Sophisticated imitation in cyclic games[*]

Josef Hofbauer[1] **and Karl H. Schlag**[2]

[1] Department of Mathematics, University of Vienna, Strudlhofgasse 4, A-1090 Wien, Austria
(e-mail: josef.hofbauer@univie.ac.at)
[2] Economics Department, European University Institute, Via dei Roccettini 9,
50016 San Domenico di Fiesole, Italy (e-mail: schlag@iue.it)

Abstract. Individuals belonging to two large populations are repeatedly randomly matched to play a cyclic 2×2 game such as Matching Pennies. Between matching rounds, individuals sometimes change their strategy after observing a finite sample of other outcomes within their population. Individuals from the same population follow the same behavioral rule. In the resulting discrete time dynamics the unique Nash equilibrium is unstable. However, for sample sizes greater than one, we present an imitation rule where long run play cycles closely around the equilibrium.

Key words: Strictly improving – Matching pennies game – Replicator dynamics – Limit cycle – Discretization

JEL-classification: C79

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 a 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 predicted. This phenomenon first appeared when Selten (1980) showed that Evolutionarily Stable Strategies of truly asymmetric contests cannot be mixed.

More recently, focus has shifted away from static concepts and now lies more on analyzing the population dynamics directly. However, none of the multi-population models based on boundedly rational individual behavior with limited individual information and memory have been able to predict mixed equilibria. In fact, most of them generate the same population dynamics, namely the *standard replicator dynamics* known from population biology (Björnerstedt and Weibull, 1995; Börgers and Sarin, 1997; Gale et al., 1995; Posch, 1997; Schlag, 1998).

Only models in which individuals have substantial information or memory are known to predict mixed equilibria. Björnerstedt and Weibull (1996) present a model where individuals imitate others but need to know the average payoff of each strategy being played in the population in order to know when to imitate. This generates the *adjusted replicator dynamics* for which it is known that the mixed equilibrium of Matching Pennies is asymptotically stable (Maynard Smith, 1982, Appendix J). *Fictitious play* (Brown, 1951) and the *continuous best response dynamics* (Hofbauer, 1995a) will also predict the mixed equilibrium in Matching Pennies, and more generally in games that are equivalent to zero-sum games. Under fictitious play, individuals remember all previous moves of their opponent, while the continuous time best response dynamics models a scenario in which each individual knows the current distribution of pure strategies in the opponent population(s).

The instability of interior equilibria under the more realistic scenarios of information and memory have biased the evolutionary game theory literature to focus on pure strategy equilibria and on set-valued solution concepts (cf. Ritzberger and Weibull, 1995). However these approaches are of little use in Matching Pennies: they either select the empty set or the entire strategy space. One might then come to the conclusion that only individuals with substantial information and/or extensive memory will learn to play a mixed equilibrium. We show that this conclusion is not true; individuals only need to observe two other individuals, to remember own outcome in the previous round and then to update appropriately — imitating others in a sophisticated manner.

We consider discrete time evolutionary dynamics in cyclic 2×2 games such as Matching Pennies where it is assumed that individuals play pure strategies. We will analyze the properties of *discrete time aggregate monotone dynamics* (Samuelson and Zhang, 1992) both in general and specifically when they can be generated from the following micro model. Assume that individuals from two large (i.e., infinite) populations are repeatedly randomly matched to play a two person game such as a cyclic 2×2 game. Some individuals receive the opportunity to change their strategy between rounds. Each of these observes a random sample of the previous play and payoff of n other individuals from the population. n will be called the sample size.

In this micro model, we assume that individuals belonging to the same population use the same behavioral rule. We restrict attention to individual behavior (what to play next) that is only a function of the information in the sample and of the own choice and payoff in the previous round. An example of such a behavioral rule is the *sequentially evaluated Proportional Observation Rule* (SPOR$_n$,

Schlag, 1999). Given this imitation rule, an individual considers each individual in his sample separately, switching to the strategy observed with a probability that is proportional to the observed payoff. This seemingly unintuitive behavior, to ignore own payoff, is uniquely selected by Schlag (1999) who models optimal individual boundedly rational behavior. We review the underlying arguments from a new perspective in the appendix.

We find that when all individuals use $SPOR_n$, then the resulting population dynamics is an *aggregate payoff monotone dynamics* — a class of evolutonary dynamics closely related to the replicator dynamics. This fact was previously only known for $n = 1$ and $n = 2$ (Schlag, 1998, 1999). Interestingly enough, the dynamics generated by $SPOR_n$ tends to Maynard Smith's adjusted replicator dynamics when the sample size n grows to infinity.

Before analyzing these dynamics, which are technically more involved as they are in discrete time, we consider as a benchmark the usual continuous time version which arises in the limit as the fraction of sampling individuals tends to zero. Pollock and Schlag (1999) show that payoff monotonicity (which is considerably more general than aggregate payoff monotonicity) and sample size one induces continuous time dynamics that are — at least for 2×2 games — still very similar to the standard replicator dynamics. This enables them to show that for cyclic 2×2 games all solutions are periodic and cycle around the equilibrium, compare also Hofbauer (1995b). In particular, this means that none of the behavioral rules based on a single observation predicts play of the mixed equilibrium.

We find that a particular class of continuous time aggregate monotone dynamics, such as those induced by $SPOR_n$ for $n \geq 2$, make the mixed equilibrium globally asymptotically stable. The condition for this result is that their scaling factors are strictly decreasing in the current average population payoffs. Thus, $SPOR_n$ ($n \geq 2$) is the first example of an individual rule that induces a payoff monotone dynamics that predicts play of a mixed equilibrium in the long run. This is not a complete coincidence even though $SPOR_n$ was developed for its local performance. Behavioral rules with sample size one selected in (Schlag, 1998) lead to the standard replicator dynamics which has closed orbits in the continuous version. $SPOR_n$ is selected in (Schlag, 1999) as it leads to a stronger increase in play of the current best response than under the selected single sample rules, thus pushing the dynamics towards the mixed equilibrium.

It is important to consider the continuous time dynamics only as a benchmark as almost all realistic learning processes take place in discrete time. Hence we return to the discrete time dynamics. For sample size one, we follow Pollock and Schlag (1999) to show that trajectories starting close to, but not in the mixed equilibrium spiral to the border. Our investigation shows that the mixed equilibrium remains unstable under any aggregate payoff monotone dynamic, much in the spirit of Eshel and Akin (1983) and Crawford (1985). However, although the population state is always pushed away from the mixed equilibrium, we find that the discrete time trajectories of those dynamics that lead to play of the mixed equilibrium in continuous time (such as $SPOR_n$ for $n \geq 2$) will stay

close if the fraction γ of individuals updating their play in each round is small.[1]
Consequently, mixed strategy equilibria need not be ruled out as predictors under
simple but sophisticated rules of behavior. At the same time we cannot expect
the long run population state to be exactly in the mixed Nash equilibrium. While
larger sample sizes under SPOR$_n$ tend to decrease the distance of long run play
from the equilibrium, we show that this distance is bounded away from 0 for
any given γ.

Some intuition may be helpful. The condition of aggregate monotonicity
induces a specific structure of the linearized dynamics near the equilibrium.
Movement within one population only depends on play in the other which causes
trajectories to cycle in continuous time and spiral out in discrete time. This
instability of the equilibrium under the linearized dynamics cannot be overturned
by considering the explicit nonlinear dynamics. However, differences between
the linearized and the actual dynamics arise when the state is not close to the
equilibrium. These differences depend on the rule used. The trajectories of the
discrete time dynamics approximate those of the continuous time dynamics when
γ is small. As the continuous time dynamics under SPOR$_n$ for $n \geq 2$ lead
trajectories into the equilibrium, their discrete time dynamics will lead trajectories
close to the equilibrium when γ is small. Our calculations reveal that $\gamma \leq \frac{n-1}{2n}$
already suffices to lead trajectories under SPOR$_n$ away from the boundary.

The paper is organized as follows. Section 2 introduces how payoffs are
realized and how individuals can learn. SPOR$_n$ is introduced and the associated
population dynamics are derived. Section 3 defines cyclic games and introduces
aggregate monotone dynamics. While continuous time dynamics are analyzed
in Section 4, Section 5 contains the more interesting and relevant discrete time
analysis. It is divided into subsections concerned with dynamics close to the
boundary and close to the equilibrium before considering the specific dynamics
induced by SPOR$_n$. Section 6 contains the conclusion and the appendix contains
more details on the foundations of SPOR$_n$.

2 Playing and updating

Consider a two person normal form (or *bimatrix*) game Γ in which player k
chooses a pure strategy from $S^k = \{1, \ldots, n_k\}$ where the payoff $\pi^k(i,j)$ that
player k receives after play $(i,j) \in S^1 \times S^2$ is contained in a bounded open
interval (α^k, ω^k). Let $\Delta(S^k)$ 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
$\mathbf{q} \in \Delta(S^2)$. $\mathbf{e}_r \in \Delta(S^k)$ denotes the choice of pure strategy r with probability
one.

In the following we describe a model of boundedly rational individuals play-
ing this game. Assume that individuals have fixed roles in the game Γ. They
belong to two infinite populations, one corresponding to each player in the game.

---

[1] In the paper, we allow for different fractions $\gamma_k \in (0,1]$ in each population k, $k = 1, 2$.

In a sequence of rounds, individuals from each population are randomly matched to play the game.[2]

A *state* $(\mathbf{p}, \mathbf{q}) \in \Delta\left(S^1\right) \times \Delta\left(S^2\right)$ 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 $(\mathbf{p}, \mathbf{q})$.

Between rounds, each individual belonging to population k independently receives the opportunity to change or update his play with probability γ_k. With probability $1 - \gamma_k$ he must play the same pure strategy again. We will refer to $1 - \gamma_k$ as the *level of inertia* in population k.

An updating individual gathers information by sequentially and independently *sampling n* individuals from the same population and observes the (pure) strategy used and the payoff obtained in the previous round by each of these individuals $(n \in \mathbb{N}, 0 < \gamma_k \leq 1, k = 1, 2)$. Sampling probabilities are modelled as if populations were finite and each individual not yet observed is sampled equally likely. Formally, the probability that an individual from population 1 with the opportunity to update after round t observes an individual who obtained the payoff $\pi^1 (i, j)$ is set equal to $p_i^t q_j^t$. The case $n = 1$ is called *single sampling*, $n = 2$ *double sampling*.

The individual's decision what to play in the next round as a function of his previous information will be called his *behavioral rule*. We will restrict attention to simple behavioral rules that only depend on information gathered since the previous round and not on previous information or play nor on the number of the round. For example, a behavioral rule under single sampling 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 obtained payoff y. More generally, a behavioral rule for sample size n is a map

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

Throughout we will be concerned with common updating where all individuals belonging to the same population follow the same behavioral rule. In this way, the population dynamics given individual behavior is uniquely described by the change of the state over time, as given by recurrence equations

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

for appropriately chosen functions $f_i, g_j : \Delta\left(S^1\right) \times \Delta\left(S^2\right) \to \mathbb{R}$ $(i \in S^1, j \in S^2)$, such that $(\mathbf{p}^t, \mathbf{q}^t) \mapsto (\mathbf{p}^{t+1}, \mathbf{q}^{t+1})$ defines a map from $\Delta(S^1) \times \Delta(S^2)$ into itself.

---

[2] As the populations are infinite, this means that an individual in population 1 is matched with an individual playing j in population 2 with probability q_j in state $(\mathbf{p}, \mathbf{q})$.

2.1 An important example

Consider the following simple algorithm to construct a behavioral rule for general sample sizes n using a single sampling rule F. 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 rule F. In the following we will demonstrate this procedure and the resulting dynamic equations in more detail for the Proportional Observation Rule.

The *Proportional Observation Rule* $\tilde{F}$ (POR, Schlag, 1999) for an individual in population k is defined by

$$\tilde{F}(i,x,i,y)_i = 1$$

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

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 all individuals use the respective POR the population state changes according to (a discrete version of) the *standard replicator dynamics* of Taylor (1979) (with step-sizes γ_k, see Schlag, 1998):

$$p_i^{t+1} = p_i^t + \frac{\gamma_1}{\omega^1 - \alpha^1}\left[\pi^1\left(i, \mathbf{q}^t\right) - \pi^1\left(\mathbf{p}^t, \mathbf{q}^t\right)\right] \cdot p_i^t$$

$$q_j^{t+1} = q_j^t + \frac{\gamma_2}{\omega^2 - \alpha^2}\left[\pi^2\left(\mathbf{p}^t, j\right) - \pi^2\left(\mathbf{p}^t, \mathbf{q}^t\right)\right] \cdot q_j^t \qquad (2)$$

Next we sequentially evaluate POR to obtain a double sampling rule $\tilde{F}^{(2)}$ (which we also will denote by SPOR$_2$) for $n = 2$. It follows, for $i \neq j$ and $r \neq j$ that

$$\tilde{F}^{(2)}(i,x,j,y,j,z)_j = \frac{y - \alpha^k}{\omega^k - \alpha^k} + \frac{\omega^k - y}{\omega^k - \alpha^k}\frac{z - \alpha^k}{\omega^k - \alpha^k}$$

$$\tilde{F}^{(2)}(i,x,r,y,j,z)_j = \frac{z - \alpha^k}{\omega^k - \alpha^k}$$

$$\tilde{F}^{(2)}(i,x,j,y,r,z)_j = \frac{y - \alpha^k}{\omega^k - \alpha^k}\frac{\omega^k - z}{\omega^k - \alpha^k}.$$

When all individuals use the SPOR$_2$ applied to the appropriate payoff range $\left(\alpha^k, \omega^k\right)$ then we obtain the population dynamics (see Schlag, 1999)

$$p_i^{t+1} = p_i^t + \frac{\gamma_1}{\omega^1 - \alpha^1}\left(1 + \frac{\omega^1 - \pi^1\left(\mathbf{p}^t, \mathbf{q}^t\right)}{\omega^1 - \alpha^1}\right)\left[\pi^1\left(i, \mathbf{q}^t\right) - \pi^1\left(\mathbf{p}^t, \mathbf{q}^t\right)\right]p_i^t$$

$$q_j^{t+1} = q_j^t + \frac{\gamma_2}{\omega^2 - \alpha^2}\left(1 + \frac{\omega^2 - \pi^2\left(\mathbf{p}^t, \mathbf{q}^t\right)}{\omega^2 - \alpha^2}\right)\left[\pi^2\left(\mathbf{p}^t, j\right) - \pi^2\left(\mathbf{p}^t, \mathbf{q}^t\right)\right]q_j^t. \quad (3)$$

When verifying (3), it is helpful to consider switching behavior even when an individual observes someone using the same strategy. Then the additional term

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

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.

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

$$p_i^{t+1} - p_i^t = \frac{\gamma_1}{\omega^1 - \alpha^1} \left[1 + \frac{\omega^1 - \pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right)}{\omega^1 - \alpha^1} + \ldots + \left(\frac{\omega^1 - \pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right)}{\omega^1 - \alpha^1}\right)^{n-1}\right]$$

$$\cdot \left[\pi^1 \left(i, \mathbf{q}^t\right) - \pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right)\right] \cdot p_i^t$$

$$q_j^{t+1} - q_j^t = \frac{\gamma_2}{\omega^2 - \alpha^2} \left[1 + \frac{\omega^2 - \pi^2 \left(\mathbf{p}^t, \mathbf{q}^t\right)}{\omega^2 - \alpha^2} + \ldots + \left(\frac{\omega^2 - \pi^2 \left(\mathbf{p}^t, \mathbf{q}^t\right)}{\omega^2 - \alpha^2}\right)^{n-1}\right]$$

$$\cdot \left[\pi^2 \left(\mathbf{p}^t, j\right) - \pi^2 \left(\mathbf{p}^t, \mathbf{q}^t\right)\right] \cdot q_j^t. \tag{4}$$

The population dynamics induced by SPOR_n satisfies

$$\pi^1 \left(\mathbf{p}^{t+1}, \mathbf{q}^t\right) - \pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right) = \sum_{i \in S^1} \left(p_i^{t+1} - p_i^t\right) \left[\pi^1 \left(i, \mathbf{q}^t\right) - \pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right)\right]$$

$$= \frac{\gamma_1}{\omega^1 - \alpha^1} \left[1 + \frac{\omega^1 - \pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right)}{\omega^1 - \alpha^1} + \ldots + \left(\frac{\omega^1 - \pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right)}{\omega^1 - \alpha^1}\right)^{n-1}\right]$$

$$\cdot \sum_{i \in S^1} \left[\pi^1 \left(i, \mathbf{q}^t\right) - \pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right)\right]^2 \cdot p_i^t \geq 0$$

where $\pi^1 \left(\mathbf{p}^{t+1}, \mathbf{q}^t\right) > \pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right)$ whenever $\pi^1 \left(i, \mathbf{q}^t\right) > \pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right)$ holds for some i with $p_i^t > 0$. This property of SPOR_n is called *strictly improving* and is the starting point that leads Schlag (1999) to argue (also from an individual standpoint) that this rule is the best sequentially evaluated single sampling rule for $n = 2$. In the appendix we rephrase the essential arguments in a new light and add why sampling more is better under this rule (Proposition 6).

In the limit $n \to \infty$, (4) converges to (a discretization of) the so-called *adjusted replicator dynamic* introduced by Maynard Smith (1982, Appendix J):

$$p_i^{t+1} = p_i^t + \gamma_1 \cdot p_i^t \cdot \frac{\pi^1 \left(i, \mathbf{q}^t\right) - \pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right)}{\pi^1 \left(\mathbf{p}^t, \mathbf{q}^t\right) - \alpha^1}$$

$$q_j^{t+1} = q_j^t + \gamma_2 \cdot q_j^t \cdot \frac{\pi^2 \left(\mathbf{p}^t, j\right) - \pi^2 \left(\mathbf{p}^t, \mathbf{q}^t\right)}{\pi^2 \left(\mathbf{p}^t, \mathbf{q}^t\right) - \alpha^2}. \tag{5}$$

3 Cyclic 2×2 games and aggregate monotonicity

In the remaining sections we restrict attention to cyclic 2×2 games. We call the bimatrix game Γ with $|S^1| = |S^2| = 2$ a *cyclic 2×2 game* if the four pure states form a best response cycle, or equivalently, if it has a unique Nash equilibrium $E = (\hat{\mathbf{p}}, \hat{\mathbf{q}})$ where E is in the interior of $\Delta(S^1) \times \Delta(S^2) = [0, 1]^2$. Examples include the buyer–seller game (Friedman, 1991; see also Pollock and Schlag, 1999), the model by Cressman et al. (1998) of crime deterrence, 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 can lead large populations (close) to a completely mixed Nash equilibrium.

General discrete time dynamics as given in (1) (not necessarily derived from the sampling scenario in Section 2) are *aggregate monotone regular selection dynamics* (Samuelson and Zhang, 1992) if there exist strictly positive continuous functions ϕ^k such that (1) can be rewritten as:

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

The dynamics (4) induced by SPOR_n belongs to this class, with the factors ϕ^k depending on the state $(\mathbf{p}, \mathbf{q})$ only through the average payoffs: $\phi^k = \Phi^k \circ \pi^k$ for some functions Φ^k. As a benchmark we will also consider the continuous time version of (7) that results, after replacing $t + 1$ in (7) by $t + \Delta t$ and setting $\gamma_1 = \gamma_2 = \Delta t$, as the limit $\Delta t \to 0$, and is given by the system of differential equations

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

The first aggregate monotone 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

$$
\phi^k(\mathbf{p}, \mathbf{q}) = \frac{1}{C_k + \pi^k(\mathbf{p}, \mathbf{q})},
\tag{9}
$$

the C_k being positive constants (standing for background fitness). However, no convincing derivation has been given for this choice. As seen from (5) in Section 2, this choice arises now from the imitation rule SPOR_n in the limit $n \to \infty$.

4 The continuous time limit

It is popular in evolutionary game theory to simplify analysis by taking appropriate limits and transforming discrete dynamics into continuous time dynamics. For us this approach helps understanding the discrete time dynamics in the next section.

It is well known that the trajectories of the continuous time version (8) of the population dynamics (2) for POR $\equiv$ SPOR$_1$ form closed orbits around E. Hofbauer (1995b) and Pollock and Schlag (1999) extend this to more general imitation rules with $n = 1$. Our next result shows that the mixed equilibrium can be the unique predictor when $n \geq 2$. This happens for aggregate monotone dynamics (8) where the multipliers are strictly decreasing functions of the average population payoff.

Theorem 1 *If $\phi^k \equiv \Phi^k \circ \pi^k$ for some strictly decreasing strictly positive functions Φ^k, as under SPOR$_n$ for $n \geq 2$, or Maynard Smith's choice (9), then the mixed equilibrium E is globally asymptotically stable for (8).*

While the proof is more technical we can provide some intuition behind this result for SPOR$_n$. Trajectories under the continuous best response dynamics converge to E (Brown, 1951; Hofbauer, 1995a) while SPOR$_1$ just barely does not increase best responses enough as it induces the standard replicator dynamics which have closed orbits around E. Proposition 6 in the appendix implies that SPOR$_n$ for $n \geq 2$ leads in almost all states to a stronger increase of the best response in 2×2 games than SPOR$_1$.

Proof. It was shown in Hofbauer and Sigmund (1988, p. 282), that for Maynard Smith's choice (9), after dividing the vectorfield (8) by the common factor $\prod p_i \prod q_j$ (which has only the effect of a change of time scale), the divergence is negative and hence the flow on int $\Delta(S^1) \times \Delta(S^2)$ decreases volume. The same calculation actually applies to strictly decreasing multiplier functions Φ^k.[3] For 2×2 games this means that the flow of (8) is area decreasing in $(0, 1)^2$. The theorems of Poincaré–Bendixson and Bendixson–Dulac (see Hofbauer and Sigmund, 1988) imply then the global asymptotic stability of the mixed equilibrium E for cyclic 2×2 games. $\square$

5 The discrete time dynamics

Next we consider the discrete time aggregate monotone dynamics (7). Pollock and Schlag (1999) suggest that for $n = 1$ all orbits (except the equilibrium E) converge to the boundary (for every $\gamma_k \in (0, 1]$), see Proposition 4 below for more details. This implies that most of the time, all individuals in the same population are playing the same pure (non-equilibrium) strategy. Thus, we can only expect predictive power of E if $n \geq 2$.

---

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

Before going into details we briefly summarize our results. While the Nash equilibrium E remains locally unstable for any aggregate monotone dynamics, we get more positive results for the dynamics (4) induced by SPOR$_n$. The proportion γ_k of individuals in each population receiving information plays an important role. For a given sample size $n \geq 2$, if too many individuals receive information (i.e., if γ_k are too large) then the boundary remains attracting. However, if $\gamma_k \leq \frac{n-1}{2n}$ then the boundary is repelling in all cyclic 2×2 games and long run play is closer to the equilibrium than under a single sample. For Matching Pennies games we derive the much better correct bound in (19). If γ_k is sufficiently small then the long run states form a limit cycle which is close to E. Simulations are added in Section 5.3 to exemplify the specific relations. There we observe that a limit cycle is reached whenever the boundary is repelling.

5.1 Behavior near the boundary

In the following we present sufficient conditions for determining the behavior of the dynamics close to the boundary of the square $\{(\mathbf{p}, \mathbf{q}) \,|\, p_1 \cdot p_2 \cdot q_1 \cdot q_2 = 0\}$.

At the corner $(\mathbf{e}_i, \mathbf{e}_j)$ we find $(\mathbf{e}_r - \mathbf{e}_i, \mathbf{0})$ to be the eigenvector corresponding to the eigenvalue

$$\lambda_{ij \to rj} := 1 + \gamma_1 \left[\pi^1(r,j) - \pi^1(i,j)\right] \phi^1(i,j) \tag{10}$$

and $(\mathbf{0}, \mathbf{e}_l - \mathbf{e}_j)$ the eigenvector corresponding to the eigenvalue

$$\lambda_{ij \to il} := 1 + \gamma_2 \left[\pi^2(i,l) - \pi^2(i,j)\right] \phi^2(i,j) \,. \tag{11}$$

Notice that $\lambda_{ij \to rj}$ is the growth rate p_r^{t+1}/p_r^t when almost all individuals in population one play i and the rest play r and all individuals in population two play j. Since payoffs to player k are from the open interval (α^k, ω^k), the eigenvalues (10, 11) are positive for (4) and (5) and all $0 \leq \gamma_k \leq 1$.

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

$$\pi^1(2,1) < \pi^1(1,1), \quad \pi^2(1,1) < \pi^2(1,2),$$
$$\pi^1(1,2) < \pi^1(2,2), \quad \pi^2(2,2) < \pi^2(2,1). \tag{12}$$

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

$$\rho = \rho_{11 \to 12} \cdot \rho_{12 \to 22} \cdot \rho_{22 \to 21} \cdot \rho_{21 \to 11} \tag{13}$$

with (β, δ denoting pairs of numbers belonging to $\{1, 2\}$)

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

Note that ρ is the product of logs of 'outgoing' eigenvalues divided by the product of logs of 'incoming' eigenvalues around the cycle (12). ρ is well-defined since the eigenvalues (10, 11) are different from 1 whenever $\gamma_k > 0$. The following is a discrete time version of a related result for differential equations, see e.g. Hofbauer and Sigmund (1988, p. 213f).

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

Proposition 1 *Assume that the multipliers in the aggregate monotone dynamics (7) depend only on average payoffs, i.e., have the form $\phi^k \equiv \Phi^k \circ \pi^k$ and let $\psi^k(x) := 1/\Phi^k(x)$. 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 $(\mathbf{p}^0, \mathbf{q}^0)$, 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 $\rho_{12\to22} > 1$, the other three factors in (13) being analogous. Inserting (10) into (14) we obtain

$$\rho_{12\to22} = \frac{\log(1 + \gamma_1(\pi^1(2,2) - \pi^1(1,2))\Phi^1\left(\pi^1(1,2)\right))}{|\log(1 + \gamma_1(\pi^1(1,2) - \pi^1(2,2))\Phi^1\left(\pi^1(2,2)\right))|} \qquad (15)$$

Denote $a := \pi^1(2,2) - \pi^1(1,2) > 0$ and $b := \pi^1(1,2)$. Then the inequality $\rho_{12\to22} > 1$ can be equivalently reformulated as

$$
\begin{aligned}
\rho_{12\to22} = \frac{\log(1 + \gamma_1 a\Phi^1(b))}{-\log(1 - \gamma_1 a\Phi^1(b+a))} &> 1 \\
\log(1 + \gamma_1 a\Phi^1(b)) &> -\log(1 - \gamma_1 a\Phi^1(b+a)) \\
(1 + \gamma_1 a\Phi^1(b))(1 - \gamma_1 a\Phi^1(b+a)) &> 1 \qquad (16) \\
\Phi^1(b) - \Phi^1(b+a) &> \gamma_1 a\Phi^1(b)\Phi^1(b+a) \\
\psi^1(b+a) - \psi^1(b) &> \gamma_1 a
\end{aligned}
$$

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

A similar calculation shows:

Proposition 2 *Under the assumptions of Proposition 1, 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.*

5.2 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, $\gamma = diag(\gamma_1, \gamma_2)$, and J the linearization of the corresponding differential equation (8) at E. For 2×2 games, (8) can be written (with $p = p_1$ and $q = q_1$) as

$$\begin{aligned}
\dot{p} &= p(1-p)\big(\pi^1(1, \mathbf{q}) - \pi^1(2, \mathbf{q})\big)\phi^1\left(\mathbf{p}^t, \mathbf{q}^t\right) \\
\dot{q} &= q(1-q)\big(\pi^2(\mathbf{p}, 1) - \pi^2(\mathbf{p}, 2)\big)\phi^2\left(\mathbf{p}^t, \mathbf{q}^t\right).
\end{aligned} \tag{17}$$

Given an interior equilibrium $E = (\hat{p}, \hat{q})$ of a 2×2 game, (17) can be written in the form

$$\dot{p} = (q - \hat{q})c_1(p, q) \qquad \dot{q} = (\hat{p} - p)c_2(p, q) \tag{18}$$

where the factors $c_i(p, q)$ (which are essentially the ϕ^i) have constant sign on $(0, 1)^2$. For a cyclic 2×2 game (12) both $c_i(p, q)$ are positive. From (18) one can read off the linearization at E, $J = \begin{pmatrix} 0 & c_1 \\ -c_2 & 0 \end{pmatrix}$, with $c_i = c_i(\hat{p}, \hat{q})$ for short. Hence the eigenvalues of J are the imaginary numbers $\pm i \sqrt{c_1 c_2}$, showing the oscillatory behavior of the solutions near E. Hence the eigenvalues of the discrete model (7) at E are given by $1 \pm i \sqrt{\gamma_1 \gamma_2 c_1 c_2}$. Since their absolute value is larger than 1 we obtain (compare also Eshel and Akin, 1983; Crawford, 1985):

Proposition 3 *E is a spiral repeller for (7) for every choice of the multiplier functions ϕ^k and of the step sizes $\gamma_k > 0$.*

We provide some geometric intuition for this result. Aggregate monotonicity causes there to be almost no change of frequencies in population 1 (2) when q_1 (p_1) is near the equilibrium value $\hat{q}_1$ ($\hat{p}_1$). This eliminates own population effects, nearby the equilibrium the dynamics in one population are mainly driven by the frequencies in the other population. This causes first order effects (determined by the linearization) of the continuous dynamics near the equilibrium to lead to closed orbits. The discrete dynamics moves tangent to the trajectories of the continuous dynamics, so in each round the first order effect of the dynamics crosses orbits of the continuous dynamics, moving outwards away from the equilibrium. Thus, the first order effects of the discrete dynamics spiral out which carries over to the actual discrete dynamics to yield our result that E is a spiral repeller. In other words, the first order effects of an aggregate monotone dynamics in a cyclic game have the same properties as the dynamics of the standard replicator dynamic: in the continuous version E is neutrally stable, in the discrete version it is a spiral repeller.

The above is a purely local result, however. If the resulting dynamics is aggregate monotone with multipliers Φ^k that are decreasing functions, as under $SPOR_n$ or the Maynard Smith dynamics (5), we can say more. Recall the global stability result for the continuous dynamics given in Theorem 1. Using a perturbation argument from numerical dynamics, namely the upper-semicontinuity of an asymptotically stable set (see e.g. Stuart, 1994), there is an attractor $A(\gamma)$ in

the discrete time model (7), with strictly decreasing Φ^k, which converges to E, as $\gamma = \gamma_1 = \gamma_2 \to 0$. In other words, for small γ, the attractor[4] of (7) is close to the equilibrium E. Furthermore, it follows from the investigation in Hofbauer and Iooss (1984) that for small γ, this attractor $A(\gamma)$ of (7) is a stable 'limit cycle', i.e., a closed invariant curve surrounding the equilibrium (most likely with an irrational rotation–like dynamics on it).

5.3 Application to the SPOR$_n$ imitation dynamics

Now we apply the above results to the imitation models SPOR$_n$ 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 2 implies that the boundary is attracting. In fact it is globally attracting:

Proposition 4 *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 $\gamma_k \in (0, 1]$).*

Proof. If we incorporate the factors $\frac{\gamma_k}{\omega^k - \alpha^k}$ into the payoff functions π^k then the increment of the map (2) is precisely the vector field (8) or (17), with $\phi^k \equiv 1$, of the standard replicator dynamics. For cyclic 2×2 games this has a constant of motion of the form $P = p_1^{\beta_1} p_2^{\beta_2} q_1^{\beta_3} q_2^{\beta_4}$ (with $\beta_i > 0$), see Hofbauer and Sigmund (1988). Since this function P is strictly concave, its tangent vector field (17) points everywhere outwards. Therefore P decreases monotonically along each trajectory of (2) (except E and on the boundary) and tends to 0. This shows that the boundary is a global attractor for (2). □

In the case $n = 2$, i.e., for the aggregate monotone dynamics (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^k and ω^k resp. Hence $\frac{1}{4} < \psi'(x) < 1$.

More generally, for finite $n \geq 2$, a longer calculation shows $\frac{n-1}{2n} < \psi'(x) < 1$.

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 1 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:[5]

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

---

[4] General theory implies also that the basin of attraction of $A(\gamma)$ contains most of the state space, e.g. $(\delta, 1 - \delta)^2 \setminus E$ with $\delta = \delta(\gamma) \to 0$ as $\gamma \to 0$. It is very likely that $A(\gamma)$ attracts *all* interior orbits except E. However, the technicalities of such a proof are beyond the scope of this paper.

[5] 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 $n \geq 2$, as we have seen in Section 4.
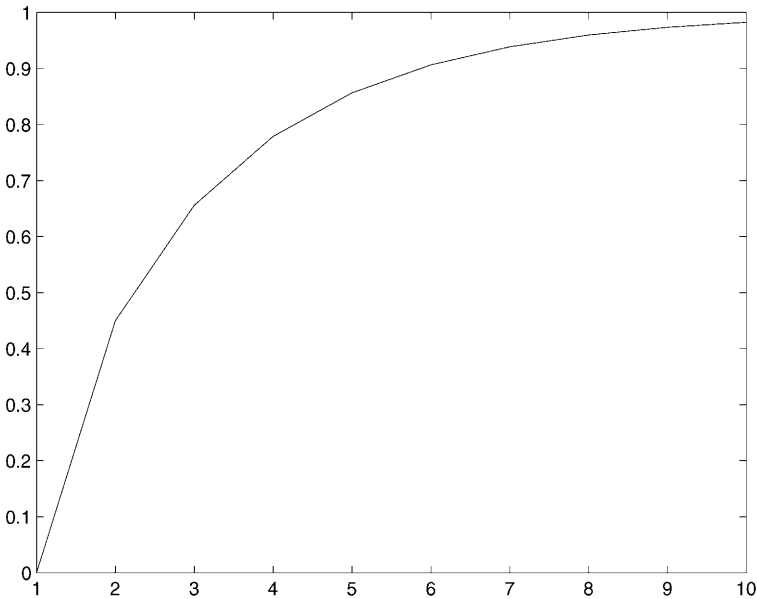
Fig. 1. The critical value $\gamma(n)$ as a function of sample size n for Matching Pennies with $(\mu, \nu) = (1, 2)$ and $(\alpha, \omega) = (0, 3)$

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 $0 < \gamma_k \leq 1$.
3. For $\gamma_k \leq \frac{n-1}{2n}$ 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 $\gamma(n) < \gamma_k \leq 1$, the boundary is attracting.
5. For $0 < \gamma_k < \gamma(n)$, the interior attractor $A(n, \gamma)$ 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.
6. For a given cyclic 2×2 game and given $0 < \gamma_1, \gamma_2 < 1$, there is a finite $n(\gamma)$ such that for $n \geq n(\gamma)$ the boundary is repelling. The attractors $A(n, \gamma)$ converge[6] to the interior attractor $A(\infty, \gamma)$ of (5). Hence, the attractors $A(n, \gamma)$ are bounded away from E, uniformly in n.

Statements 1–3 were shown above. Statement 4 follows from the monotone dependence of (13, 14, 15) on the step-sizes γ_k and the fact that $\psi'(x) \to 1$ as $n \to \infty$ uniformly on each compact subinterval of the open interval (α^k, ω^k). Statement 5 follows from the discussion at the end of Section 5.2. Statement 6 follows again from the upper semicontinuity of asymptotically stable sets (here $A(\infty, \gamma)$) and the robustness of the repellors E and the boundary against perturbations (from (5) to (4) with large n).

---

[6] In the sense that every neighbourhood of $A(\infty, \gamma)$ contains $A(n, \gamma)$ for all large enough n
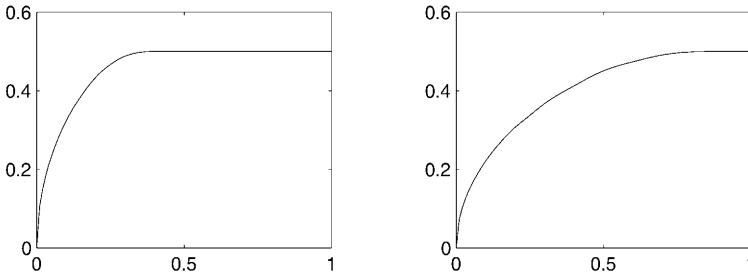
Fig. 2. Distance of the limit cycle from E as a function of γ in Matching Pennies with $(\mu, \nu) = (1, 2)$ and $(\alpha, \omega) = (0, 3)$; left: $n = 2$, right: $n = \infty$

Following Theorem 2, 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):

$$\gamma(n) = \frac{\psi_n(\nu) - \psi_n(\mu)}{\nu - \mu} = \frac{1}{\nu - \mu} \left[\frac{\nu - \alpha}{1 - \left(\frac{\omega - \nu}{\omega - \alpha}\right)^n} - \frac{\mu - \alpha}{1 - \left(\frac{\omega - \mu}{\omega - \alpha}\right)^n} \right]. \quad (19)$$

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 (measured on the ray $q = \frac{1}{2}, p > \frac{1}{2}$) against the step-size γ. In the left diagram we use $n = 2$, for which the boundary is reached for $\gamma \approx .4$ [7] while the right diagram shows the graph for the limit case $n = \infty$, for which the boundary is reached only at $\gamma = 1$.

---

[7] This value is taken from numerical simulations, while the precise value follows from (19) 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.

6 Conclusion

In cyclic 2×2 games, there is a basic difference in the long run dynamics of POR$\equiv$SPOR$_1$ on the one hand and SPOR$_n$ for $n \geq 2$ on the other hand. The discrete time 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 if 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 fractions γ_k of individuals receiving information between rounds goes to zero.

In this sense the interior 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 $n = 1$.

Differences among the SPOR$_n$ for different sample sizes $n \geq 2$ are less severe. SPOR$_2$ clearly has the advantage that it relies on the least observations and hence can be argued to be least costly. On the other hand, we find that sampling more is better. For interior states where players are not indifferent, current best responses increase stronger in each round under higher sample sizes (see Proposition 6 on the last page of the Appendix). Thus, it is very plausible that higher sample sizes should lead long run play closer to the equilibrium. However, there is a limit to what larger sample sizes can achieve as play is uniformly bounded away from the equilibrium for all sample sizes (see Theorem 2 (6)). An interesting topic for future research is to assume that sampling is costly, that individuals have different sample sizes n and that individuals using SPOR$_n$ not only imitate successful strategies but also the associated sample size.

SPOR$_n$ is a sophisticated imitation rule. Here, sophistication is in the sense that the probabilities of switching under POR are chosen to ensure the strictly improving condition for SPOR$_n$. 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 yield an aggregate monotone dynamics and stabilize the Nash equilibrium too. We cited Pollock and Schlag (1999) to show that sample sizes larger than one are necessary. Probabilistic strategy selection also remains a necessary ingredient. It is easily verified that deterministic imitation rules such as the popular, simple, and seemingly more intuitive, imitation rules *Imitate The Best* (Axelrod, 1984) and *Imitate Best Average* (Ellison and Fudenberg, 1993) do not induce a payoff monotone dynamics (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. A similar result also holds when individuals play a best response to a finite sample

of previous strategies among their potential opponents. This rule fails to respond sufficiently accurately to the frequencies of strategies played in the opponent population (see Pollock and Schlag, 1999).

Some sophistication on the 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 diploid inheritance, again leads to a stable limit cycle in the games[8] considered here, see Maynard Smith and Hofbauer (1987), or Hofbauer and Sigmund (1988, p. 312 ff).

Some comments on the feasible set of behavioral rules may be added. Individuals do not remember occurrences prior to the previous round. All individuals belonging to the same population use the same rule. While an investigation of weaker assumptions seems interesting, both of these assumptions are necessary if the dynamics should take the standard form of evolutionary dynamics. These dynamics are autonomous with the state of the process being the vector of frequencies of play. Another important assumption is that individuals do not know which round they are in. This generates stationary behavioral rules. If non-stationary rules were to be allowed then it is conceivable that some lead to play of the Nash equilibrium in the long run. We believe that this is true for the non-stationary rule that specifies for round t to use $SPOR_n$ ($n \geq 2$) with probability $1/t$ and to keep the previous strategy otherwise.[9]

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 (2000) 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.

Despite our positive results, stability of interior equilibria or even convergence of learning processes should not be generally expected in any game that has a unique equilibrium, even under sufficiently sophisticated learning rules, such as fictitious play or $SPOR_n$, and even with continuous time (Shapley, 1964; Gaunersdorfer and Hofbauer, 1995; Hofbauer and Swinkels, 1995). However, cyclic games such as Matching Pennies are sufficiently simple. Individuals have an incentive to take advantage of cyclic patterns and to invest in more information or in more memory if necessary. Hence, when payoffs in such games are important, then we expect individuals to choose sophisticated behavior that leads to (or approximates behavior of) the Nash equilibrium. To explicitly model

---

[8] Cyclic 2×2 games arise in biology as the 'battle of the sexes' game introduced by Dawkins as a model of the parental investment conflict, see Maynard Smith (1982).

[9] General theorems like in Garay and Hofbauer (1997) imply that such a trajectory converges either to E or to the boundary. That the latter does not occur seems rather plausible, but it requires a proof.

evolution of rules in this context is an interesting topic of future research with useful results already contained in (Pollock and Schlag, 1999).

Appendix

A Sophisticated imitation

In the following we will briefly review the properties of $SPOR_1$ and $SPOR_2$ in a new light using a domination argument to then argue why $SPOR_n$ can be considered the best sequentially evaluated rule.

A.1 Strictly improving behavior

Schlag (1998) develops a theory of optimal behavior for individuals that are boundedly rational in the following sense. Each individual only knows his own strategy set together with the bounded interval containing his payoffs. Whenever an individual changes his choice then he only uses information he has obtained since the previous round.

Various scenarios are introduced to motivate the following selection among behavioral rules. The individual searches for a rule F that, when used by all others in his population, *always* causes the average payoff in his population to increase over time should the individuals in the other population not change their play. This increase should be strict whenever not all strategies played in population k achieve the same expected payoff. Formally this means (for an individual of population 1) that the induced dynamics satisfy

$$
\begin{aligned}
\pi^1\left(\mathbf{p}^{t+1}\left(F\right), \mathbf{q}^t\right) &\geq \pi^1\left(\mathbf{p}^t, \mathbf{q}^t\right) \text{ for all } \left(\mathbf{p}^t, \mathbf{q}^t\right) \in \Delta\left(S^1\right) \times \Delta\left(S^2\right) \\
\text{where} \quad &> \quad \text{holds if there exists } i \qquad\qquad (20) \\
&\text{s.t. } \pi^1\left(i, \mathbf{q}^t\right) > \pi^1\left(\mathbf{p}^t, \mathbf{q}^t\right) \text{ and } p_i^t > 0 .
\end{aligned}
$$

Notice that we wrote $\mathbf{p}^{t+1}(F)$ instead of $\mathbf{p}^{t+1}$ to emphasize that dynamics in population 1 are driven by the behavioral rule F. Given the individual's knowledge, the above condition is required for any round t, any state $(\mathbf{p}^t, \mathbf{q}^t)$ in round t, and game Γ with strategy sets S^1 and S^2 that generates payoffs for player k in (α^k, ω^k). A behavioral rule for an individual belonging to population 1 with this property is called *strictly improving*.

When $|S^1| = 2$ then (21) implies that the proportion of individuals in population 1 playing a best response to the population state of the last round always increases.

It is easily shown that rules of the general form presented in Section 2 which are strictly improving are imitating.[10] Of course, not all imitating rules are strictly improving.

---

[10] The proof for $n = 1, 2$ contained in (Schlag, 1999) extends immediately to general sample sizes n.

Consider now the population dynamics (4) under $SPOR_n$. Then

$$\pi^1\left(\mathbf{p}^{t+1}, \mathbf{q}^t\right) - \pi^1\left(\mathbf{p}^t, \mathbf{q}^t\right) = \sum_{i \in S^1}\left(p_i^{t+1} - p_i^t\right)\left[\pi^1\left(i, \mathbf{q}^t\right) - \pi^1\left(\mathbf{p}^t, \mathbf{q}^t\right)\right]$$

$$= \frac{\gamma_1}{\omega^1 - \alpha^1}\left[1 + \frac{\omega^1 - \pi^1(\mathbf{p}^t, \mathbf{q}^t)}{\omega^1 - \alpha^1} + .. + \left(\frac{\omega^1 - \pi^1(\mathbf{p}^t, \mathbf{q}^t)}{\omega^1 - \alpha^1}\right)^{n-1}\right]$$

$$\cdot \sum_{i \in S^1}\left[\pi^1\left(i, \mathbf{q}^t\right) - \pi^1\left(\mathbf{p}^t, \mathbf{q}^t\right)\right]^2 \cdot p_i^t \geq 0 .$$

which leads to the following result [see also Schlag 1998 and 1999 for $n = 1, 2$].

Proposition 5 *$SPOR_n$ generates a strictly improving rule for any sample size n.*

A.2 A case for optimality

As demonstrated above, $SPOR_n$ is strictly improving. In the following we will demonstrate the unique properties of $SPOR_n$. While we will be relying on many results obtained in other papers, the presentation below will put these in new light.

Consider two behavioral rules F and G for player one. We say that F *dominates* G if the terms in (20) are always larger under F than under G, with strict inequality in some cases, i.e., if

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

for any feasible payoff distributions in Γ and any state $\left(\mathbf{p}^t, \mathbf{q}^t\right)$ in round t, and not "$\equiv$". Notice that a strictly improving rule dominates the rule "never switch strategies". 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 $\mathscr{D}$ if $F \in \mathscr{D}$ and if there is no rule $G \in \mathscr{D}$ that dominates F.

The properties of $SPOR_n$ for single (Schlag, 1998) and double (Schlag, 1999) sampling, stated in terms of dominance, are as follows.

(Single sampling) $SPOR_1 \equiv POR$ is undominated among the strictly improving single sampling behavioral rules. Any other single sampling rule with this property[11] induces the same population adjustment (2) as POR.

(Double sampling) $SPOR_2$ is the *unique* rule that is undominated among the sequentially evaluated 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.

---

[11] Such as the Proportional Imitation Rule and the Proportional Reviewing Rule (Schlag, 1998).

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$_2$. In fact, it turns out to be difficult to generate interesting strictly improving rules through sequential evaluation. In particular, the Proportional Imitation Rule, selected in Schlag (1998) as the "best" single sampling rule does not generate a strictly improving double sampling rule when it is evaluated sequentially.

Can alternative, possibly more complicated, methods for constructing a double sampling rule outperform SPOR$_2$? Strictly improving double sampling rules that are not dominated by SPOR$_2$ exist (Schlag, 1999). 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 (20), SPOR$_n$ performs better the larger the sample size. Hence,

Proposition 6 *SPOR$_n$ dominates SPOR$_m$ for all $n > m \geq 1$.*

References

Axelrod RM (1984) The evolution of cooperation. Basic Books, New York

Björnerstedt J, Weibull J (1996) Nash equilibrium and evolution by imitation. In: Arrow K, Colombatto E (eds) The rational foundations of economic behaviour. Macmillan, London

Börgers T, Sarin R (1997) Learning through reinforcement and replicator dynamics. Journal of Economic Theory 77: 1–14

Brown GW (1951) Iterative solution of games by fictitious play. In: Koopmans TC (ed) Activity analysis of production and allocation, pp 374–376. Wiley, New York

Crawford VP (1985) Learning behavior and mixed-strategy Nash equilibrium. Journal of Economic Behavior Organization 6: 69–78

Cressman R, Morrison WG, Wen J-F (1998) On the evolutionary dynamics of crime. Canadian Journal of Economics 31(5): 1101–1112

Ellison G, Fudenberg D (1995) Word-of-mouth communication and social learning. The Quarterly Journal of Economics 440: 93–125

Eshel I, Akin E (1983) Coevolutionary instability of mixed Nash solutions. Journal of Mathematical Biology 18: 123–133

Friedman D (1991) Evolutionary games in economics. Econometrica 59: 637–66

Gale J, Binmore KG, Samuelson L (1995) Learning to be imperfect: the Ultimatum Game. Games and Economic Behaviour 8: 56–90

Garay B, Hofbauer J (1997) Chain recurrence and discretization. Bulletin Australian Mathematical Society 55: 63–71

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

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

Hofbauer, J (1995b): Imitation dynamics for games. Mimeo, University of Vienna

Hofbauer J, Iooss G (1984) A Hopf bifurcation theorem for difference equations approximating a differential equation. Monatshefte für Mathematik 98: 99–113

Hofbauer J, Sigmund K (1988) The theory of evolution and dynamical systems. Cambridge University Press, Cambridge

Hofbauer J, Swinkels J (1995) A universal Shapley example. Mimeo

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

Maynard Smith J, Hofbauer J (1987) The "battle of the sexes": a genetic model with limit cycle behavior. Theoretical Population Biology 32: 1–14

Oechssler J, Schlag KH (2000) Loss of commitment? An evolutionary analysis of Bagwell's example. International Game Theory Review (to appear)

Pollock G, Schlag KH (1999) Social roles as an effective learning mechanism. Rationality and Society 11(4): 371–397

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

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

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

Schlag KH (1998) Why imitate, and if so, how? A bounded rational approach to multi-armed bandits. Journal of Economic Theory 78: 130–156

Schlag KH (1999) Which one should I imitate? Journal of Mathematical Economics 31: 493–522

Selten R (1980) A note on evolutionarily stable strategies in asymmetric animal conflicts. Journal of Theoretical Biology 84: 93–101

Shapley L (1964) Some topics in two–person games. Annals of Mathematical Studies 5: 1–28

Stuart AM (1994) Numerical analysis of dynamical systems. In: Iserles A (ed) Acta Numerica 1994, pp 467–572. Cambridge University Press, Cambridge

Taylor PD (1979) Evolutionarily stable strategies with two types of players. Journal of Applied Probability 16: 76–83

Weibull J (1995) Evolutionary game theory. MIT Press, Cambridge, MA