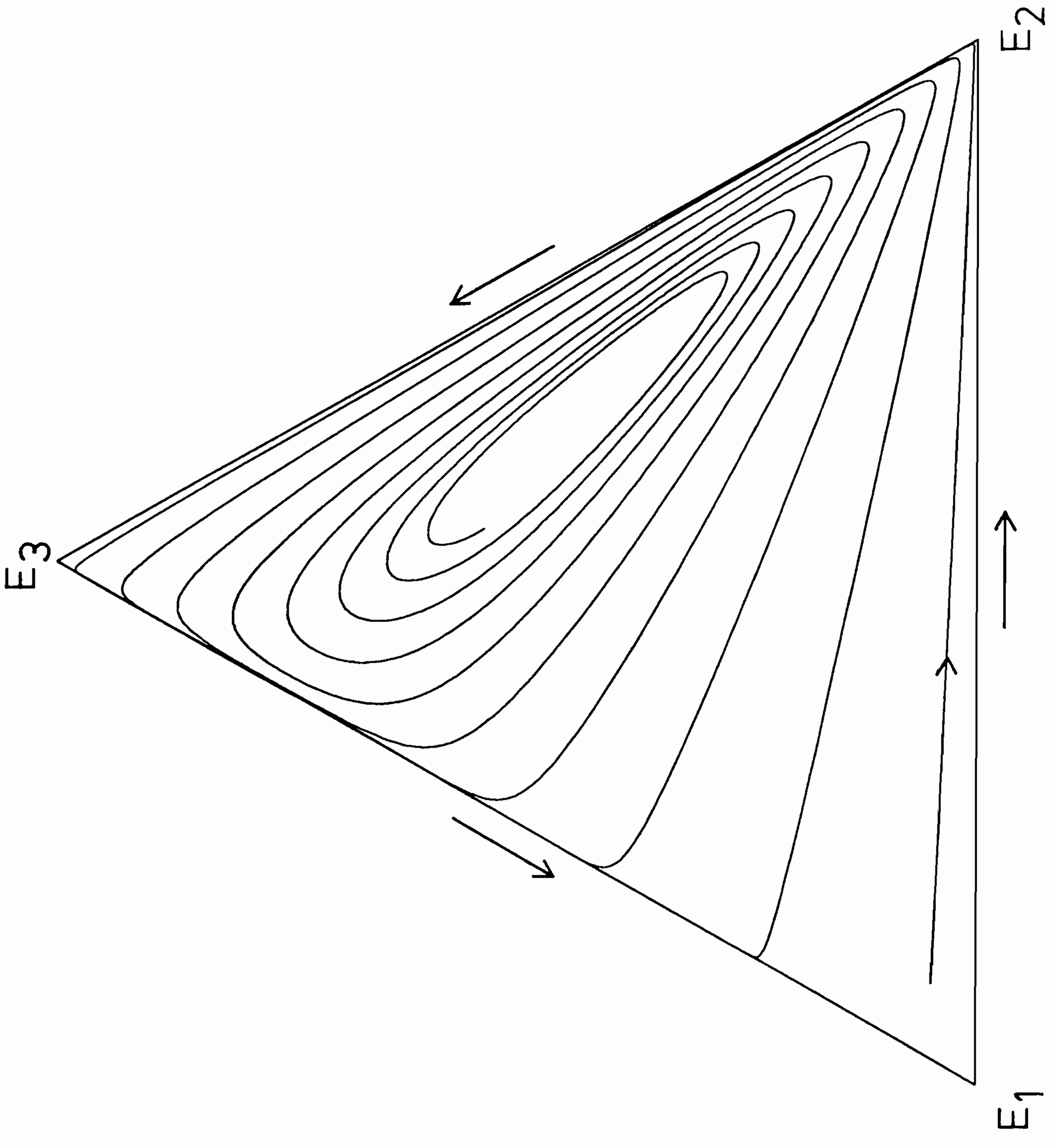

Fig.4: Phase portrait of (3) with $E_1 = (0.40, 0.75, 0.75)$, $E_2 = (0.40, 0.75, 0.25)$, $E_3 = (0.40, 0.95, 0.25)$. All interior orbits converge to $\underline{\hat{\mathbf{x}}}$. The boundary is a repellor.

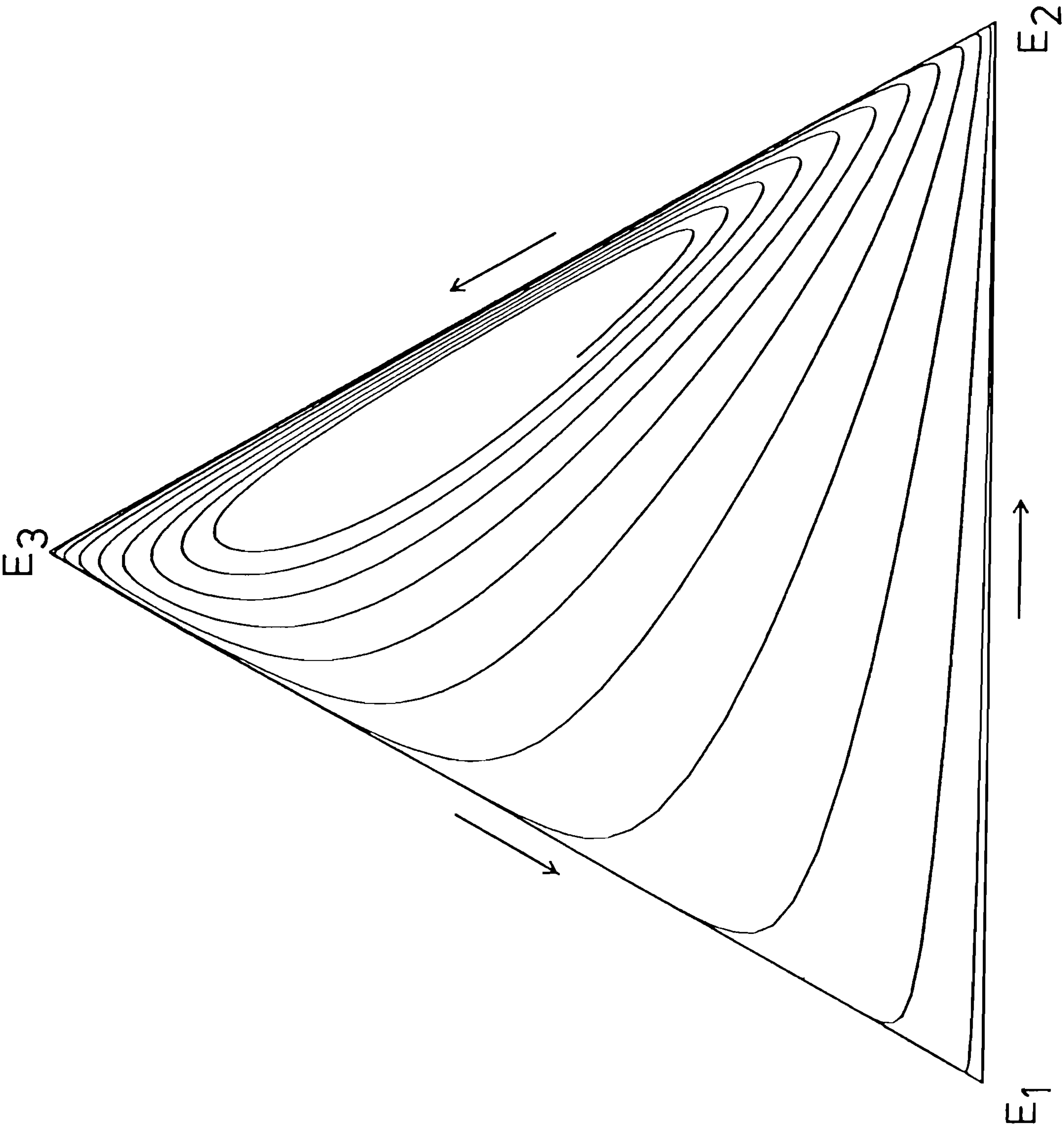

Fig.5: Phase portrait of (3) with $E_1 = (0.75, 0.75, 0.75)$, $E_2 = (0.75, 0.75, 0.25)$,$E_3 = (0.54, 0.95, 0.30)$. All interior orbits converge to the boundary.

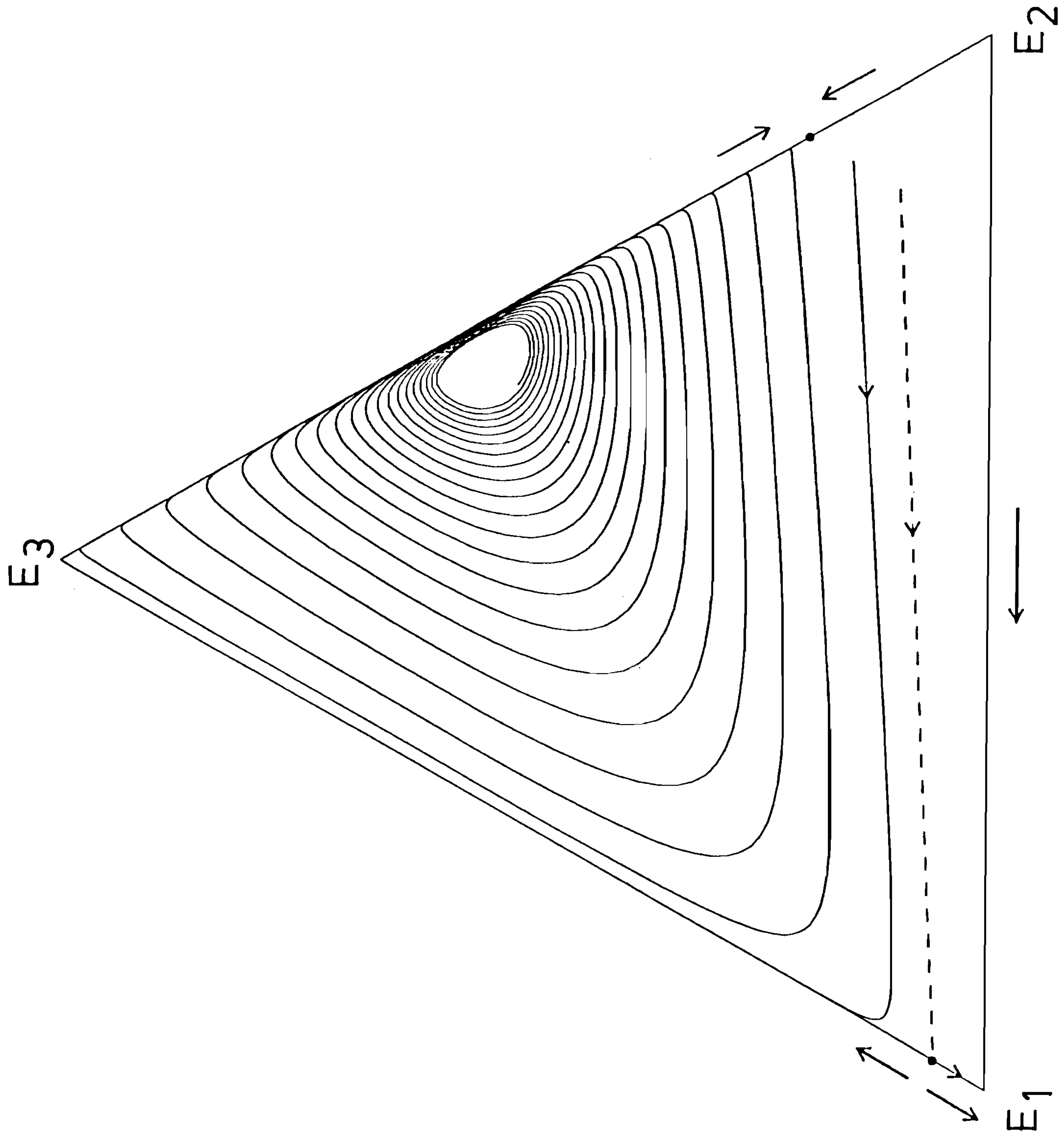

Fig.6: Phase portrait of (3) with $E_1 = (0,0,0)$, $E_2 = (1,1,1)$, $E_3 = (0.9,1,0)$. The interior is divided into two basins of attraction.

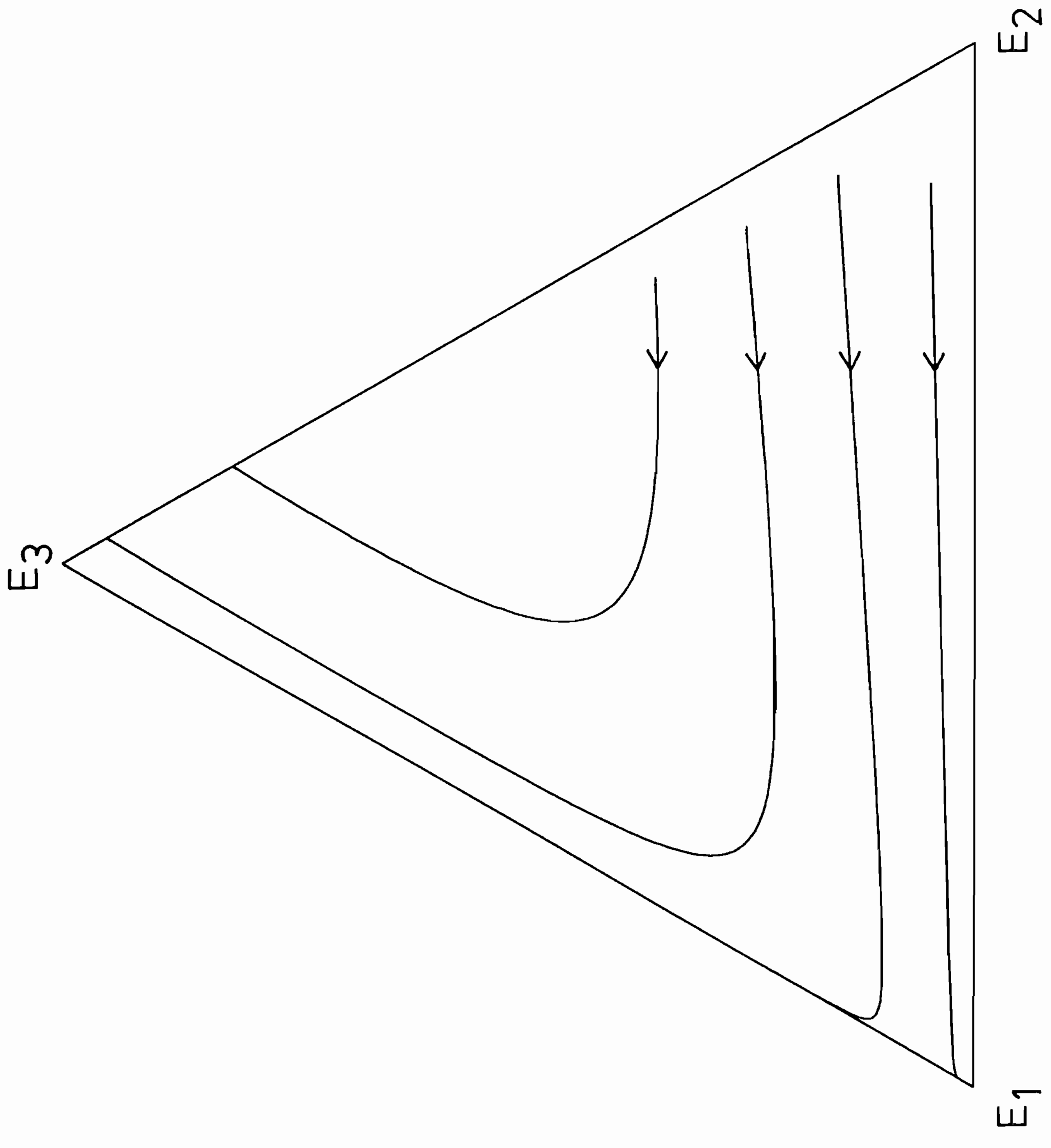

Fig.7: Phase portrait of (3) with $E_1 = (0,0,0)$, $E_2 = (1,1,1)$,$E_3 = (1,1,0)$. Some orbits converge to E_1, the others to the edge E_2E_3.

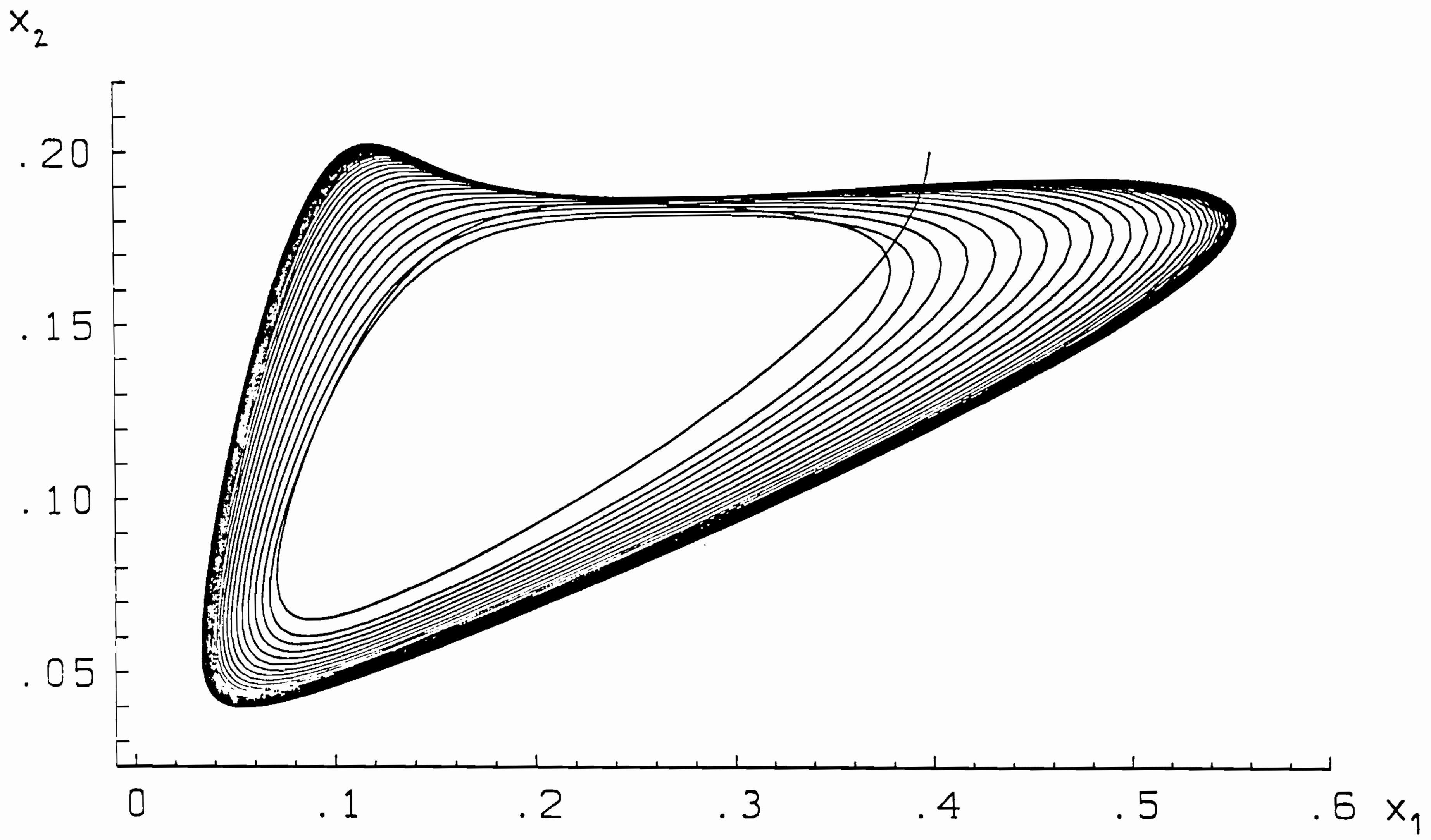

Fig.8: Projection on the x_1-x_2-plane of an orbit of (3) with $E_1 = (0.75, 0.75, 0.75)$,$E_2 = (0.75, 0.75, 0.25)$, $E_3 = (0.40, 1.00, 0.30)$,$E_4 = (0.70, 1.00, 0.00)$. The orbit converges to a limit cycle.