

INTERIM REPORT IR-97-073 / November

Equal Pay for all Prisoners / The Logic of Contrition

Maarten C. Boerlijst (mcb@einstein.zoo.ox.ac.uk)

Martin A. Nowak (novak@zoo.ox.ac.uk)

Karl Sigmund (ksigmund@esi.ac.at)

Approved by
Ulf Dieckmann (dieckman@iiasa.ac.at)
Project Coordinator, *Adaptive Dynamics Network*



The Adaptive Dynamics Network at IIASA fosters the development of new mathematical and conceptual techniques for understanding the evolution of complex adaptive systems.

Focusing on these long-term implications of adaptive processes in systems of limited growth, the Adaptive Dynamics Network brings together scientists and institutions from around the world with IIASA acting as the central node.

Scientific progress within the network is reported in the IIASA Studies in Adaptive Dynamics series.

THE ADAPTIVE DYNAMICS NETWORK

The pivotal role of evolutionary theory in life sciences derives from its capability to provide causal explanations for phenomena that are highly improbable in the physicochemical sense. Yet, until recently, many facts in biology could not be accounted for in the light of evolution. Just as physicists for a long time ignored the presence of chaos, these phenomena were basically not perceived by biologists.

Two examples illustrate this assertion. Although Darwin's publication of "The Origin of Species" sparked off the whole evolutionary revolution, oddly enough, the population genetic framework underlying the modern synthesis holds no clues to speciation events. A second illustration is the more recently appreciated issue of jump increases in biological complexity that result from the aggregation of individuals into mutualistic wholes.

These and many more problems possess a common source: the interactions of individuals are bound to change the environments these individuals live in. By closing the feedback loop in the evolutionary explanation, a new mathematical theory of the evolution of complex adaptive systems arises. It is this general theoretical option that lies at the core of the emerging field of adaptive dynamics. In consequence a major promise of adaptive dynamics studies is to elucidate the long-term effects of the interactions between ecological and evolutionary processes.

A commitment to interfacing the theory with empirical applications is necessary both for validation and for management problems. For example, empirical evidence indicates that to control pests and diseases or to achieve sustainable harvesting of renewable resources evolutionary deliberation is already crucial on the time scale of two decades.

The Adaptive Dynamics Network has as its primary objective the development of mathematical tools for the analysis of adaptive systems inside and outside the biological realm.

IIASA STUDIES IN ADAPTIVE DYNAMICS

- No. 1 Metz JAJ, Geritz SAH, Meszéna G, Jacobs FJA, van Heerwaarden JS:
Adaptive Dynamics: A Geometrical Study of the Consequences of Nearly Faithful Reproduction.
IIASA Working Paper WP-95-99.
van Strien SJ, Verduyn Lunel SM (eds.): *Stochastic and Spatial Structures of Dynamical Systems*, KNAW Verhandelingen, North Holland, Amsterdam, pp. 183-231 (1996).
- No. 2 Dieckmann U, Law R:
The Dynamical Theory of Coevolution: A Derivation from Stochastic Ecological Processes.
IIASA Working Paper WP-96-01.
Journal of Mathematical Biology (1996) 34, 579–612.
- No. 3 Dieckmann U, Marrow P, Law R:
Evolutionary Cycling of Predator-Prey Interactions: Population Dynamics and the Red Queen.
Journal of Theoretical Biology (1995) 176, 91–102.
- No. 4 Marrow P, Dieckmann U, Law R:
Evolutionary Dynamics of Predator-Prey Systems: An Ecological Perspective.
IIASA Working Paper WP-96-02.
Journal of Mathematical Biology (1996) 34, 556–578.
- No. 5 Law R, Marrow P, Dieckmann U:
On Evolution under Asymmetric Competition.
IIASA Working Paper WP-96-03.
Evolutionary Ecology (1997) 11, 485–501.
- No. 6 Metz JAJ, Mylius SD, Dieckmann O:
When Does Evolution Optimise? On the Relation between Types of Density Dependence and Evolutionarily Stable Life History Parameters.
IIASA Working Paper WP-96-04.
- No. 7 Ferrière R, Gatto M:
Lyapunov Exponents and the Mathematics of Invasion in Oscillatory or Chaotic Populations.
Theoretical Population Biology (1995) 48, 126–171.
- No. 8 Ferrière R, Fox GA:
Chaos and Evolution.
Trends in Ecology and Evolution (1995) 10, 480–485.
- No. 9 Ferrière R, Michod RE:
The Evolution of Cooperation in Spatially Heterogeneous Populations.
IIASA Working Paper WP-96-29.
American Naturalist (1996) 147, 692–717.

- No. 10 Van Dooren TJM, Metz JAJ:
Delayed Maturation in Temporally Structured Populations with Non-Equilibrium Dynamics.
IIASA Working Paper WP-96-70.
Journal of Evolutionary Biology (1997) in press.
- No. 11 Geritz SAH, Metz JAJ, Kisdi E, Meszéna G:
The Dynamics of Adaptation and Evolutionary Branching.
IIASA Working Paper WP-96-77.
Physical Review Letters (1997) 78, 2024–2027.
- No. 12 Geritz SAH, Kisdi E, Meszéna G, Metz JAJ:
Evolutionarily Singular Strategies and the Adaptive Growth and Branching of the Evolutionary Tree.
IIASA Working Paper WP-96-114.
Evolutionary Ecology (1997) in press.
- No. 13 Heino M, Metz JAJ, Kaitala V:
Evolution of Mixed Maturation Strategies in Semelparous Life-Histories: the Crucial Role of Dimensionality of Feedback Environment.
IIASA Working Paper WP-96-126.
Philosophical Transactions of the Royal Society of London Series B (1997) in press.
- No. 14 Dieckmann U:
Can Adaptive Dynamics Invade?
IIASA Working Paper WP-96-152.
Trends in Ecology and Evolution (1997) 12, 128–131.
- No. 15 Meszéna G, Czibula I, Geritz SAH:
Adaptive Dynamics in a Two-Patch Environment: a Simple Model for Allopatric and Parapatric Speciation.
IIASA Interim Report IR-97-01.
Journal of Biological Systems (1997) in press.
- No. 16 Heino M, Metz JAJ, Kaitala V:
The Enigma of Frequency-Dependent Selection.
IIASA Interim Report IR-97-61.
- No. 17 Heino M:
Management of Evolving Fish Stocks.
IIASA Interim Report IR-97-62.
- No. 18 Heino M:
Evolution of Mixed Reproductive Strategies in Simple Life-History Models.
IIASA Interim Report IR-97-63.
- No. 19 Geritz SAH, van der Meijden E, Metz JAJ:
Evolutionary Dynamics of Seed Size and Seedling Competitive Ability.
IIASA Interim Report IR-97-71.
- No. 20 Galis F, Metz JAJ:
Why are there so many Cichlid Species? On the Interplay of Speciation and Adaptive Radiation.
IIASA Interim Report IR-97-72.
Trends in Ecology and Evolution (1997) in press.

- No. 21 Boerlijst MC, Nowak MA, Sigmund K:
Equal Pay for all Prisoners. / The Logic of Contrition.
IIASA Interim Report IR-97-73.
- No. 22 Law R, Dieckmann U:
Symbiosis without Mutualism and the Merger of Lineages in Evolution.
IIASA Interim Report IR-97-74.
- No. 23 Klinkhamer PGL, de Jong TJ, Metz JAJ:
Sex and Size in Cosexual Plants.
IIASA Interim Report IR-97-78.
Trends in Ecology and Evolution (1997) 12, 260–265.

Issues of the IIASA Studies in Adaptive Dynamics series can be obtained free of charge. Please contact:

Adaptive Dynamics Network
International Institute for Applied Systems Analysis
Schloßplatz 1
A-2361 Laxenburg
Austria

Telephone +43 2236 807, Telefax +43 2236 71313, E-Mail adn@iiasa.ac.at,
Internet <http://www.iiasa.ac.at/Research/ADN>

Contents

Equal Pay for all Prisoners	1
The Logic of Contrition	4
1 Introduction	4
2 Preliminaries on the Repeated Prisoner’s Dilemma	5
3 In Search of Stability	6
4 Stochastic Strategies with Standing	10
5 Numerical Simulations	13
6 The Alternating PD	16
7 Discussion	17

Abstract

This report deals with two questions concerning the emergence of cooperative strategies in repeated games. The first part is concerned with the Perfect Folk Theorem and presents a vast class of equilibrium solutions based on Markovian strategies. Simple strategies, called equalizers, are introduced and discussed: if players adopt such strategies, the same payoff results for every opponent. The second part analyzes strategies implemented by finite automata. Such strategies are relevant in an evolutionary context; an important instance is called Contrite Tit For Tat. In populations of players adopting such strategies, Contrite Tit For Tat survives very well—at least as long as errors are restricted to mistakes in implementation ('the trembling hand'). However, this cooperative strategy cannot persist if mistakes in perception are included as well.

About the Authors

Maarten Boerlijst
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3PS, UK

Martin Nowak
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3PS, UK

Karl Sigmund
Institut für Mathematik
Universität Wien
Strudlhofgasse 4
A-1090 Vienna, Austria
and
Adaptive Dynamics Network
International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

Acknowledgment

Financial support from the Wellcome Trust (MAN) and the Austrian Forschungsförderungsfonds (KS) is gratefully acknowledged.

Equal Pay for all Prisoners

Maarten C. Boerlijst

Martin A. Nowak

Karl Sigmund

By prisoners we mean, of course, players of the well-known Prisoner's Dilemma game (to be described presently). We shall show that there exist simple strategies for the infinitely iterated Prisoner's Dilemma that act as *equalizers* in the sense that all co-players receive the same payoff, no matter what their strategies are like.

The Prisoner's Dilemma game, a favorite with game theorists, social scientists, philosophers, and evolutionary biologists, displays the vulnerability of cooperation in a minimalistic model (see [1] to [5]). The two players engaged in this game can choose whether to cooperate or to defect. If both defect, they gain 1 point each; if both cooperate, they gain 3 points; but if one player defects and the other does not, then the defector receives 5 points and the other player only 0. The right move is obviously to defect, no matter what the other player does. As a result, both players earn 1 point instead of 3.

But if the same two players repeat the game very frequently, there exists no strategy that is best against all comers. The diversity of strategies is staggering. If we simulate on a computer populations of strategies evolving under a mutation-selection regime (with mutation introducing new strategies and selection weening out those with lowest payoff), we observe a rich variety of evolutionary histories frequently leading to cooperative regimes dominated by strategies like Pavlov (cooperate whenever the opponent's move, in the previous round, matched yours) or Generous Tit For Tat (always reciprocate your opponent's cooperative move, but reciprocate only two-thirds of the defections). Remarkably, all strategies of the iterated Prisoner's Dilemma, which can be very complex and make up a huge set, obtain the same payoff against some rather simple equalizer strategies.

More generally, let us consider a two-player game where both players have the same two strategies and the same payoff matrix. We denote the first strategy (row 1) by **C** (for 'cooperate') and the second (row 2) by **D** (for 'defect') and write the payoff matrix as

$$\begin{array}{c|cc}
 & \text{Opponent} & \\
 & \mathbf{C} & \mathbf{D} \\
 \hline
 \text{You} & \mathbf{C} & R, R \quad S, T \\
 & \mathbf{D} & T, S \quad P, P
 \end{array} \tag{1}$$

Such games include the Prisoner's Dilemma, where $T > R > P > S$, and the Chicken game, where $T > R > S > P$. (In the Prisoner's Dilemma case, R stands for the *reward* for mutual cooperation, P is the *penalty* for mutual defection, T is the *temptation* payoff for unilaterally defecting and S the *sucker* payoff for being exploited.)

Let us assume that the game is repeated infinitely often. A *strategy* in such a supergame is a program telling the player in each round whether to play **C** or **D**. The program may be history-dependent and stochastic: it specifies at every step the probability for playing

C, depending on what happened so far. If A_n is the payoff in the n -th round, the expected long-run average payoff for a player is given by

$$\lim_{N \rightarrow \infty} \frac{A_1 + \cdots + A_N}{N} \quad (2)$$

provided it exists. It need not always exist: think of two players cooperating in the first 10 rounds, defecting in the next 100 rounds, then cooperating in the following 1000 rounds, etc.

Memory-one strategies are particularly simple. Such a strategy is given by the probability to play **C** in the first round, and a quadruple $\mathbf{p} = (p_R, p_S, p_T, p_P)$, where p_i denotes the probability that the player plays **C** after having experienced outcome $i \in \{R, S, T, P\}$ in the previous round. Some of the most successful strategies belong to this class, including Generous Tit For Tat $(1, 1/3, 1, 1/3)$ and Pavlov $(1, 0, 0, 1)$.

Theorem: *If $\max(S, P) < \min(R, T)$, then there exist, for every value π between these numbers, memory-one strategies \mathbf{p} such that every opponent obtains the long-run average payoff π against a player using such a strategy. The vector \mathbf{p} is given by*

$$(1 - (R - \pi)a, 1 - (T - \pi)a, (\pi - S)a, (\pi - P)a) \quad (3)$$

where a is any real number such that $1/a \geq \max(T - \pi, R - \pi, \pi - S, \pi - P)$.

Proof: The condition on a guarantees that the p_i are probabilities. Let us denote by $q_i(n)$ the conditional probability that the opponent plays **C** in the following round, given that the n -th round resulted in outcome i , and by $s_i(n)$ the probability that the outcome in the n -th round is i . By conditioning on round n , we obtain:

$$s_R(n+1) = s_R(n)q_R(n)[1 - (R - \pi)a] + s_S(n)q_S(n)[1 - (T - \pi)a] + s_T(n)q_T(n)(\pi - S)a + s_P(n)q_P(n)(\pi - P)a. \quad (4)$$

Similarly,

$$s_S(n+1) = s_R(n)(1 - q_R(n))[1 - (R - \pi)a] + s_S(n)(1 - q_S(n))[1 - (T - \pi)a] + s_T(n)(1 - q_T(n))(\pi - S)a + s_P(n)(1 - q_P(n))(\pi - P)a \quad (5)$$

Summing (4) and (5) yields the probability that you play **C** in round $n + 1$

$$s_R(n+1) + s_S(n+1) = s_R(n)[1 - (R - \pi)a] + s_S(n)[1 - (T - \pi)a] + s_T(n)(\pi - S)a + s_P(n)(\pi - P)a.$$

Hence

$$a^{-1}[s_R(n) + s_S(n) - s_R(n+1) - s_S(n+1)] = Rs_R(n) + Ss_T(n) + Ts_S(n) + Ps_P(n) - \pi[s_R(n) + s_S(n) + s_T(n) + s_P(n)]. \quad (6)$$

Since the $s_i(n)$ sum up to 1, the right-hand side is just $A_n - \pi$, where A_n is the opponent's payoff in the n -th round (we must bear in mind that one player's outcome S is the other player's outcome T). Summing up (6) for $n = 1, \dots, N$ and dividing by N , we obtain

$$\frac{1}{aN}[s_R(1) + s_S(1) - s_R(N+1) - s_S(N+1)] = \frac{A_1 + \cdots + A_N}{N} - \pi,$$

and hence

$$\lim_{N \rightarrow \infty} \frac{A_1 + \cdots + A_N}{N} = \pi.$$

A few final remarks. Two players using equalizer strategies are in *Nash equilibrium*, which means that neither has an incentive to change strategy. Nash equilibria exist for every game; for iterated games, they abound. Indeed, the so-called Folk Theorem in game theory states that every feasible pair of payoff-values exceeding the *minimax* (the highest payoff that a player can enforce, which in our case is $\max(S, P)$) can be realized by a Nash-equilibrium pair [2, p. 373]. Our theorem is related to this: the strategies are equalizers with memory one. Two players using such strategies have no reason to switch unilaterally to another strategy, since they cannot improve their payoff; however, they have no reason *not* to adopt another strategy either, since they will not be penalised. Since their opponent plays an equalizer strategy, they can switch to *any* other strategy, and not be worse off. If *both* players opt for a change, however, they are likely to end up in a non-equilibrium situation.

If a is chosen small enough, the runs of consecutive defections or cooperations can be made arbitrarily long. The condition $\min(R, T) > \max(S, P)$ and its converse are not only sufficient, but also necessary for the existence of such equalizer strategies. It is easy to construct other equalizer strategies. For example, play **C** until the opponent's mean payoff is larger than π , then play **D** until it is smaller than π , then play **C** until it is larger again, etc. However, such a strategy requires monitoring the opponent's entire payoff sequence. The point is that even within memory-one strategies, equalizers exist.

References

- Axelrod, R., *The Evolution of Cooperation*, Penguin, Harmondsworth, 1990.
- Binmore, K.G., *Fun and Games: a Text on Game Theory*, Heath and Co, Lexington, Massachusetts, 1992.
- Mesterton-Gibbons, M., *An Introduction to Game-theoretic Modelling*, Addison-Wesley, Redwood City, California, 1992.
- Nowak, M.A., R.M. May, and K. Sigmund, The Arithmetics of Mutual Aid, *Scientific American* **272** (1995). 76–81.
- Sigmund, K., *Games of Life: Explorations in Ecology, Evolution and Behaviour*, Penguin, Harmondsworth, 1995.

The Logic of Contrition

Maarten C. Boerlijst

Martin A. Nowak

Karl Sigmund

Abstract

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

1 Introduction

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

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

defection leads to a bad standing. The strategy *cTFT* begins with a cooperative move, and cooperates except if provoked.

If two *cTFT*-players engage in a repeated Prisoner’s Dilemma, and if the first player defects by mistake, he loses his good standing. In the next round, he will cooperate, whereas the other *cTFT*-player will defect without losing his good standing. Then both players will be in good standing and resume their mutual cooperation in the following round. This strategy is related to Dawkins’ (1989) Remorseful Prober, who defects once in a while but accepts retaliation in the following round without complaint.

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

On the other hand, the additional complexity of the *cTFT* strategy has its drawbacks. In particular, while *cTFT* is immune to errors in the implementation of a move, it is not immune to errors in the perception of a move. If, in a match between two *cTFT* players, one player mistakenly believes that the other is in bad standing, this leads to a sequence of mutual backbiting, just as with *TFT*. (Errors in perception – rather than implementation – have been studied in Miller (1989), Kollock (1993), Nowak et al (1995b).)

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

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

2 Preliminaries on the Repeated Prisoner’s Dilemma

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

$$T > R > P > S \tag{1}$$

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

In the iterated PD, the game is played for several rounds. We shall assume that there is a constant probability *w* for another round. The length of the game is a stochastic variable with mean value $\frac{1}{1-w}$. A strategy for the iterated PD is a program telling the player in

each round whether to chose C or D (this can be a randomised decision: cooperate with such and such a probability). If A_n is the payoff for one player in the n -th round, his expected payoff is $\sum A_n w^n$ (note that w^n is the probability that an n -th round occurs). We shall mostly be interested in large w (close to 1). Frequently, the limiting case $w = 1$ is considered (the infinitely repeated PD). In this case, the payoff is the limit of the mean $\frac{1}{n}(A_1 + \dots + A_n)$, for $n \rightarrow \infty$ (if it exists). We shall assume

$$2R > T + S \tag{2}$$

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

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

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

In a given round, a player can be in three possible states: Cg , Dg and Db : the first means that he has cooperated (which automatically entails good standing), the second that he has defected with good reason, the third that he has wantonly defected. The *state* of the game in a given round is made up of the states of the first and the second player. There are nine such combinations: (Cg, Cg) , (Cg, Dg) , (Cg, Db) , (Dg, Cg) , (Dg, Db) , (Db, Cg) , (Db, Dg) , (Db, Db) and (Dg, Dg) . It is easy to check that this last state can never be reached: we therefore omit it, and number the remaining eight states in this order.

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

3 In Search of Stability

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

It is easy to see that for the infinitely repeated Prisoner's Dilemma, i.e. for $w = 1$, there exists no ESS.

This is due simply to the fact that two strategies differing only in their first – say – three hundred moves will have exactly the same payoff.

But as shown in Sugden (1986), for $w < 1$ the strategy $cTFT$ is evolutionarily stable in a very important sense: if there is a small, but non-vanishing probability of misimplementing a move, every strategy that deviates, against a $cTFT$ -player, from what the $cTFT$ -rule would prescribe, fares less well than it would have by following this rule. Note that if there is such an error probability, every finite sequence of moves will have a positive

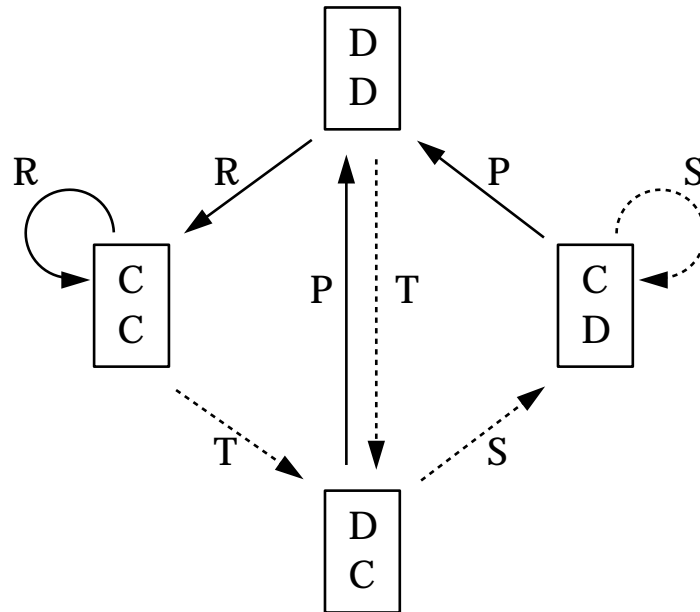


Figure 1: *PAVLOV* is an ESS if $T + wP < R + wR$. Solid lines indicate the moves specified by the *PAVLOV* strategy; dotted lines indicate the alternative moves. See text for further explanation.

probability. (See Selten (1975), Selten and Hammerstein (1984), and Boyd (1989) where the connection with Selten's concept of a perfect equilibrium is discussed.)

The basic idea of Sugden's proof allows to decide for *every* deterministic rule \hat{S} based on finitely many states whether it is evolutionarily stable in the sense defined above. Because of the error probability, every state can be reached with positive probability. Let us start in any of the possible states, assuming for the moment that no error will occur in the following rounds, and let us follow the fate of a player invading a \hat{S} population.

Since the next move of his adversary is always specified, there are only two possible states that can be reached in the next round, depending on whether the invader uses *C* or *D*. From each of these states, two states can be reached in turn, etc. Since there are only finitely many states, each branch of the game-tree must eventually return to a state it had visited before. Therefore, it is possible to compute the payoff along every branch, discounting by the factor w at every step.

One of the two branches issued from each state describes what happens if I use \hat{S} myself. If this always yields the highest payoff, and no alternative does, then \hat{S} is evolutionarily stable, provided the probability for mistakes in implementation is sufficiently small.

In Fig. 1 we check this for *PAVLOV*. Two arrows issue forth from each state, depending on whether the invader plays *C* or *D* against his \hat{S} -adversary. The vertices of the graphs describe the invader's state in the first (or upper) position, and the state of his opponent in the second (or lower) position. The arrows describe the possible transitions, which only depend of my choice, since the opponent's moves are specified by \hat{S} . The solid arrow indicates the move the invader would choose if he were also a \hat{S} -strategist. We see in Fig.1 that *PAVLOV* is an ESS if and only if $T + wP < R + wR$, as has been shown by Harrington and Axelrod (1995). The critical decision occurs when we are in (D, D) or (C, C) and I have to decide whether to get two *R*'s in succession, or a *T* followed by a *P*. On the other hand, Fig. 2 shows that *TFT* is never an ESS: in (C, D) , my best move leads to (C, C) .

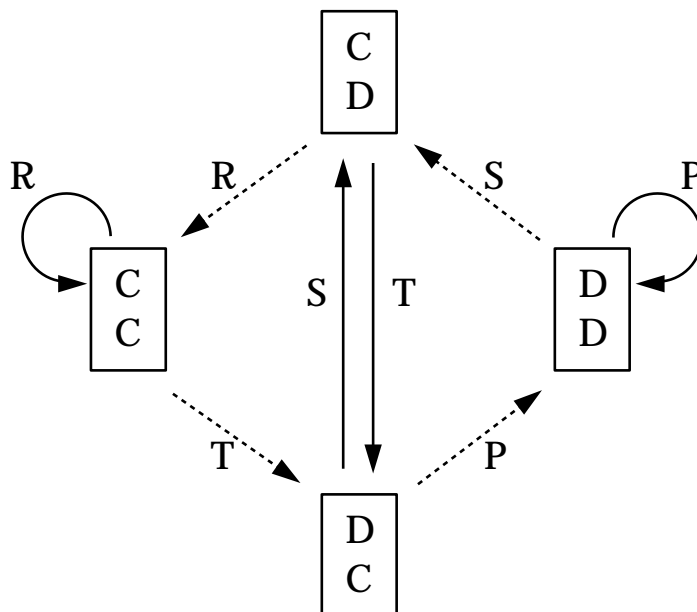


Figure 2: *TFT* is no ESS.

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

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

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

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

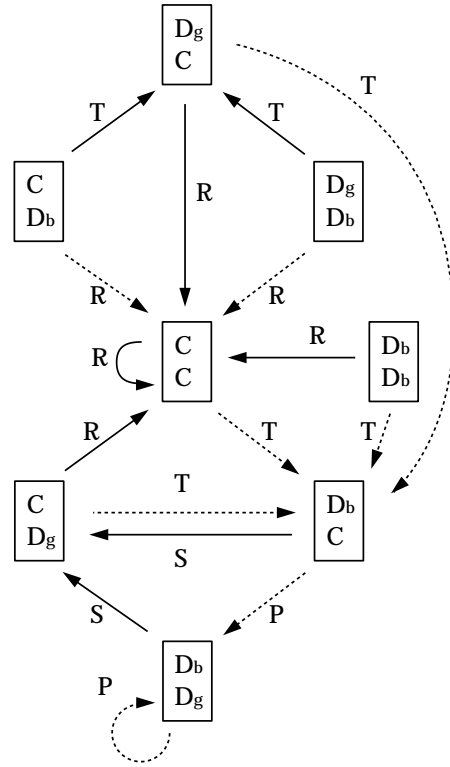


Figure 3: *cTFT* is an ESS.

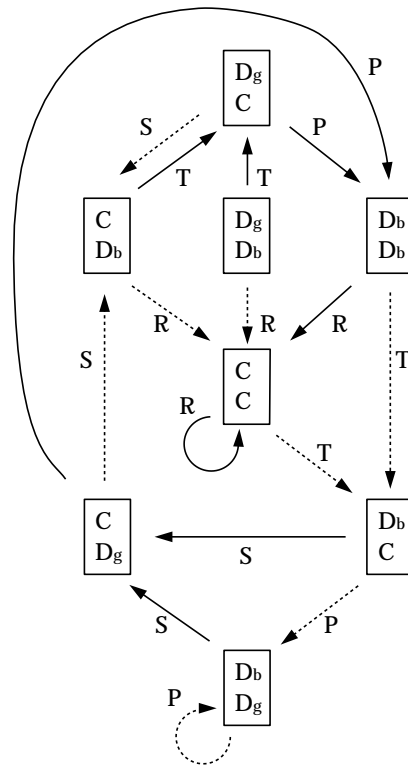


Figure 4: *REMORSE* is an ESS if $T + wP > R + wR$.

The average payoff, after the mistake, is $\frac{T+S}{2}$, which is less than R .

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

<i>Cg</i>	<i>Cg</i>	<i>Dg</i>	<i>Db</i>	<i>Cg</i>	<i>Dg</i>	<i>Dg</i>	...
<i>Cg</i>	<i>Db*</i>	<i>Cg</i>	<i>Db</i>	<i>Db</i>	<i>Db</i>	<i>Cg</i>	...
<i>Cg</i>	<i>Cg</i>	<i>Db</i>	<i>Db</i>	<i>Cg</i>	<i>Db</i>	<i>Db</i>	...
<i>Cg</i>	<i>Cg</i>	<i>Cg</i>	<i>Dg</i>	<i>Dg</i>	<i>Db</i>	<i>Cg</i>	...

The average payoff, after the mistake, is $\frac{T+S+2P}{4}$, which is less than R .

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

<i>C</i>	<i>C</i>	<i>D</i>	<i>D</i>	<i>C</i>	...
<i>C</i>	<i>D*</i>	<i>C</i>	<i>D</i>	<i>C</i>	...
<i>C</i>	<i>C</i>	<i>D</i>	<i>D</i>	<i>C</i>	...
<i>C</i>	<i>C</i>	<i>C</i>	<i>D</i>	<i>C</i>	...

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

4 Stochastic Strategies with Standing

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

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

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

$$\mathbf{T} = \begin{bmatrix} p_1 p'_1 & p_1(1-p'_1) & (1-p_1)p'_1 & (1-p_1)(1-p'_1) \\ p_2 p'_3 & p_2(1-p'_3) & (1-p_2)p'_3 & (1-p_2)(1-p'_3) \\ p_3 p'_2 & p_3(1-p'_2) & (1-p_3)p'_2 & (1-p_3)(1-p'_2) \\ p_4 p'_4 & p_4(1-p'_4) & (1-p_4)p'_4 & (1-p_4)(1-p'_4) \end{bmatrix} \quad (3)$$

(Note that p_2 is matched with p'_3 and vice versa; one player's S is the other player's T). If \mathbf{p} and \mathbf{p}' are in the interior of the strategy cube, then all entries of this stochastic matrix are

strictly positive, and hence there exists a unique stationary distribution $\mathbf{s} = (s_1, s_2, s_3, s_4)$ such that $p_i^{(n)}$, the probability to be in state i in the n -th round, converges to s_i for $n \rightarrow \infty$ ($i = 1, 2, 3, 4$). The components s_i are strictly positive and sum up to 1. They denote the asymptotic frequencies of R, S, T and P . The stochastic vector \mathbf{s} is a left eigenvector of \mathbf{T} for the eigenvalue 1, i.e. satisfies $\mathbf{s} = \mathbf{s}\mathbf{T}$.

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

$$A(\mathbf{p}, \mathbf{p}') = R s_1 + S s_2 + T s_3 + P s_4. \quad (4)$$

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

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

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

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

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

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

$$\begin{bmatrix} q_1 q'_1 & 0 & q_1(1-q'_1) & 0 & 0 & (1-q_1)q'_1 & 0 & (1-q_1)(1-q'_1) \\ q_2 q'_4 & 0 & q_2(1-q'_4) & 0 & 0 & (1-q_2)q'_4 & 0 & (1-q_2)(1-q'_4) \\ q_3 q'_6 & 0 & q_3(1-q'_6) & (1-q_3)q'_6 & (1-q_3)(1-q'_6) & 0 & 0 & 0 \\ q_4 q'_2 & 0 & q_4(1-q'_2) & 0 & 0 & (1-q_4)q'_2 & 0 & (1-q_4)(1-q'_2) \\ q_5 q'_7 & 0 & q_5(1-q'_7) & (1-q_5)q'_7 & (1-q_5)(1-q'_7) & 0 & 0 & 0 \\ q_6 q'_3 & q_6(1-q'_3) & 0 & 0 & 0 & (1-q_6)q'_3 & (1-q_6)(1-q'_3) & 0 \\ q_7 q'_5 & q_7(1-q'_5) & 0 & 0 & 0 & (1-q_7)q'_5 & (1-q_7)(1-q'_5) & 0 \\ q_8 q'_8 & 0 & q_8(1-q'_8) & 0 & 0 & (1-q_8)q'_8 & 0 & (1-q_8)(1-q'_8) \end{bmatrix} \quad (5)$$

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

$$R s_1 + S(s_2 + s_3) + T(s_4 + s_6) + P(s_5 + s_7 + s_8) . \quad (6)$$

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

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

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

$$\frac{3}{7}R + \frac{2}{7}(S + T), \quad (8)$$

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

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

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

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

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

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

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

(where a and b are real parameters such that all q_i lie between 0 and 1) every strategy obtains the same payoff, namely $P + \pi$. This can be shown by a computation similar to that in Boerlijst et al (1996), but considerably more tedious. For $a = b$ we obtain the (p_1, \dots, p_4) -strategies described in Boerlijst et al (1996).

5 Numerical Simulations

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

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

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

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

High temptation ($T = 5.5$): At high temptation we find the ESS's: *ALLD*, *GRIM*, *cTFT* and *REMORSE*. Simulations starting with just one of these strategies show that populations of *ALLD* and *GRIM* do not persist for a long time, whereas populations consisting of *cTFT* and *REMORSE* do persist. This still holds if w is slightly smaller than 1. The apparent contradiction that an ESS population can be invaded by mutants can be explained by the fact that in our model the score of a newly introduced mutant is (marginally) influenced by the mutant playing against itself. We argue that ESS's that are not stable against such small perturbations are structurally unstable: biologically, we assume that mutant strategies invade in small clusters, or clones.

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

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

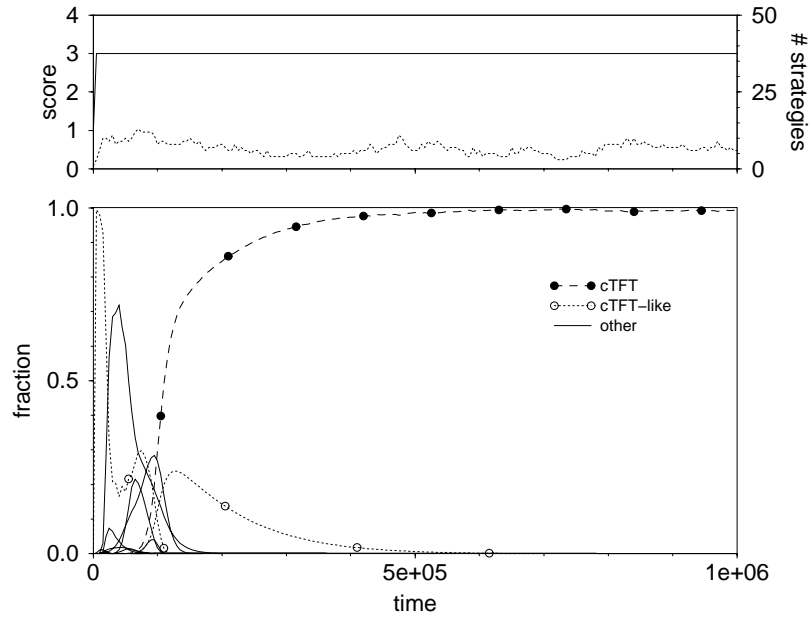


Figure 5a: Settling in *cTFT*. Evolution of a population of strategies starting from pure *ALLD* with high temptation to defect ($T = 5.5$). In the upper panel the solid line indicates the average population score whereas the dotted line indicates the number of different strategies.

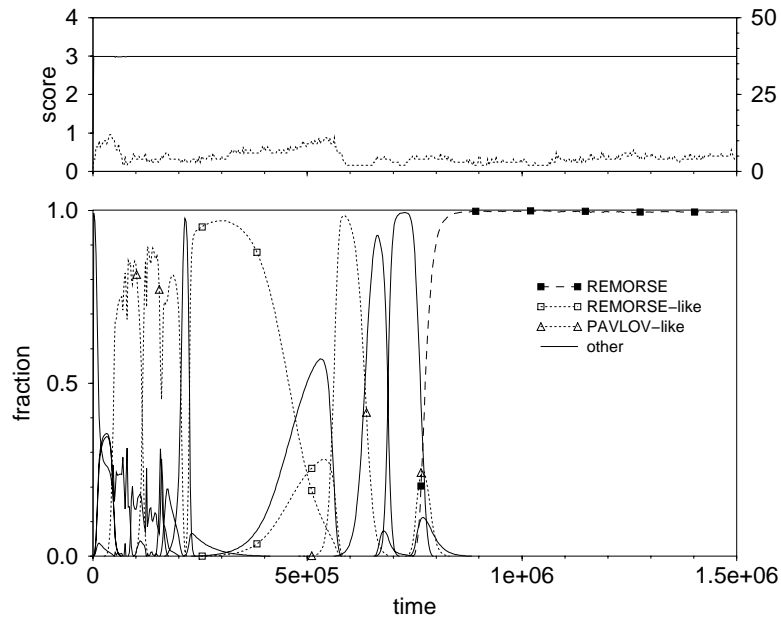


Figure 5b: Settling in *REMORSE*. After Lindgren (1991). Evolution of a population of strategies starting from pure *ALLD* with high temptation to defect ($T = 5.5$). In the upper panel the solid line indicates the average population score whereas the dotted line indicates the number of different strategies.

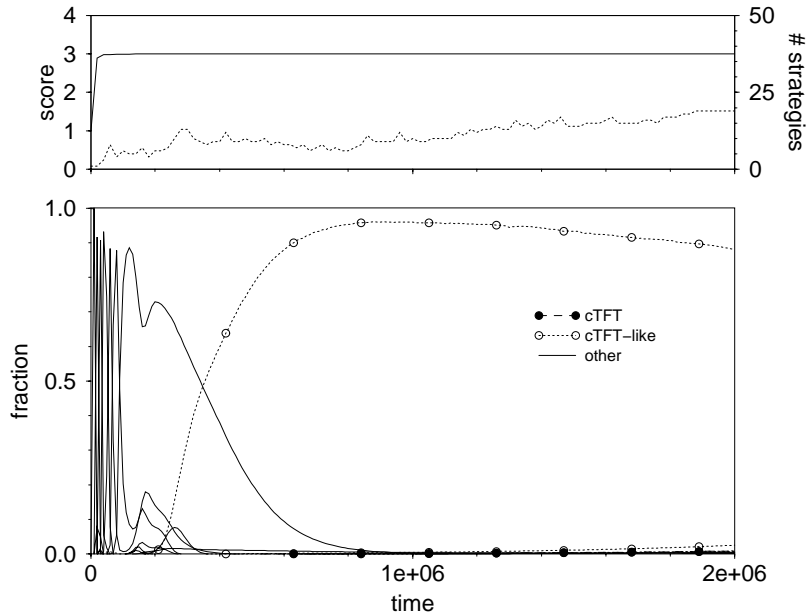


Figure 6: Simulation settling in a mixture of $cTFT$ -like strategies.

drift and the accumulation of neutral mutants. Note that pure $cTFT$ is also present in Fig. 6, but it fails to dominate the population.

To explore the basins of attraction of the ESS's we ran 100 simulations starting from *ALLD*: 68 ended in $cTFT$ -like mixtures, 11 ended in pure $cTFT$, 15 ended in *REMORSE*-like mixtures, and 6 ended in pure *REMORSE*. It seems that competition is decided on the base of which strategy first exceeds a certain threshold. The fact that there are more neutral mutants around $cTFT$ than around *REMORSE* explains the bias towards the former strategy. Simulations starting from 100 random mutants show similar statistics.

Low temptation ($T = 3.5$): Known ESS's at low temptation are *ALLD*, *GRIM*, $cTFT$, *PAVLOV* and *WEAKLING*. Again, *ALLD* and *GRIM* are easily invaded, whereas the other strategies persist. Starting 100 simulations from *ALLD* we get 63 $cTFT$ -like mixtures, 8 pure $cTFT$, 17 *PAVLOV*-like mixtures, 3 pure *PAVLOV*, 6 *WEAKLING*-like mixtures, and 3 times pure *WEAKLING*. The dynamics resembles that as described for high temptation. Fig. 7 shows a simulation that ends in pure *WEAKLING*. It can be seen that the appearance of *WEAKLING*-like strategies causes a drop in the score. Pure *WEAKLING* will slowly outcompete the other *WEAKLING*-like strategies, and the population stays fixed in a sequence of alternating mutual cooperation and defection, giving a score of $(R + P)/2$. Only 9 out of 100 simulations end in this non-cooperative mode, *PAVLOV*(-like) and $cTFT$ (-like) populations both reach a score close to R .

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

To conclude, we see that the addition of a standing in the Prisoner's Dilemma facilitates the evolution of cooperation. Populations with random mutations in most cases

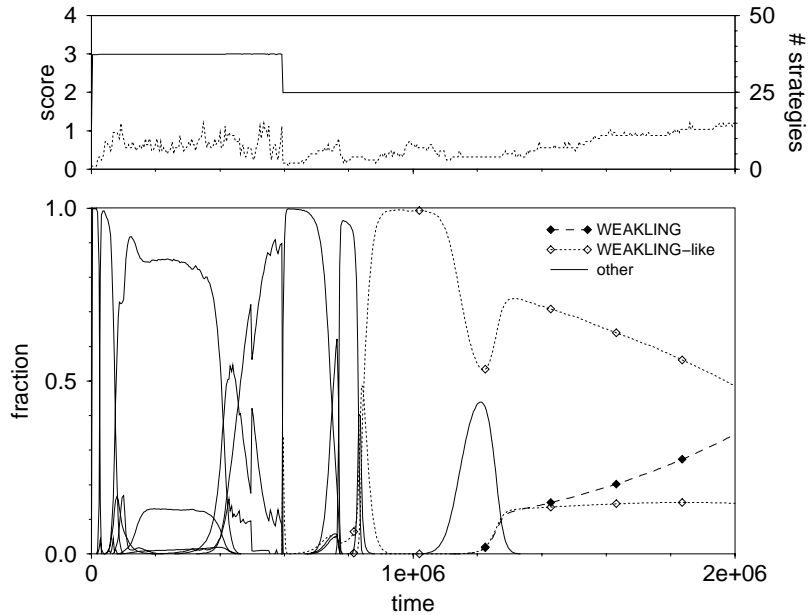


Figure 7: Evolution of a population of strategies starting from pure *ALLD* with low temptation to defect ($T = 3.5$).

quickly adapt to a cooperative mode, and only rarely the population is trapped in the *WEAKLING* strategy. Surprisingly, this suboptimal trapping is only observed in situations with low temptation to defect.

6 The Alternating PD

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

$$\mathbf{T} = \begin{bmatrix} q_1 q'_1 & q_1(1-q'_1) & 0 & 0 & (1-q_1)q'_2 & (1-q_1)(1-q'_2) \\ q_2 q'_5 & q_2(1-q'_5) & (1-q_2)q'_6 & (1-q_2)(1-q'_6) & 0 & 0 \\ q_3 q'_1 & q_3(1-q'_1) & 0 & 0 & (1-q_3)q'_2 & (1-q_3)(1-q'_2) \\ q_4 q'_5 & q_4(1-q'_5) & (1-q_4)q_6 & (1-q_4)(1-q'_6) & 0 & 0 \\ q_5 q'_1 & q_5(1-q'_1) & 0 & 0 & (1-q_5)q'_2 & (1-q_5)(1-q'_2) \\ q_6 q'_3 & q_6(1-q'_3) & 0 & 0 & (1-q_6)q'_4 & (1-q_6)(1-q'_4) \end{bmatrix} \quad (11)$$

If \mathbf{s} , again, denotes the stationary vector, then the payoff for the \mathbf{q} -player is

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

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

7 Discussion

All strategies considered in this paper can be implemented by finite automata. For the extensive theory in this field, we refer to Binmore and Samuelson (1992). One might ask whether the *cTFT*-strategy can be implemented by a strategy uniquely based on a finite (but possibly very long) memory of the moves of the two players, and not using the notion of standing. This however is not the case. If, for instance, a sequence of alternating defections occurs, only the player that started to defect will have a bad standing. The next move is not specified by a finite memory of previous moves in case the initial defection happened prior to the memorised moves.

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

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

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

Here is, as an intriguing example, the strategy *Prudent-PAVLOV* (*pPAVLOV*). This strategy follows in most cases the *PAVLOV*-strategy, as the name suggests. However, after any defection it will only resume cooperation after two rounds of mutual defection. This is achieved by normally playing defections with standing D_1 , and only playing D_0 after a mutual defection or an erroneous defection. Suppose that two *pPAVLOV*s are

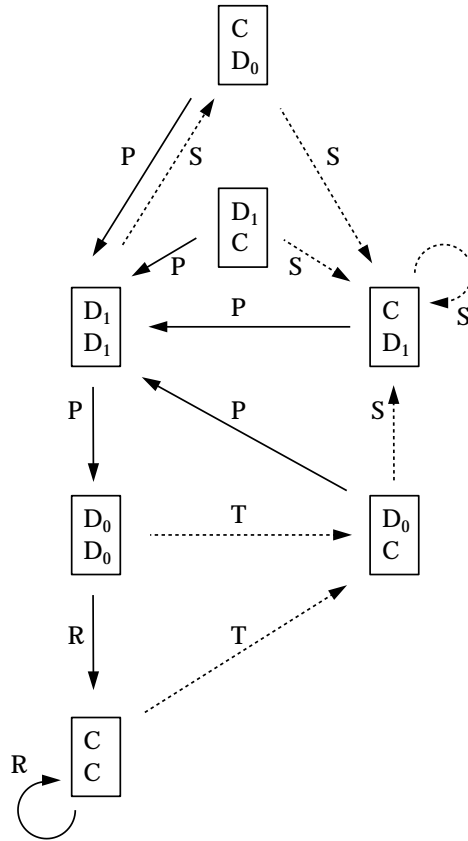


Figure 8: *pPAVLOV* is an ESS if $T + wP + w^2P < R + wR + w^2R$.

engaged in a match. They usually both cooperate. If one defects by mistake, the state is (C, D_0) . In the next round, the state is (D_1, D_1) ; in the second-next round, it is (D_0, D_0) , and hereafter mutual cooperation is resumed. This strategy, which depends only on the previous round, acts to all purposes like Lindgren's (1991) memory two strategy. An erroneous defection against its like entails two rounds of mutual defection, and then leads back to mutual cooperation. An *AllC*-opponent will be exploited ruthlessly; but against an *AllD* opponent, *pPAVLOV* will be suckered every third round. It is easy to see (cf. Fig.8) that this strategy is an ESS whenever

$$R + wR + w^2R > T + wP + w^2P \tag{13}$$

and numerical simulations show that it attracts very well.

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

C	C	D_1	D_1	D_0	C	\dots
C	D^*	C	D	D	C	\dots
						\dots
C	C	D	D	D	C	\dots
C	C	C	D_1	D_0	C	\dots

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

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

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

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

References

- Axelrod, R. (1984) *The Evolution of Cooperation*, Basic Books, New York (reprinted 1989 in Penguin, Harmondsworth).
- Axelrod, R. and Hamilton, W.D. (1981), The evolution of cooperation, *Science* **211**, 1390-6.
- Axelrod, R. (1987), The evolution of strategies in the iterated Prisoner's Dilemma, in Davis, D. (ed.) *Genetic Algorithms and Simulated Annealing*, Pitman, London.
- Axelrod, R. and Dion, D. (1988), The further evolution of cooperation, *Science* **242**, 1385-90.
- Bendor, J. (1993), Uncertainty and the evolution of cooperation. *Journal of Conflict Resolution* **37**, 709-34.
- Bendor, J. R.M. Kramer and S. Stout (1991) When in doubt: Cooperation in a noisy Prisoner's Dilemma. *Journal of Conflict Resolution* **35**, 691-719.
- Binmore, K.G. and Samuelson, L. (1992), Evolutionary stability in repeated games played by finite automata, *Journal of Economic Theory* **57**, 278-305.
- Boerlijst, M.C., Nowak, M.A. and Sigmund, K. (1996), Equal Pay for all Prisoners, to appear.
- Boyd, R. (1988), Is the repeated Prisoner's Dilemma a good model of reciprocal altruism?, *Ethology and Sociobiology* **9**, 278-305.
- Boyd, R. (1989), Mistakes Allow Evolutionary Stability in the Repeated Prisoner's Dilemma Game, *J. Theor. Biol.* **136**, 47-56.

- Dawkins, R. (1989) *The Selfish Gene*, Oxford Univ. Press (New Edition).
- Frean, M. (1994), The Prisoner's Dilemma without synchrony, *Proc. R. Soc. Lond. B* **257**, 75-79.
- Fudenberg, D. and E. Maskin (1990), Evolution and cooperation in noisy repeated games. *American Economic Review* **80**, 274-9.
- Harrington, J.C. and R. Axelrod (1995) Tit For Tat with Contrition is an evolutionarily stable strategy in the iterated Prisoner's Dilemma with noise. To appear.
- Hofbauer, J. and Sigmund, K. (1988), *The Theory of Evolution and Dynamical Systems*, Cambridge UP.
- Kollock, P. (1993), An eye for an eye leaves everyone blind: Cooperation and accounting systems. *American Sociological Review* **58**, 768-86.
- Kraines, D. and Kraines, V.(1988), Pavlov and the Prisoner's Dilemma, *Theory and Decision* **26**, 47-79
- Lindgren, K. (1991) Evolutionary phenomena in simple dynamics, in *Artificial life II* (ed. C.G. Langton et al), Santa Fe Institute for Studies in the Sciences of Complexity, Vol. X, 295-312.
- May, R.M. (1987), More evolution of cooperation, *Nature* **327**, 15-17.
- Maynard Smith, J. (1982) *Evolution and the theory of games*, Cambridge UP.
- Miller, J.H. (1989) The coevolution of automata in the Repeated Prisoner's Dilemma, working paper of the Santa Fe Institute Economics Research Program.
- Molander, P (1985), The optimal level of generosity in a selfish, uncertain environment, *J. Conflict resolution* **29**, 611-618.
- Müller, U. (1987), Optimal retaliation for optimal cooperation, *J. Conflict Resolution* **31**, 692-724
- Nowak. M.A. (1990), Stochastic strategies in the Prisoner's Dilemma, *Theor. Pop. Biol.* **38**, 93-112.
- Nowak, M.A, May, R.M. and Sigmund, K, (1995a), The arithmetics of mutual help, *Sci. Am.* **272(6)**, 76-81.
- Nowak, M.A. and Sigmund, K, (1990), The evolution of reactive strategies in iterated games, *Acta Applicandae Math.* **20**, 247-265.
- Nowak, M.A. and Sigmund, K. (1992), Tit for tat in heterogeneous populations, *Nature* **355**, 250-2.
- Nowak, M.A. and Sigmund, K. (1993a), Chaos and the evolution of cooperation, *Proc. Nat. Acad. Science USA* **90**, 5091-94.
- Nowak, M.A. and Sigmund, K. (1993b), Win-stay, lose-shift outperforms tit-for-tat, *Nature*, **364**, 56-8.
- Nowak, M.A. and Sigmund, K. (1994), The alternating Prisoner's Dilemma, *J. of Theoretical Biology* **168**, 219-26.
- Nowak, M.A. and Sigmund, K. (1995), Invasion dynamics for the finitely repeated Prisoner's Dilemma, *Games and Economic Behaviour*, to appear.
- Nowak, M.A., Sigmund, K. and El-Sedy, E. (1995b), Automata, repeated games, and noise, *J. Math. Biol.* **33**, 703-32.

- Selten, R. (1975), Re-examination of the perfectness concept for equilibrium points in extensive games, *International Journal of Game Theory* **4**, 25-55.
- Selten, R. and Hammerstein, P. (1984), Gaps in Harley's argument on evolutionarily stable learning rules and in the logic of tit for tat, *The Behavioural and Brain Sciences* **7**, 115-6
- Sigmund, K. (1995), *Games of Life*, Penguin, Harmondsworth.
- Sugden, R. (1986), *The economics of rights, co-operation and welfare*, Blackwell, Oxford.
- Wu, J. and R. Axelrod (1995), How to cope with noise in the iterated Prisoner's Dilemma, *Journal of Conflict resolution* **39**, 183-9.
- Young, H. (1993), The evolution of conventions, *Econometrica* **61**, 57-84.