WORKING   PAPERS

Martin Posch

WIN STAY, LOSE SHIFT OR IMITATION – ONLY THE CHOICE OF
PEERS COUNTS

July 2001

Working Paper No: 0109

DEPARTMENT OF ECONOMICS

UNIVERSITY   OF   VIENNA

WIN STAY, LOSE SHIFT OR IMITATION – ONLY THE CHOICE OF PEERS COUNTS

Martin Posch

Department of Medical Statistics, University of Vienna, Schwarzspanierstr. 17, A-1180 Vienna, Austria, E-mail: Martin.Posch@univie.ac.at

Abstract

Win Stay, Lose Shift as well as imitation strategies for iterated games rely on an aspiration level. With both learning rules a move is repeated unless the pay-off fell short of the aspiration level. I investigate social adaptation mechanisms for the aspiration level and their impact on the efficiency of learning in a large population of agents that repeatedly play one round of a symmetric 2×2 game against randomly chosen opponents. It turns out that if the aspiration level is given by the last payoff of the current opponent the population receives the maximal symmetric payoff of the game in the long run. If the aspiration level is determined by independently chosen agents the outcome is related to the evolutionarily stable strategies. This holds for win stay, lose shift as well as for imitation based learning. These results suggest that the choice of peers can be crucial for the efficiency of learning.

Keywords: Games, Learning Rules, Imitation, Aspiration Level, Aspiration Adaptation, Cooperation, Prisoner's Dilemma

1. Introduction

Satisfaction with an achievement is not independent of the achievements of our peers. What counts is where we stand in comparison to others and not the absolute value of a success. The social psychologists Thibaut and Kelley (1959) already noted that aspiration levels, defined as the lowest satisfying outcomes, are not constant but subject to learning. The aspiration level (which they call comparison level) will adapt to salient outcomes a person is confronted with:

> *With variations in the particular instigations or reminders that are present (e.g. the particular persons brought to mind for comparison), there appear to be corresponding fluctuations in the comparison level.*
>
> Thibaut and Kelley (1959, p. 98)

Which are the persons brought to mind for comparison? These can be randomly chosen agents or somehow salient individuals as e.g. persons we interact with. Consider the situation where agents are in every round newly randomly matched in pairs to play one round of a symmetric 2×2 game. The agents update their actions if their last payoff fell short of their aspiration level. I investigate two updating mechanisms: a simple win stay, lose shift behavior and a more sophisticated imitation rule.

In this paper I show that the choice of peers for comparison is crucial for the efficiency of learning: If the aspiration level is given by the last payoff of the current opponent the population will in the limit receive the maximal symmetric payoff of the game. If, in contrast, the aspiration level is determined by the last payoff of independently randomly chosen agents, the limit payoff is related to the

evolutionarily stable strategies of the one stage game. These findings hold for the simple win stay, lose shift rule as well as for imitation learning.

As a special case I study the Prisoner's Dilemma. Here all players defect in the limit if the aspiration level is determined by the payoff of independently chosen agents but cooperate if the aspiration level is given by the payoff of the current opponent. But is this cooperation stable if there are also players present that always defect? It turns out that some cooperation persists as long as these defectors are not in the majority. This holds for the win stay, lose shift as well as for the imitation rule. The latter however is able to keep a higher level of cooperation. To investigate the robustness of both aspiration based learning rules I study a selection dynamics, assuming that rules leading to a higher payoff spread faster in the population than those with a lower payoff. For a large class of Prisoner's Dilemma games the aspiration based learning rules cannot be invaded by always defecting players. For some Prisoner's Dilemma games the imitating rule (but not the win stay, lose shift rule) can even invade a population of always defecting players.

Recently several authors investigated learning dynamics based on endogenous aspiration levels for repeated games, where the same two players repeatedly play a 2×2 game. They assumed either that the aspiration level of a player depends on the payoffs he himself received in the past (Karandikar, Mookherjee, Ray & Vega-Redondo 1998, Posch 1999, Posch, Pichler & Sigmund 1999, Börgers & Sarin 2000) or on an average population payoff (Dixon 2000, Oechssler 2000).

The approach closest to this paper is probably by Palomino & Vega-Redondo (1999) who also consider a setting where the players are newly matched in every round. The main difference to my approach is that Palomino and Vega-Redondo assume that the aspiration levels of the players depend on the average population payoff, which is assumed to be common knowledge. In contrast, in the model presented here also the formation of the aspiration level is based solely on "local" information. Interestingly, in the setup of Palomino and Vega-Redondo only partial cooperation emerges for the Prisoner's Dilemma.

Closely related to the discrete imitation model studied in this paper, where independently, randomly chosen agents that achieved a higher payoff are imitated, are the continuous imitation dynamics (Schlag 1998, Hofbauer 1995, Björnerstedt & Weibull 1996, Weibull 1995, Hofbauer & Sigmund 1998). Schlag (1998) showed that imitating the better with independent matching is not optimal in the sense that there are situations where the expected payoff of a player may decrease when following this rule. He proposes a stochastic imitation rule where the probability to imitate a player is proportional to the payoff difference. Hofbauer and Schlag (2000) show that for cyclic games this imitation rule may result in cycling learning dynamics. Another related approach is by Vega-Redondo (1997) who considers an N-player oligopoly model where the same N players repeatedly interact. He proves that imitating strategies of more successful opponents leads to convergence to the Walrasian equilibrium.

The paper is organized as follows. In Section 2 the learning rules and matching schemes are introduced. In Section 3 I analyze the learning dynamics for homogeneous populations of win stay, lose shift and imitating players that update their actions if they achieved less than their current opponent did. In Section 4 the independent matching rule is studied where the players compare their payoffs to independently chosen agents. In Section 5 I consider mixed populations for the Prisoner's Dilemma where also always defecting players are present and study the evolution of these strategies under selection. Finally, in the discussion I explore generalizations to symmetric games with more than two actions and players.

2. THE SHIFT IF BETTER AND IMITATE IF BETTER STRATEGIES FOR SYMMETRIC 2×2 GAMES

Consider a large population of players that are at each round randomly matched in pairs to play one round of a symmetric 2×2 game. In the first round the players choose an action at random such that each action is chosen with positive probability. In all following rounds they update their actions using the aspiration based learning rules described below.

2.1. **Independent Matching and Competitor Matching.** In the *independent matching* case (Schlag 1998) every player first randomly samples an agent and updates his own action if his own last payoff is smaller than the opponent's last payoff. Next, each player independently samples another agent and plays one round of the game using the current action. In the *competitor matching* case only one agent is sampled in each round. The players update their actions if their

last payoff falls short of the last payoff of their opponent and then play against the same opponent one round of the game.

2.2. **Shift if Better and Imitate if Better Updating.** I consider two updating rules: one is shifting, i.e. the players just switch to the other action. The second is imitation, i.e. the players copy the strategy of the player they sampled to compare their payoff to. Obviously, imitation is a more sophisticated strategy that requires information about the action the opponent used in the last round. I call the first learning rule *shift if better* (SiB) and the second *imitate if better* (IiB). The latter term is also due to Schlag (1998).

2.3. **Symmetric 2×2 Games.** Denoting the two actions by **C** and **D** the payoff matrix for the game is given by

$$
\begin{array}{c|cc}
 & \mathbf{C} & \mathbf{D} \\
\hline
\mathbf{C} & R & S \\
\mathbf{D} & T & P
\end{array},
$$

i.e. playing **C** against **C** leads to the payoff R, **D** against **C** to T etc. The learning rules SiB and IiB depend only on the rank ordering of the payoffs. Considering only the generic situation where all payoff values are distinct there are 24 different rank orderings and the only symmetric payoffs are R and P. If we assume additionally that $R > P$ (which is no restriction of generality: otherwise just re-label **C** and **D**), they are reduced to 12 rank orderings (Nowak, Sigmund & El-Sedy 1995, Colman & Stirk 1998). I normalize the games by setting $R = 1$ and $P = 0$. Then each rank ordering corresponds to a region in the S-T plane as sketched in Figure 1 (cf. Posch *et al.* (1999)). The 12 classes of games correspond to very different strategic situations. A widely used model for altruism is e.g. the

Prisoner's Dilemma game, where $T > 1 > 0 > S$. Here the two options **C** and **D** are interpreted as *cooperation* and *defection*.
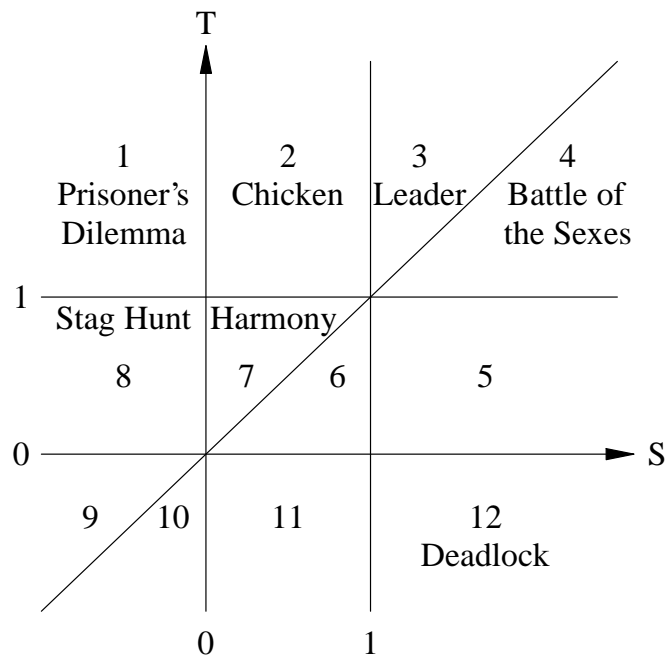

FIGURE 1. The 12 games in the S-T plane.

2.4. **An Example.** Let us consider the SiB and IiB strategies with competitor matching for the Prisoner's Dilemma game. Whenever a SiB player got a lower payoff than his opponent he will switch to the other action. For an IiB player this only holds if the agents additionally played different actions in the former round. E.g., if a SiB player received payoff S and his opponent 1, he will defect in the next round while an IiB player will cooperate; if a SiB or IiB player received payoff 0 and his opponent 1, they will both cooperate in the next round. Table 1 gives the full definition of the SiB and IiB strategies for the Prisoner's Dilemma game. For all other symmetric 2×2 games the strategies can be defined analogously.

	SiB						IiB			
	T	1	0	S			T	1	0	S
T	**D**	**D**	**D**	**D**		T	**D**	**D**	**D**	**D**
1	D	**C**	**C**	**C**		1	D	**C**	**C**	**C**
0	C	C	**D**	**D**		0	$\underline{D}$	C	**D**	**D**
S	D	D	D	**C**		S	D	$\underline{C}$	D	**C**

TABLE 1. The SiB and IiB strategies with competitor matching for the Prisoner's Dilemma. The table gives the action of a SiB (resp. IiB) player that obtained in the last round the row-payoff and his opponent the column-payoff. The grayed entries denote cases where the row player is unsatisfied and updates his action. Note that only the underlined actions in the matrix for IiB players differ from the ones for SiB players.

3. COMPETITOR MATCHING: THE SIB AND THE IIB STRATEGIES IN HOMOGENEOUS POPULATIONS

3.1. **The SiB strategy.** Let r_n, s_n, t_n, p_n denote the relative frequencies of SiB players receiving payoffs $1, S, T,$ and 0 in round n. For a population consisting entirely of SiB players the relations $t_n = s_n$ and $r_n = 1 - p_n - 2\,s_n$ hold for all n. Thus, the dynamics of the four frequencies can be reduced to (s_n, p_n). Note that at most one of the two matched players switches his action. They will play two different actions only if they played the same action in the past round but got different payoffs. This occurs if a player with payoff 1 meets a player with payoff S or if a player with payoff 0 meets a player with payoff T. Thus, for all games the frequency of players receiving S in the next round is given by

$$s_{n+1} = s_n\,r_n + t_n\,p_n$$

$$(1) \qquad\qquad = s_n(1 - 2\,s_n).$$

Hence, for all initial conditions $\lim_{n\to\infty} s_n = 0$. Next, consider the dynamics of p_n for the case of the Prisoner's Dilemma, where $T > 1 > 0 > S$. A pair of SiB players receives the payoff 1 if the players got the payoffs $(1,1), (S,S), (1,0)$ or $(0,1)$ in the last round. As noted above, they receive S (resp. T) after the payoffs $(1,S), (0,T)$ (resp. $(S,1), (T,0)$). It follows that they receive 0 after all other possible outcomes. Thus, the law of motion is given by

$$p_{n+1} = p_n^2 + 2r_n t_n + 2p_n s_n + 2s_n t_n + t_n^2$$

$$(2) \qquad\qquad = p_n^2 + 2s_n - s_n^2.$$

I showed above that $\lim_{n\to\infty} s_n = 0$. Assuming for the moment that $s_n = 0$ for large n it easily follows that $\lim_{n\to\infty} p_n = 0$ as long as $p_1 < 1$ (see Appendix A for a rigorous proof without the above assumption). Thus, for the Prisoner's Dilemma game it follows that $\lim_{n\to\infty} r_n = 1$

Also for all other games pairs of players that received payoffs $(1,1), (S,S), (1,0), (0,1)$ will receive 1 in the next round. Players with payoffs $(1,S), (0,T), (S,1), (T,0)$ receive S or T. Thus, for all symmetric 2×2 games the fraction p_{n+1} of players receiving 0 in round $n+1$ can be estimated from above:

$$p_{n+1} \leq p_n^2 + 2r_n t_n + 2p_n s_n + 2s_n t_n + t_n^2$$

$$= p_n^2 + 2s_n - s_n^2 =: f(p_n, s_n) \quad \text{(say)}.$$

From above it follows that for the recursion $q_1 = p_1$, $q_n = f(q_n, s_n)$ (which corresponds to the Prisoner's Dilemma) we have $\lim_{n\to\infty} q_n = 0$. Since $f(x,y)$ is monotonically increasing in x on $[0,1]$ we get $p_n \leq q_n$ for all $n > 1$ and thus $\lim_{n\to\infty} p_n = 0$. Hence, if $p_1 < 1$ then $\lim_{n\to\infty} r_n = 1$. Since we assume that in

the first round both actions are played with positive probability this gives the following result:

Proposition 1. *For all symmetric* 2×2 *games a population of SiB players receives in the limit the maximal symmetric payoff.*

In the case of the Prisoner's Dilemma cooperation emerges. However, as follows from the proof of Proposition 1, the Prisoner's Dilemma is somehow the "worst case": here convergence is the slowest.

3.2. **The IiB strategy.** Let again r_n, s_n, t_n, and p_n denote the relative frequencies of IiB players receiving payoffs $1, S, T$, and 0 in round n. If two IiB players are matched they both play the same action: if their payoffs were different, the player with the lower payoff imitates the other, if their payoffs were equal they already played the same action. Thus, for a population consisting entirely of IiB players independent of the initial moves for all $n > 1$ we have $s_n = t_n = 0$ and $p_n = 1 - r_n$. A pair of IiB players gets 1 in the next round if the players received the payoffs $(1,1), (1,0)$, or $(0,1)$ in the last round. Only if both players received 0 they also receive 0 in the next round. Hence, the law of motion is given by

$$p_{n+1} = p_n^2.$$

A straightforward calculation gives the following result:

Proposition 2. *For all symmetric* 2×2 *games a population of IiB players receives in the limit the maximal symmetric payoff.*

Thus, the limit outcome is the same as for the SiB rule. However, the convergence rates differ dramatically: for the SiB rule the convergence rate is slower than geometric while for the IiB rule it is faster than geometric.

4. Independent Matching: Keeping Up with Anybody

For the behavior of both the SiB and the IiB strategy, it was essential to assume that they play the game with the same agent they compare their payoff to. Assuming independent matching the learning dynamics changes completely. I illustrate the long run behavior of a homogeneous populations of SiB resp. IiB players using this matching scheme for the 12 symmetric 2×2 games.

Let c_n denote the fraction of players that played the action $\mathbf{C}$ in the last round. Then $r_n = c_n^2$ receive payoff 1, $s_n = c_n (1 - c_n)$ payoff S, and so on. Thus, it suffices to study the dynamics of c_n. This will depend on the payoff ordering. Consider e.g. the Prisoner's Dilemma. Here, for the SiB rule the law of motion is given by

$$c_{n+1} = c_n \left(1 - 2 c_n^2 + 2 c_n^3\right).$$

Thus, $\lim_{n \to \infty} c_n = 0$. Similarly for the IiB rule the law of motion is given by

$$c_{n+1} = c_n^2 \left(3 - 4 c_n + 2 c_n^2\right).$$

Again, $\lim_{n \to \infty} c_n = 0$. However, as for competitor matching convergence for the IiB rule is faster than geometric while for the SiB rule it is slower than geometric. Thus, with both, the SiB and the IiB rule, all players will use strategy $\mathbf{D}$ in the limit. As shown in Section 3, if they use the same person for the updating of

their strategy as for playing, both rules use **C** in the limit. Hence, the matching rule is crucial for the outcome.

The results for all 12 games are given in Table 2. The limit outcome of the SiB and IiB rule with independent matching is closely related to the pure and mixed *evolutionarily stable strategies* of the game. A game has a *pure* evolutionarily stable strategy if either a population playing always **C** or a population playing always **D** cannot be invaded by any other strategy that occurs with small frequency. For games with a *mixed* evolutionarily stable strategy a stochastic strategy that plays the actions **C** and **D** with certain probabilities (which depend on the actual payoff values) cannot be invaded.

As listed in Table 2 for the IiB rule all pure evolutionarily stable strategies are also limit outcomes for the IiB rule. For the SiB rule this relation holds with three exceptions: for the games 8 and 9 the frequency of playing **C** always stays equal to the initial frequency; in game 10, where **C** and **D** are evolutionarily stable, the population plays always **C** in the limit, regardless of the initial conditions. Thus, here an equilibrium selection takes place: the game dynamics selects the so called Stackelberg equilibrium (The Stackelberg solution is the strategy which optimises the payoff under the assumption that the reply is optimal from the co-player's view, see Colman and Stirk, 1998).

For games that have a mixed evolutionarily stable strategy populations of IiB and populations of SiB players will play a mixed strategy in the limit. As the dynamics for both the IiB and SiB rule depend only on the payoff ordering also

this mixed strategy depends only on the payoff ordering and will in general not coincide with the evolutionarily stable strategy.

Games	SiB	IiB	ESS
1	0	0	0
2,3	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{S}{S+T-1}$
4	$\frac{1}{2}(-1+\sqrt{5})$	$\frac{1}{\sqrt{2}}$	$\frac{S}{S+T-1}$
5,6,7,11,12	1	1	1
8,9	fixed point line	$0 \leftarrow \frac{1}{2} \rightarrow 1$	$0,1$
10	1	$0 \leftarrow 1 - \frac{1}{\sqrt{2}} \rightarrow 1$	$0,1$

TABLE 2. The limits of the fraction of players playing strategy C for the SiB and IiB rule if they independently match an agent for the updating of the strategy and another agent to play the next round. The last column lists the pure and mixed evolutionary stable strategies of the respective games. The notation $0 \leftarrow x \rightarrow 1$ denotes a situation where the limit for initial values c_1 greater x is 1 and for initial values smaller than x is 0.

5. COMPETITOR MATCHING: KEEPING UP WITH JAILBIRDS

In this section I investigate the robustness of the SiB and IiB rule for the Prisoner's Dilemma game with competitor matching. From above it follows that both rules, when applied in a homogeneous environment, lead to a cooperating population, regardless of the initial moves. How will these strategies perform in a rougher environment where also players that always defect are present? I first study the game dynamics in a population where a fraction δ plays the strategy All D, i.e. these agents always play **D**. Then I consider a selection dynamics for populations of All D and SiB as well as All D and IiB players. Let $t_n^{\mathbf{D}}, p_n^{\mathbf{D}}$ denote the frequency of players that use the strategy All D and experience the outcomes T and 0. Note that these players never experience 1 or S.

5.1. **The SiB strategy.** Let r_n, s_n, t_n, p_n denote the fractions of players that use the SiB rule and receive the payoffs $1, S, T$, and 0. The six dimensional dynamics $(r_n, s_n, t_n, \ p_n, t_n^{\mathbf{D}}, p_n^{\mathbf{D}})$ for a population of SiB and ALL D players reduces to a three dimensional system since $\delta = t_n^{\mathbf{D}} + p_n^{\mathbf{D}}$, $s_n = t_n + t_n^{\mathbf{D}}$, and the sum of all frequencies is 1:

$$
\begin{aligned}
r_{n+1} &= r_n^2 + 2 r_n \, p_n + s_n^2 \\[4pt]
&= s_n^2 + 2\, r_n \, (1 - \delta - s_n - t_n) - r_n^2 \\[4pt]
s_{n+1} &= r_n \, s_n + t_n \, p_n + r_n \, p_n^{\mathbf{D}} + p_n \, t_n^{\mathbf{D}} \\[4pt]
&= \delta \, (r_n - s_n) - s_n^2 + r_n \, t_n + s_n \, (1 - r_n - t_n) \\[4pt]
t_{n+1} &= r_n \, s_n + t_n \, p_n \\[4pt]
&= r_n \, s_n + t_n \, (1 - \delta - r_n - s_n - t_n)
\end{aligned}
$$

(3)

For all $\delta \in [0, 1]$ the state $(0, 0, 0)$ is a fixed point for this system. The Jacobian at this fixed point has eigenvalues $(1 - \delta, 1 - \delta, 2 - 2\delta)$. Thus, for $\delta > 1/2$ the point $(0, 0, 0)$ is asymptotically stable; for $\delta < 1/2$ it is unstable. I computed all fixed points of (3) numerically for a fine grid of δ-values using the Mathematica (Wolfram 1996) function Solve: for $\delta \geq 1/2$ $(0, 0, 0)$ is the only nonnegative fixed point; for $\delta < 1/2$ there is another fixed point $(\bar{r}(\delta), \bar{s}(\delta), \bar{t}(\delta))$ which is locally asymptotically stable and − as numerical simulations suggest − even globally asymptotically stable. As shown in Section 3.1 $(\bar{r}(0), \bar{s}(0), \bar{t}(0)) = (1, 0, 0)$. Figure 2a shows for given δ the fractions of SiB and All D players that receive the payoffs $1, S$, and T in the limit. The SiB rule performs strikingly well. If more than half of the population consists of defectors $(\delta > 1/2)$, it also defects
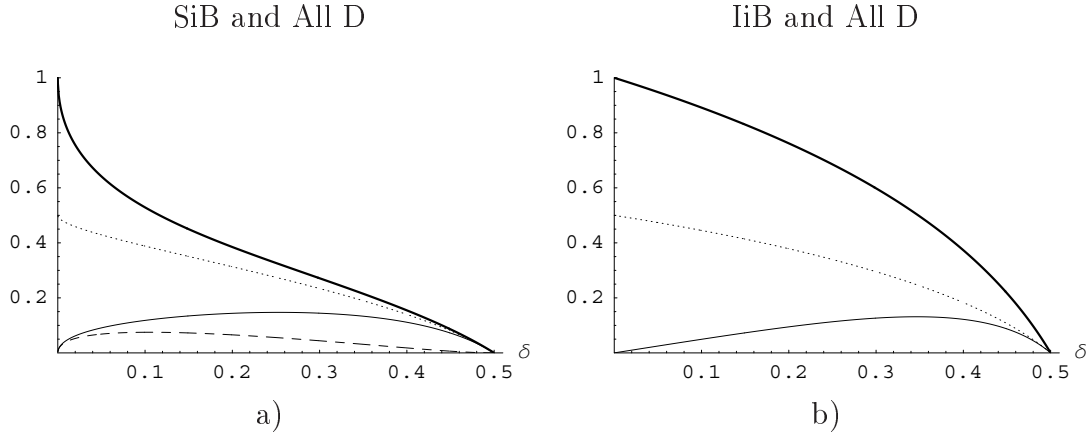
SiB and All D IiB and All D

a) b)

Figure 2. The fraction of All D players and a) IiB resp. b) SiB players receiving the payoffs $1, S$, and T in a population with a fraction δ of All D players and $1 - \delta$ SiB (resp. IiB) players: fraction of SiB (resp. IiB) players receiving 1 thick line, S thin line, T dashed line, and the fraction of All D players receiving T dotted line.

in every round. If there are no defectors at all it cooperates in the limit. If the frequency of defectors is somewhere between 0 and 1/2 the limit outcome is mixed. But the frequency of SiB players that mutually cooperate is always larger than the frequency of players receiving S. The higher the number of defectors the less cooperation. If the fraction of All D players tends to zero the All D players receive alternatingly the payoffs 0 and T: an All D player that received T in the last round cannot exploit a SiB player unless the latter received 0 in the last round, which is very unlikely if δ is low. For small values of δ the SiB players receive S mainly against another SiB player, while for larger values of δ they receive S mainly against an All D player.

Evolution. Next I study the evolution of a population of SiB and All D players. For simplicity I assume that in each generation the agents play an infinite number

of rounds and consider only the average payoff the players receive at the asymptotically stable states. The reproductive success of the players is given by their average payoff such that selection is acting via a monotone selection dynamics (Hofbauer & Sigmund 1998), i.e. strategies with a higher payoff spread faster than those with a lower payoff. For a population with two strategies this implies that the strategy with the higher payoff spreads, while the other decreases. The dynamics is given by a differential equation

$$\dot{\delta} = \delta \, g(\delta),$$

where $g(\delta)$ is a continuously differentiable function such that $g(1) = 0$ and $g(\delta)$ has the same sign as the difference of the average payoffs of All D and SiB players. This difference is given by

$$\frac{T \, \bar{t}^D(\delta)}{\delta} - \frac{\bar{r}(\delta) + \bar{s}(\delta) \, S + \bar{t}(\delta) \, T}{1 - \delta}.$$

Thus, if $T < c_1(\delta) + c_2(\delta) \, S$, where $c_1(\delta) = \delta \, \bar{r}(\delta)/[(1 - \delta) \, \bar{s}(\delta) - \bar{t}(\delta)]$ and $c_2(\delta) = \delta \, \bar{s}(\delta)/[(1 - \delta) \, \bar{s}(\delta) - \bar{t}(\delta)]$, the SiB players will have a higher average payoff than the All D players. The functions $c_1(\delta), c_2(\delta)$ are plotted in Figure 3. For $\delta \to 0$ the inequality reduces to $T < 2$. With increasing δ the set of games where SiB can spread becomes smaller. For $\delta \to 1/2$ the condition becomes $T = 1$, $S = 0$ which is no longer a Prisoner's Dilemma game. Thus, for every Prisoner's Dilemma game All D can spread if its frequency is below but sufficiently close to $1/2$. If $\delta > 1/2$ both strategies receive the average payoff 0 and thus none of the strategies can spread.

From the monotonicity of c_1 and c_2 in δ follows that the sets of games where SiB can spread are nested for growing δ. Thus, if for fixed S, T and some $\bar{\delta}$

SiB   IiB

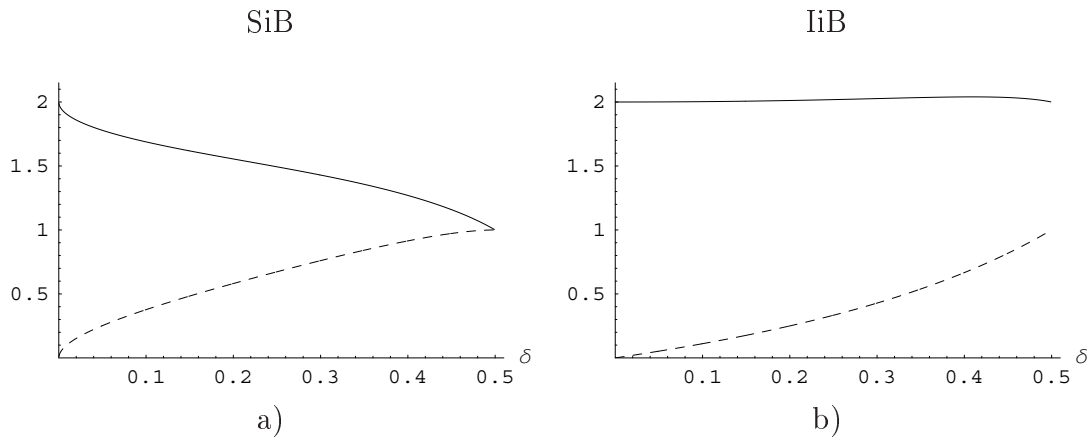a)   b)

FIGURE 3.   A fraction δ of All D players cannot invade a) a SiB
population if the payoff values satisfy $T < c_1(\delta) + c_2(\delta)\,S$, where
$c_1(\delta)$ is given by the solid and $c_2(\delta)$ by the dashed line and b) an
IiB population if the payoff values satisfy $T < c_3(\delta) + \delta/(1-\delta)\,S$,
where $c_3(\delta)$ is given by the solid and $\delta/(1-\delta)$ by the dashed line.

the strategy SiB can spread, then it spreads also for all $\delta < \bar{\delta}$. Hence, an orbit

starting at $\bar{\delta}$ leads to the fixed point 0 and SiB will come to fixation. If in contrary

for some $\bar{\delta}$ the strategy All D can spread, then it spreads also for all $\delta \in [\bar{\delta}, 1/2)$

and an orbit starting at $\bar{\delta}$ converges to $1/2$. Summarizing, we get

Proposition 3. *For all Prisoner's Dilemma games there is a continuum of fixed*

points on the interval $[1/2, 1]$. *If $T > 2$ then all orbits starting in $(0, 1/2)$ converge*

to $1/2$. If $T < 2$ there exists a fixed point δ^* such that all orbits starting in

$(\delta^*, 1/2)$ converge to $1/2$ and all orbits starting in $(0, \delta^*)$ converge to 0 (see Figure
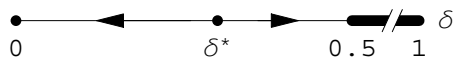
4).

FIGURE 4.   The phase portrait of the selection dynamics of a
population of SiB and All D players for Prisoner's Dilemma games
with $T < 2$. The interval $[1/2,1]$ consists of of fixed points.

Thus, for no Prisoner's Dilemma game All D can come to fixation. For Prisoner's Dilemma games with $T < 2$ the point 0 is asymptotically stable and thus a population of SiB players cannot be invaded by All D. Note that these are the games for which the win stay, lose shift rule Pavlov in the repeated Prisoner's Dilemma game is an evolutionarily stable strategy (Boerlijst, Nowak & Sigmund 1997, Leimar 1997).

5.2. **The IiB strategy.** Let r_n, s_n, t_n, p_n denote the fractions of players that use the IiB rule and receive the payoffs $1, S, T$, and 0. Since an IiB player never receives the payoff T we have $t_n = 0$ for all $n > 1$ and thus also $s_n = t_n^{\mathbf{D}}$. Additionally, $p_n^{\mathbf{D}} = \delta - t_n^{\mathbf{D}}$ and the sum of all frequencies is equal to 1. Thus, the dynamics can be reduced to two dimensions:

$$
\begin{aligned}
r_{n+1} &= (r_n + s_n)^2 + 2 r_n\, p_n \\
&= r_n(2 - 2\,\delta - r_n) + s_n^2 \\
s_{n+1} &= r_n\, p_n^{\mathbf{D}} \\
&= r_n(\delta - s_n)
\end{aligned}
$$

The state $(0,0)$ is a fixed point for this system for all $\delta \in [0,1]$. The Jacobian at this fixed point has eigenvalues $(0, 2 - 2\delta)$. Thus, for $\delta > 1/2$ the point $(0,0)$ is asymptotically stable, for $\delta < 1/2$ it is unstable. I found numerically that for $\delta \geq 1/2$ $(0,0)$ is the only non negative fixed point. For $\delta < 1/2$ there is another fixed point $(\bar{r}(\delta), \bar{s}(\delta))$ which is asymptotically stable and – as simulations suggest

– even globally asymptotically stable[1]. As shown in Section 3.2, $(\bar{r}(0), \bar{s}(0)) = (1, 0)$.

Figure 2b shows the fractions of IiB and All D players that receive in the limit the payoffs $1, S$, and T. The IiB players can maintain a higher level of cooperation than SiB as long as the frequency of defectors is less than $1/2$. If more than half of the population consists of defectors also the IiB players will always defect in the limit.

Evolution. Again, I consider a monotone selection dynamics and study the payoff difference between IiB and All D players given by

$$(4) \qquad \frac{T \, \bar{t}^D(\delta)}{\delta} - \frac{\bar{r}(\delta) + \bar{s}(\delta) \, S}{1 - \delta}.$$

If $T < c_3(\delta) + \delta/(1-\delta) \, S$, where $c_3(\delta) = \delta \, \bar{r}(\delta)/[(1-\delta) \, \bar{s}(\delta)]$, IiB players receive a higher average payoff than All D players. The function $c_3(\delta)$ is plotted in Figure 3b. It has a local maximum at $\delta = 0.41$ where it takes the value 2.04. Thus, the sets of games where IiB spreads are not nested for growing δ as it is the case for the SiB rule. This leads to a more complex dynamics.

For all Prisoner's Dilemma games with $T < 2$ IiB gets a higher payoff than All D if the fraction of All D players is sufficiently small. This coincides with the result for the SiB rule. However, for a population of IiB players there are Prisoner's Dilemma games where All D never spreads: since $c_3(\delta) \geq 2$ and $\delta/(1-\delta) \leq 1$ for all $\delta \in (0, 1/2)$ in all Prisoner's Dilemma games with $T - S < 2$ the IiB strategy receives a higher payoff than All D. Since $\lim_{\delta \to 1/2} c_3(\delta) = 2$, for all games where

---

[1]For the corresponding differential equation it is easily seen that the divergence is strictly negative. Thus, for the continuos system all orbits converge to a fixed point.

$T - S > 2$ All D receives a higher payoff than the IiB strategy if the fraction of All D players is close to but smaller than $1/2$.

For games where $T < 2$ and $T - S > 2$ there is at least one fixed point in the interval $(0, 1/2)$, where IiB and All D receive the same payoff. If $T \in (1.9901, 2.04)$ and S is close to 0 there are up to three fixed points. There, I numerically analyzed the respective bifurcation diagrams (see Appendix B).

Proposition 4. *For all Prisoner's Dilemma games the interval $[1/2, 1]$ contains a continuum of fixed points. The dynamics in $[0, 1/2]$ is classified in Figure 5: for games in region d) all paths in $[0,1/2]$ converge to $1/2$ while for games in region e) where $T - S < 2$ they converge to 0. For games in the remaining regions the dynamics is more complex and specified in Figure 5.*
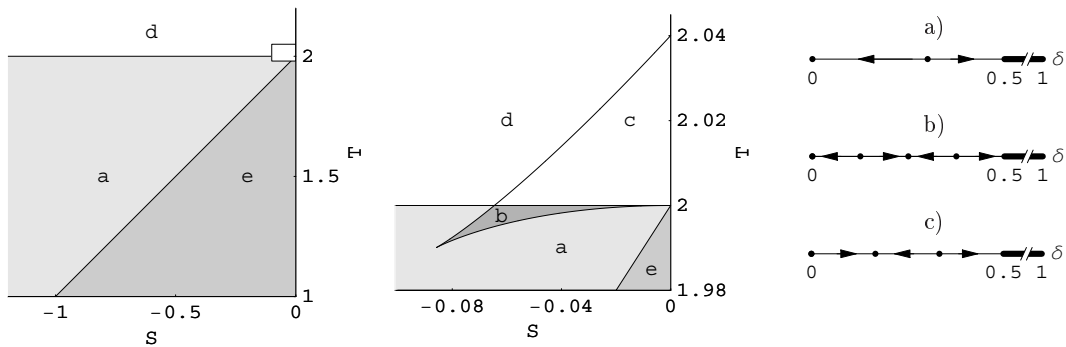

FIGURE 5.   The phase portrait of the selection dynamics of a population of SiB and All D players for Prisoner's Dilemma games with $T < 2$. The interval $[1/2,1]$ consists of fixed points.

If $T - S < 2$ the IiB strategy can even invade a population of All D players: for $\delta > 1/2$ there is a continuum of fixed points and thus the IiB strategy does as well as All D . Hence, IiB can spread by random drift. If it crosses the 50% mark by chance, IiB receives a higher payoff than All D and will come to fixation.

6. Discussion

Using the last payoff of the current opponent as aspiration level turns out to be a very efficient strategy for symmetric 2×2 games. For all games the SiB and the IiB rule with competitor matching receive in the limit the maximal symmetric payoff. An immediate question is if there is an analogous result for symmetric $k \times k$ games with $k > 2$. The SiB strategy can be adapted easily for this situation: here the players switch to one of the other actions if their payoff in the last round is lower than the one of the opponent. Which of the other actions they choose is random. An analogous result to Proposition 1 would imply that the fraction of players that receive the highest payoff on the diagonal of the payoff matrix converges to 1. Numeric investigations show, however, that this is not the case for the SiB rule. Here the limit outcome is always mixed, i.e. the players do not coordinate on a single action pair. In contrast, the IiB strategy is also successful for these games: after the first round a population of IiB players receives only payoff values that lie on the diagonal of the payoff matrix. An analogous argument as in Proposition 2 shows that a homogeneous population of IiB players coordinates in the limit on the highest payoff on the diagonal of the payoff matrix. A corresponding result for symmetric n person games, where the opponent with the highest payoff is imitated, can be proven similarly. Thus, imitating the current opponent if he was more successful appears to be a very powerful learning principle.

Also reciprocal strategies as Tit for Tat (Axelrod 1984) for the repeated Prisoner's Dilemma or the discriminating strategy in the indirect reciprocity model Nowak and Sigmund (1998a, 1998b) are imitating learning rules: they just copy the

last move of the opponent without taking into account the opponents success. These learning rules perform well if the players start with the "right" action (which is **C** in this case) and if either no errors occur or only a small number of rounds are played. Since these reciprocal learning rules are independent of the game structure it depends solely on the initial actions if they receive the Pareto optimal payoff. If the game structure changes, they cannot adapt to the new situation.

This paper illustrates that "social" strategies that are guided by the payoff of the opponent may lead to a very efficient outcome. For 2×2 games even the simple SiB rule succeeds. It turns out, however, that the choice of the matching rule is crucial. Anyhow, following the arguments of Thibaut and Kelley (1959) discussed in the introduction, it seems to be less plausible to assume independent matching than competitor matching: players we strategically interact with are surely more salient to us than randomly chosen individuals.

Appendix A

I first show for the dynamics (1),(2) the convergence of p_n by constructing a Lyapunov function. Let $1/2 > \alpha > 0$ and set $v_\alpha(p) = \min(\max(\alpha, p), 1 - \alpha)$. Then there exists an n_0 such that for all $n > n_0$ we have

$$(5) \qquad\qquad v_\alpha(p_{n+1}) - v_\alpha(p_n) \leq 0.$$

Fix some initial conditions p_1, s_1. Let $\varepsilon_n = 2s_n - (s_n)^2$ and choose an n_0 such that $\varepsilon_{n_0} < \alpha - \alpha^2$. Then the following proves (5) for all $n > n_0$: a) $p_n \leq \alpha \Rightarrow p_{n+1} \leq \alpha$, b) $\alpha < p_n < 1 - \alpha \Rightarrow p_{n+1} - p_n \leq 0$. a) follows since $p_{n+1} \leq p_n^2 + \varepsilon_{n_0} \leq \alpha$. To show b) note that the increment of $p_{n+1} - p_n$ given by $g(p_n, \varepsilon_n)$, where $g(p, \varepsilon) = p^2 + \varepsilon - p$ has zeroes at $q_1(\varepsilon), q_2(\varepsilon) = 1/2(1 \pm \sqrt{1 - 4\varepsilon})$. Between these zeroes it is negative. A straightforward calculation shows that for $\varepsilon < \alpha - \alpha^2$ we have $q_1(\varepsilon) < \alpha$ and $q_2(\varepsilon) > 1 - \alpha$.

Thus, v_α is a Lyapunov function. Since α can be chosen arbitrarily small it follows that p_n converges to one of the fixed points of the system of difference equations (1), (2) given by $(\bar{s}_1, \bar{p}_1) = (0, 0)$, $(\bar{s}_2, \bar{p}_2) = (0, 1)$. The latter is however unstable: assume that $\lim_{n \to \infty} p_n = 1$. Then $\lim_{n \to \infty} r_n = 0$. However, for $r_n < 0.2$

$$r_{n+1} - r_n = r_n^2 + s_n^2 + 2\, r_n\, p_n - r_n = (s_n - 2\, r_n)^2 + r_n - 5\, r_n^2 > 0,$$

which gives a contradiction. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\;\Box$

APPENDIX B

To determine the different phase portraits we first set the term (4) equal to 0 and solve the resulting equation for S. This leads to a function $\sigma(T, \delta)$ which gives for each T and δ the S value such that ALL D and IiB get the same average payoff. If S is larger (smaller) than this value IiB receives a higher (lower) payoff than ALL D. Figure 6 shows the function $\sigma(T, \delta)$ for several values of T. Each horizontal line in the (δ, S) plane corresponds to the phase portrait of a 2×2 game. Fix an S and T. Then for all δ values such that $S > \sigma(T, \delta)$ IiB will spread (and the dynamics moves on the corresponding line to the left), while

for all δ such that $S < \sigma(T, \delta)$ ALL D spreads (and the dynamics moves to the right). The intersections of $\sigma(T, \delta)$ with the horizontal line $y = S$ are the fixed points. To determine the different regions in Figure 5 I computed for each T the local extrema of the function $\sigma(T, \delta)$ from which then intervals for S are derived where the number of fixed points (i.e. intersections of $\sigma(T, \delta)$ with the horizontal line $y = S$) is constant.
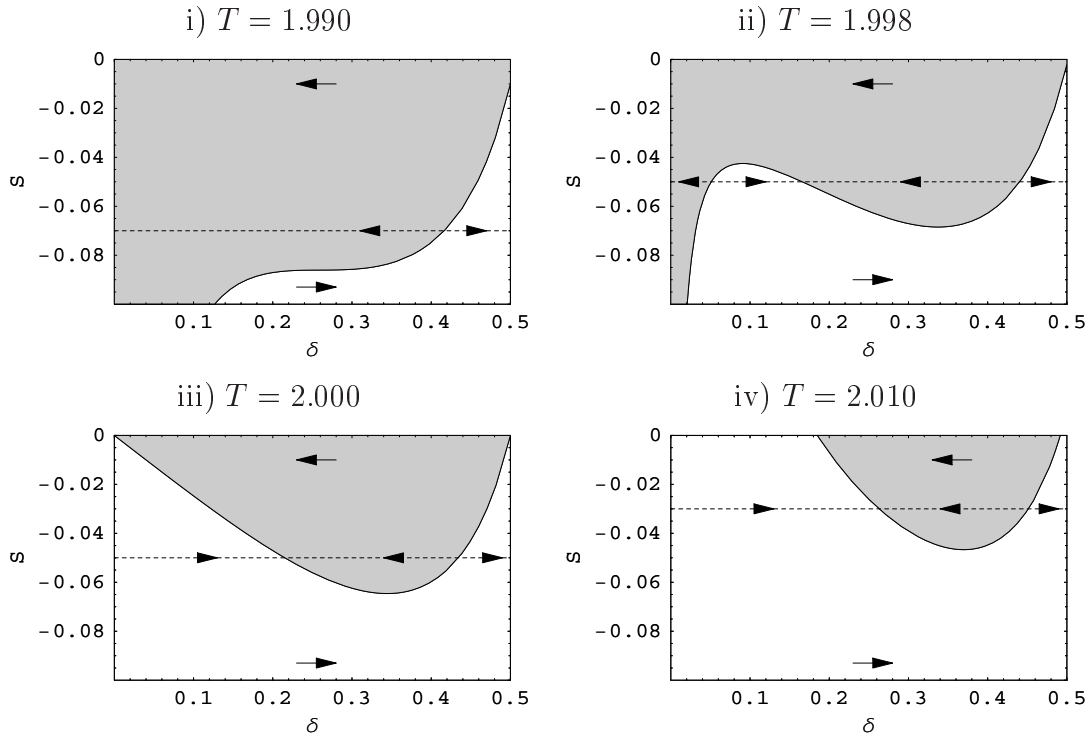

FIGURE 6. Dynamics of a population of IiB and All D players with competitor matching. In the grayed region the IiB players receive a higher payoff than the All D players, while in the white region the All D players succeed. Thus, in the gray region the dynamics points to the left and in the white region to the right. The dashed line in each graph shows a phase portrait for the specified value of T and the value S on the y-axis. For δ values greater $1/2$ both strategies receive the same payoff and there is a continuum of fixed points.

References

Axelrod, R. 1984 *The Evolution of Cooperation*, New York: Basic Books.

Björnerstedt, J. & Weibull, J. 1996 Nash equilibrium and evolution by imitation, in *The Rational Foundations of Economic Behaviour* (eds. K. Arrow & E. Colombatto), pp. 155–171, New York: MacMillan.

Boerlijst, M. C., Nowak, M. A. & Sigmund, K. 1997 The logic of contrition. *J. Theor. Biol.* **185**, 281–293.

Börgers, T. & Sarin, R. 2000 Naive reinforcement learning with endogenous aspirations. *Int. Econ. Rev.* **41**, 921–950.

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

Dixon, H. 2000 Keeping up with the joneses: Competition and the evolution of collusion. *J. Econ. Behav. Org.* **43**, 223–238.

Hofbauer, J. 1995 Imitation dynamics for games, Department of Mathematics, University of Vienna.

Hofbauer, J. & Schlag, K. H. 2000 Sophisticated imitation in cyclic games. *J. Evol. Econ.* **10**, 507–521.

Hofbauer, J. & Sigmund, K. 1998 *Evolutionary Games and Population Dynamics*, Cambridge: Cambridge UP.

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

Leimar, O. 1997 Repeated games: a state space approach. *J. Theor. Biol.* **184**, 471–98.

Nowak, M. A. & Sigmund, K. 1998a The dynamics of indirect reciprocity. *J. Theor. Biol.* **194**, 561–574.

Nowak, M. A. & Sigmund, K. 1998b Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577.

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

Oechssler, J. 2000 Cooperation as a result of learning with aspiration levels, Department of Economics, University of Bonn.

Palomino, F. & Vega-Redondo, F. 1999 Convergence of aspirations and (partial) cooperation in the Prisoner's Dilemma. *Int. J. Game Theory* **28**, 465–488.

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., Pichler, A. & Sigmund, K. 1999 The efficiency of adapting aspiration levels. *Proc. R. Soc. Lond. B* **266**, 1427–1435.

Schlag, K. 1998 Why imitate, and if so, how? A bounded rational approach to multi-armed bandits. *J. Econ. Theory* **78**, 130–156.

Thibaut, J. W. & Kelley, H. H. 1959 *The Social Psychology of Groups*, New York: Wiley.

Vega-Redondo, F. 1997 The evolution of Walrasian behavior. *Econometrica* **65**, 375–384.

Weibull, J. W. 1995 *Evolutionary Game Theory*, Massachusets: MIT Press.

Wolfram, S. 1996 *The Mathematica Book*, Melbourne, Australia: Wolfram Media/Cambridge University Press, 3rd edn.