THE ROYAL SOCIETY

The efficiency of adapting aspiration levels

Martin Posch[1], **Alexander Pichler**[2] **and Karl Sigmund**[2,3]

[1] *Universität Wien, Institut für Medizinische Statistik, Schwarzspanierstrasse 17, 1090 Wien, Austria*
[2] *Universität Wien, Institut für Mathematik, Strudlhofgasse 4, 1090 Wien, Austria*
[3] *International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria*

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 pay-off 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 optimizing? Analytical results for 2 × 2 games are presented. They suggest that in a large variety of social interactions, self-centred rules (based uniquely on one's own pay-off) cannot suffice.

Keywords: games; satisficing; learning rules; natural selection

1. INTRODUCTION

In a game theory without rationality (see Rapoport 1984), players are not assumed to be able to fully understand the situation in which they are engaged. 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 of 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:

> '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.' (Thorndike 1911, p. 244)

The wide range of validity of this principle was soon recognized (see for example, Hoppe 1931; Rescorla & Wagner 1972). In the hands of H. 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 optimizing.

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 germ line has done the learning already and the result is encoded in the genome. 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 & Kraines 1988; Nowak & Sigmund 1993). PAVLOV is a win–stay, lose–shift rule with an aspiration level lying somewhere between the two highest and the two lowest pay-offs. Is there any reason to assume that selection or learning will adapt the aspiration level precisely to this interval? 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 pay-off which they obtain in each round.

In § 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 § 3–5, we turn to the simplest games, symmetrical 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 optimizing?

Received 16 March 1999 *Accepted* 8 April 1999

In this paper, our approach will be based on analytical methods. We restrict our attention to deterministic win–stay, lose–shift strategies based on switching to the alternative option if, and only if, the pay-off 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 pay-off 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 (1999).

2. GAMES AGAINST NATURE

Consider a two-armed bandit. Pulling one lever yields pay-off R, pulling the other yields pay-off P, with $P<R$. Let a be the aspiration level of a player. The player will repeat the former action if the pay-off 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 $\epsilon \rightarrow 0$ (that is, we compute the outcome for given $\epsilon>0$ and then let ϵ converge to zero). We assume that the game consists of a large number of rounds, and that the pay-off for the repeated game is given by the limit-in-the-mean (LIM) of the pay-off per round (i.e. $\lim(p_1+\ldots+p_N)/N$ for $N\rightarrow\infty$, where p_n is the pay-off in round n). If $a>R$, the player will switch after every round, and obtain as LIM pay-off $(R+P)/2$. If $a\leqslant P$, the player will always be satisfied, switch only by mistake, and then repeat the new action until the next mistake occurs. Again the LIM pay-off is $(R+P)/2$. For $P<a\leqslant R$, the player will always pull the R-lever, except by mistake; after an erroneous P, the player will switch back to R. The LIM pay-off 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 pay-off is converted into reproductive fitness, and that like begets like. This yields the replicator equation

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

where f_i is the LIM pay-off for strategy i and $\bar{f}=\sum x_k f_k$ is the average LIM pay-off in the population (see Hofbauer & Sigmund 1998). The dynamics on the corresponding unit simplex S_3 lead to the extinction of the 'wrong' aspiration levels: x_2 converges to unity. 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 pay-off obtained in the previous round. More precisely, if a_n is the aspiration level and p_n the pay-off in the nth 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, however, 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.

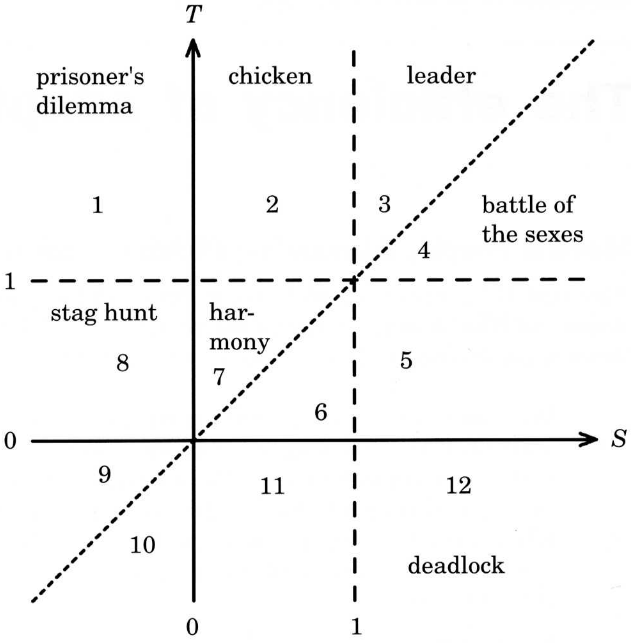

Figure 1. A partitioning of the (S, T) plane which displays the 12 symmetrical 2 × 2 games.

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 pay-off 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 moves towards $(R+P)/2$ and eventually enters the interval $]P,R]$. Once there, it converges towards R. The direction of change defines 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 symmetrical, i.e. that the two players are interchangeable. The pay-off matrix is

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

i.e. R is the pay-off 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 pay-off values are

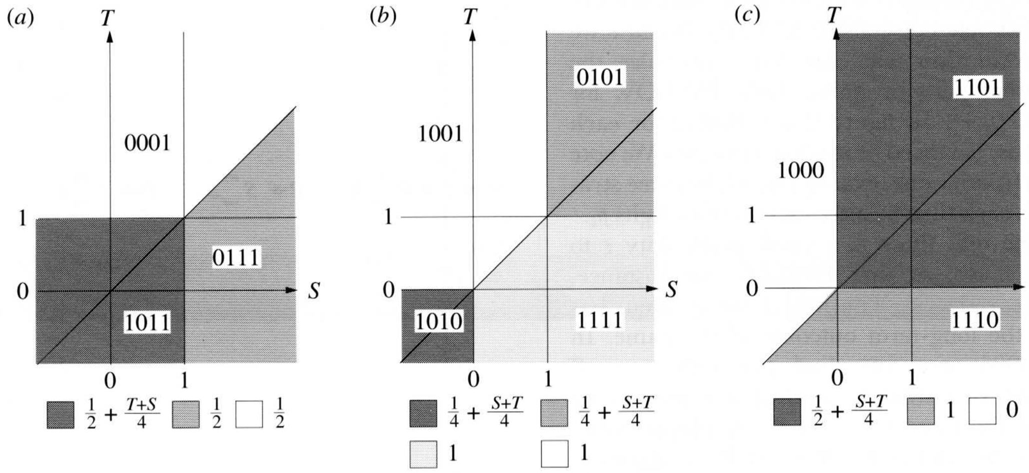

Figure 2. A description of the (*a*) ambitious, (*b*) balanced, (*c*) modest win–stay, lose–shift strategies corresponding to the different 2×2 games. The figure displays the corresponding (p_R, p_S, p_T, p_P) coding (see text), and the LIM pay-off for a player using this strategy against a player using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

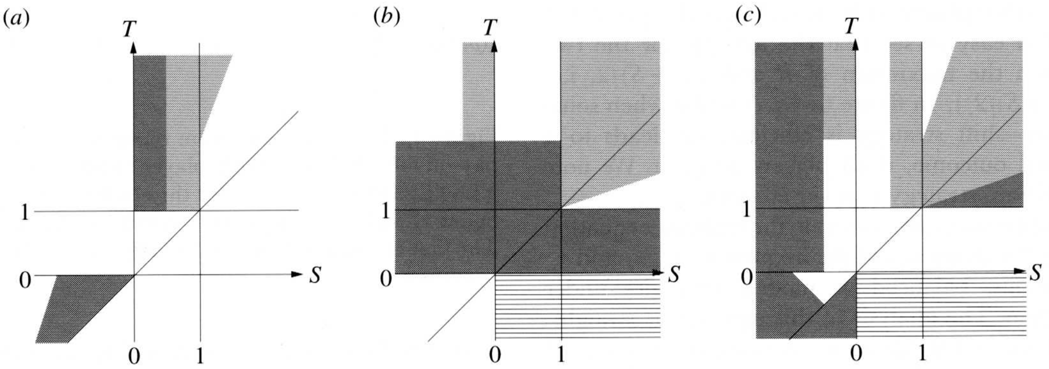

Figure 3. When does selection among the different win–stay, lose–shift strategies lead to the (*a*) ambitious (*b*) balanced or (*c*) modest strategy? The dark shading describes the (S, T) region where a monomorphic population using this strategy can emerge, and the light grey describes that region where selection leads to a stable bimorphic population, with a well-defined fraction using this strategy. In the striped region (cases 11 and 12 for 3*b*,*c*), 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.

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; 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 normalize the values such that $R = 1$ 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 figure 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 & Selten (1988), Van Damme (1991) and Samuelson (1997). In the games 1, 5, 6, 7, 11 and 12, both players have a dominant strategy (which yields a higher pay-off than the alternative, irrespective of the other player's choice); the games 6, 7, 8, 9, 10 and 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 optimizes the pay-off 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 & Stirk 1998.)

The four pay-off values divide the real line into five 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 of 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 of 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 $k \in \{R, S, T, P\}$. Because we consider only deterministic win–stay, lose–shift rules, the p_k values are either zero or unity. Thus PAVLOV, for instance, is $(1, 0, 0, 1)$. In figure 2 we display for each game the ambitious, balanced or modest strategies. We note that in crossing a frontier line, exactly one of the three strategies is modified, each time by altering two of its digits p_k.

We now assume that there is a small probability ϵ to misimplement 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 LIM pay-off obtained by using one strategy against a player using another, for the limiting case $\epsilon \rightarrow 0$. A player using PAVLOV obtains, for instance, $(R+S+P)/3$ against a player using the BULLY strategy $(0, 0, 0, 1)$, respectively pay-off R against another PAVLOV player (with our normalization, this becomes $(1+S)/3$ respectively 1).

An outcome is pareto-optimal if no other combination of strategies offers an improvement (i.e. a higher LIM pay-off) for either player without reducing the pay-off of the other. 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 figure 6*a* 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, because most of the 12 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: (i) Which aspiration levels get selected? (ii) When is the outcome efficient?

Concerning (i), 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 uses one and part another interval, with well-defined frequencies of the two types. In most cases, the attractor consists of one type only. In figure 3*a*–*c*, we have darkly shaded 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 (ii), we refer to figure 6*b*. We denote in dark grey the area in the (S,T)-plane where selection always leads to a pareto-optimal outcome, and in light 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.

We mention in this context that there exists, for this type of repeated games, a variant of evolutionary stability which is called limit evolutionarily stable strategy (ESS); see Leimar (1997). If everyone in the population is using such a strategy, then a deviation will be penalized at some state of the game (including states that are only reached by mistake). It is easy to check that an ambitious strategy is never a limit ESS. Among the modest strategies, GRIM $(1, 0, 0, 0)$ is always a limit ESS, and no other strategy is. Among the balanced strategies, $(1, 1, 1, 1)$ is always a limit ESS, PAVLOV $(1, 0, 0, 1)$ if and only if $T<2$, and the other two strategies are never a limit ESS. We emphasize that a limit ESS need not be an attractor for the replicator dynamics (for instance, GRIM can be invaded by PAVLOV if $T<2$ and $-1/2<S<0$).

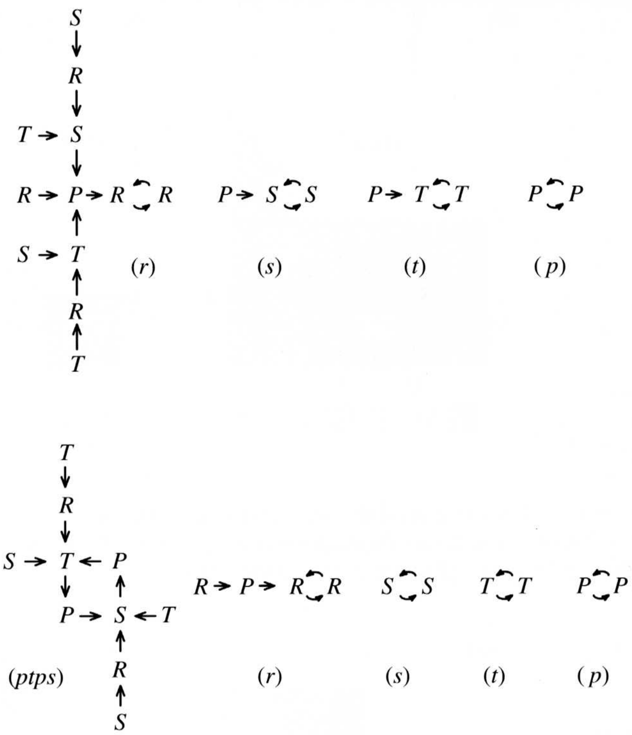

Figure 4. The transitions, from round to round, in the pay-off for a YESTERDAY player against another YESTERDAY player, (*a*) for the chicken game (case 2 in figure 1) and (*b*) the prisoner's dilemma (case 1 in figure 1). The first transition is assumed to be given as the initial condition.

4. THE STRATEGY YESTERDAY

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

Let us consider this for the chicken game (number 2 in our notation). Figure 4*a* shows how the game develops. For any initial condition, one of the four stationary states r, s, t or p is 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 $R \rightarrow R$ to $R \rightarrow T$ or to $R \rightarrow S$ (but not to $R \rightarrow P$). Because this leads, after three rounds, back to r, and because 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 $S \rightarrow R$ or to $S \rightarrow P$, 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 $P \rightarrow S$ or $P \rightarrow T$, 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 pay-off (defined as the LIM of the pay-off 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 the chicken game, 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 pay-off values. In addition to the four steady states r, s, t and p, we now find a cycle of period four, namely $T \rightarrow P \rightarrow S \rightarrow P \rightarrow T \rightarrow \ldots$, which we call *ptps*. In figure 4*b*, we display the transitions.

From the steady states r, s, t and p, every misimplementation leads to *ptps*. Errors occurring within the cycle have a more varied outcome. A misimplementation turns $S \rightarrow P$ either into $S \rightarrow S$ or into $S \rightarrow T$, and hence leads with equal probability either into the steady state s or back into *ptps*. Similarly, mistakes turn $T \rightarrow P$ with equal probability either into the steady state t or back into *ptps* again, whereas they turn $P \rightarrow S$ and $P \rightarrow T$ 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 pay-off per round is $(2R + 3S + 3T + 6P)/14$, which is considerably lower than the pareto-optimal outcome.

One can similarly compute the pay-off for YESTERDAY against itself for each of the remaining games. The result is shown in figure 6*c*. 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 pay-off 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 pay-off is not necessarily the pareto optimum R; rather, it is the maximin solution (which is P in case 9).

Figure 6*c* displays the games for which YESTERDAY is efficient.

5. THE STRATEGY FARAWAY

A very large updating factor (an α-value close to unity) often seems inefficient. Small α-values promise to do better. Numerical simulations show that we can approximate convex updating with very small α-values (infinitesimally slow updating) by the following continuous time dynamics. The aspiration levels of the two players at time t are denoted by $a_{\mathrm{I}}(t)$ respectively $a_{\mathrm{II}}(t)$. The two corresponding axes are divided by the pay-off values R, S, T and P into five intervals each, and the $(a_{\mathrm{I}}, a_{\mathrm{II}})$-plane therefore into 25 regions. In each of these regions, the win–stay, lose–shift strategies of both players are well defined and lead to LIM pay-offs $P_{\mathrm{I}}(a_{\mathrm{I}}, a_{\mathrm{II}})$ and $P_{\mathrm{II}}(a_{\mathrm{I}}, a_{\mathrm{II}})$. If we assume now that the aspiration levels are steadily updated in direction of the LIM pay-off actually achieved, we obtain

$$\begin{aligned} \dot{a}_{\mathrm{I}} &= P_{\mathrm{I}}(a_{\mathrm{I}}, a_{\mathrm{II}}) - a_{\mathrm{I}}, \\ \dot{a}_{\mathrm{II}} &= P_{\mathrm{II}}(a_{\mathrm{I}}, a_{\mathrm{II}}) - a_{\mathrm{II}}. \end{aligned} \tag{5}$$

This yields dynamics in the $(a_{\mathrm{I}}, a_{\mathrm{II}})$-plane which, as they describe the trait values of the two players, are somewhat related to adaptive dynamics (see Metz *et al.* (1996)), although they describe individual learning rather than evolution.

We shall only sketch the mathematical basis of this model (see Posch & Sigmund (1999) for details). The orbits of (5) are piecewise linear. The vector field can be discontinuous on the boundaries of the 25 regions. A standard way to handle such a differential equation is to transform it into a differential inclusion

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

where $F(a_{\mathrm{I}}^*, a_{\mathrm{II}}^*)$ is the smallest convex set containing all limit values of the right-hand side of equations (5), for $(a_{\mathrm{I}}, a_{\mathrm{II}}) \rightarrow (a_{\mathrm{I}}^*, a_{\mathrm{II}}^*)$. 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, because all orbits end up there. The dynamics are

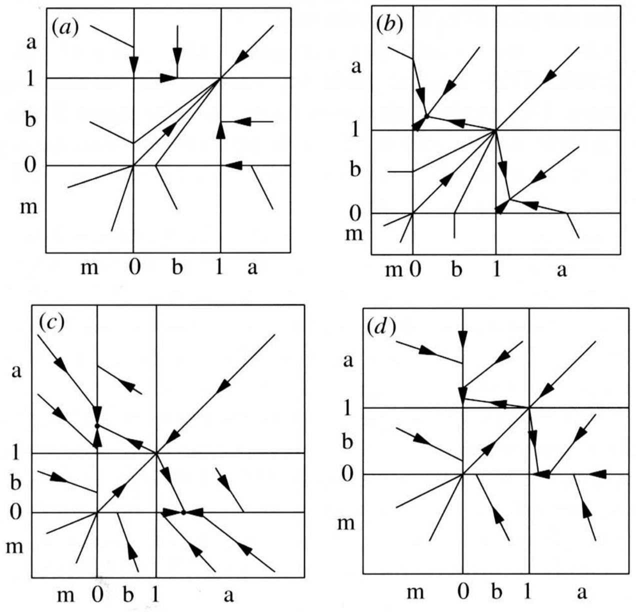

Figure 5. Different parameter values for the prisoner's dilemma lead to different dynamical outcomes for two players using FARAWAY as an updating strategy. (*a*) $T<2$, (*b*) $T>2$ and $S>-1$, (*c*) $T>2$, $S<-1$, and $(S-2)/T>(T-2)/(S+1)$, and (*d*) $T>2$, $S<-1$, and $(S-2)/T<(T-2)/(S+1)$. At the asymmetrical attractors for (*b*), one player is ambitious and experiences the pay-off sequence *tprtpr*. . . , the other is balanced and experiences *sprspr*. . . .

symmetrical in $(a_{\mathrm{I}}, a_{\mathrm{II}})$ and it suffices to study the regions where $a_{\mathrm{I}} \leqslant a_{\mathrm{II}}$. Hence, we have to consider only six regions. In each rectangle, the pay-off values $(P_{\mathrm{I}}, P_{\mathrm{II}})$ are constant. All orbits in that rectangle point towards $(P_{\mathrm{I}}, P_{\mathrm{II}})$.

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. There, random shocks will push them into the rectangle $b \times b$. Here, 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 more complicated and we have to distinguish four cases (see figure 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 (figure 5*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 $((1+S)/3, (1+T)/3)$ (figure 5*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 $(S-2)/T>(T-2)/(S+1)$, there is an attractor point $(0,[T(1+T)-S(1+S)]/(3T-2-2S))$ on the boundary of $b \times a$ and $m \times a$ to which all orbits in $b \times a$ converge (figure 5*c*). If $(S-2)/T<(T-2)/(S+1)$ the orbits at this boundary point downwards and will eventually reach the rectangle $b \times b$, where they converge to (1, 1). Only an error pushes them back to $b \times a$ (figure 5*d*). Hence, if the probability for errors is low, the players cooperate most of the time.

Thus, in figure 5*a,d* FARAWAY leads to co-operation. However, only in figure 5*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 in figure 5*b,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.

A similar analysis can be performed for the chicken game (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 figure 6*d* the area where FARAWAY is efficient is shaded.

6. DISCUSSION

The problem of adapting the aspiration level has intrigued psychologists and economists alike (see for example, Thibaut & Kelley (1959), Sauermann & Selten (1962), Weber (1976) and Tietz (1997)). In order to obtain analytical 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 pay-off 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 x between aspiration level and pay-off. It is reasonable to assume that this function is monotonically increasing, but our restriction to the step function $f(x)=0$ for $x \leqslant 0$ 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 refer only to Bush & Mosteller (1951), Staddon (1983), Kraines & Kraines (1988), Gilboa & Schmeidler (1995), Wedekind & Milinski (1996), Posch (1997), Fudenberg & Levine (1998), and Young (1999).

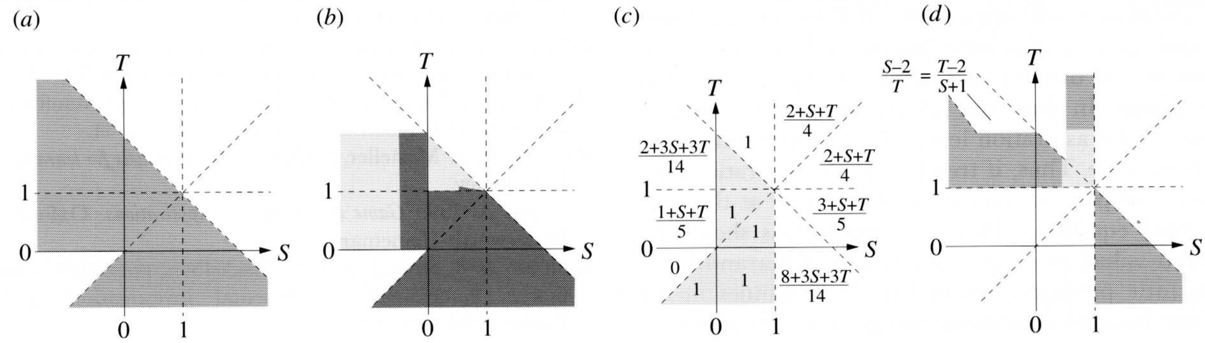

Figure 6. Efficiency. (*a*) The shaded region describes the (S, T) values for which some win–stay, lose–shift strategy is efficient. (*b*) In the dark region, selection among the different win–stay, lose–shift strategies always leads to the fixation of an efficient strategy, whereas in the light grey region some, but not all, initial conditions lead to such an outcome. (*c*) This displays the pay-off values obtained by one YESTERDAY player against another. The light grey 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 light grey region some, but not all, initial conditions lead to it. In some regions, the two FARAWAY players end up with different aspiration levels. The chicken game is the only one for which both YESTERDAY and FARAWAY are efficient.

We note in this context that within the class of deterministic memory-one strategies (those for which p_k is zero or unity, up to the error probability), the highest pay-off 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 out of two FARAWAY players may end up with LIM pay-off 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, because escalated contests are avoided.

We must also stress that in the games we have considered (both against nature and 2 × 2) the pay-off was a deterministic function of the outcome. This excludes important situations such as 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

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

which is obviously not efficient. Interestingly, however, this comes very close to what untutored players actually do (Estes' law or the matching rule; see, for example, 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 pay-offs experienced so far (see Karandikar *et al.* (1998), Kim (1999) and Pazgal (1999)). 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 usually requires more insight into the structure of the interaction than can be achieved by updating strategies implemented by purely self-centred robots.

There is obviously no reason to assume that our parameterization 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, because they reflect whether to escalate a conflict or not; in contrast, 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 the win–stay, lose–shift type).

In the prisoner's dilemma game, for instance, YESTERDAY obtains against PAVLOV the same pay-off as PAVLOV against itself, namely R (this can easily be checked by the same method as in §4). Because YESTERDAY obtains against itself a lower pay-off, 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 pay-off 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, for example, for the prisoner's dilemma in the cases (*b*) and (*c*) discussed above (see Figure 5). This contrasts with the asymptotic results of Karandikar *et al.* (1998). These authors study a related win–stay, lose–shift rule, which however is stochastic. Karandikar *et al.* show that for all games with $T \geqslant S$ and $S < 0$ (i.e. games 1, 8 and 4), the players will obtain pay-off 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 $(a_{\mathrm{I}}, a_{\mathrm{II}})$-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-values the asymptotic results of Karandikar *et al.* have little predictive power for the dynamics in the 'short' run, because 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 emphasized 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 (i) the ability to recognize very specific types of interaction, and to adopt strategies which are hand-tailored for them, and (ii) the emergence of an understanding based on more than just registering the own pay-off 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 pay-off 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 pay-off R, except in case 9, a coordination game, in which case they obtain the maximin P. (Using the same method as in §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 curious 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 the Jones'. But clearly YESTERMAX requires a substantial cognitive ability; to monitor the co-player's pay-off and to compare it with one's own implies a high degree of empathy.

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; see Alexander (1987) or de Waal (1996).

This project has been supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung P11144.

REFERENCES

Alexander, R. D. 1987 *The biology of moral systems*. New York: Aldine de Gruyter.

Binmore, K. G. 1992 *Fun and games: a text on game theory*. Lexington, MA: Heath & Co.

Bomze, I. 1995 Lotka–Volterra equation and replicator dynamics: new issues in clarification. *Biol. Cyber.* **72**, 447–453.

Bögers, T. & Sarin, R. 1997 Learning through reinforcement and replicator dynamics. *J. Econ. Theory* **77**, 1–14.

Bush, R. R. & Mosteller, F. 1951 *Stochastic models for learning*. New York: Wiley.

Colman, A. 1995 *Game theory and its applications*. Oxford, UK: Butterworth-Heinemann.

Colman, A. & Stirk, J. A. 1998 Stackelberg reasoning in mixed-motive games: an experimental investigation. *J. Econ. Psychology* **19**, 279–293.

de Waal, F. 1996 *Good natured: the origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.

Filippov, A. 1988 *Differential equations with discontinuous right-hand sides*. Amsterdam: Kluwer.

Fudenberg, D. & Levine, D. K. 1998 *Theory of learning in games*. Cambridge, MA: MIT Press.

Gilboa, I. & Schmeidler, D. 1995 Case-based decision theory. *Q. J. Econ.* **110**, 605–639.

Harsanyi, J. C. & Selten, R. 1988 *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press.

Hofbauer, J. & Sigmund, K. 1998 *Evolutionary games and population dynamics*. Cambridge University Press.

Hoppe, F. 1931 Erfolg und Misserfolg. *Psychol. Forsch.* **14**, 1–62.

Karandikar, R., Mookherjee, D., Ray, D. & Vega-Redondo, F. 1998 Evolving aspirations and cooperation. *J. Econ. Theory* **80**, 292–331.

Kelley, H. H., Thibaut, J. W., Radloff, R. & Mundy, D. 1962 The development of cooperation in the minimal social situation. *Psychol. Monogr.* **76**, no. 19.

Kim, Y. 1999 Satisficing and optimality in 2 × 2 common interest games. *Econ. Theory* **13**, 365–375.

Kraines, D. & Kraines, V. 1988 Pavlov and the Prisoner's dilemma. *Theor. Decision* **26**, 47–79.

Leimar, O. 1997 Repeated games: a state space approach. *J. Theor. Biol.* **184**, 471–498.

Maynard Smith, J. 1982 *Evolution and the theory of games*. Cambridge University Press.

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

Nowak, M. A. & Sigmund, K. 1993 Win–stay, lose–shift outperforms tit for tat. *Nature* **364**, 56–58.

Nowak, M. A., Sigmund, K. & El-Sedy, E. 1995 Automata, repeated games and noise. *J. Math. Biol.* **33**, 703–722.

Pazgal, A. 1999 Satisficing leads to cooperation in mutual interest games. *Int. J. Game Theory* **26**, 439–453.

Pichler, A. 1998 *Repeated games with failure in implementation or perception*. MSc thesis, University of Vienna.

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

Posch, M. 1999 Win–stay, lose–shift strategies for repeated games—memory length, aspiration levels, and noise. *J. Theor. Biol.* **198**, 183–195.

Posch, M. & Sigmund, K. 1999 An adaptive dynamics for the aspiration level. Unpublished manuscript, University of Vienna.

Radner, R. 1975 Satisficing. *J. Math. Econ.* **2**, 253–262.

Rapoport, A. 1984 Game theory without rationality. *Behav. Brain Sci.* **7**, 114–115.

Rapoport, A., Guyer, M. J. & Gordon, D. G. 1976 *The 2 × 2 game*. Ann Arbor, MI: University of Michigan Press.

Rescorla, R. A. & Wagner, A. R. 1972 A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In *Classical conditioning*, vol. 2 (ed. A. Black & W. R. Prokasy). New York: Appleton–Century–Crofts.

Samuelson, L. 1997 *Evolutionary games and equilibrium selection.* Cambridge, MA: MIT Press.

Sauermann, H. & Selten, R. 1962 Anspruchsanpassungstheorie der Unternehmung. *Z. Ges. Staatswiss.* **118**, 577–597.

Simon, H. A. 1955 A behavioural model of rational choice. *Q. J. Econ.* **69**, 99–118.

Simon, H. A. 1957 *Models of man.* New York: Wiley.

Simon, H. A. 1962 Theories of decision making in economics and behavioural science. *Am. Econ. Rev.* **49**, 253–283.

Staddon, J. E. R. 1983 *Adaptive behaviour and learning.* New York: Cambridge University Press.

Thibaut, J. W. & Kelley, H. H. 1959 *The social psychology of groups.* New York: Wiley.

Thorndike, E. L. 1911 *Animal intelligence.* New York: Macmillan.

Tietz, R. 1997 Adaptation of aspiration levels—theory and experiment. In *Understanding strategic interaction. Essays in honor of Reinhard Selten* (ed. W. Albers, W. Güth, P. Hammerstein, B. Moldovanu & E. van Damme), pp. 345–364. Berlin and Heidelberg: Springer.

Van Damme, E. 1991 *Stability and perfection of Nash equilibria.* Berlin: Springer.

Weber, H.-J. 1976 Theory of adaptation of aspiration levels in a bilateral decision setting. *Z. Ges. Staatswiss.* **132**, 582–591.

Wedekind, C. & Milinski, M. 1996 Human cooperation in the simultaneous and the alternating Prisoner's dilemma: Pavlov versus generous tit-for-tat. *Proc. Natl Acad. Sci. USA* **93**, 2686–2689.

Winter, J. 1971 Satisficing, selection, and the innovative remnant. *Q. J. Econ.* **85**, 237–261.

Young, H. P. 1999 *Individual strategy and social structure.* Princeton University Press.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.