Sophisticated Imitation in Cyclic Games*

Josef Hofbauer
Mathematics
University of Vienna
Strudlhofgasse 4,
A-1090 Wien, Austria

Karl H. Schlag
Economic Theory III
University of Bonn
Adenauerallee 24-26
D-53113 Bonn, Germany

13 February 1998

Abstract

Consider a large population of individuals that are repeatedly randomly matched to play a cyclic 2×2 game such as Matching Pennies with fixed roles assigned in the game. Some learn by sampling previous play of a finite number of other individuals in the same role. We analyze population dynamics under optimal boundedly rational behavior (in the sense of Schlag, 1998c). We find that long run play is close to the Nash equilibrium (when few individuals receive information) if and only if the sample size is greater than one.

*Financial support from the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn and the Austrian FWF, project P10717-MAT is gratefully acknowledged.

1 Introduction

Evolutionary Game Theory has provided dynamic models of replication, imitation and learning for analyzing change of play in games. The interpretation of a mixed strategy as polymorphic population state in which each individual chooses a pure strategy is particularly appealing. However, in multi-population models where individuals do not interact with their kin, mixed population states are rarely stable. This phenomenon first appeared when Selten (1980) showed that ESS of truly asymmetric contests can not be mixed. With increasing interest in learning and imitation, models of individual behavior have been developed that induce the replicator dynamic known from population biology (Björnerstedt and Weibull, 1995; Börgers and Sarin, 1997; Gale et al., 1995; Posch, 1997; Schlag, 1998c). With this connection, features known for the replicator dynamic can be used to predict behavior among interacting and learning individuals. However, the derivations mentioned lead to the so-called *standard replicator dynamic* (Taylor, 1979) where mixed strategies fail to be asymptotically stable (Hofbauer and Sigmund, 1988). The alternative *adjusted* version due to Maynard Smith (1982) for which asymptotically stable states can be interior (such as in Matching Pennies) is motivated from replicators. The only individual learning model known that generates this dynamic (Björnerstedt and Weibull, 1996) requires individual knowledge of population averages.[1] The instability of interior equilibria under the standard replicator dynamic has biased predictions of learning models towards pure strategies and set-valued solution concepts (Ritzberger and Weibull, 1995, Proposition 5; Oechssler and Schlag, 1997, Proposition 2[2]). Up to now, only models of learning that heavily utilize information (e.g., best response dynamics, Hofbauer, 1995; imitation based on current population averages, Björnerstedt and Weibull, 1996) or memory (fictitious play, Brown, 1951) have been able to stabilize interior equilibria (e.g. in Matching Pennies). Of course, stability of interior equilibria or even convergence of learning processes should not generally be expected in all games (Shapley, 1964; Gaunersdorfer and Hofbauer, 1995). However, when play cycles in simple games such as Matching Pennies, then an individual has an incentive to invest in more information or more memory. Hence,

[1] Specifically, the individual imitation rule is a function of the average payoff of each strategy present among the (infinitly many) individuals belonging to the same player role.

[2] See also the conclusion of our paper.

convergence to (or approximating behavior of) the Nash equilibrium in Matching Pennies remains a test for the sophistication of individual behavioral rules.

Schlag (1998c) has analyzed optimal boundedly rational learning when an individual observes previous play of one other individual in the same player role (single sampling). Here optimal behavior is imitative and leads to the standard replicator dynamic. Recently, Schlag (1998b) extended this approach to samples of size two (double sampling) where optimal individual behavior is again imitative, now leading to an *aggregate monotone dynamic* (Samuelson and Zhang, 1992). More specific properties of the resulting dynamic have not been considered as of yet. In this paper, we extend the optimal behavioral rule found under double sampling to general sample sizes and then analyze population dynamics in cyclic 2×2 games (i.e., generalized Matching Pennies). In particular, we analyze the discrete dynamic directly instead of making the common shortcut of reverting to a continuous time approximation. Our results summarized in section 3 show a discontinuous jump in results between single sampling and sample sizes greater than one. We find long run states situated close to the interior mixed Nash equilibrium if and only if the sample size is larger than one. On the side we obtain that population dynamics are approximated by the *adjusted replicator dynamic* of Maynard Smith (1982) when the sample size is large.

Thus we find that mixed strategies can be stabilized after all in learning models based on limited information provided play of at least two individuals in the same role is observable and individual behavior is imitative and sufficiently sophisticated.

2 Sophisticated imitation

Consider a two person normal form (or *bimatrix*) game Γ in which player k chooses a pure strategy from S^k and play $(i, j) \in S^1 \times S^2$ yields independent stochastic payoffs for player k contained in a bounded interval $\left[\alpha^k, \omega^k\right]$. Let $\pi^k(i, j)$ denote the expected payoff to player k when player one chooses pure strategy i and player two pure strategy j. Let $\Delta\left(S^k\right)$ denote the set of mixed strategies of player k and let $\pi^k(\mathbf{p}, \mathbf{q}) = \sum_{i,j} \pi^k(i, j)\, p_i q_j$ denote the expected payoff of player k when player 1 chooses the mixed strategy $\mathbf{p} \in \Delta(S^1)$ and player 2 the mixed strategy $\mathbf{q} \in \Delta(S^2)$.

In the following we describe a model of boundedly rational individuals learning how to play this game. The model and results in this section for sample sizes one and two are taken from (Schlag , 1998b, 1998c). Assume that individuals have fixed roles in the game Γ. They belong to two countably infinite populations, one corresponding to each player in the game. In a sequence of rounds, individuals from each population are randomly matched to play the game.

Between rounds, a fixed proportion γ_k of individuals belonging to population k are given the opportunity to receive some information about the play of others. An individual receives information by *sampling* n individuals at random from the same population and observes the last strategy used and payoff attained in the previous round by each individual in this sample ($n \in \mathbb{N}$, $0 < \gamma_k \leq 1$, $k = 1, 2$). The case $n = 1$ is called *single sampling*, $n = 2$ *double sampling.* After each round, each individual uses a *behavioral rule* (that is a function of his knowledge and previous experience and information) to determine which strategy to play in the next round. A *state* $(\mathbf{p}, \mathbf{q}) \in \Delta(S^1) \times \Delta(S^2)$ describes the proportions of pure strategies used in each population. Thus, the average payoff in population k is equal to $\pi^k(\mathbf{p}, \mathbf{q})$ in state $(\mathbf{p}, \mathbf{q})$. Individual behavior induces a population dynamic $(\mathbf{p}^t, \mathbf{q}^t)_{t=0,1,2,..}$ as a function of the initial population state $(\mathbf{p^0}, \mathbf{q^0})$ where

$$\begin{aligned} p_i^{t+1} &= p_i^t + \gamma_1 \cdot f_i\left(\mathbf{p}^t, \mathbf{q}^t\right) \\ q_j^{t+1} &= q_j^t + \gamma_2 \cdot g_j\left(\mathbf{p}^t, \mathbf{q}^t\right) \end{aligned}$$

for appropriately chosen continuous maps $f_i, g_j : \Delta(S^1) \times \Delta(S^2) \to L_0(S^1) \times L_0(S^2)$ ($i \in S^1$, $j \in S^2$), with $L_0(S^k) = \{(\xi_i)_{i \in S^k} : \xi_i \in R, \sum_{i \in S^k} \xi_i = 0\}$.

Schlag (1998c) develops a theory of optimal behavior for individuals that are boundedly rational in the following sense. Initially an individual only knows the own strategy set together with the bounded interval containing his payoffs. After each round an individual forgets all occurrences before this round. In round 0 there is an arbitrary initial state $(\mathbf{p^0}, \mathbf{q^0})$ in the interior of $\Delta(S^1) \times \Delta(S^2)$. It is assumed that individuals belonging to the same population use the same behavioral rule. This generates a population dynamic $(\mathbf{p}^t, \mathbf{q}^t)_{t=0,1,2,..}$. Then a behavioral rule for individuals in population k is called *strictly improving* if the average payoff in this population increases over time whenever play in the other population does not change, this increase being strict whenever not all strategies played in population k achieve the same expected payoff. When $\left|S^k\right| = 2$ then this means that the proportion of individuals in population k playing a best response to the population state of the last round always increases. Motivated by the assumption that the individual has no information about the specific payoff distributions generated in Γ other than the payoff interval $\left[\alpha^k, \omega^k\right]$ this condition must hold in any such game. Formally, the conditions for a rule to be used in population one to be strictly improving are that, in any game Γ with strategy sets S^1 and S^2 that generates stochastic payoffs for player one in $[\alpha^1, \omega^1]$,

$$\begin{aligned}\pi^1\left(\mathbf{p}^{t+1}, \mathbf{q}^t\right) &\geq \pi^1\left(\mathbf{p}^t, \mathbf{q}^t\right) \text{ for all } \left(\mathbf{p}^t, \mathbf{q}^t\right) \in \Delta\left(S^1\right) \times \Delta\left(S^2\right) \\ \text{where } &> \text{ holds if } \pi^1\left(i, \mathbf{q}^t\right) > \pi^1\left(\mathbf{p}^t, \mathbf{q}^t\right) \text{ for some } i \text{ with } p_i^t > 0. \quad (1)\end{aligned}$$

In the following we use a simple algorithm to construct a strictly improving behavioral rule for general sample sizes n. Consider first single sampling. Due to the limitations on individual memory, the behavioral rule of an individual who observed a sample of size one can be described as a map

$$F : S^k \times \left[\alpha^k, \omega^k\right] \times S^k \times \left[\alpha^k, \omega^k\right] \to \Delta\left(S^k\right)$$

where $F(i, x, j, y)_r$ is the probability of playing strategy r in the next round after playing strategy i and receiving payoff x and observing an individual who used strategy j and attained payoff y. We will assume that individuals that do not receive additional information through a sample do not change their strategy.[3]

The *Proportional Observation Rule* $F^{(1)}$ (POR, Schlag, 1998b) is defined by

$$\begin{aligned}F^{(1)}(i, x, j, y)_j &= \frac{y - \alpha^k}{\omega^k - \alpha^k} \\ F^{(1)}(i, x, j, y)_i &= 1 - F^{(1)}(i, x, j, y)_j, \, i \neq j.\end{aligned}$$

Notice that POR is an *imitation rule* since it either prescribes to switch to a strategy sampled or not to switch at all. POR specifies to imitate the observed individual with a probability that is proportional to the payoff of the sampled individual and independent of own payoff. When both populations use the respective POR the population state changes according to (a discrete version of) the replicator dynamics of Taylor (1979) (with stepsizes γ_k, see Schlag, 1998c):

$$\begin{aligned}p_i^{t+1} &= p_i^t + \frac{\gamma_1}{\omega^1 - \alpha^1}\left[\pi^1\left(i, \mathbf{q}^t\right) - \pi^1\left(\mathbf{p}^t, \mathbf{q}^t\right)\right] \cdot p_i^t \\ q_j^{t+1} &= q_j^t + \frac{\gamma_2}{\omega^2 - \alpha^2}\left[\pi^2\left(\mathbf{p}^t, j\right) - \pi^2\left(\mathbf{p}^t, \mathbf{q}^t\right)\right] \cdot q_j^t \quad (2)\end{aligned}$$

[3] In fact, it can be shown that strictly improving rules must have this property.

When concerned with larger samples we will limit attention to behavioral rules that can be constructed from a single sampling rule F using the following simple procedure. Apply the rule F in sequence to each individual in the sample, replacing own strategy and payoff by observed strategy and payoff whenever the rule F prescribes to switch strategies. The strategy left with after applying this procedure to each individual in the sample is the strategy to be played in the next round. This generates a behavioral rule for sample size n we will refer to as *sequentially evaluating* the single sampling rule F. In the following we will demonstrate this procedure and the resulting dynamic equations in more detail for POR. Sequentially evaluating POR when $n = 2$ leads to a double sampling rule $F^{(2)}$. It follows, for $|\{i, j, r\}| = 3$, that[4]

$$\begin{aligned} F^{(2)}(i,x,j,y,r,z)_j &= \frac{y-\alpha^k}{\omega^k-\alpha^k}\frac{\omega^k-z}{\omega^k-\alpha^k} \\ F^{(2)}(i,x,j,y,r,z)_r &= \frac{z-\alpha^k}{\omega^k-\alpha^k} \\ F^{(2)}(i,x,j,y,r,z)_i &= 1-F^{(2)}(i,x,j,y,r,z)_j - F^{(2)}(i,x,j,y,r,z)_r . \end{aligned}$$

This leads to the dynamic (Schlag, 1998b)

$$\begin{aligned} p_i^{t+1} &= p_i^t + \frac{\gamma_1}{\omega^1-\alpha^1}\left(1+\frac{\omega^1-\pi^1(\mathbf{p}^t,\mathbf{q}^t)}{\omega^1-\alpha^1}\right)\left[\pi^1(i,\mathbf{q}^t)-\pi^1(\mathbf{p}^t,\mathbf{q}^t)\right]p_i^t \\ q_j^{t+1} &= q_j^t + \frac{\gamma_2}{\omega^2-\alpha^2}\left(1+\frac{\omega^2-\pi^2(\mathbf{p}^t,\mathbf{q}^t)}{\omega^2-\alpha^2}\right)\left[\pi^2(\mathbf{p}^t,j)-\pi^2(\mathbf{p}^t,\mathbf{q}^t)\right]q_j^t . \end{aligned} \tag{3}$$

The trick simplifying the calculations is to consider switching behavior even when an individual observes someone using the same strategy. Then the additional term

$$\frac{\omega^1-\pi^1(\mathbf{p}^t,\mathbf{q}^t)}{\omega^1-\alpha^1} = \sum_{i\in S^1} p_i^t\left(1-\frac{\pi^1(i,\mathbf{q}^t)-\alpha^1}{\omega^1-\alpha^1}\right)$$

as compared to the single sampling case (2) is the probability that an individual does not switch after evaluating POR to the first individual in the sample.

[4] $F^{(2)}(i,x,j,y,r,z)_v$ is the probability of playing strategy v in the next round after playing strategy i and receiving payoff x and sequentially evaluating POR first to an individual who used strategy j and attained payoff y and then to one who used strategy r and attained payoff z in the last round.

For general n, if each individual uses the rule derived by sequentially evaluating POR (which we will call SPOR_n) we obtain the following dynamic

$$\begin{aligned}
p_i^{t+1}-p_i^t &= \frac{\gamma_1}{\omega^1-\alpha^1}\left[1+\frac{\omega^1-\pi^1\left(\mathbf{p}^t,\mathbf{q}^t\right)}{\omega^1-\alpha^1}+\ldots+\left(\frac{\omega^1-\pi^1\left(\mathbf{p}^t,\mathbf{q}^t\right)}{\omega^1-\alpha^1}\right)^{n-1}\right] \\
&\quad\cdot\left[\pi^1\left(i,\mathbf{q}^t\right)-\pi^1\left(\mathbf{p}^t,\mathbf{q}^t\right)\right]\cdot p_i^t \\
q_j^{t+1}-q_j^t &= \frac{\gamma_2}{\omega^2-\alpha^2}\left[1+\frac{\omega^2-\pi^2\left(\mathbf{p}^t,\mathbf{q}^t\right)}{\omega^2-\alpha^2}+\ldots+\left(\frac{\omega^2-\pi^2\left(\mathbf{p}^t,\mathbf{q}^t\right)}{\omega^2-\alpha^2}\right)^{n-1}\right] \\
&\quad\cdot\left[\pi^2\left(\mathbf{p}^t,j\right)-\pi^2\left(\mathbf{p}^t,\mathbf{q}^t\right)\right]\cdot q_j^t\,. \qquad (4)
\end{aligned}$$

Thus,

$$\begin{aligned}
\pi^1\left(\mathbf{p}^{t+1},\mathbf{q}^t\right)-\pi^1\left(\mathbf{p}^t,\mathbf{q}^t\right) &= \textstyle\sum_{i\in S^1}\left(p_i^{t+1}-p_i^t\right)\left[\pi^1\left(i,\mathbf{q}^t\right)-\pi^1\left(\mathbf{p}^t,\mathbf{q}^t\right)\right] \\
&= \tfrac{\gamma_1}{\omega^1-\alpha^1}\left[1+\tfrac{\omega^1-\pi^1\left(\mathbf{p}^t,\mathbf{q}^t\right)}{\omega^1-\alpha^1}+..+\left(\tfrac{\omega^1-\pi^1\left(\mathbf{p}^t,\mathbf{q}^t\right)}{\omega^1-\alpha^1}\right)^{n-1}\right] \\
&\quad\cdot\textstyle\sum_{i\in S^1}\left[\pi^1\left(i,\mathbf{q}^t\right)-\pi^1\left(\mathbf{p}^t,\mathbf{q}^t\right)\right]^2\cdot p_i^t\geq 0\,.
\end{aligned}$$

which leads to the following result (see also Schlag, 1998b, 1998c for $n=1,2$).

Proposition 1 *$SPOR_n$ generates a strictly improving rule for any sample size n.*

In the limit $n\to\infty$, (4) converges to (a discretization of) the version of the replicator dynamic introduced by Maynard Smith (1982, Appendix J):[5]

$$\begin{aligned}
p_i^{t+1} &= p_i^t+\frac{\gamma_1}{\pi^1(\mathbf{p}^t,\mathbf{q}^t)-\alpha^1}p_i^t\cdot\left[\pi^1\left(i,\mathbf{q}^t\right)-\pi^1\left(\mathbf{p}^t,\mathbf{q}^t\right)\right] \\
q_j^{t+1} &= q_j^t+\frac{\gamma_2}{\pi^2(\mathbf{p}^t,\mathbf{q}^t)-\alpha^2}q_j^t\cdot\left[\pi^2\left(\mathbf{p}^t,j\right)-\pi^2\left(\mathbf{p}^t,\mathbf{q}^t\right)\right] \qquad (5)
\end{aligned}$$

2.1 Optimality of SPOR_n

SPOR_n defines an imitation rule that is strictly improving. In fact, any strictly improving rule is imitating.[6] Moreover, SPOR_n has unique optimality properties among the strictly improving rules as we will see below. Consider two behavioral

[5] We note that any of the maps $\left(\mathbf{p}^t,\mathbf{q}^t\right)\longmapsto\left(\mathbf{p}^{t+1},\mathbf{q}^{t+1}\right)$ given by (2, 3, 4, 5) defines a map from $\Delta\left(S^1\right)\times\Delta\left(S^2\right)$ into itself. Thus, these expressions generate what we know from their derivation, namely, a well defined dynamics on $\Delta\left(S^1\right)\times\Delta\left(S^2\right)$.

[6] The proof for $n=1,2$ contained in (Schlag, 1998b) extends immediately to general sample sizes n.

rules F and G for player one. We say that F *dominates* G if the terms in (1) are always larger under F than under G, with strict inequality in some cases, i.e., if

$$\pi^1\left(\mathbf{p}^{t+1}(F), \mathbf{q}^t\right) \geq \pi^1\left(\mathbf{p}^{t+1}(G), \mathbf{q}^t\right)$$

for any payoff distributions in Γ and any state $(\mathbf{p}^t, \mathbf{q}^t)$ in round t, and not "$\equiv$", where $\mathbf{p}^{t+1}(H)$ describes the proportions used in round $t+1$ in population one when all individuals in population one use rule H. Notice that a rule F dominates the rule "never switch strategies" if and only if F is strictly improving. The notion of domination is a very stringent condition and hence only defines a partial order on the set of behavioral rules. In the following we will point out some rules that are best according to dominance in a given set of rules. We will say that F is *undominated* in the set of rules $\mathcal{D}$ if $F \in \mathcal{D}$ and if there is no rule $G \in \mathcal{D}$ that dominates F.

The properties of SPOR_n for single (Schlag, 1998c) and double (1998b) sampling, stated in terms of dominance, are as follows.

(Single sampling) $\text{SPOR}_1 \equiv$POR is undominated among the strictly improving single sampling behavioral rules. Any other single sampling rule with this property[7] induces the same population adjustment (2) as POR.

(Double sampling) SPOR_2 is the *unique* rule that is undominated among the sequentially evaluated single sampling rules. SPOR_2 dominates POR. In fact, in 2×2 games, SPOR_2 is undominated among the double sampling rules that dominate POR. Any other double sampling rule with this property leads to the same population adjustment (3) as SPOR_2.

Some notes are in place. The bounded rationality assumptions underlying this model make it natural to restrict attention to simple rules of behavior such as through sequential evaluation of a single sampling rule. Two questions arise.

Why use POR in this construction? The fact that POR prescribes to forget own previous payoff is counterintuitive. However, reformulating the uniqueness statement for double sampling made above, any other single sampling rule either does not generate a strictly improving double sampling rule under sequential evaluation or is dominated by SPOR_2.

Can alternative, possibly more complicated, methods for constructing a double sampling rule outperform SPOR_2? Strictly improving double sampling rules

[7]such as the Proportional Imitation Rule and the Proportional Reviewing Rule (Schlag, 1998c).

that are not dominated by SPOR_2 exist (Schlag, 1998b). However, such rules do not dominate POR (or any other single sampling rule that is undominated among the strictly improving single sampling rules). Restricting attention to rules that dominate POR reflects the condition that optimal behavior under multiple sampling should outperform optimal behavior under single sampling. In this class, SPOR_2 performs best.

As for behavioral rules based on sample sizes $n > 2$ general results on optimality have not been derived. Of course, in want of one procedure that works well for all sample sizes, it is natural to employ SPOR_n. Notice that, following (1), SPOR_n performs better the larger the sample size. Formally,

Proposition 2 *$SPOR_n$ dominates $SPOR_m$ for all $n > m \geq 1$.*

3 Cyclic games and summary of main results

We call the bimatrix game Γ with $|S^1| = |S^2| = 2$ a *cyclic* 2×2 *game* if it has a unique Nash equilibrium E where E is in the interior of $\Delta(S^1) \times \Delta(S^2)$. Examples include the buyer-seller game by Friedman (1991, see also Schlag, 1998a), Dawkins' battle of the sexes game (see Maynard Smith, 1982, or Hofbauer and Sigmund, 1988) and generalized Matching Pennies, the latter given by the following normal form

$$\begin{array}{ccc} & 1 & 2 \\ 1 & \nu,\mu & \mu,\nu \\ 2 & \mu,\nu & \nu,\mu \end{array} \tag{6}$$

where $\mu \neq \nu$ and $E = \left(\left(\frac{1}{2},\frac{1}{2}\right),\left(\frac{1}{2},\frac{1}{2}\right)\right)$. Cyclic games are a simple testing ground for whether bounded rational learning rules such as SPOR_n that we derived from local criteria can lead large populations (close) to a completely mixed Nash equilibrium.

In the following we will summarize our results (described in detail in Section 5) on the dynamics induced by SPOR_n in cyclic 2×2 games. Foundations for these results are derived for general aggregate monotone dynamics in Section 4. Under single sampling, trajectories spiral out to the boundary which means that individuals in the same population are playing the same pure (non-equilibrium) strategy most of the time. While the Nash equilibrium E remains unstable for larger sample sizes, results now also depend on the proportion γ_k of individuals in each population receiving information. For a given sample size n, if too many individuals receive information then the boundary remains attracting. However, if $\gamma_k < 1/4$ then the boundary is repelling in all cyclic 2×2 games. Moreover, when γ_k is sufficiently small then the long run states form a limit cycle which is close to E. Simulations are added in Section 5 to exemplify the specific relations.

4 Aggregate monotone dynamics (discrete time) for cyclic 2×2 games

Consider now a general aggregate monotone dynamics in discrete time for a two person game:

$$\begin{aligned} p_i^{t+1} &= p_i^t[1 + \gamma_1(\pi^1(i, \mathbf{q}^t) - \pi^1(\mathbf{p}^t, \mathbf{q}^t))\phi^1(\pi^1(\mathbf{p}^t, \mathbf{q}^t))] \\ q_j^{t+1} &= q_j^t[1 + \gamma_2(\pi^2(\mathbf{p}^t, j) - \pi^2(\mathbf{p}^t, \mathbf{q}^t))\phi^2(\pi^2(\mathbf{p}^t, \mathbf{q}^t))] \end{aligned} \tag{7}$$

with

$$\pi^k(\mathbf{p}, \mathbf{q}) = \sum_{i,j} \pi^k(i, j) p_i q_j$$

the payoff function for player $k = 1, 2$, $\phi^k : R \to R_+$ the (continuous, positive-valued) multiplier functions and $0 < \gamma_k \leq 1$ the step sizes. As $\gamma_1 = \gamma_2 \to 0$, (7) turns into the differential equation

$$\begin{aligned} \dot{p}_i &= p_i(\pi^1(i, \mathbf{q}) - \pi^1(\mathbf{p}, \mathbf{q}))\phi^1(\pi^1(\mathbf{p}, \mathbf{q})) \\ \dot{q}_j &= q_j(\pi^2(\mathbf{p}, j) - \pi^2(\mathbf{p}, \mathbf{q}))\phi^2(\pi^2(\mathbf{p}, \mathbf{q})). \end{aligned} \tag{8}$$

Such game dynamics have been called *aggregate monotone* (Samuelson and Zhang (1992), Weibull (1995)). The first such dynamics, besides the standard replicator dynamics of Taylor (1979), for which the factors ϕ^k are identical 1, was suggested by Maynard Smith (1982, Appendix J), with

$$\phi^k(u) = \frac{1}{C_k + u}, \tag{9}$$

the C_k being positive constants (standing for background fitness). However, no convincing derivation has been given for this choice. As seen in section 2, this choice arises now from the imitation rule SPOR$_n$ in the limit $n \to \infty$.

4.1 Behavior near the boundary

At F_{ij} the eigenvalue in direction k for player 1 (= geometric rate of in(de-)crease of strategy k) is given by

$$\frac{p_k^{t+1}}{p_k^t}|_{F_{ij}} = 1 + \gamma_1(\pi^1(k,j) - \pi^1(i,j))\phi^1(\pi^1(i,j)) =: \lambda_{ij\to kj} \tag{10}$$

and in direction l for player 2 by

$$\frac{q_l^{t+1}}{q_l^t}|_{F_{ij}} = 1 + \gamma_2(\pi^2(i,l) - \pi^2(i,j))\phi^2(\pi^2(i,j)) =: \lambda_{ij\to il} \tag{11}$$

Consider now a cyclic 2×2 bimatrix game as defined in Section 3. Assume that the best reply cycle runs clockwise as

$$F_{11} \to F_{12} \to F_{22} \to F_{21} \to F_{11}. \tag{12}$$

Then the boundary of the unit square is invariant under the dynamics (7) and forms a *heteroclinic cycle*, i.e., a closed loop of stable/unstable manifolds of saddle points. The stability of this heteroclinic cycle is determined by the quantity

$$\rho = \rho_{11\to12}\rho_{12\to22}\rho_{22\to21}\rho_{21\to11} \tag{13}$$

with (α, β denoting pairs of pure strategies)

$$\rho_{\alpha\to\beta} = \frac{\log \lambda_{\alpha\to\beta}}{|\log \lambda_{\beta\to\alpha}|}. \tag{14}$$

Note that ρ is the product of logs of 'outgoing' eigenvalues divided by the product of logs of 'incoming' eigenvalues around the cycle (12). The following is a discrete time version of a related result for differential equations, see e.g. Hofbauer and Sigmund (1988, p. 213f).

Lemma 3 *If $\rho > 1$ (resp. $\rho < 1$) then the boundary of the square is repelling (resp. attracting) for the dynamics (7).*

Applying this stability criterion to (7) yields

Proposition 4 *Let $\psi^k(x) := 1/\phi^k(x)$, and suppose that $\frac{d\psi^k(x)}{dx} > \gamma_k$ holds in the payoff range of player k for $k = 1, 2$. Then $\rho > 1$ and the boundary of the square is repelling: There exists $\delta > 0$ such that for each completely mixed initial condition $(\mathbf{p}, \mathbf{q}) \in (0,1)^2$, there exists t_0 such that $p_i(t) > \delta$ and $q_i(t) > \delta$ for $i = 1, 2$ and $t \geq t_0$.*

This means that each pure strategy will be used with a certain positive probability $\delta > 0$ after time t_0.

Proof. We show that $\rho_{12\to 22} > 1$, the other three factors in (13) being analogous. Inserting (10) into (14) we obtain

$$\rho_{12\to 22} = \frac{\log(1+\gamma_1(\pi^1(2,2)-\pi^1(1,2))\phi^1(\pi^1(1,2)))}{|\log(1+\gamma_1(\pi^1(1,2)-\pi^1(2,2))\phi^1(\pi^1(2,2)))|} \tag{15}$$

Denote $a := \pi^1(2,2) - \pi^1(1,2) > 0$ and $b := \pi^1(1,2)$. Then the inequality $\rho_{12\to 22} > 1$ can be equivalently reformulated as

$$\rho_{12\to 22} = \frac{\log(1+\gamma_1 a\phi^1(b))}{-\log(1-\gamma_1 a\phi^1(b+a))} \quad > \quad 1 \tag{16}$$

$$\log(1+\gamma_1 a\phi^1(b)) \quad > \quad -\log(1-\gamma_1 a\phi^1(b+a)) \tag{17}$$

$$(1+\gamma_1 a\phi^1(b))(1-\gamma_1 a\phi^1(b+a)) \quad > \quad 1 \tag{18}$$

$$\phi^1(b) - \phi^1(b+a) \quad > \quad \gamma_1 a\phi^1(b)\phi^1(b+a) \tag{19}$$

$$\psi^1(b+a) - \psi^1(b) \quad > \quad \gamma_1 a \tag{20}$$

By the mean value theorem, this last inequality holds whenever $(\psi^1)'(x) > \gamma_1$ for all x in the payoff range. □

A similar calculation shows:

Proposition 5 *If* $\frac{d\psi^k(x)}{dx} < \gamma_k$ *holds in the payoff range of player* k *for* $k = 1, 2$ *then* $\rho < 1$ *and the boundary of the square is locally asymptotically stable.*

4.2 Behavior near the interior equilibrium

Linearizing (7) at an interior equilibrium E leads to a matrix of the form $I + \gamma J$ with I the identity matrix, $\gamma = \mathrm{diag}(\gamma_1, \gamma_2)$, and J the linearization of the corresponding differential equation (8) at E. Because of the bipartite structure of bimatrix games, the latter takes the form (see Hofbauer and Sigmund, pp. 142 and 274)

$$J = \begin{bmatrix} 0 & C \\ D & 0 \end{bmatrix} \tag{21}$$

and hence its eigenvalues occur in pairs $\pm\lambda$. In particular, for the equilibrium $E = (\hat{p}, \hat{q})$ of a cyclic 2×2 game, (8) can be written in the form

$$\dot{p} = (q - \hat{q})c_1(p, q) \qquad \dot{q} = (\hat{p} - p)c_2(p, q) \tag{22}$$

with $c_i > 0$, and the eigenvalues of its linearization turn out to be $\pm i\sqrt{c_1 c_2}$, in agreement with the oscillatory behavior of the solutions near E. Hence the eigenvalues of the discrete model (7) near E are given by $1 \pm i\sqrt{\gamma_1\gamma_2 c_1 c_2}$. Since their absolute value is larger than 1, E is a spiral repeller for (7) for *every* choice of the multiplier functions ϕ^k and step sizes $\gamma_k > 0$.

If the multipliers ϕ^k are decreasing functions, as in all imitation models in section 2, more can be said. It was shown in Hofbauer and Sigmund (1988, p. 282), that for Maynard Smith's choice (9), the flow generated by (8) decreases a certain volume form. The same proof actually applies to strictly decreasing [8] multiplier functions ϕ^i. For cyclic 2×2 games this result shows the global asymptotic stability of the mixed equilibrium E for (22). Using a perturbation argument from numerical dynamics, namely the upper-semicontinuity of an asymptotically stable set, see e.g. Stuart (1994), the global attractor of the discrete time model (7), with strictly decreasing ϕ^k, converges[9] to E, as $\gamma \to 0$. In other words, for small γ_k, the attractor is close to the equilibrium E. Actually it follows from the investigation in Hofbauer and Iooss (1984) that for small γ, this attractor of (7) consists of a stable 'limit cycle', i.e., a closed invariant curve surrounding the equilibrium (most likely with an irrational rotation-like dynamics on it).

5 Application to the SPOR$_n$ imitation dynamics

Now we apply the above results to the imitation models introduced in section 2. In the case $n = 1$, i.e., for (2), the multipliers are simply constants $\phi^k = (\omega^k - \alpha^k)^{-1}$. Hence $\psi' = 0$ and proposition (5) implies that the boundary is attracting. In fact it is globally attracting:

Theorem 6 *For cyclic 2×2 game all orbits (except the equilibrium) of the discrete version of the replicator dynamics (2) converge to the boundary (for every $\gamma_k \in (0, 1]$).*

[8] Compare also the proof of proposition 5 in Ritzberger and Weibull (1995, p. 1396) for an opposite result for increasing multiplier functions.

[9] One has to take care also that no invariant set bifurcates from the boundary. This can be shown in a similar way as Lemma 1, using average Liapunov functions. One can construct a uniform zone of repulsion near the boundary for small γ. However, the details of this proof are beyond the scope of this paper.

Proof. The standard continuous time replicator dynamics has a constant of motion for cyclic 2×2 games of the form $p_1^{\beta_1}p_2^{\beta_2}q_1^{\beta_3}q_2^{\beta_4}$ (with $\beta_i > 0$), see Hofbauer and Sigmund (1988). Since this function is concave, and the map (2) points into the same direction as the vector field (8),[10] it decreases monotonically along orbits of (2) and tends to 0. This shows that the boundary is a global attractor for (2). □

In the case $n = 2$, i.e., for (3), the multipliers ϕ^k are given by $\phi^k(x) = \frac{1}{\omega^k-\alpha^k}\left(1+\frac{\omega^k-x}{\omega^k-\alpha^k}\right)$. Hence $\psi(x) = \frac{(\omega^k-\alpha^k)^2}{2\omega^k-\alpha^k-x}$ and $\psi'(x) = \frac{(\omega^k-\alpha^k)^2}{(2\omega^k-\alpha^k-x)^2}$. Since ψ is convex, ψ' attains its minimum and maximum value at the ends of its domain, i.e., α and ω resp. Hence $\frac{1}{4} \le \psi'(x) \le 1$.

More generally, for finite $n \ge 2$, a longer calculation shows $\frac{n-1}{2n} \le \psi'(x) \le 1$.[11]

In the limiting case $n \to \infty$, for the map (5) we have $\phi(x) = (x-\alpha^k)^{-1}$, hence $\psi'(x) = 1$. Then proposition (4) shows that the boundary is repelling for every $\gamma^k < 1$.

Summarizing, we obtain the following results for the dynamics (4) resulting under the use of SPOR_n for given n:[12]

Theorem 7 *1. The Nash equilibrium is unstable for every cyclic 2×2 game.*

2. For $\gamma_1 = \gamma_2 = 1$ the boundary is attracting for every cyclic 2×2 game and for every n. The same holds for $n = 1$ and $0 < \gamma_k \le 1$.

3. For $\gamma_k < \frac{n-1}{2n}$ the boundary is repelling for every cyclic 2×2 game.

4. For any given cyclic 2×2 game, and any n, there is a value $\gamma(n) \in [\frac{n-1}{2n}, 1)$ with $\gamma(n) \to 1$ as $n \to \infty$ such that for $0 < \gamma_k < \gamma(n)$, the boundary is repelling, while for $\gamma(n) < \gamma_k \le 1$, the boundary is attracting.

5. For $0 < \gamma_k < \gamma(n)$, the global attractor is a closed, annulus–shaped set, disjoint from E and the boundary. If $\gamma_k = \gamma c_k$ (with c_k positive constants) then for small $\gamma > 0$, this attractor is a smooth closed invariant curve close to E.

[10] For this we have to incorporate the factors $\frac{\gamma_k}{\omega^k-\alpha^k}$ into the payoff functions π^k.

[11] Notice that the lower bound converges to $1/2$ as $n \to \infty$. If the value α^k can be attained by the payoff function π^k, then the above estimate for $\psi'(x)$ is best possible. However if there is an $\varepsilon > 0$ such that $\pi^k \ge \alpha^k + \varepsilon$ then the lower bound converges instead to 1, as $n \to \infty$. For a typical behavior see (23). The reason for this decrepancy is the non-uniformity of convergence of the multiplier functions near $\pi^k = \alpha^k$ in the limit $n \to \infty$.

[12] Under continuous dynamics (8) obtained through taking the limit $\gamma_1 = \gamma_2 \to 0$, trajectories cycle in closed orbits around E when $n = 1$ whereas E is globally attracting for $n \ge 2$, as we have seen in section 4.2.

Statement 4 follows from the monotone dependence of (13, 14, 15) on the stepsize γ and the fact that $\psi'(x) \to 1$ as $n \to \infty$ uniformly on each compact subinterval of the *open* interval (α, ω). Statement 5 follows from the discussion at the end of section 4.2.

Following Theorem 7, when $\gamma_k < \gamma(n)$ then all long run outcomes will be bounded away both from E and from the boundary. General results for the location and the form of this attracting set are not known. Only for sufficiently small γ there are rigorous results: In this case it is a smooth curve surrounding E, and the distance of this 'limit cycle' from the equilibrium behaves like a constant (depending on n) times $\sqrt{\gamma}$, so it increases faster than linear (see Hofbauer and Iooss (1984) for details). Numerical simulations suggest that the attractor remains a smooth curve for larger values of γ until it merges the boundary heteroclinic cycle. No 'fat' annuli have been observed in these simulations. Figure 2 shows the behavior of the distance of these curves from the equilibrium.

For a Matching Pennies game (6) with payoffs $\mu < \nu$ in the interval (α, ω), the critical values $\gamma(n)$ follow from the calculation in (16):

$$\gamma(n) = \frac{\psi_n(\nu) - \psi_n(\mu)}{\nu - \mu} = \frac{1}{\nu - \mu}\Big[\frac{\nu - \alpha}{1 - (\frac{\omega-\nu}{\omega-\alpha})^n} - \frac{\mu - \alpha}{1 - (\frac{\omega-\mu}{\omega-\alpha})^n}\Big]. \tag{23}$$

Figure 1 shows a plot of these values $\gamma(n)$ against $n = 1, \ldots, 10$ for $(\mu, \nu) = (1, 2)$ and payoff range $[\alpha, \omega] = [0, 3]$. The figure indicates that for $n \geq 2$ the range of values γ below the curve (for which there exists a limit cycle) is substantial and quickly grows towards 1.

Figure 2: For Matching Pennies with $(\mu, \nu) = (1, 2)$ and payoff range $[\alpha, \omega] = [0, 3]$ we plot the distance of the attracting limit cycle from the Nash equilibrium against the stepsize γ. In the left diagram we use $n = 2$, for which the boundary is reached for $\gamma \approx .4$ [13] while the right diagram shows the graph for the limit case $n = \infty$, for which the boundary is reached only at $\gamma = 1$.

[13] This value is taken from numerical simulations, while the precise value follows from (23) as $\frac{9}{20} = 0.45$. This discrepancy reflects the extremely quick approach to the boundary: the distance of the limit cycle to the boundary seems to be a flat function (all derivatives 0) at the critical value of γ.

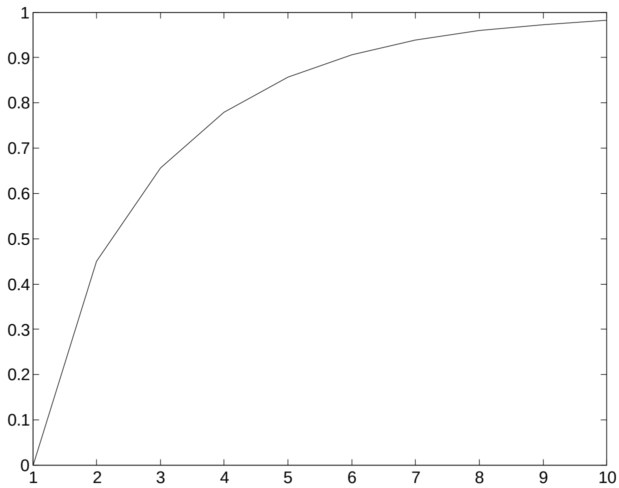

Figure 1: The critical value $\gamma(n)$ as a function of sample size n for Matching Pennies with $(\mu, \nu) = (1, 2)$ and $[\alpha, \omega] = [0, 3]$.

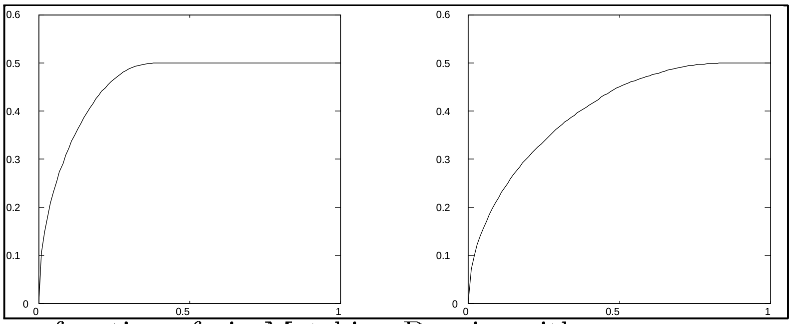

Figure 2: Distance from E as a function of γin Matching Pennies with $(\mu, \nu) = (1, 2)$and $[\alpha, \omega] = [0, 3]$; left: $n = 2$, right: $n = \infty$.

6 Conclusion

There is a basic difference in the long run dynamics of POR $=$ SPOR$_1$ on the one hand and SPOR$_n$ for $n \geq 2$ on the other hand in cyclic 2×2 games. The dynamics resulting from single sampling diverges to the boundary, leading to a strange aperiodic oscillatory behavior, with geometrically increasing sojourn times near the pure states, similar to the behavior of fictitious play in Shapley's example. In practical terms this would mean that almost surely one of these pure (non-equilibrium) states is reached and the population is stuck there because it is an absorbing state for imitators. It would need some best reply players to get away again.

In contrast, if two or more individuals are sampled, and the γ_k are not too large, then the population dynamics runs into a limit cycle repeating behavior in a periodic or quasi-periodic fashion. Moreover, this limit cycle shrinks to the Nash equilibrium when the proportion γ_k of individuals receiving information between rounds of playing the game goes to zero.

In this sense Nash equilibrium (for Matching Pennies like games) can still be justified by evolutionary arguments based on the SPOR_n model for $n \geq 2$, while it cannot for $n = 1$.

SPOR_n is a sophisticated imitation rule. Here, sophistication is in the sense that the probabilities of switching of POR are chosen in a particular way to ensure the strictly improving condition. Strictly improving requires "nice" behavior in a large variety of environments and one might wonder which alternative behavior selected specifically for a given cyclic game might stabilize the Nash equilibrium too. In the following we argue that both sample sizes larger than one and probabilistic strategy selection remain necessary ingredients. Indeed, divergence under single sampling is more general than appears at first sight. Schlag (1998a) shows that trajectories converge to the boundary under single sampling (as in Theorem 6) whenever the conditions on strictly improving stated in (1) hold in the given cyclic 2×2 game. The popular, simple, and seemingly more intuitive, imitation rules Imitate The Best (Axelrod, 1984) and Imitate Best Average (Ellison and Fudenberg, 1993) generally perform worse than SPOR_n in cyclic 2×2 games. For these deterministic behavioral rules it is easily shown that the Nash equilibrium even fails to be a rest point for almost all cyclic 2×2 games. This is due to the insensitivity of these rules to small payoff changes. This result and intuition also applies when individuals play a best response to a finite sample of previous strategies among their potential opponents (see also Schlag, 1998b, 1998a).

Some sophistication on part of the players is required to lead them away from unreasonable behavior and bring them closer to Nash equilibrium. It may be worth pointing out another case, where sophistication has a stabilizing effect: In biology, (2) and related (discrete time) models are based on the not always realistic assumption of haploid, asexual individuals which reproduce by cloning. Taking account of the genetic structure, namely diploidity, again leads to a stable limit cycle in the games [14] considered here, see Maynard Smith and Hofbauer

[14] Cyclic 2×2 games arise in biology as the 'battle of the sexes' game introduced by Dawkins

(1987), or Hofbauer and Sigmund (1988, p. 312 ff).

Finally, notice that the stabilizing behavior of double (or more) sampling in cyclic 2×2 games also has an important impact on the following "larger" game. Oechssler and Schlag (1997) find for a simple extensive-form game that the backwards induction outcome ceases to be stable under (single sampling) sophisticated imitation once observations are imperfect. Our analysis can be used to show that stability is recovered under double (or more) sampling.

7 References

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

J. Björnerstedt, and J. Weibull (1996): Nash equilibrium and evolution by imitation. In: *The Rational Foundations of Economic Behaviour* (K. Arrow and E. Colombatto, eds.). Macmillan.

Börgers, T. and R. Sarin (1997): Learning thorugh reinforcement and replicator dynamics. *J. Econ. Theory* **77**, 1-14.

Brown G. W. (1951): Iterative solution of games by fictitious play. In: *Activity Analysis of Production and Allocation* (T.C. Koopmans, ed), pp. 374–376. Wiley, New York.

Ellison, G. and D. Fudenberg (1995): Word-of-mouth communication and social learning, *Quart. J. Econ.* **440**, 93-125.

Friedman, D. (1991): Evolutionary games in economics. *Econometrica* **59**, 637-66.

Gale, J., K. G. Binmore, and L. Samuelson (1995): Learning to be imperfect: the Ultimatum Game. *Games Econ. Beh.* **8**, 56-90.

Gaunersdorfer, A. and J. Hofbauer (1995): Fictitious play, Shapley polygons, and the replicator equation. *Games and Economic Behaviour* **11**, 279–303.

Hofbauer, J. (1995): Stability for the best response dynamics. Mimeo, University of Vienna.

Hofbauer, J. and G. Iooss (1984): A Hopf bifurcation theorem for difference equations approximating a differential equation. *Monatsh. Math.* **98**, 99-113.

Hofbauer, J. and K. Sigmund (1988): *The Theory of Evolution and Dynamical Systems.* Cambridge University Press.

as a model of the parental investment conflict, see Maynard Smith (1982).

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

Maynard Smith, J. and J. Hofbauer (1987): The "battle of the sexes": a genetic model with limit cycle behavior. *Theor. Pop. Biol.* **32**, 1–14.

Oechssler, J. and K. H. Schlag (1997): Loss of commitment? An evolutionary analysis of Bagwell's example. Mimeo, University of Bonn.

Posch M. (1997): Cycling in a stochastic learning algorithm for normal form games. *J. Evolutionary Economics* **7**, 193–207.

Ritzberger, K. and J. W. Weibull (1995): Evolutionary selection in normal-form games. *Econometrica* **63**, 1371–1399.

Samuelson, L. and J. Zhang (1992): Evolutionary stability in asymmetric games. *J. Economic Theory* **57**, 363–391.

Schlag K. H. (1998a): Social roles as an effective learning mechanism. Mimeo, University of Bonn.

Schlag K. H. (1998b): Which one should I imitate? *J. Mathematical Economics* (forthcoming).

Schlag K. H. (1998c): Why imitate, and if so, how? A bounded rational approach to multi-armed bandits. *J. Economic Theory* (forthcoming).

Selten R. (1980): A note on evolutionarily stable strategies in asymmetric animal conflicts. *J. Theor. Biol.* **84**, 93-101.

Shapley L. (1964): Some topics in two–person games. *Ann. Math. Studies* **5**, 1–28.

Stuart A. (1994): Numerical analysis of dynamical systems. *Acta Numerica* **4**, 467–572.

Taylor, P. D. (1979): Evolutionarily stable strategies with two types of players. *J. Appl. Prob.* **16**, 76–83.

Weibull, J. (1995): *Evolutionary Game Theory.* Cambridge, MA: MIT Press.