IIASA

International Institute for Applied Systems Analysis • A-2361 Laxenburg • Austria
Tel: +43 2236 807 • Fax: +43 2236 71313 • E-mail: info@iiasa.ac.at • Web: www.iiasa.ac.at

INTERIM REPORT IR-98-103 / December

The Efficiency of Adapting Aspiration Levels

Martin Posch (Martin.Posch@univie.ac.at)
Alexander Pichler (apichler@radon.mat.univie.ac.at)
Karl Sigmund (ksigmund@esi.ac.at)

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

IIASA STUDIES IN ADAPTIVE DYNAMICS NO. 33

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-95-099.
van Strien SJ, Verduyn Lunel SM (eds.): Stochastic and Spatial Structures of Dynamical Systems, Proceedings of the Royal Dutch Academy of Science (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-96-001.
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-96-002.
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-96-003.
Evolutionary Ecology (1997) 11, 485–501.

No. 6 Metz JAJ, Mylius SD, Diekmann O:
When Does Evolution Optimise? On the Relation between Types of Density Dependence and Evolutionarily Stable Life History Parameters.
IIASA Working Paper WP-96-004.

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-96-029.
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-96-070.
Journal of Evolutionary Biology (1998) 11, 41–62.

No. 11 Geritz SAH, Metz JAJ, Kisdi E, Meszéna G:
The Dynamics of Adaptation and Evolutionary Branching.
IIASA Working Paper WP-96-077.
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 (1998) 12, 35–57.

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) 352, 1647–1655.

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-97-001.
Journal of Biological Systems (1997) 5, 265–284.

No. 16 Heino M, Metz JAJ, Kaitala V:
The Enigma of Frequency-Dependent Selection.
IIASA Interim Report IR-97-061.
Trends in Ecology and Evolution (1998) in press.

No. 17 Heino M:
Management of Evolving Fish Stocks.
IIASA Interim Report IR-97-062.
Canadian Journal of Fisheries and Aquatic Sciences (1998) in press.

No. 18 Heino M:
Evolution of Mixed Reproductive Strategies in Simple Life-History Models.
IIASA Interim Report IR-97-063.

No. 19 Geritz SAH, van der Meijden E, Metz JAJ:
Evolutionary Dynamics of Seed Size and Seedling Competitive Ability.
IIASA Interim Report IR-97-071.

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-97-072.
Trends in Ecology and Evolution (1998) 13, 1–2.

No. 21 Boerlijst MC, Nowak MA, Sigmund K:
Equal Pay for all Prisoners. / The Logic of Contrition.
IIASA Interim Report IR-97-073.
AMS Monthly (1997) 104, 303–307.
Journal of Theoretical Biology (1997) 185, 281–294.

No. 22 Law R, Dieckmann U:
Symbiosis without Mutualism and the Merger of Lineages in Evolution.
IIASA Interim Report IR-97-074.
Proceedings of the Royal Society of London Series B (1998) 265, 1245–1253.

No. 23 Klinkhamer PGL, de Jong TJ, Metz JAJ:
Sex and Size in Cosexual Plants.
IIASA Interim Report IR-97-078.
Trends in Ecology and Evolution (1997) 12, 260–265.

No. 24 Fontana W, Schuster P:
Shaping Space: The Possible and the Attainable in RNA Genotype-Phenotype Mapping.
IIASA Interim Report IR-98-004.

No. 25 Kisdi E, Geritz SAH:
Adaptive Dynamics in Allele Space: Evolution of Genetic Polymorphism by Small Mutations in a Heterogeneous Environment.
IIASA Interim Report IR-98-038.

No. 26 Fontana W, Schuster P:
Continuity in Evolution: On the Nature of Transitions.
IIASA Interim Report IR-98-039.
Science (1998) 280, 1451–1455.

No. 27 Nowak MA, Sigmund K:
Evolution of Indirect Reciprocity by Image Scoring. / The Dynamics of Indirect Reciprocity.
IIASA Interim Report IR-98-040.
Nature (1998) 393, 573–577.

No. 28 Kisdi E:
Evolutionary Branching Under Asymmetric Competition.
IIASA Interim Report IR-98-045.

No. 29 Berger U:
Best Response Adaptation for Role Games.
IIASA Interim Report IR-98-086.

No. 30 Van Dooren TJM:
The Evolutionary Ecology of Dominance-Recessivity
IIASA Interim Report IR-98-096.

No. 31 Dieckmann U, O'Hara B, Weisser W:
The Evolutionary Ecology of Dispersal.
IIASA Interim Report IR-98-108.
Trends in Ecology and Evolution (1999) in press.

No. 32 Sigmund K:
Complex Adaptive Systems and the Evolution of Reciprocation.
IIASA Interim Report IR-98-100.

No. 33 Posch M, Pichler A, Sigmund K:
The Efficiency of Adapting Aspiration Levels.
IIASA Interim Report IR-98-103.

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

Abstract

Win-stay, lose-shift strategies in repeated games are based on an aspiration level. A move is repeated if and only if the outcome, in the previous round, was satisficing in the sense that the payoff was at least as high as the aspiration level. We investigate the conditions under which adaptive mechanisms acting on the aspiration level (selection, for instance, or learning) can lead to an efficient outcome; in other words, when can satisficing become optimising? Analytic results for 2×2-games are presented. They suggest that in a large variety of social interactions, self-centered rules (based uniquely on one's own payoff) cannot suffice.

About the Authors

Martin Posch
Institut für Medizinische Statistik
Schwarspanierstrasse 17
A-1090 Vienna, Austria

Alexander Pichler
Institut für Mathematik
Universität Wien
Strudlhofgasse 4
A-1090 Vienna, Austria

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

The Efficiency of Adapting Aspiration Levels

Martin Posch
Alexander Pichler
Karl Sigmund

1 Introduction

In a *game theory without rationality* (see Rapoport, 1984), players are not assumed to be able to fully understand the situation they are engaged in. Their moves are based on knee-jerk rules rather than on strategic analysis. Possibly the simplest of such rules is the win-stay, lose-shift principle, which consists in repeating an action if it proved successful, and in switching to another action if not. Suppose that we were playing a machine with two levers, one resulting in a positive, the other in a negative outcome. The win-stay, lose-shift principle would result in our repeating the action with the positive outcome; if we erroneously tried the wrong action, we would switch back, in the next round, to the right action. Many experiments have shown that such a behaviour, or some approximation of it, is widespread among human and animal actors. Interestingly, this crudest form of a learning rule works even in situations involving several agents, as in the so-called *minimal social situation* (Colman, 1995).

The win-stay, lose-shift principle was originally formulated by Thorndike (1911):

Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction are more firmly connected with the situation; those which are accompanied or closely followed by discomfort have their connection with the situation weakened.

The wide range of validity of this principle was soon recognised (see, e.g., Hoppe, 1931, Rescorla and Wagner, 1972). In the hands of Herbert Simon, satisfaction-seeking behaviour became a leading contender for explaining social and economic decision making (see Simon, 1955, 1957, 1962; Winter, 1971; Radner, 1975). A considerable amount of empirical evidence suggests that the behaviour of individuals and firms aims at *satisficing*, rather than optimising.

But when do we feel satisfied? In certain situations (as when foraging for food, or for sex) our body knows. In other situations, we have to find out. We may feel pleased if we pulled a lever which delivers one dollar, but not if we are told that the alternative would have delivered ten. In such a situation, we must learn what to aim for; whereas in the foraging case, our genome has done the learning already. Natural selection operating in a population, or a learning rule based on individual trial and error, can cause an *adaptation* of the aspiration level.

It is easy to see how selection, or learning rules, lead to an optimal aspiration level when playing *against nature*. We are interested in exploring how adaptation works when playing *against other players*. In the repeated Prisoner's Dilemma game, for instance, a strategy called PAVLOV does very well (see Kelley et al, 1962, Colman, 1995, Kraines and Kraines, 1988, and Nowak and Sigmund, 1993). PAVLOV is a win-stay, lose-shift rule with an aspiration level lying somewhere between the two highest and the two lowest payoffs. Is there any reason to assume that selection, or learning, will adapt the aspiration level precisely to this interval? And how would such adaptive mechanisms fare in other games? We will assume that our players are 'blind robots' without any knowledge of the structure of the iterated game, except that they have two options. They need not even be aware of the existence of another player. Their only information is the payoff which they obtain in each round.

In section 2, we shall briefly discuss some mechanisms for adapting the aspiration level, studying first the action of selection, and then two particularly simple learning rules, which are extremal cases of convex updating of the aspiration level, called YESTERDAY and FARAWAY. In sections 3 to 5, we turn to the simplest games, symmetric games between two players having two strategies each. We examine whether adaptive mechanisms lead to an efficient outcome for such 2×2 games. This is one aspect of a larger question, namely: when is satisficing optimising?

In this paper, our approach will be based on analytic methods. We restrict our attention to *deterministic* win-stay, lose-shift strategies based on switching to the alternative option if and only if the payoff from the previous round falls below the aspiration level. (In Thorndike's formulation, win-stay, lose-shift is a *stochastic* rule: the difference between aspiration level and actual payoff only affects the *propensity* to switch.) For a simulation-based exploration of win-stay, lose-shift strategies with longer memory sizes we refer to Posch (1998).

2 Games against Nature

Consider a two-armed bandit. Pulling one lever yields payoff R, pulling the other P, with $P < R$. Let a be the aspiration level of a player. The player will repeat the former action if the payoff was at least a, and switch to the other action otherwise. With some probability $\epsilon > 0$ this action is misimplemented. For simplicity, we shall only consider the limiting case $\epsilon \to 0$. We assume that the game consists of a large number of rounds, and that the payoff for the repeated game is given by the limit-in-the-mean (l.i.m.) of the payoff per round (i.e. $\lim(p_1 + ... + p_N)/N$ for $N \to \infty$, where p_n is the payoff in round n). If $a > R$, the player will switch after every round, and obtain as l.i.m. payoff $(R+P)/2$. If $a \leq P$, the player will always be satisfied, switch only by mistake, and then repeat the new action till the next mistake occurs. Again the l.i.m. payoff is $(R+P)/2$. For $P < a \leq R$, the player will always pull the R-lever, except by mistake; after an erroneous P, the player will switch back to R. The l.i.m. payoff is R.

How does selection act on the frequencies x_1, x_2 and x_3 of the three strategies corresponding to the intervals $]-\infty, P]$, $]P, R]$ and $]R, +\infty[$ of possible aspiration levels? We shall assume that payoff is converted into reproductive fitness, and that like begets like. This yields the replicator equation

$$\dot{x}_i = x_i(f_i - \bar{f}) \qquad (1)$$

where f_i is the l.i.m. payoff for strategy i and $\bar{f} = \sum x_k f_k$ is the average l.i.m. payoff in the population (see Hofbauer and Sigmund, 1998). The dynamics on the corresponding unit simplex S_3 leads to the extinction of the 'wrong' aspiration levels: x_2 converges to 1. In this sense, selection yields an aspiration level a in $]P, R]$.

What about learning? Conceivably the simplest way in which experience can affect a player's aspiration level consists in *convex updating*, by taking into account the payoff obtained in the previous round. More precisely, if a_n is the aspiration level and p_n the payoff in the n-th round, then $a_n = (1-\alpha)a_{n-1} + \alpha p_{n-1}$ for some fixed $\alpha \in]0, 1[$. If the aspiration level is initially higher than R, then the player will restlessly switch between the two possible actions, and a_n will steadily decrease until it is lower than R. If, on the other hand, a_n is lower than P, then the player will repeat the previous action. If this action happens to yield R, the aspiration level will soon be between R and P. If the action yields P, then a_n approaches P from below. A mistake in implementation will eventually bring it into the 'right' interval. Once there, it will converge towards R from below. An eventual mistake in implementation happening now will not cause a_n to leave the interval $]P, R]$ and will immediately be corrected.

When players play each other (rather than a two-armed bandit), convex updating can lead to complex outcomes. We shall therefore restrict attention to two updating rules which represent two instructive extremal cases. With YESTERDAY, $\alpha = 1$ i.e. a_n is just p_{n-1}, the payoff obtained in round $n-1$. Even if a player starts with the P-lever, the first mistake will lead to the R-lever. The player then stays with this option: any further mistake will immediately be corrected.

FARAWAY is the opposite case, in some sense. Of course $\alpha = 0$ means no updating at all, which is uninteresting. Instead of this, we shall assume that the aspiration level is slowly, but continuously modified towards the long-run average. This means that if the aspiration level is in $]-\infty, P]$ or $[R, +\infty[$, it steadily inches towards $(R+P)/2$ and eventually enters the interval $]P, R]$. Once there, it converges towards R. The direction of change defines a dynamics leading asymptotically towards R, which is just 'right.

3 2×2-games

The simplest non-trivial games involve two players with two options each, which we call **C** and **D**. We shall assume that the game is symmetric, i.e. that the two players are interchangeable. The payoff matrix is

$$\begin{pmatrix} R & S \\ T & P \end{pmatrix} \tag{2}$$

i.e. R is the payoff for using **C** against a player also using **C**, S for using **C** against **D** etc. We consider only the generic situation where the four payoff values are pairwise distinct. There are then 12 different rank orderings. They correspond to very different strategic situations, see for example Rapoport et al (1976), Binmore (1992) or Colman (1995). It is no restriction of generality to assume $R > P$ (if this does not hold, we just interchange **C** and **D**) and to normalise the values such that

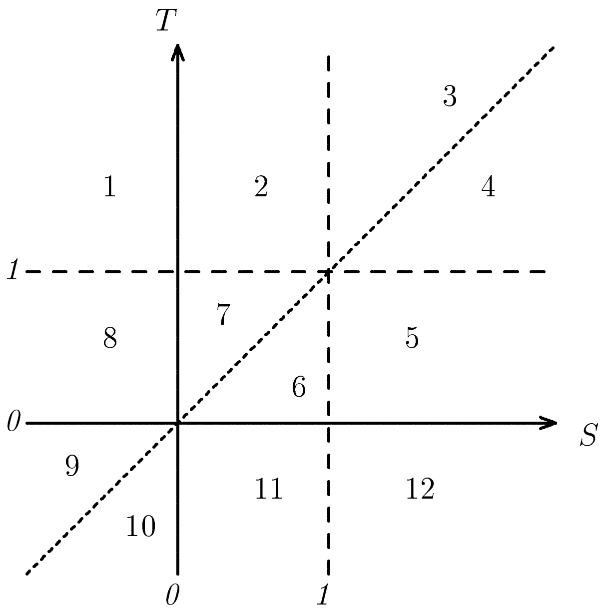


Figure 1: A partitioning of the (S,T)-plane which displays the 12 symmetric 2×2-games.

$R = 1$ and $P = 0$. Each game, then, corresponds to a point in the (S,T)-plane, and the 12 rank orderings correspond to 12 planar regions, see Fig.1.

For the Prisoner's Dilemma, for instance, we have $T > 1$ and $S < 0$, for the Chicken game (also known as Hawk-Dove) $T > 1 > S > 0$ etc. For the issue of equilibrium selection in such games, we refer to Harsanyi and Selten (1988), van Damme (1991) and Samuelson (1997). In the games 1 and 5, 6, 7, 11, 12, both players have a dominant strategy (which yields a higher payoff than the alternative, irrespective of the other player's choice); the games 6, 7, 8, 9, 10, 11 are common interest games (the best outcome for one player is also best for the other – namely R); and the union of these games, i.e. all except 2, 3 and 4, are Stackelberg-soluble. (The Stackelberg solution is the strategy which optimises the payoff under the assumption that the reply is optimal from the co-player's view. The game is Stackelberg-soluble if, when both players adopt their Stackelberg solution, none can do better by deviating unilaterally, see Colman and Stirk, 1998).

The four payoff values divide the real line into 5 intervals. All aspiration levels in the same interval define the same win-stay, lose-shift strategy. The two unbounded intervals correspond to strategies which are unaffected by the co-player. They consist in switching to the other option in every round (this will be called NO SATISFACTION), or in sticking with one option until a mistake leads to the alternative (this is called LET IT BE). The three bounded intervals correspond (in ascending order) to aspiration levels which are *modest, balanced*, or *ambitious.* For both the Prisoner's Dilemma and the Chicken game, for instance, a balanced aspiration level lies in $[0, 1[$ and corresponds to the strategy PAVLOV. This strategy consists in playing **C** if and only if the co-player used the same option, in the previous round, as one did oneself.

We may describe each strategy based on the outcome of the previous round by a quadruple (p_R, p_S, p_T, p_P) where p_k is the probability of using **C** after having experienced in the previous round outcome $k \in \{R, S, T, P\}$. Since we consider only deterministic win-stay, lose-shift rules, the p_k are either 0 or 1. Thus PAVLOV, for instance, is $(1, 0, 0, 1)$. In Fig.2 we display for each game the ambitious, balanced or modest strategies. We note that in crossing a frontier line, exactly one strategy is modified, each time by altering two of its digits p_k.

We now assume that there is a small probability ϵ to mis-implement a move, so that PAVLOV, for instance, becomes $(1-\epsilon, \epsilon, \epsilon, 1-\epsilon)$. The initial move, then, has no influence on the long-term outcome of the game. In (Nowak et al, 1995) one can find the l.i.m. payoff obtained by using one strategy against a player using another, for the limiting case $\epsilon \to 0$. A player using PAVLOV obtains, for instance, $(R + S + P)/3$ against a player using the BULLY strategy $(0, 0, 0, 1)$, resp. payoff R against another PAVLOV player (with our normalisation, this becomes $(1+S)/3$ resp. 1).

An outcome is Pareto-optimal if no other choice of strategies can lead to an improvement (i.e. a higher l.i.m. payoff) for *both* players. It is easy to see that the average for the two players is then the maximum of R and $(T + S)/2$, i.e. $\max\{(1, (T+S)/2)\}$. In Fig. 6a we describe when *some* win-stay, lose-shift strategy is efficient, i.e. leads to a Pareto-optimal outcome, if all players adopt it. We note that the ambitious strategy is never efficient.

For any given game, one can set up the replicator equation (1) describing the dynamics of the frequencies x_a, x_b and x_m of the ambitious, balanced or modest strategies under natural selection. The analysis of this equation is straightforward, but somewhat laborious, since most of the twelve types of game give rise, depending on the parameters S and T, to several different long-term behaviours (see Pichler, 1998, based on Bomze, 1995). We add that no attractor can be invaded by the win-stay, lose-shift strategies NO SATISFACTION $(0, 0, 1, 1)$ or LET IT BE $(1, 1, 0, 0)$.

We do not describe all 37 cases, but concentrate on the following issues: (a) which aspiration levels get selected? and (b) when is the outcome efficient?

Concerning (a), the three aspiration intervals never coexist. At least one is always eliminated. Two intervals can, in some instances, stably coexist, in the sense that the dynamics lead to a bimorphic population, part of which use one and part another interval, with well-defined frequencies of the two types. In most cases, the attractor consists of one type only. In Fig.3 a-c, we have shaded in black the areas where an aspiration range is stably adopted by the whole population, and in grey the areas where it is part of a bimorphism (a stable mixture where a fraction of the population adopts it). We note that bistable situations (where the initial condition influences the outcome) are not rare.

Concerning (b), we refer to Fig. 6b. We denote in black the area in the (S, T)-plane where selection always leads to a Pareto optimal outcome, and in grey the zone where some initial conditions $(x_a, x_b, x_m) \in S_3$ lead to Pareto-optimality and others do not. We note that only for a part of the games of type 1, an unstable efficient outcome exists.

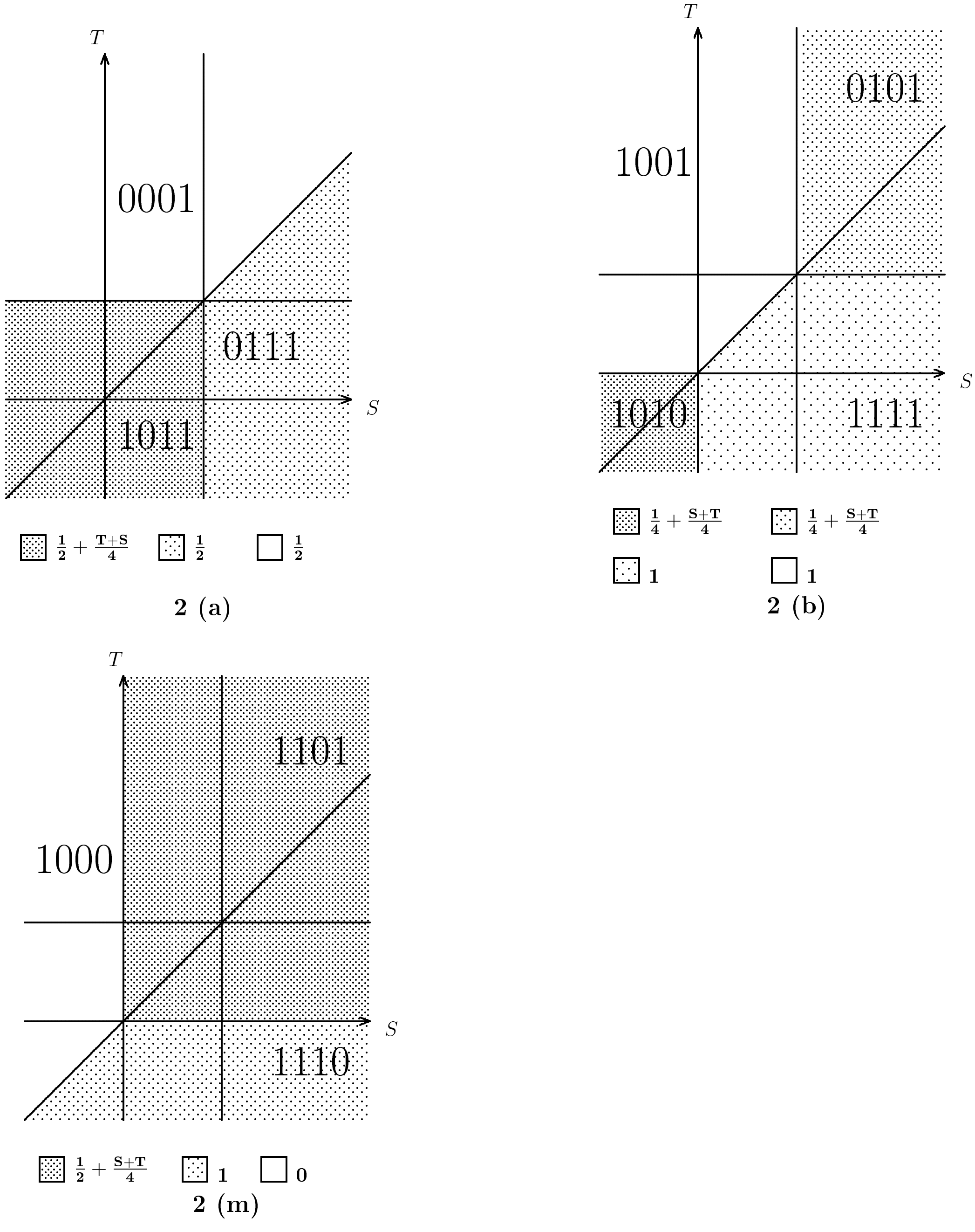
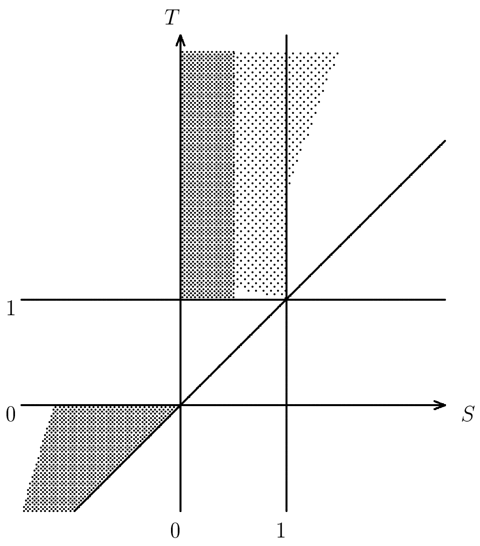
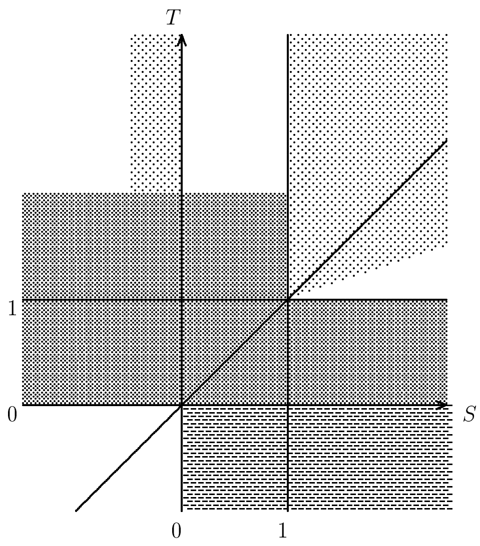


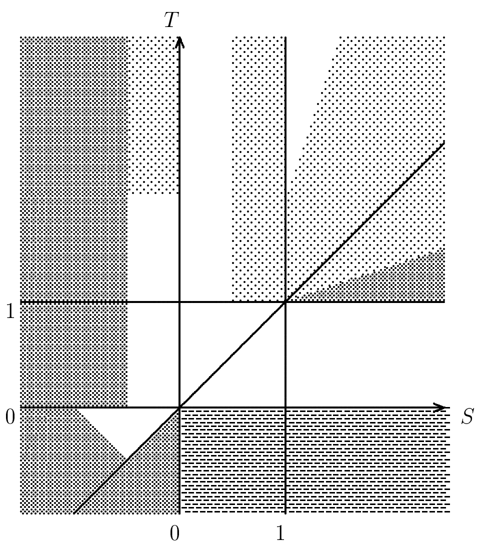
Figure 2: A description of the (a) ambitious, (b) balanced, (m) modest win-stay, lose-shift strategies corresponding to the different 2×2 games. The figure displays the corresponding (p_R, p_S, p_T, p_P)-coding (see text), and the l.i.m. payoff for a player using this strategy against a player using the same strategy.



3 (a)



3 (b)



3 (m)

Figure 3: When does selection among the different win-stay, lose-shift strategies lead to the (a) ambitious (b) balanced or (m) modest strategy? The dark shading describes the (S, T)-region where a monomorphic population using this strategy can emerge, and the light grey describes that region where selection leads to a stable bimorphic population, with a well-defined fraction using this strategy. In the dark grey region (cases 11 and 12 for b and 3c), selection leads to a mixed population in which both the balanced and the modest strategies coexist in a mixture which depends on the initial condition.

4 The strategy YESTERDAY

YESTERDAY repeats the previous move if and only if it obtained a payoff at least as good as in the round before. Let us compute the average payoff between two YESTERDAY players. As soon as the initial condition, i.e. the transition from the first round to the next, is given (for instance $T \to T$ or $T \to R$) all further transitions are specified. Obviously, players experiencing the same outcome in two consecutive moves (for instance $T \to T$) will not shift to another move (except by mistake, but we shall ignore this for the moment). This yields four stationary states, namely $r : R \to R \to R \to ...$, and similarly s, t and p. Furthermore, since $P < R$ by convention, the transition $P \to R$ and $R \to P$ must be followed by the stationary state r. The other transitions depend on the rank ordering of the payoff values.

Let us consider this for the Chicken game (number 2 in our notation). Fig. 4a shows how the game develops. For any initial condition, one of the four stationary states r, s, t or p are reached.

We allow now for misimplementing a move with a small probability ϵ. In the stationary state r, for instance, one of the players can mistakenly play **D** instead of **C** (we assume that both players are equally likely to get their next move wrong, and we neglect the possibility that both players make a mistake in the same round, an event occuring with probability ϵ^2). Thus a mistake can lead from $R \to R$ to $R \to T$ or to $R \to S$ (but not to $R \to P$). Since this leads, after three rounds, back to r, and since we may neglect the possibility that two mistakes occur within three rounds (which again has a probability proportional to ϵ^2), a mistake leads from r back to r. Similarly, a mistake in s leads to $S \to R$ or to $S \to P$, and hence after two or four rounds yields the steady state r. The same happens if a mistake occurs when in state t. But a mistake in p leads with equal probability to $P \to S$ or $P \to T$, and from there to the steady states s or t.

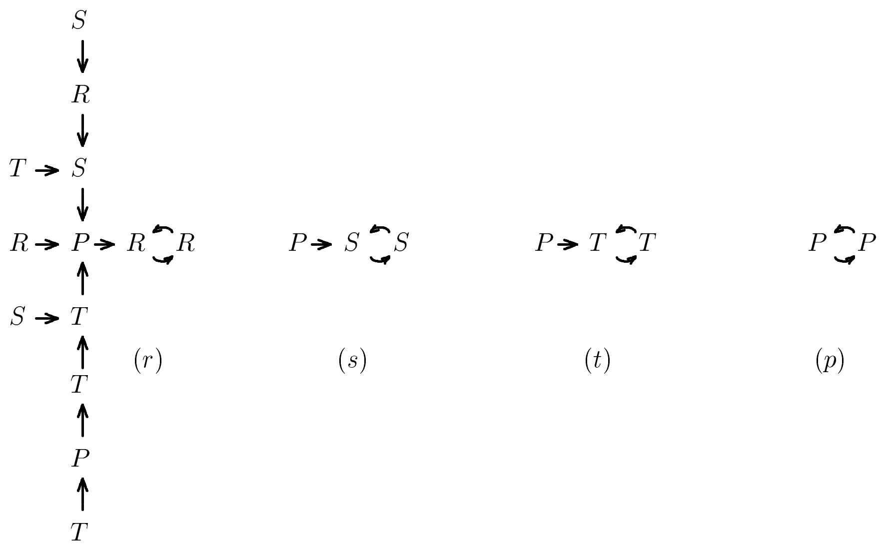
Thus errors in implementation can be described by a Markov chain having as states r, s, t and p (in this order), and as transition matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \end{pmatrix}. \tag{3}$$

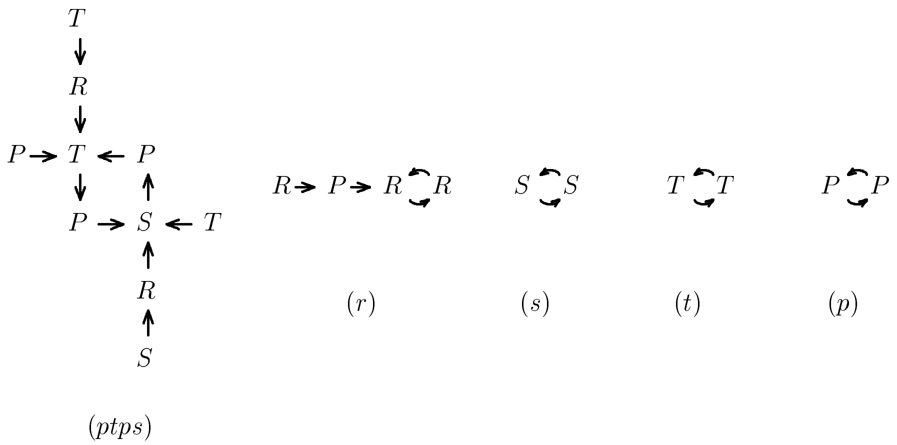
This matrix has a unique stationary distribution π, given by $\pi_r = 1$ and $\pi_s = \pi_t = \pi_p = 0$. It follows that if two players using YESTERDAY play a repeated Chicken game, their payoff (defined as limit-in-the-mean of the payoff per round) is R, which is an eminently sensible outcome. Both players cooperate (i.e. do not escalate the conflict).

If we consider, instead of Chicken, the Prisoner's Dilemma game (number 1 in our notation), we find a very different outcome, in spite of the fact that only P and S have been permuted in the rank ordering of payoff values. In addition to the four steady states r, s, t, p, we now find a cycle of period four, namely $T \to P \to S \to P \to T \to ...$, which we call *ptps*. In Fig. 4b, we display the transitions.

From the steady states r, s, t and p, every mis-implementation leads to *ptps*. Errors occuring within the cycle have a more varied outcome. A mis-implementation



4a



4b

Figure 4: The transitions, from round to round, in the payoff for a YESTERDAY player against another YESTERDAY player, (a) for Chicken (case 2 in Fig.1) and (b) the Prisoner's Dilemma (case 1 in Fig.1). The first transition is assumed to be given as initial condition.

turns $S \to P$, either into $S \to S$ or into $S \to T$, and hence leads with equal probability either into the steady state s or back into *ptps*. Similarly, mistakes turn $T \to P$ with equal probability either into the steady state t or back into *ptps* again, whereas they turn $P \to S$ and $P \to T$ into r or p. The transition matrix between the steady states r, s, t, p and *ptps* (in this order) is given by

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1/4 & 1/8 & 1/8 & 1/4 & 1/4 \end{pmatrix}. \tag{4}$$

The unique stationary distribution π is given by $\frac{1}{14}(2,1,1,2,8)$ and the mean payoff per round is $(2R+3S+3T+6P)/14$, which is considerably lower than the Pareto-optimal outcome.

One can similarly compute the payoff for YESTERDAY against itself for each of the remaining games. The result is shown in Fig. 6c. The game 8 (the *stag hunt* game) admits the cycle *ptps* and the two games 5 and 12 admit the cycle *rsrt*. All other games have only the steady states r, s, t and p.

Among the games where **C** is the dominating solution, (i.e. where $T < 1$ and $S > 0$) YESTERDAY always leads to the corresponding outcome r, except for games 12 and 5. These happen to be precisely the two cases where $(T+S)/2$ can be larger than R. The payoff achieved is actually a convex combination of these two values.

Another interesting point concerns games 9 and 10. In these games, players have to coordinate their strategies, and this is actually achieved by YESTERDAY. However, the payoff is not necessarily the Pareto optimum R; rather, it is the maximin solution (which is P in case 9).

Fig. 6c displays the games for which YESTERDAY is efficient.

5 The strategy FARAWAY

A very large updating factor (an α close to 1) seems often inefficient. Small α's promise to do better. Numerical simulations show that we can approximate convex updating with very small α (infinitesimally slow updating) by the following continuous time dynamics. The aspiration levels of the two players at time t are denoted by $a_I(t)$ resp. $a_{II}(t)$. The two corresponding axes are divided by the payoff values R, S, T and P into five intervals each, and the (a_I, a_{II})-plane therefore into twenty-five regions. In each of these regions, the win-stay, lose-shift strategies of both players are well defined and lead to l.i.m. payoffs $P_I(a_I, a_{II})$ and $P_{II}(a_I, a_{II})$. If we assume now that the aspiration levels are steadily updated in direction of the l.i.m. payoff actually achieved, we obtain

$$\dot{a}_I = P_I(a_I, a_{II}) - a_I$$
$$\dot{a}_{II} = P_{II}(a_I, a_{II}) - a_{II}. \tag{5}$$

This yields a dynamics in the (a_I, a_{II})-plane which, as it describes the *trait values* of the two players, is somewhat related to adaptive dynamics (cf. Metz et al, 1996), although it describes individual learning rather than evolution.

We shall only sketch the mathematical basis of this model (see Posch and Sigmund, 1998, for details). The orbits of (5) are piecewise linear. The vector field can be discontinuous on the boundaries of the twenty-five regions. A standard way to handle such a differential equation is to transform it into a differential inclusion

$$(\dot{a}_I, \dot{a}_{II}) \in F(a_I, a_{II}) \tag{6}$$

where $F(a_I^*, a_{II}^*)$ is the smallest convex set containing all limit values of the right hand side of (5), for $(a_I, a_{II}) \to (a_I^*, a_{II}^*)$. Such a differential inclusion has at least one solution, see Filippov (1988).

It is easy to see that we can restrict our attention to the bounded intervals of the aspiration levels, namely m, b, and a, since all orbits end up there. The dynamics is symmetric in (a_I, a_{II}) and it suffices to study the regions where $a_I \le a_{II}$. Hence, we have to consider only six regions. In each rectangle, the payoff values (P_I, P_{II}) are constant. All orbits in that rectangle point towards (P_I, P_{II}).

Let us describe this in case 1, which includes the Prisoner's Dilemma. In the rectangle $m \times m$ (where both players use the modest strategy) the orbits point towards $(0, 0)$, which is the upper right corner. Similarly in the rectangle $b \times b$ (where both players use the balanced strategy, i.e. PAVLOV), all orbits point towards the upper right corner, namely $(1, 1)$. From the rectangle $a \times a$ the orbits point towards $(1/2, 1/2)$ and thus lead into $b \times b$ or $b \times a$. In $m \times b$ the orbits point towards $((1 + 2T)/5, (1 + 2S)/5)$ and hence lead either into $m \times m$ or $b \times b$. In $m \times a$ the orbits point to $(T/2, S/2)$ and thus lead into $b \times a$ or $m \times b$. Hence, eventually the dynamics leads to the rectangles $b \times b$ (and thus to (1,1)) or $b \times a$.

In $b \times a$ the orbits point towards $((1 + S)/3, (1 + T)/3)$. This is where things can get sticky and we have to distiguish four cases (see Fig. 5).

For $T < 2$ the orbits point downwards into the rectangle $b \times b$ such that (1,1) becomes an attractor. Thus, the aspirations ultimately converge to $(1, 1)$ and the players cooperate (case a).

If $T > 2$ the orbits starting at the lower edge of the rectangle $b \times a$ point upwards and thus (1,1) is no longer attracting. If additionally $S > -1$ the point $((1 + S)/3, (1 + T)/3)$ lies in the rectangle $b \times a$ and hence becomes an attractor. Thus, all orbits in $b \times a$ converge to $((1 + S)/3, (1 + T)/3)$ (case b). If instead $S < -1$, all orbits in $b \times a$ lead into the rectangle $m \times a$. The orbits in $m \times a$ in turn lead into $b \times a$. Thus, they converge to the boundary of the rectangles $m \times a$ and $b \times a$. There the dynamics can lead up or down: if $\frac{S-2}{T} > \frac{T-2}{S+1}$, there is an attractor point $(0, \frac{T(1+T)-S(1+S)}{3T-2-2S})$ on the boundary of $b \times a$ and $m \times a$ to which all orbits in $b \times a$ converge (case c). If $\frac{S-2}{T} < \frac{T-2}{S+1}$ the orbits at the boundary point downwards and will eventually reach the rectangle $b \times b$, where they converge to $(1, 1)$. Only an error pushs them back to $b \times a$ (case d). Hence, if the probability for errors is low, the players cooperate most of the time.

Thus, in cases (a) and (d) FARAWAY leads to cooperation. However, only in case (a) (1,1) is an attractor. Note that this is exactly the parameter range for which the Pavlov strategy is evolutionarily stable (see Leimar 1997). For the parameter ranges (b) and (c) there are two attracting fixed points for the aspiration levels where the agents switch actions every round, and thereby achieve a Pareto optimal outcome.

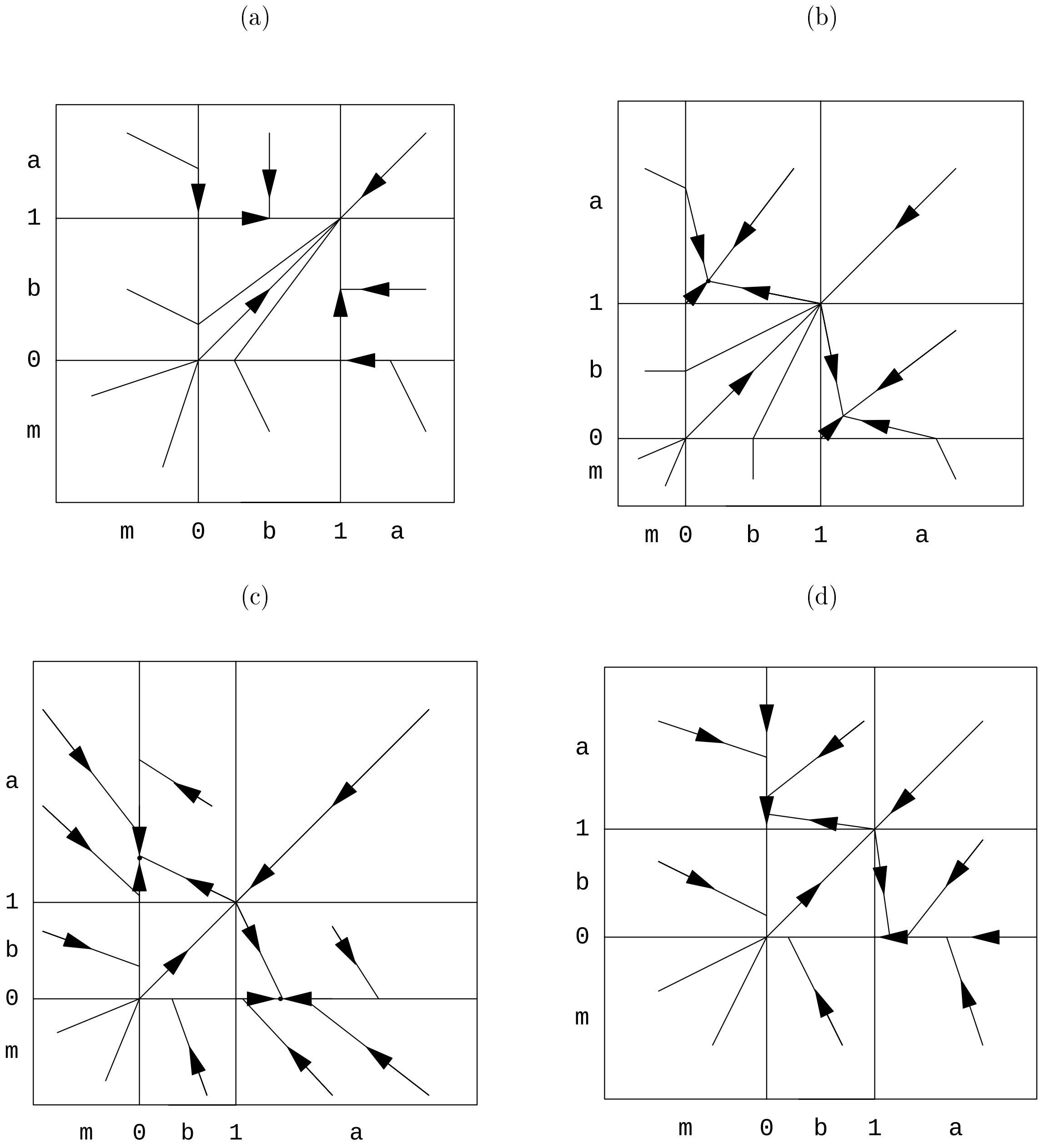


Figure 5: Different parameter values for the Prisoner's Dilemma lead to different dynamical outcomes for two players using FARAWAY as updating strategy. Case (a) $T < 2$, (b) $T > 2$ and $S > -1$, (c)$T > 2, S < -1$, and $(S-2)/T > (T-2)/(S+1)$, and (d) $T > 2, S < -1$, and $(S-2)/T < (T-2)/(S+1)$.

A similar analysis can be performed for Chicken (case 2). Again, slow updating leads to many different outcomes. Only for $S < 1/2$ and $T < 2$ will all orbits converge to $(1,1)$. For $S < 1/2$ and $T > 2$, the point $((1+S)/3,(1+T)/3)$ will be an attractor in $b \times a$ where the players switch actions every round; for $S > 1/2$ the points (T,S) and (S,T) are attractors (if $T < 2$ the point $(1,1)$ will also be an attractor). In these cases the initial aspiration levels determine which equilibrium gets selected.

In Fig. 6d the area where FARAWAY is efficient is shaded.

6 Discussion

The problem of adapting the aspiration level has intrigued psychologists and economists alike (see e.g. Thibaut and Kelley, 1959, Sauermann and Selten, 1962, Weber, 1976, Tietz, 1997). In order to obtain analytic results, we have concentrated on a particularly simple setting. Our agents are robots with minimal cognitive abilities. They use 'hard-wired' deterministic win-stay, lose-shift rules based on a specific aspiration level and on the payoff obtained in the previous round. These are severe restrictions, and we must discuss how much they affect the conclusions.

The lack of stochasticity in the switching rule is certainly a serious drawback. In more general win-stay, lose-shift rules, the propensity to switch from one option to the other is a function of the difference between aspiration level and payoff. It is reasonable to assume that this function is monotonically increasing, but our restriction to the step function $f(x) = 0$ for $x \leq 0$ and $f(x) = 1$ for $x > 0$ is certainly too narrow. Often, it pays to display a certain degree of frustration tolerance, i.e. not always to switch after an unsatisfactory outcome, but only with a certain probability. There is a huge literature on stochastic decision rules, we only refer to Bush and Mosteller (1951), Staddon (1983), Kraines and Kraines (1988), Stephens and Clements (1996), Wedekind and Milinski (1996), Posch (1997).

We note in this context that within the class of deterministic memory-one strategies (those for which p_k is 0 or 1, up to the error probability), the highest payoff achievable by the whole population, in case $R < (T+S)/2$, is $(2R+T+S)/4$ (see Nowak et al, 1995). Hence our win-stay, lose-shift rules can never be efficient in this case, whereas *stochastic* memory one strategies, for instance $(1/2, 0, 1, 1/2)$, can. We stress that for the Chicken game with $S > 1/2$, one of two FARAWAY players may end up with l.i.m. payoff T, the other with S. In this case the outcome is Pareto-optimal: but the two players will converge to different roles, one dominating the other. This is a good outcome for the entire population, since escalated contests are avoided.

We must also stress that in the games we have considered (both against nature and 2×2) the payoff was a deterministic function of the outcome. This excludes important situations like the binary choice model (a stochastic two-armed bandit whose left lever yields one dollar with probability p, and whose right lever yields one dollar with probability q). In that case, a deterministic win-stay, lose-shift rule leads to pulling the left lever with probability

$$\frac{1-q}{(1-p)+(1-q)}$$

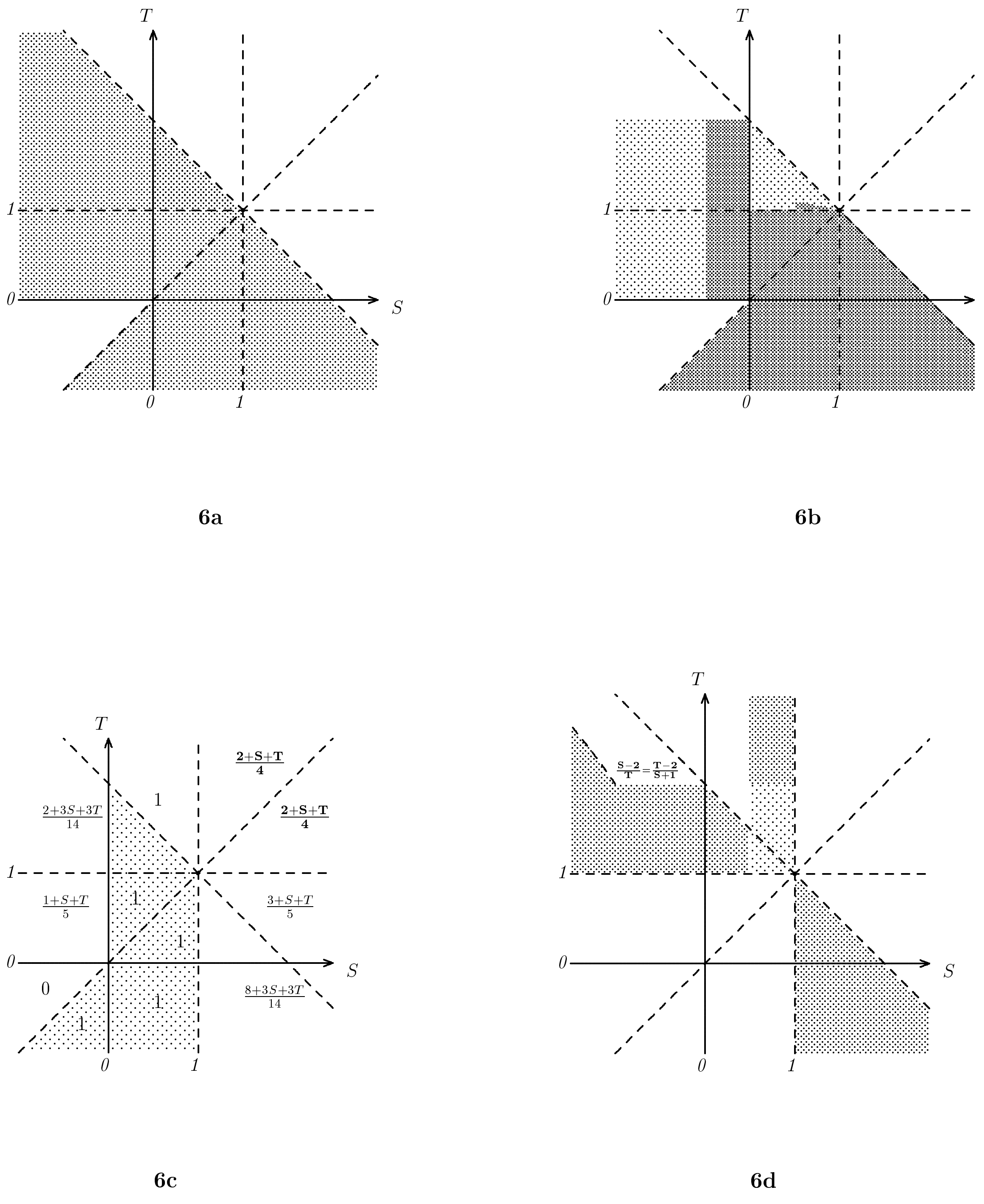


Figure 6: Efficiency. In (a) the shaded region describes the (S, T)-values for which *some* win-stay, lose-shift strategy is efficient. (b) In the dark region selection among the different win-stay, lose-shift strategies always leads to the fixation of an efficient strategy, whereas in the grey region some but not all initial conditions lead to such an outcome; (c) This displays the payoff values obtained by one YESTERDAY player against another. The shaded region describes the (S, T)-values for which the outcome is efficient; (d) In the dark region the adaption of the aspiration level by two FARAWAY players always leads to a Pareto-optimal outcome, whereas in the grey region some but not all initial conditions lead to it.

which is obviously not efficient. Interestingly, however, this comes very close to what untutored players actually do (Estes' law or the matching rule, see e.g. Colman, 1995), although these players do *not* adhere to a deterministic win-stay, lose-shift rule.

Furthermore, we have concentrated on *deterministic* updating. In general, updating strategies for repeated games are defined by algorithms specifying the aspiration level as a (possibly stochastic) function of the initial level and the payoffs experienced so far. We have only considered some extreme cases, which can be treated analytically. We believe nevertheless that our results also carry over to more realistic situations. In particular, whereas almost every updating procedure works well in deterministic games against nature, it offers *no general recipe* in dealing with stochastic effects or the interdependence of several players.

In many cases (such as in the minimal social situation, or the iterated Prisoner's Dilemma), having the right aspiration level leads to a good outcome. But finding this aspiration level through trial and error requires usually more insight into the structure of the interaction than can be achieved by updating strategies implemented by purely self-centered robots.

There is obviously no reason to assume that our parameterisation of the (S,T)-plane reflects in any way the relative importance of the 12 different game-theoretic situations. Some interactions, for instance Chicken games, are likely to occur in most social groups, since they reflect whether to escalate a conflict or not; on the other hand, it is hotly debated whether the Prisoner's Dilemma game is often found in real world situations. It seems plausible that for games which occur frequently, selection leads to the evolution of specific strategies (which may or may not be of win-stay, lose-shift type).

In the *Prisoner's Dilemma* game, for instance, YESTERDAY obtains against PAVLOV the same payoff as PAVLOV against itself, namely R (this can easily be checked by the same method as in section 4). Since YESTERDAY obtains against itself a lower payoff, it follows that PAVLOV dominates YESTERDAY. Having the 'right' aspiration level *a priori* turns out, not surprisingly, to be better than adapting it from round to round. This contest is unfair, of course, if we assume that there is no way of knowing in advance the payoff structure of the game encountered. But for particularly relevant games, knowledge could be hard-wired into an innate response.

We note that for many games, FARAWAY leads to outcomes where the agents switch their actions again and again as e.g. for the Prisoner's Dilemma in the cases b and c discussed above. This contrasts with the asymptotic results in Karandikar et al (1999). These authors study a related win-stay, lose-shift rule, which however is stochastic. Karandikar et al show that for all games with $T \geq S$ and $S < 0$ (i.e. games 1,8, and 4), the players will obtain payoff R most of the time, in the limiting case of infinitesimally slow updating. This is mainly due to the fact that players do not always shift after a failure. In that case all regions in the (a_I, a_{II}) plane where the aspiration levels change periodically are left in finite time. Thus, if trembles in the aspirations are very rare, the process stays most of the time at the vicinity of pure equilibria. However, simulations show that for small α the asymptotic results of Karandikar et al have little predictive power for the dynamics in the 'short' run, since aspirations can get stuck for hundreds of thousands of rounds close to equilibria where players switch actions again and again (see Posch, 1998).

We have emphasised the *efficiency* (or inefficiency) of learning rules. This issue is distinct from the evolutionary stability of such rules (see Maynard Smith, 1982, and for a notion more appropriate to repeated games, Leimar, 1997). Nevertheless our results make it seem doubtful that deterministic learning rules which are valid for a wide range of games will evolve. We believe that selection, in the realm of social interactions, favours

(a) the ability to recognise very specific types of interaction, and to adopt strategies which are hand-tailored for them, and

(b) the emergence of an understanding based on more than just registering the own payoff sequence.

Let us explain this last point. We have seen, for instance, that YESTERDAY excels only for a rather restricted range of games. This is in stark contrast with the strategy YESTERMAX, where players use as aspiration level in round n the maximum of their own *and their co-player's* payoff in round $n-1$. If both players use YESTERMAX, they always have the same aspiration level (clearly), and it can easily be shown that they always end up obtaining payoff R, except in case 9, a coordination game, in which case they obtain the maximin P. (Using the same method as in section 4, one can easily show that here are only two attractors for the transition chains, namely r and p, and that mistakes both in r and in p always lead to r, with the odd exception of case 9, when they always lead to p.) This is a remarkable performance, showing that, oddly enough, envy is often an efficient impulse. Indeed, YESTERMAX is just a trite instance of the principle of 'keeping up with Jones'. But clearly YESTERMAX requires a substantial cognitive ability; to monitor the co-player's payoff and to compare it with one's own implies a high degree of emphaty.

The view that even the simplest repeated games require a strategic understanding agrees well with the currently favoured opinion that the major selective stimulus for the evolution of intelligence comes, not from games against nature (like optimal foraging or anti-predator behaviour), but from the demands of social interactions (cf. Alexander, 1987, or de Waal, 1996).

References

Alexander, R.D. (1987) *The Biology of Moral Systems*, Aldine de Gruyter, New York.

Binmore, K.G. (1992) *Fun and Games: a Text on Game Theory*, Heath and Co, Lexington, Mass.

Bomze, I. (1995) Lotka-Volterra equation and replicator dynamics: new issues in clarification. *Biol. Cybernetics* **72**, 447-53.

Bush, R.R. and Mosteller, F. (1951) *Stochastic Models for Learning*, Wiley, New York.

Colman, A. (1995) *Game Theory and its Applications*, Butterworth-Heinemann, Oxford

Colman, A. and Stirk, J.A. (1998) Stackelberg reasoning in mixed-motive games: an experimental investigation, *J. Econ. Psychology* **19**, 279-93

de Waal, F. (1996) *Good natured: the origins of right and wrong in humans and other animals*, Harvard UP, Cambridge, Mass.

Filippov, A. (1988) *Differential Equations with Discontinuous Right-hand Sides*, Kluwer, Amsterdam

Harsanyi, J.C. and Selten, R. (1988) *A General Theory of Equilibrium Selection in Games*, M.I.T. Press, Cambridge Mass.

Hofbauer, J. and Sigmund, K. (1998), *Evolutionary Games and Population Dynamics*, Cambridge UP.

Hoppe, F. (1931) Erfolg und Misserfolg, *Psychologische Forschung* **14**, 1-62.

Karandikar, R., Mookherjee, D., Ray, D. and Vega-Redondo, F. (1999), Evolving aspirations and cooperation, to appear in *J. Economic Theory*

Kelley, H.H., Thibaut, J.W., Radloff, R. and Mundy, D. (1962) The development of cooperation in the minimal social situation, *Psychological Monographs* **76**, whole No. 19.

Kraines, D. and Kraines, V. (1988) Pavlov and the Prisoner's Dilemma, *Theory and Decision* **26**, 47-79.

Leimar, O. (1997), Repeated games: a state space approach, *Journ. Theor. Biol.*, **184**, 471-98.

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

Metz, J.A.J., Geritz, S.A., Meszena, G., Jacobs, F.J.A., and van Heerwarden, S.J. (1996), Adaptive dynamics: a geometrical study of the consequences of nearly faithful replication. In S.J. Van Strien and S.M.Verduyn Lunel (eds), *Stochastic and spatial structures of dynamical systems*, 183-231. Amsterdam, North Holland.

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

Nowak, M.A., Sigmund, K. and El-Sedy, E. (1995), Automata, repeated games and noise, *J. Math. Biol* **33**, 703-22.

Pichler, A. (1998) *Repeated Games with Failure in Implementation or Perception*, Master's Thesis, Univ. of Vienna

Posch, M. (1997) Cycling in a stochastic learning algorithm for normal form games, *J. Evolutionary Econ.* **7**, 193

Posch, M. (1998) Win-stay, lose-shift, an elementary behavioural principle for repeated games, submitted to *J. Theor. Biol.*

Posch, M. and Sigmund, K. (1998), An adaptive dynamics for the aspiration level, mimeo.

Radner, R. (1975) Satisficing, *J. Math. Econ.* **2**, 253-62

Rapoport, A. (1984) Game theory without rationality, *Behavioural and Brain Sciences* **7**, 114-15.

Rapoport, A., Guyer, M.J. and Gordon, D.G. (1976) *The* 2×2 *game*, Univ. of Michigan Press, Ann Arbor

Rescorla, R.A. and Wagner, A.R. (1972) A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement,. In *Classical Conditioning II*, eds A. Black and W.R. Prokasy, New York. Appleton-Century-Crofts.

Samuelson, L. (1997), *Evolutionary Games and Equilibrium Selection*, MIT Press, Cambridge, Mass.

Sauermann, H. und Selten, R. (1962) Anspruchsanpassungstheorie der Unternehmung, *Zeitschrift für die gesamte Staatswissenschaft* **118**, 577-597.

Simon, H.A. (1955) A behavioural model of rational choice, *Quart. J. Economics* **69**, 99-118

Simon, H.A. (1957) *Models of Man*, Wiley, New York

Simon, H.A. (1962) Theories of Decision Making in Economics and Behavioural Science, *The American Economic Review* **49**, 253-83.

Staddon, J.E.R. (1983) *Adaptive Behaviour and Learning*, New York, Cambridge UP.

Stephens, D.W. and Clements, K.C. (1998), Game theory and Learning, in *Game Theory and Animal Behaviour*, eds. L.A. Dugatkin and H.K. Reeve, Oxford UP, Oxford.

Tietz, R. (1997), Adaptation of aspiration levels – theory and experiment, in W. Albers et al, *Understanding Strategic Interaction. Essays in Honor of Reinhard Selten*, Springer, Berlin-Heidelberg

Thibaut, J.W. and Kelley, H.H. (1959) *The Social Psychology of Groups*, Wiley, New York

Thorndike, E.L. (1911) *Animal Intelligence*, New York, Hafner Publishing

van Damme, E. (1991) *Stability and Perfection of Nash Equilibria*, Springer, Berlin

Weber, H.-J. (1976) Theory of Adaptation of Aspiration Levels in a Bilateral Decision Setting, *Zeitschrift für die gesamte Staatswissenschaft* **132**, 582-91.

Wedekind, C. and Milinski, M. (1996), Human cooperation in the simultaneous and the alternating Prisoner's Dilemma: Pavlov versus generous tit-for-tat. *Proc. Nat. Acad. Sci. USA*, **93**, 2686-89.

Winter, J. (1971) Satisficing, Selection, and the Innovative Remnant, *Quart. J. Econ.* **85**, 237-61.