Evolution in Repeated Prisoner's Dilemma under Perturbed Best-Reply Dynamics - Marius-Ionut Ochea[1] - extended abstract

1 Introduction

Studying evolutionary dynamics on iterated prisoner's dilemma (IPD) demands a selection out of the vast set of the repeated game strategies. Brandt and Sigmund (2006) discuss one such particular ecology consisting of unconditional cooperators (AllC), unconditional defectors (AllD) and reactive players (TFT) and show that Replicator Dynamics exhibits, among others, a rock-scissors-paper pattern of cyclic behavior. However, such replicator cycles are not robust as small perturbations lead to extinction of all but one of the strategies. In this paper, we first aim to extend this strategy space with two repeated strategies that seem to receive less attention in the evolutionary IPD game literature: the error-proof, "generous" tit-for-tat which, with a certain probability, re-establishes cooperation after a (possibly by mistake) defection of the opponent and the penitent, "stimulus-response" (thus dubbed Pavlov) strategy that resets cooperation after the opponent punished for defection. Second, we contrast Replicator Dynamics behavior with a perturbed version of the Best-Reply dynamics, the Logit Dynamics, allowing for an imperfect switching towards a myopic best reply to the existing strategies distribution. A bifurcation analysis with respect to various model parameters is needed in order to reveal qualitative changes in the set of long-run (non)cooperative behaviors. Preliminary results show that the Logit Dynamics on the resulting 4x4 and 5x5 IPD games displays *stable co-existence* of repeated strategies but subjected to perpetual oscillations or even chaotic patterns in the distribution of IPD strategies.

2 An Evolutionary Iterated PD game

We consider a standard 2**x**2 Prisoner's Dilemma stage game where b stands for the benefits of cooperation and c for the costs associated with cooperative behavior:

$$\begin{bmatrix} C/D & C & D \\ C & b-c, b-c & -c, b \\ D & b, -c & 0, 0 \end{bmatrix} ; b, c > 0$$

At each time t the state of the play between two opponents is: $\Omega = \{CC, CD, DC, DD\}$.Agents interact repeatedly and the strategy set available for this iterated PD is limited to stochastic *memory-one* strategies (as in, for instance, Kraines&Kraines (2000)). Play starts with a first random move C or D and then proceeds with playing C with probability (r, s, t, p) conditional on

[1]PhD student, Center for Nonlinear Dynamics in Economics and Finance(CeNDEF), Dept. of Quantitative Economics, University of Amsterdam, the Netherlands; e-mail: m.i.ochea@uva.nl; phone: + 31 20 525 7356

realized state at time $t-1$ being CC, CD, DC, DD, respectively. Deterministic agents are particular limits in this stochastic strategy space: unconditional cooperators $AllC-(1,1,1,1)$, unconditional defectors $AllD-(0,0,0,0)$, conditional cooperators "Tit-for-Tatters" $TFT-(1,0,1,0)$, generous cooperators "Generous-Tit-for-Tat" $GTFT-(1,n,1,n)$,penitent or Pavlov players "Win-StayLoseShift" $WSLS-(1,0,0,1)$ and so on. Stochastic agents are $\varepsilon-$perturbations of the deterministic ones, where ε has the natural interpretation of mistakes or errors in implementation/execution of the deterministic strategies. Following Kraines&Kraines (2000) the iterated Prisoner's Dilemma game between two stochastic agents $S_1=(r,s,t,p)$ and $S_2=(x,y,z,w)$ leads to a Markov chain on states CC, CD, DC, DD with transition probabilities given by:

$$M=\begin{bmatrix} rx & sz & ty & pw \\ r(1-x) & s(1-z) & t(1-y) & p(1-w) \\ (1-r)x & (1-s)z & (1-t)y & (1-p)w \\ (1-r)(1-x) & (1-z)(1-s) & (1-t)(1-y) & (1-p)(1-w) \end{bmatrix}$$

One can show that, for strictly positive perturbation parameters, this process is ergodic, i.e. there is positive probability of escaping from any of the states in Ω.Therefore, it has a stationary invariant distribution given by the eigenvector of M corresponding to an eigenvalue 1 (Kemeny and Snell (1976)). The invariant distribution represents the average time the play between two stochastic agents spends in each state in Ω and, without discounting future payoffs, it enables computation of *average* expected payoff resulted from the interaction of two *repeated* game strategies. One such payoff matrix is constructed below for an ecology consisting in the stochastic versions of the following strategies: $\{AllD, TFT, GTFT, WSLS, AllC\}$

$$M=\begin{bmatrix} r/c & AllD & TFT & GTFT & WSLS & AllC \\ AllD & \varepsilon(b-c) & m_{12} & m_{13} & \frac{1}{2}b-c\varepsilon & b-b\varepsilon-c\varepsilon \\ TFT & m_{21} & \frac{1}{2}b-\frac{1}{2}c & m_{23} & \frac{1}{2}b-\frac{1}{2}c & m_{25} \\ GTFT & m_{31} & m_{32} & \frac{n}{n+\varepsilon}(b-c) & m_{34} & m_{35} \\ WSLS & b\varepsilon-\frac{1}{2}c & \frac{1}{2}b-\frac{1}{2}c & m_{43} & m_{44} & b-\frac{1}{2}c-b\varepsilon \\ AllC & b\varepsilon-c+c\varepsilon & m_{52} & m_{53} & m_{54} & (1-\varepsilon)(b-c) \end{bmatrix}$$

where $m_{ij}'s$ are complicated algebraic expressions of the stage game (b,c) errors (ε) and 'generosity' (n) parameters (see appendix).

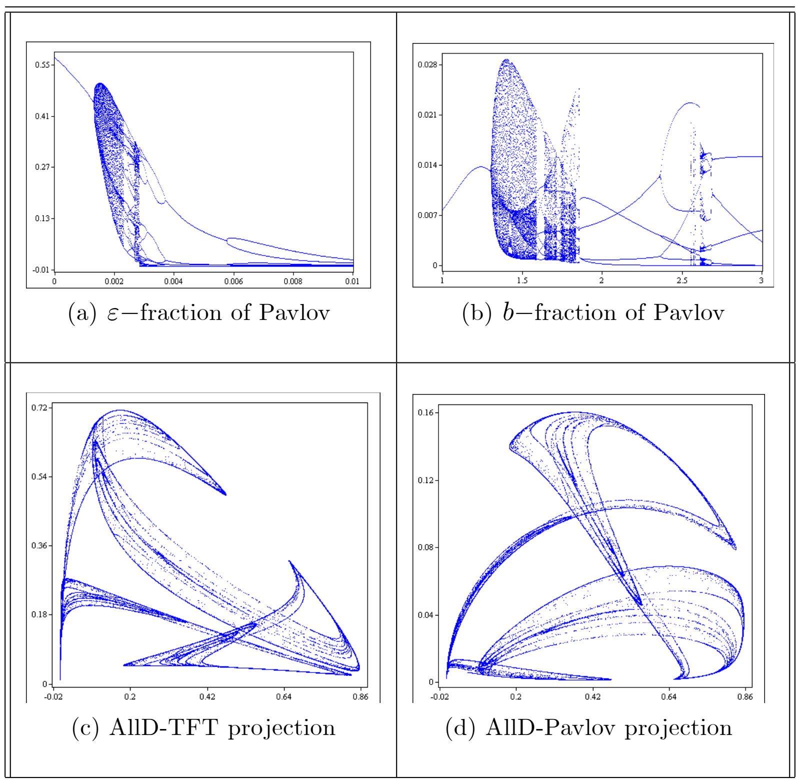
Players are allowed to switch their repeated game strategies based on realized, past *average* performance. Thus, at time t they will play the best-response to time $t-1$ distribution of (repeated) strategies with the logistic probability:

$$\dot{x}_{i,t+1}=\frac{e^{\beta(M\mathbf{x})_{i,t}}}{\sum\limits_{i=1}^{5} e^{\beta(M\mathbf{x})_{i,t}}}, \sum_{i=1}^{5} x_{i,t}=1 \tag{1}$$

with β denoting responsiveness to payoff differences between alternative strategies and $x_{i,t}'s$ the time t fractions of $ALLD, TFT, GTFT, WSLS, ALLC$ players, respectively.

3 Preliminary results

The evolution of discrete-time nonlinear system (1) can be first scrutinized by computer simulations. Figure 1 below displays evolution and bifurcation of Pavlov strategy fractions as different model parameters are varied. For instance, panel (a) shows a route by which a stable interior equilibrium loses stability via a (supercritical) Neimark-Sacker bifurcation. A stable invariant curve is born as the error parameter ε hits the threshold of 0.0018 and thus stable oscillations in the frequencies of all repeated strategies emerge. As we continue increasing the mistake probability a chaotic region is encountered (see panels (c), (d) for two dimensional projections of such a strange attractor for $\varepsilon = 0.004$). Increasing ε even further, but still below a 'reasonable' mistake probability of 0.01, the system sets in a two-cycle region created via a Period-Doubling bifurcation. A similar scenario is observed when the benefit of cooperation parameter is varied. Interestingly, low b/c ratios promotes stability of an interior equilibrium where the 5 repeated game strategies co-exist while increasing temptation to defect b generates perpetual cycles in the fraction of Pavlov strategists.



(a) $\varepsilon-$fraction of Pavlov

(b) $b-$fraction of Pavlov

(c) AllD-TFT projection

(d) AllD-Pavlov projection

Fig. 1. (a)-(b) bifurcations of Pavlov (WSLS) fractions with respect to the error in implementation ε and benefit of cooperation b parameters. Panels (c)-(d) display two-dimensional projections of a strange attractor for $\varepsilon = 0.004$. The remaining parameters are fixed to $n = 0.3, \beta = 14$.

The visual analysis of these bifurcation diagrams could be confirmed by a rigorous continuation procedure, i.e. the computation of curves of equilibria along with their detected codimension I bifurcations as one parameter of interest is varied. One such equilibrium is continued below in Figure 2, panel (a) as the benefit of cooperation b is increased. Two singularities are detected in the meaningful region $[b > c = 1]$ of b : a Neimark-Sacker point at $b \approx 3.25$ and a Period-Doubling at $b \approx 3.29$. These two points are next 'continued' with respect to another parameter as depicted in panels (b)-(d) with codimension II singularities[2] detected along them. Such analysis is instrumental to reveal that the behavior envisaged by simulations in Figure 1 is not a mere oddity but it is both *generic* in the space of repeated Prisoner's Dilemma games (as parametrized by the b/c ratio) and *robust* to the perturbation ε and generosity n parameters.

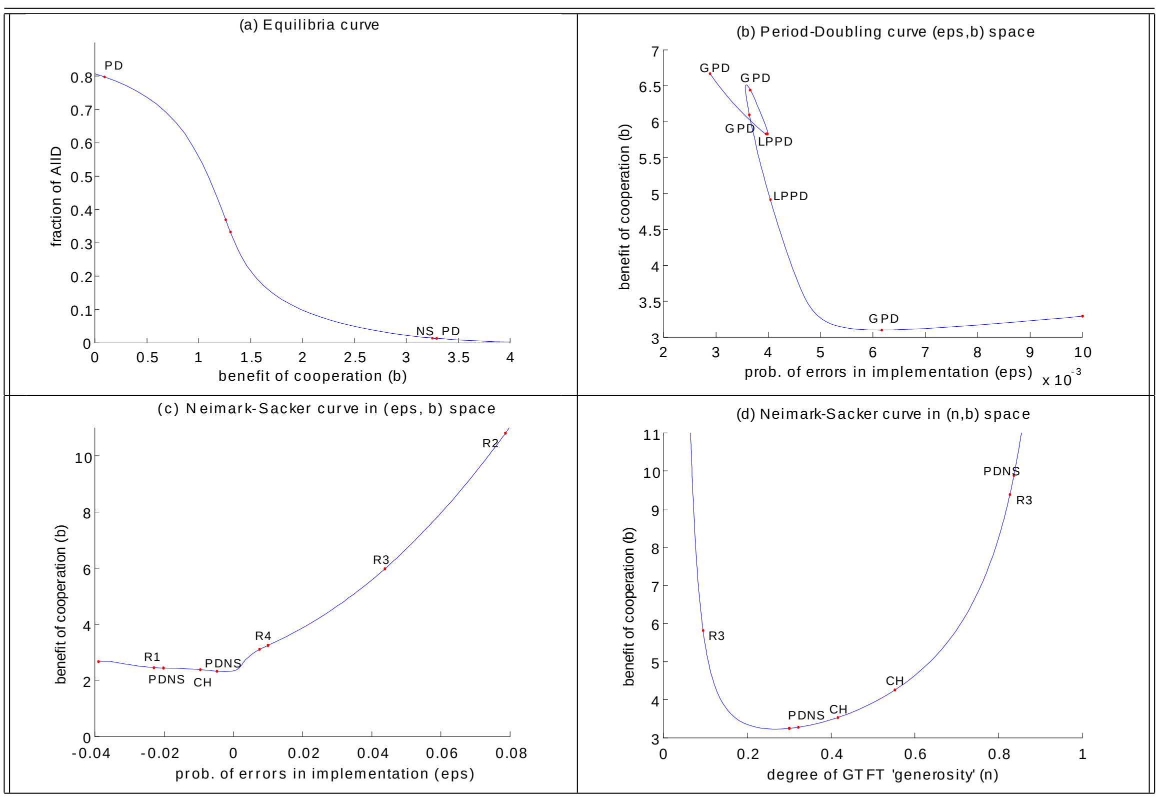


Fig. 2. Continuation of a fixed point (Panel (a)), a Period-Doubling point (Panel (b)) and a Neimark-Sacker point (Panels (c),(d))

[2] e.g. Panel (b) at a LPPD point a *fold* bifurcation of the two-cycle occurs when *two* additional *two-cycles* are born, one stable and one unstable. Panel (d) a Chenciner CH codim II bifurcation gives rise to much more complicated patterns of behavior (see Kuznetsov(1995)).

4 References

Kemeny J. G., J. L Snell (1975). *Denumerable Markov chains.* Springer-Verlag, New-York.

Kraines D., V. Kraines (2000). Natural selection of memory-one strategies for the iterated prisoner's dilemma, Journal of Theoretical Biology 203, 335-355.

Kraines D., V. Kraines (1995). Evolution of learning among Pavlov strategies, J. Conflict Resolution 39, 439-466.

Kuznetsov, Y. A., (1995). *Elements of applied bifurcation theory.* Applied Mathematical Sciences, vol. 112, Springer-Verlag, Berlin and New York, 1995.

Sigmund, K., H. Brandt (2006). The good, the bad and the discriminator - Errors in direct and indirect reciprocity, Journ. Theor. Biol. 239, 183-194

Sigmund K., M.A. Nowak (1989). Game dynamical aspects of the Prisoner's Dilemma, Applied Math. and Comp. 30, 191 - 213.

Sigmund K., M.A. Nowak (1993). Chaos and the evolution of cooperation, Proc. Nat. Acad. Science USA 90, 5091 - 5094.

Sigmund K., M.A. Nowak (1993). Win-stay, lose-shift outperforms tit-for-tat, Nature 364, 56 - 58.

5 Appendix

$m_{13} = b\varepsilon - c\varepsilon - b\varepsilon^2 + bn - bn\varepsilon; m_{23} = -\frac{c\varepsilon - b\varepsilon + b\varepsilon^2 - bn + cn + bn\varepsilon - 2cn\varepsilon}{n + 3\varepsilon - 2n\varepsilon - 2\varepsilon^2}$

$m_{25} = b - c - b\varepsilon + 2c\varepsilon - 2c\varepsilon^2; m_{31} = b\varepsilon - c\varepsilon + c\varepsilon^2 - cn + cn\varepsilon$

$m_{32} = \frac{b\varepsilon - c\varepsilon + c\varepsilon^2 + bn - cn - 2bn\varepsilon + cn\varepsilon}{n + 3\varepsilon - 2n\varepsilon - 2\varepsilon^2}; m_{35} = b - c - b\varepsilon + 2c\varepsilon - c\varepsilon^2 - cn\varepsilon$

$m_{34} = -m_{341}/m_{342}$

$m_{341} = 3b\varepsilon - 3c\varepsilon - bn^2 + cn^2 - 5b\varepsilon^2 + 2b\varepsilon^3 + 8c\varepsilon^2 - 7c\varepsilon^3 + 2c\varepsilon^4 + bn - cn - 4bn^2\varepsilon^2 +$
$+ 2cn^2\varepsilon^2 - 6bn\varepsilon + 3cn\varepsilon + 10bn\varepsilon^2 + 4bn^2\varepsilon - 4bn\varepsilon^3 - 6cn\varepsilon^2 - 3cn^2\varepsilon + 4cn\varepsilon^3$

$m_{342} = 4n^2\varepsilon^2 - 4n^2\varepsilon + n^2 + 8n\varepsilon^3 - 16n\varepsilon^2 + 8n\varepsilon - n + 4\varepsilon^4 - 12\varepsilon^3 + 15\varepsilon^2 - 7\varepsilon$

$m_{43} = m_{431}/m_{342}$

$m_{431} = 3c\varepsilon - 3b\varepsilon + bn^2 - cn^2 + 8b\varepsilon^2 - 7b\varepsilon^3 + 2b\varepsilon^4 - 5c\varepsilon^2 + 2c\varepsilon^3 - bn + cn + 2bn^2\varepsilon^2 -$
$- 4cn^2\varepsilon^2 + 3bn\varepsilon - 6cn\varepsilon - 6bn\varepsilon^2 - 3bn^2\varepsilon + 4bn\varepsilon^3 + 10cn\varepsilon^2 + 4cn^2\varepsilon - 4cn\varepsilon^3$

$m_{44} = (b - c)\left(1 - 4\varepsilon^3 + 6\varepsilon^2 - 3\varepsilon\right); m_{52} = b - c - 2b\varepsilon + c\varepsilon + 2b\varepsilon^2$

$m_{53} = b - c - 2b\varepsilon + c\varepsilon + b\varepsilon^2 + bn\varepsilon; m_{54} = b - c - 2b\varepsilon + c\varepsilon + 2b\varepsilon^2$