



# HIER

## Harvard Institute of Economic Research

Discussion Paper Number 2143

Heterogeneous Beliefs and Local Information  
in Stochastic Fictitious Play

by

Drew Fudenberg  
and  
Satoru Takahashi

November 2007

HARVARD UNIVERSITY  
Cambridge, Massachusetts

This paper can be downloaded without charge from:  
<http://www.economics.harvard.edu/journals/hier2007>

The Social Science Research Network Electronic Paper Collection:  
<http://ssrn.com/abstract=1030328>

# Heterogeneous Beliefs and Local Information in Stochastic Fictitious Play\*

Drew Fudenberg  
Harvard University  
dfudenberg@harvard.edu

Satoru Takahashi  
Princeton University  
satorut@princeton.edu

October 15, 2007

## Abstract

Stochastic fictitious play (SFP) assumes that agents do not try to influence the future play of their current opponents, an assumption that is justified by appeal to a setting with a large population of players who are randomly matched to play the game. However, the dynamics of SFP have only been analyzed in models where all agents in a player role have the same beliefs. We analyze the dynamics of SFP in settings where there is a population of agents who observe only outcomes in their own matches and thus have heterogeneous beliefs. We provide conditions that ensure that the system converges to a state with homogeneous beliefs, and that its asymptotic behavior is the same as with a single representative agent in each player role.

---

\*We thank Michel Benaïm and Josef Hofbauer for very helpful comments and suggestions.

# 1 Introduction

Consider a situation where two agents are playing a known two-player game in strategic form. The standard practice in most of economic theory, and almost all of its applications, is to assume that the outcome of the game will be one of its Nash equilibria, yet rationality alone (and even common knowledge of rationality) is not enough to imply that equilibrium analysis is valid. To provide a rationale for Nash equilibrium, the literature on learning in games proposes that Nash equilibrium should be thought of as the result of a non-equilibrium dynamic process of learning or adaptation. Non-equilibrium dynamics can be used to discriminate between equilibria of a given game, and to predict that equilibrium is more likely to be observed in some games than in others.

This paper contributes to the study of one particular type of non-equilibrium dynamics, namely the process known as stochastic (or smooth) fictitious play, henceforth “SFP.” SFP is an example of a belief-based learning process, in which agents form beliefs about the play of their opponents, and then update the beliefs on the basis of their observations. One important aspect of this model is that the agents act like Bayesians facing a fixed but unknown distribution over opponents’ play, so that once they have a sufficiently large sample, their beliefs are close to the empirical distribution that they have observed. A second key aspect of this model is that agents do not try to influence the future play of their current opponents. Instead, they take this distribution as exogenous, and choose their actions on the basis of the expected payoffs in the current play of the game; in particular, each agent plays a “smoothed best response” to the anticipated distribution of opponent’s strategies in the current play. Thus the agents are engaged in many plays of a stage game and not a single play of a repeated game.

To motivate the assumption that agents treat opponent’s play as exogenous, and thus the use of fictitious play as a descriptive model, Fudenberg and Kreps [1993] appeal to a setting with a large population of agents who are randomly matched to play the game, with each agent observing only the outcomes in his own matches, as is common in experimental tests of game theory.<sup>1</sup> In this “local information” setting, the assumption that each agent

---

<sup>1</sup>The original work on fictitious play with exact best responses (starting with Brown [1951]) viewed it as describing a pre-play thought process as opposed to real-time learning, so the issue of agents influencing each other’s future play did not arise. Moreover, this literature treated the deterministic two-cycle sequence  $(A, B, A, B, \dots)$  as equivalent to a

considers only the current payoff when choosing his action can be justified by the idea that in a large population of agents, the agent is unlikely to interact with the same partner, or anyone else who has interacted with that partner, for a long time, so the discounted value of any effect of current play on opponents' future actions is small.<sup>2</sup> However, the dynamics of SFP have only been analyzed in “global information” models where all agents in a given role have the same observations and the same beliefs. This includes models with one agent in each player role (Fudenberg and Kreps [1993], Benaïm and Hirsch [1999], Hofbauer and Sandholm [2002], Hofbauer and Hopkins [2005]), models with a continuum of agents, who behave as if they observe the outcomes of all matches (Hopkins [1999a], Ellison and Fudenberg [2000]), and models with a single population playing a symmetric game and using common beliefs (Hopkins [1999b], Hofbauer and Sandholm [2002]).

To see why it is important to allow for heterogeneous beliefs, recall that Fudenberg and Levine [1993b] point out that rational Bayesian agents can maintain heterogeneous beliefs about play off the equilibrium path in an extensive-form game, and that Fudenberg and Levine [1993a, 1997] show that this sort of heterogeneity can lead to new equilibrium outcomes and help explain the results of some game theory experiments. In this paper, we consider play of a strategic-form game, with the chosen strategies observed at the end of each round, so the issue of off-path beliefs does not arise, but even so it is easy to see that heterogeneous beliefs matter in some sorts of non-equilibrium processes: Consider for example a process where agents are randomly paired each period, and then each agent plays a best response to his observation in the previous period. In this case, with an even number of agents in a symmetric anti-coordination game with two actions  $A$  and  $B$ , the global-information process oscillates between the extremes “all  $A$ ” and “all  $B$ ,” or possibly is absorbed at a state near the mixed equilibrium. In contrast, for any  $k$ , the local-information process can oscillate between “ $k$   $A$  and the rest  $B$ ” and “ $k$   $B$  and the rest  $A$ .”

These examples show that heterogeneous beliefs can persist in some models, which raises the question of whether heterogenous beliefs can persist when agents behave as in SFP, and whether past results on the convergence and stability properties of SFP extend to finite population settings with initially

---

50-50 randomization, even though the two sequences are very different from the perspective of an agent trying to predict his opponent's next play.

<sup>2</sup>Implicit here is the idea that the population size is large compared to the discount factor. See Ellison [1997] for some calibrations of just what “large” means.

heterogeneous beliefs.

We address this question in three related models, all of which follow SFP for the behavior of individual agents, but differ in the details of the matching structure, specifically how many agents are drawn from the society in each period, and whether an agent and his partner are chosen from the same population or two different populations. In the first model, the society consists of two populations with equal size, and, in each period, every agent in one population plays the stage game with a randomly chosen partner from another population. The second model is similar except that the society consists of a single population with an even number of agents. These are models with “synchronous clocks,” meaning that each agent plays every period, so that the sample size of every agent is the same, as is the case in most laboratory experiments. Our third model, and most original, model has a single population and “asynchronous clocks”: Each period, two agents are randomly drawn to play the stage game, so that some agents end up playing more often than others. For this model, we generalize from a uniform matching protocol to one where the probability that a given pair of agents  $i, j$  interact can depend on their indices; this lets us provide the first analysis of SFP on networks.

When agents have a long memory (as in SFP), and are all observing draws at the same frequency from a common (possibly time varying) distribution, it is fairly intuitive that they should eventually have the same beliefs, so that the set of asymptotic outcomes should be the same as in the “unitary,” one-agent-per-role models. It is less obvious that this is the case when the distribution faced by different agents can differ, as it does in a one-population model, or when the distribution is cycling and different agents observe it at different frequencies; we show that this is the case provided that the number of agents,  $M$ , is sufficiently large compared to the slope of the smooth best response function and the matching protocol is stationary. (The reason that the population size  $M$  matters here is that agents do not observe and respond to their own play, so that the distribution faced by any two agents can differ by  $1/(M - 1)$ .) We also extend many of the past results on convergence and stability with one agent per player role to the case of heterogeneous beliefs, even with asynchronous clocks. One new issue that arises with heterogeneous agents is that the perturbations used to generate the smoothing may also be heterogeneous; to handle this possibility, we show that the average of smoothed best responses is also a smoothed best response.

Our analysis follows past work on SFP in using the techniques of stochas-

tic approximation to relate the behavior of a discrete-time stochastic system of interest to that of a deterministic system in continuous time. Because this approach focuses on the asymptotic behavior of the system, it does not provide very precise estimates of the effect of the initial condition of the system. To explore the impact of initial conditions, we present some simulations to illustrate the role of the intensity of prior beliefs on the eventual long-run outcome. Our first set of simulations concerns a pure coordination game where both actions are equally efficient, and suppose that a majority of the agents have prior beliefs that make them start out playing action  $A$  with high probability. In this case, the system is very likely to converge to a state where agents expect action  $A$  and play  $A$  with probability near 1. However, if the agents who initially expect action  $B$  have a much stronger conviction about their forecasts (more concentrated priors) than those who take action  $A$ , then the system has a very high probability of convergence to the state where all agents place probability near 1 on action  $B$ . This phenomenon is observed both in local- and global-information models, but the impact is stronger in the local one.<sup>3</sup> We also present some simulations on an anti-coordination game that suggest that our result on the minimum population size for conformity is tight.

## 2 Preliminaries

### 2.1 Stage Game

Consider a symmetric  $m \times m$  game  $G = (S, u)$ , where  $S$  is a finite set of pure strategies, and  $u: S^2 \rightarrow \mathbb{R}$  is a utility function.<sup>4</sup> For each  $(s, s') \in S^2$ , a player gets payoff  $u(s, s')$  if he and his opponent choose strategies  $s$  and  $s'$ , respectively. Let  $\Sigma$  be the set of mixed strategies,  $\{\sigma \in \mathbb{R}^S \mid \sigma(s) \geq 0, \sum_{s \in S} \sigma(s) = 1\}$ , where  $\mathbb{R}^S$  is endowed with the sup-norm:  $|x| = \max_{s \in S} |x(s)|$  for each  $x \in \mathbb{R}^S$ . We extend the domain of  $u$  to  $\Sigma^2$  by the expected utility hypothesis:  $u(\sigma, \sigma') = \sum_{s, s' \in S} u(s, s')\sigma(s)\sigma'(s')$  for each  $(\sigma, \sigma') \in \Sigma^2$ . In the standard abuse of notation, we let  $s$  denote both a pure strategy and the

---

<sup>3</sup>This set of simulations illustrates a key difference between belief-based learning models such as SFP, and evolutionary models such as the replicator dynamic: With SFP, the initial actions played are not a sufficient statistic for the future evolution of the system.

<sup>4</sup>The results in Subsection 3.1 extend to asymmetric games, but since symmetry is important in Subsections 3.2 and 3.3, we impose it now.

mixed strategy that puts probability 1 on  $s$ .

A *best response* to strategy  $\sigma'$  is a mixed strategy  $\sigma$  that maximizes  $u(\sigma, \sigma')$ . A *smoothed best response function*  $\overline{BR}: \Sigma \rightarrow \Sigma$  is a twice continuously differentiable function (hence “smooth”) that is bounded away from 0: regardless of the opponent’s strategy  $\sigma'$ ,  $\overline{BR}(\sigma')$  assigns at least some minimum probability to every  $s \in S$ . Since  $\overline{BR}$  is continuously differentiable on a compact domain, it is Lipschitz continuous. Let  $K > 0$  be a Lipschitz constant for  $\overline{BR}$ , i.e.,

$$|\overline{BR}(\sigma) - \overline{BR}(\sigma')| \leq K|\sigma - \sigma'|$$

for all  $(\sigma, \sigma') \in \Sigma^2$ . For example, the *logit best response*

$$\overline{BR}(\sigma')(s) = \frac{\exp(\beta u(s, \sigma'))}{\sum_{\tilde{s}} \exp(\beta u(\tilde{s}, \sigma'))}$$

with  $\beta > 0$  is smoothed.

In the processes we study, agents use smoothed best responses to their beliefs. The results hold for any strictly positive twice continuously differentiable functions, but the intended interpretation relies on the idea that the smoothed best responses are an approximation of the exact best responses, as for example in the logit best response with large  $\beta$ . Fudenberg and Kreps [1993] generate smoothed best responses from i.i.d. payoff perturbations in the spirit of Harsanyi’s [1973] purification theorem:

$$\overline{BR}(\sigma')(s) = P(\max_{s'}(u(s', \sigma') + \varepsilon(s')) = s). \quad (\text{SP})$$

They point out that fixed points of the smoothed best responses are then Bayesian Nash equilibria of the incomplete information game, and call these fixed points *Nash distributions*; Nash distributions converge to the Nash equilibria of the unperturbed game as random components converge to 0 in probability (Hofbauer and Sandholm [2002, Proposition 3.1]).

Smoothed best responses can also be generated from maximizing a perturbed utility function that is non-linear in probabilities and penalizes pure strategies: set

$$\overline{BR}(\sigma') = \arg \max_{\sigma} (u(\sigma, \sigma') + v(\sigma)), \quad (\text{DP})$$

where  $v$  is negative definite (so that the perturbed utility function is strictly concave), and its gradient becomes arbitrarily large near the boundary of

the simplex (so that the argmax assigns strictly positive probability to all pure strategies). If  $v$  is bounded and  $\sup_{\sigma, \sigma'} |v(\sigma) - v(\sigma')|$  becomes small, fixed points of  $\overline{BR}$  become close to the Nash equilibria. Any smoothed best response function generated by an additive random utility model can be derived from this sort of deterministic non-linear perturbation (Hofbauer and Sandholm [2002, Theorem 2.1]).

## 2.2 Stochastic Fictitious Play

The model of stochastic fictitious play has two components, namely the specification of how beliefs are formed and the specification of how beliefs are used to determine behavior. In SFP, each agent's predictions about the play of his opponents has a simple form derived from Bayesian updating, where the agent believes that he is facing i.i.d. draws from a fixed but unknown distribution of play, and the agent has a Dirichlet prior over that distribution.<sup>5</sup> This leads to a very simple functional form for the forecasted distribution of play in each period.

Specifically, each agent  $i$  has an exogenous initial weight function  $\kappa_{i,0}: S \rightarrow \mathbb{R}_+$  with  $\kappa_{i,0}(s) > 0$  for some  $s \in S$ ; this corresponds to the parameters of the agent's Dirichlet prior. The agent updates this weight by adding 1 to the weight of each opponent strategy each time it is played, so that  $\kappa_{i,n+1}(s) = \kappa_{i,n}(s) + I(s_{-i,n+1} = s)$ ;  $\bar{\kappa}_{i,0} = \sum_{s \in S} \kappa_{i,0}(s)$  measures the strength of the agent's prior. The agent's beliefs about his opponent's play at period  $n$  correspond to the relative weights on the various strategies: the probability the agent assigns to the opponent playing  $s$  is

$$\theta_{i,n}(s) = \frac{\kappa_{i,n}(s)}{\bar{\kappa}_{i,n}},$$

where  $\bar{\kappa}_{i,n} = \sum_{s \in S} \kappa_{i,n}(s)$ . This is the expected value of the opponent's mixed strategy with respect to the Dirichlet distribution corresponding to  $\kappa_{i,n}$ . This updating rule is asymptotically empirical in the sense that beliefs converge to the empirical distribution along every sample path.<sup>6</sup>

---

<sup>5</sup>Stationarity is a reasonable first hypothesis in many situations. However, we might expect agents to eventually reject it given sufficient evidence to the contrary. In particular, if the system in which the agents are learning converges to a cycle, then the assumption that agents ignore the cycle and continue to believe that the environment is stationary may not be plausible.

<sup>6</sup>Fudenberg and Kreps [1993] say that a rule with this property is "asymptotically

We consider situations where the agents are playing a known two-player game in strategic form. That is, they know the strategy space and the payoff function. All that is unknown is how the other side is going to play. In each period, some of the agents are randomly drawn from the population, and matched with each a partner to play the stage game. Each agent believes that he is playing against a randomly drawn opponent, and that the distribution of opponents' play is constant over time. Moreover, each agent considers only the current period's expected payoff when choosing his action.

Classical or exact fictitious play is then defined as any behavior rule that assigns actions to histories by first computing  $\theta_{i,n}$  as above and then picking any strategy that is an exact best response to  $\theta_{i,n}$ , i.e., any strategy  $s \in S$  that maximizes  $u(s, \theta_{i,n})$ . Under SFP, agent  $i$  computes  $\theta_{i,n}$  in the same way, but then plays a smoothed best response to it; that is, the agent chooses strategy  $s \in S$  with probability  $\overline{BR}(\theta_{i,n})(s)$ . This smoothing of the best responses results in a learning process with more appealing properties, and avoids the nonsensical rapid cycling of exact fictitious play that was pointed out by Fudenberg and Kreps [1993]. Fudenberg and Levine [1995] point out that randomization can also be used to satisfy the (non-Bayesian) criterion of “universal consistency,” in particular the randomization that comes from maximizing perturbed payoff functions of the form (DP).

## 2.3 Dynamical Systems

Next we review some standard concepts for the study of dynamical systems; our language and definitions are taken from Conley [1976] and Benaïm [1999]. Let  $X$  be a nonempty subset of  $\mathbb{R}^m$  with the Euclidean metric  $d$ . A *semi-flow* is a continuous function  $\Phi: X \times \mathbb{R}_+ \rightarrow X$  such that  $\Phi_0(x) = x$  and  $\Phi_s(\Phi_t(x)) = \Phi_{s+t}(x)$  for every  $x \in X$  and  $s, t \in \mathbb{R}_+$ .<sup>7</sup> We extend  $\Phi$  to the subsets of  $X$  by defining  $\Phi_t(A) = \{\Phi_t(x) : x \in A\}$  and say that  $A \subset X$  is *invariant* if  $\Phi_t(A) = A$  for all  $t \in \mathbb{R}_+$ . A subset  $A$  is an *attracting set* if  $A$  is nonempty and compact, and there exists some open neighborhood  $U$  of  $A$

---

empirical,” and note that many of the properties of fictitious play extend to other asymptotically empirical rules. Diaconis and Freedman [1990] show that any Bayesian belief about a series of i.i.d. draws converges to the empirical distribution at a uniform rate if the prior is “non-doctrinaire.” Dirichlet priors are non-doctrinaire provided that all of the initial weights are positive.

<sup>7</sup>We need to consider semi-flows and not flows to accommodate the analysis of our third model.

such that  $\lim_{t \rightarrow \infty} d(\Phi_t(x), A) = 0$  uniformly in  $x \in U$ . An *attractor* is an invariant attracting set. Every attracting set contains an attractor. A *globally attracting set* is an attracting set such that  $\lim_{t \rightarrow \infty} d(\Phi_t(x), A) = 0$  for every  $x \in X$ . The  $\omega$ -*limit set* of a point  $x$  is the set  $\omega(x) = \{y \in X : \Phi_{t_k}(x) \rightarrow y \text{ for some } t_k \rightarrow \infty\}$ . The definition of  $\omega$ -limit sets extends to discrete-time sequences in the obvious way.

**Definition 1.** For a nonempty invariant set  $A$ , a  $(\delta, T)$ -*pseudo-orbit from  $a$  to  $b$  in  $A$*  is a finite series of partial trajectories,  $\{\Phi_t(x_i) : 0 \leq t \leq t_i\}$ ;  $i = 1, \dots, k-1$ ;  $t_i \geq T$ , such that  $x_i \in A$ ,  $d(x_0, a) < \delta$ ,  $d(\Phi_{t_i}(x_i), x_{i+1}) < \delta$  for all  $i = 1, \dots, k-1$ , and  $x_k = b$ . A nonempty compact invariant set  $A$  is *internally chain-transitive* if, for every  $a, b \in A$  and every  $\delta > 0$ ,  $T > 0$ , there is a  $(\delta, T)$ -pseudo-orbit in  $A$  from  $a$  to  $b$ .

Every internally chain-transitive set is connected. Moreover, a nonempty compact invariant set is internally chain-transitive if and only if the set has no proper attracting set.<sup>8</sup> The internally chain-transitive sets provide a characterization of the possible long-run behavior of the system; Benaïm [1999] has shown that this particular concept of long-run behavior is useful when working with stochastic approximation.

The stochastic approximation algorithm is a discrete-time stochastic process whose step size decreases with time, so that asymptotically the system converges to its deterministic continuous-time limit. The early work on stochastic approximation was done by Robbins and Monro [1951] and Kiefer and Wolfowitz [1952] and has since been applied and extended by a number of authors. Benaïm [1999] gives a self-contained presentation of a number of stochastic approximation results, along with some generalizations and new proofs. For our purposes, the main results are the following.

**Theorem A** (Benaïm, 1999). *Consider the discrete time process on a non-empty convex subset  $X$  of  $\mathbb{R}^m$  defined by the recursion,*

$$x_{n+1} - x_n = \frac{1}{n+1}(F(x_n) + U_{n+1} + b_{n+1}), \quad (1)$$

*and the corresponding continuous time semi-flow  $\Phi$  induced by the system of ordinary differential equations (ODEs)*

$$\dot{x}(t) = F(x(t)), \quad (2)$$

---

<sup>8</sup>Benaïm [1999, Proposition 5.3] showed that internal chain-transitivity for a compact invariant set is equivalent to admitting no proper attractor. We replace ‘‘attractor’’ in this statement by ‘‘attracting set’’ since every attracting set contains an attractor.

where

1.  $F: X \rightarrow \mathbb{R}^m$  is continuous and (2) is globally integrable,
2.  $\{U_n\}$  and  $\{b_n\}$  are stochastic processes adapted to filtration  $\{\mathcal{F}_n\}$ , i.e., for each  $n \in \mathbb{N}$ ,  $U_n$  and  $b_n$  are random variables that are measurable with respect to  $\mathcal{F}_n$ , where  $\mathcal{F}_n$  is the  $\sigma$ -algebra corresponding to the history of the system up through the end of period  $n$ ,
3.  $E(U_{n+1}|\mathcal{F}_n) = 0$  almost surely and  $E(|U_n|^2) < \infty$ ,
4.  $\lim_{n \rightarrow \infty} b_n = 0$  almost surely, and
5.  $\{x_n \mid n \geq 0\}$  is precompact in  $X$  almost surely.<sup>9</sup>

Then, with probability 1, every  $\omega$ -limit of the process  $\{x_n\}$  lies in an internally chain-transitive set for  $\Phi$ .

*Remark 1.* Intuitively, the fact that the step size is of order  $1/n$  means that the discrete-time system will converge to the continuous-time limit  $\dot{x} = F(x)$  when the perturbation terms  $\{U_n\}$  and  $\{b_n\}$  are absent. The bound on the variance of the noise, together with the fact that  $\sum_n 1/n^2 < \infty$ , ensures that the noise term is asymptotically negligible, in the spirit of the martingale version of the law of large numbers. The conditions on the noise term  $\{U_n\}$  are the usual “martingale noise” conditions for stochastic approximation; the error term  $\{b_n\}$  permits a slight generalization in that the mean of the disturbance is only 0 asymptotically. Benaïm presents more general conditions on the step size.

*Remark 2.* If  $X$  is compact and  $F$  is Lipschitz continuous, then (2) is globally integrable and  $\{x_n\}$  is precompact.

*Remark 3.* Unlike Benaïm’s original statement, we consider  $F$  to be defined only on  $X$ , as opposed to the whole space  $\mathbb{R}^m$ . Accordingly, we modify his assumption A2 in Proposition 4.2 so that  $\{x_n\}$  is not only bounded but also precompact in  $X$ .

*Proof.* Benaïm’s Proposition 4.2 and Remark 4.5 imply that with probability 1, the assumptions of Proposition 4.1 hold, i.e., with probability 1, the interpolated process (which embeds discrete time paths in continuous time)

---

<sup>9</sup>A subset of a topological space  $X$  is *precompact* in  $X$  if its closure in  $X$  is compact.

is a precompact asymptotic pseudotrajectory of the semi-flow  $\Phi$  induced by  $F$ .<sup>10</sup> Theorem 5.7 then implies that every  $\omega$ -limit of the pseudotrajectory is internally chain-transitive for  $\Phi$ .  $\square$

Theorem A means that the realization of the stochastic process (1) converges to one of the internally chain-transitive sets for the semi-flow (2) almost surely, but it does not exclude unstable equilibria from possible limit sets. This is because the theorem allows for random noise terms, but does not actually require that any noise is present. A sharper conclusion can be drawn when there is a lower bound on the amount of noise in (1). Specifically, we would expect that noise would prevent the process from converging to an unstable equilibrium, because the noise will move the process off the lower-dimensional stable manifold. To make this precise, recall that an equilibrium point  $x^*$  (i.e.,  $F(x^*) = 0$ ) is *linearly unstable* if the Jacobian matrix of  $F$  at  $x^*$ ,  $DF(x^*)$ , has some eigenvalue with a positive real part. Let  $\mathbb{R}^m = E_+ \oplus E_-$ , where  $E_+$  and  $E_-$  are the generalized eigenspaces of  $DF(x^*)$  corresponding to eigenvalues with positive and nonpositive real parts, respectively. If  $x^*$  is linearly unstable, then  $E_+$  has at least one dimension.

**Theorem B** (Brandière and Duflo, 1996). *Consider (1) on a nonempty open subset  $X$  of  $\mathbb{R}^m$ . Let  $x^*$  be a linearly unstable rest point of  $F$  and  $U_n^{(r)}$  be the projection of  $U_n$  on  $E_+$  in the directions of  $E_-$ . Assume that*

1.  $F$  is continuously differentiable and its derivative is Lipschitz continuous on a neighborhood of  $x^*$ ,
2.  $\{U_n\}$  and  $\{b_n\}$  are adapted to  $\{\mathcal{F}_n\}$ ,
3.  $E(U_{n+1}|\mathcal{F}_n) = 0$ ,  $\limsup_{n \rightarrow \infty} E(|U_{n+1}|^2|\mathcal{F}_n) < \infty$ , and  $\liminf_{n \rightarrow \infty} E(|U_{n+1}^{(r)}||\mathcal{F}_n) > 0$  almost surely, and
4.  $\sum_{n=1}^{\infty} |b_n|^2 < \infty$  almost surely.

*Then  $\lim_{n \rightarrow \infty} x_n = x^*$  with probability 0.*

---

<sup>10</sup>A continuous function  $x: \mathbb{R} \rightarrow X$  is an *asymptotic pseudotrajectory* for  $\Phi$  if

$$\lim_{t \rightarrow \infty} \sup_{0 \leq h \leq T} d(x(t+h), \Phi_h(x(t))) = 0$$

for every  $T > 0$ .

*Proof.* See Brandière and Duflo [1996]. Intuitively speaking, even if  $x_n$  is close to  $x^*$ , since  $U_{n+1}$  has sufficiently large components in the unstable directions of  $DF(x^*)$ , the process cannot stay near  $x^*$ .  $\square$

*Remark 4.*  $U_{n+1}$  is required to have a positive variance in unstable directions of  $DF(x^*)$ . Hopkins and Posch [2005] apply this version to reinforcement learning models.

*Remark 5.* A similar nonconvergence result without noise term  $\{b_n\}$  was obtained by Pemantle [1990]. See also Brandière [1998], Benaïm [1999], and Tarrès [2000] for results about nonconvergence towards hyperbolic unstable periodic orbits and other types of repelling sets, under various sets of assumptions.<sup>11</sup>

Benaïm and Hirsch [1999] show that versions of these theorems apply to the standard SFP and its continuous-time mean field

$$\begin{aligned}\dot{\theta}_1(t) &= \overline{BR}(\theta_2(t)) - \theta_1(t), & \text{(Unitary)} \\ \dot{\theta}_2(t) &= \overline{BR}(\theta_1(t)) - \theta_2(t);\end{aligned}$$

Note that the rest points of this system are exactly the equilibrium distributions.

By studying this continuous-time system, Benaïm and Hirsch [1999] show that SFP converges to distributions that approximate one of the two pure-strategy equilibria in  $2 \times 2$  coordination games, and not to approximations of the (unstable) mixed equilibrium, while play converges to the (unique) equilibrium distribution in  $2 \times 2$  games with a unique mixed-strategy equilibrium. They provide extensions to some many-player two-action games, and show that the unique equilibrium in Jordan’s [1993] three-player matching-pennies game is linearly unstable. Ellison and Fudenberg [2000] study this continuous-time system in  $3 \times 3$  games; they find that there are many games in which whether the totally mixed equilibrium is locally stable depends on the details of the way the best response functions are smoothed. Hofbauer and Sandholm [2002] prove that SFP converges to the unique equilibrium distribution in all two-player zero-sum games, and to one of the equilibrium distributions in many-player potential games if all of the rest points are hyperbolic; they also present results on supermodular games. Hofbauer and

---

<sup>11</sup>We use Brandière and Duflo [1996] rather than Benaïm [1999] or Tarrès [2000] since the boundedness of  $\{U_n\}$  assumed in the latter two papers is not satisfied in our asynchronous-clock model.

Hopkins [2005] prove convergence of perturbed best response dynamics to perturbed equilibria in all two-player games that can be rescaled to be zero-sum, and in two-player games that can be rescaled to be partnerships;<sup>12</sup> they also show that isolated interior equilibria of all generic symmetric games are linearly unstable for all small symmetric perturbations of the best response correspondence, where a “symmetric perturbation” means that the two players have the same smoothed best response functions. This instability result applies in particular to symmetric versions of the famous example of Shapley [1964], and to non-constant-sum variations of the game “rock-scissors-paper.” (The constant-sum case is one of the non-generic games where the equilibrium is stable.) The results of this paper give conditions that extend all of the above results to settings with populations of agents who observe only the outcomes in their own matches.

### 3 Stochastic Fictitious Play with Personal Histories

We now consider three versions of stochastic fictitious play with random matching, where agents observe only the outcomes of play in their own matches. Consider a society with finitely many agents. In each period, some of the agents are randomly drawn from the society, and anonymously matched with each other to play the stage game. Each agent forms a belief about the distribution of the strategies he will see, based on empirical frequencies of his past partners’ plays. Whenever he is drawn from the society, then he randomly chooses a pure strategy according to the smoothed best response to his current belief about the other agents’ strategies. The agents update their beliefs in a purely decentralized way: they do not observe outcomes of other matches, nor communicate with other agents. Also note that this learning process collapses to the standard model of stochastic fictitious play if there are only two agents in the society.

The three models we consider differ in the details of the matching structure, specifically in (a) how many agents are drawn from the society in each period, and (b) whether an agent and his partner are chosen from the same population or two different populations. In the first model, society consists

---

<sup>12</sup>These two classes are exhaustive for generic  $2 \times 2$  games but not for generic games with more actions.

of two populations with equal size, and, in each period, every agent in one population plays the stage game a randomly chosen partner from another population. In the second model, society consists of a single population with an even number of agents, and, in each period, every agent plays the stage game with his partner. These two models have “synchronous clocks,” in the sense that each agent plays every period, so that the sample size of every agent is the same, and all players have access to a common measure of time. The third model has a single population and “asynchronous clocks: Each period, two agents are randomly drawn to play the stage game. so that some agents can end up playing more often than others.

In each of these models, we will show that asymptotic behavior of the profile of players’ beliefs is well described by the corresponding “unitary” SFP with one agent in each player role. (In the second and third models, this requires the assumption that the population is sufficiently large.) More precisely, in each case we use Theorem A to show that the  $\omega$ -limits of our SFP are almost surely contained in an internally chain-transitive set of a deterministic continuous-time dynamic with heterogeneous beliefs, and that the subspace where all agents in a given population have the same beliefs is a globally attracting set for the heterogeneous system; restricted to this subspace, the dynamic is the same as the “unitary” dynamic that captures the long-run behavior of systems with a single agent in each player role. We then show that an equilibrium that is linearly unstable for the unitary system is linearly unstable for the system with heterogeneous beliefs. This lets us use Theorems A and B to argue that the SFP must converge to an internally chain-transitive set that is not a linearly unstable rest point in the unitary system. Thus, allowing for multiple agents with heterogeneity in beliefs does not change the asymptotic results derived from stochastic approximation.

The first model, with synchronous clocks and two distinct populations, is the simplest. Intuitively, since all agents are observing draws from the same distribution, a suitable version of the law of large numbers implies that the difference between two agents’ predictions about play converges to 0 almost surely. Since beliefs determine behavior and the smoothed best response function is continuous, one can show that the behavior of all agents in a given population must become the same. Instead of formalizing this intuition, our proof uses the techniques of stochastic approximation; this sacrifices a little generality but makes it compare the analysis here to that of the subsequent cases.

In the second model, with a single population, the empirical distribution

observed by one agent may be different from that observed by another even in the limit, because an agent is never matched with himself. This effect is small when the population size is large, and we can show that it disappears in the long run in a large population. (This is in essence a continuity argument, and the conclusion is false if players use (discontinuous) exact best responses.)

Adding asynchronous clocks does not change the qualitative feature of these analysis. We will explain this in the third model.

### 3.1 Synchronous Matching in Two Populations

There are two populations  $p = 1, 2$  in the society, each of which consists of  $M$  agents. In every period  $n = 1, 2, \dots$ , each agent  $i$  in population  $p$  is matched with a randomly chosen agent  $\mu_n(i)$  in population  $q = 3 - p$ ; that is, there is probability  $1/M$  of being matched with each agent in the other population, independent across periods. The matched agents do not observe others' past histories. When matched, agents  $1i$  and  $2j$  simultaneously choose strategies that we denote  $s_{1i,n}$  and  $s_{2j,n}$ ; each agent then observes the outcome of their match at the current period, but not the outcomes in other matches.

At the end of period  $n$ , agent  $i$  in population  $p$  has a belief  $\theta_{pi,n} \in \Sigma$  about his opponent's play in the next period. As in usual fictitious play, this is the average of what the agent has observed and his initial weights  $\kappa_{pi,0}$  of size  $\bar{\kappa}_{pi,0} = \sum_{s \in S} \kappa_{pi,0}(s)$ :

$$\theta_{pi,n} = \frac{1}{n + \bar{\kappa}_{pi,0}} \left( \sum_{k=1}^n s_{q\mu_k(i),k} + \kappa_{pi,0} \right).$$

Along any realization of the process, beliefs evolve according to the recursion

$$\begin{aligned} \theta_{pi,n+1} - \theta_{pi,n} &= \frac{1}{n + \bar{\kappa}_{pi,0} + 1} (s_{q\mu_{n+1}(i),n+1} - \theta_{pi,n}) \\ &= \frac{1}{n + 1} (s_{q\mu_{n+1}(i),n+1} - \theta_{pi,n} + b_{pi,n+1}), \end{aligned}$$

where we introduce the error term

$$b_{pi,n+1} = -\frac{\bar{\kappa}_{pi,0}}{n + \bar{\kappa}_{pi,0} + 1} (s_{q\mu_{n+1}(i),n+1} - \theta_{pi,n}).$$

Thus we can write the system with an equal step size for each agent although the "size" of their initial weights can differ.

Note that each agent plays a smoothed best response to his beliefs about play in the other population. Since the agents are different, they could in principle have different smoothed best response functions, corresponding to different stochastic or deterministic payoff shocks. We explore this possibility below, but for now we suppose that all agents use a common smoothed best response function.

Next we introduce noise terms  $U_{pi}$  to capture the difference between the realized play of the agent's opponent and its expected value:

$$\theta_{1i,n+1} - \theta_{1i,n} = \frac{1}{n+1} \left( \frac{1}{M} \sum_{j=1}^M \overline{BR}(\theta_{2j,n}) - \theta_{1i,n} + U_{1i,n+1} + b_{1i,n+1} \right), \quad (\text{SFP1})$$

$$\theta_{2j,n+1} - \theta_{2j,n} = \frac{1}{n+1} \left( \frac{1}{M} \sum_{i=1}^M \overline{BR}(\theta_{1i,n}) - \theta_{2j,n} + U_{2j,n+1} + b_{2j,n+1} \right),$$

where

$$U_{pi,n+1} = s_{q\mu_{n+1}(i),n+1} - \frac{1}{M} \sum_{j=1}^M \overline{BR}(\theta_{qj,n}).$$

The first step of the analysis is to show that these noise terms satisfy the conditions of Theorem A.

**Lemma 1.**  $\{U_{pi,n}\}$  and  $\{b_{pi,n}\}$  defined above satisfy the following properties:

1.  $\{U_{pi,n}\}$  and  $\{b_{pi,n}\}$  are adapted to  $\{\mathcal{F}_n\}$ ,
2.  $E(U_{pi,n+1}|\mathcal{F}_n) = 0$ ,  $|U_{pi,n}| \leq 1$ ,
3.  $|b_{pi,n}| \leq \bar{\kappa}_{pi,0}/(n + \bar{\kappa}_{pi,0})$ , and the right-hand side is square-summable.

*Proof.* 1.  $U_{pi,n}$  and  $b_{pi,n}$  are obviously measurable with respect to  $\mathcal{F}_n$ .

2. Since  $\mu_{n+1}(i)$  is uniformly drawn from  $\{1, \dots, M\}$ , and, and each player's play does not depend on the identity of his current opponent, for each possible value  $j$  of  $\mu_{n+1}(i)$ ,  $s_{qj,n+1}$  is a random variable whose distribution conditional on that the state of the system at period  $n$  and the realization of matching at period  $n+1$  is  $\overline{BR}(\theta_{qj,n})$ . Therefore,  $E(s_{q\mu_{n+1}(i),n+1}|\mathcal{F}_n) = E(E(s_{qj,n+1}|\mathcal{F}_n, \mu_{n+1}(i) = j)|\mathcal{F}_n) = E(\overline{BR}(\theta_{qj,n})|\mathcal{F}_n) = (1/M) \sum_{j=1}^M \overline{BR}(\theta_{qj,n})$ . Thus  $E(U_{pi,n+1}|\mathcal{F}_n) = 0$ .

For every  $s \in S$ ,  $-1 \leq U_{pi,n}(s) \leq 1$ , so we have  $|U_{pi,n}| \leq 1$ .

3. Since  $|s_{q\mu_n(i),n} - \theta_{pi,n-1}| \leq 1$ , we have  $|b_{pi,n}| \leq \bar{\kappa}_{pi,0}/(n + \bar{\kappa}_{pi,0})$ .  $\square$

Taking the limit as the step size goes to 0, and ignoring the perturbation terms  $\{U_n\}$  and  $\{b_n\}$ , leads to the system of ODEs:

$$\begin{aligned}\dot{\theta}_{1i}(t) &= \frac{1}{M} \sum_{j=1}^M \overline{BR}(\theta_{2j}(t)) - \theta_{1i}(t), \\ \dot{\theta}_{2j}(t) &= \frac{1}{M} \sum_{i=1}^M \overline{BR}(\theta_{1i}(t)) - \theta_{2j}(t),\end{aligned}\tag{Hetero}$$

which induces semi-flow  $\Phi$  on  $\Sigma^{2M}$ .

By Theorem A and Lemma 1, we can characterize the long run behavior of the discrete-time stochastic process  $\{(\theta_{11,n}, \dots, \theta_{2M,n})\}$  by analyzing the system (Hetero). Specifically, with probability 1, the  $\omega$ -limit of the realization of (SFP1) is internally chain-transitive under (Hetero).

Let  $H = \{(\theta_{11}, \dots, \theta_{2M}) \in \Sigma^{2M} \mid \theta_{11} = \dots = \theta_{1M}, \theta_{21} = \dots = \theta_{2M}\}$  be the subspace where all agents in a given population have the same (“homogeneous”) beliefs, just as they do in the unitary models with a single agent in each role. Then the following holds:

**Lemma 2.**  *$H$  is a globally attracting set for  $\Phi$ .*

*Proof.* For any  $p = 1, 2$ , and  $i \neq j$ , we have  $\dot{\theta}_{pi}(t) - \dot{\theta}_{pj}(t) = -(\theta_{pi}(t) - \theta_{pj}(t))$ , thus  $|\theta_{pi}(t) - \theta_{pj}(t)| \leq e^{-t} |\theta_{pi}(0) - \theta_{pj}(0)| \rightarrow 0$  as  $t \rightarrow \infty$ .  $\square$

Moreover, for any  $(\theta_{11}, \dots, \theta_{2M}) \in H$ , let us denote  $\theta_1 = \theta_{11} = \dots = \theta_{1M}$  and  $\theta_2 = \theta_{21} = \dots = \theta_{2M}$ . Then the semi-flow restricted to  $H$ ,  $\Phi|_H$ , reduces to the system (Unitary).

Let  $L$  be an  $\omega$ -limit set for (SFP1); this is a random variable whose properties depend on the particular realization of the process.

**Theorem 1.**  *$L$  is a subset of  $H$  and internally chain-transitive for  $\Phi|_H$  with probability 1.*

*Proof.* By Theorem A and Lemma 1,  $L$  is internally chain-transitive for  $\Phi$  with probability 1. Fix a probability-1 set of realizations on which  $L$  is invariant and internally chain-transitive for  $\Phi$ .

For each such realization, pick any  $\theta \in L$ . Since  $L$  is invariant,  $\Phi_t(\theta) \in L$  for all  $t \in \mathbb{R}_+$ . Also, since  $H$  is a globally attracting set by Lemma 2,  $d(\Phi_t(\theta), H) \rightarrow 0$ . Thus, by the compactness of  $L$  and  $H$ ,  $L \cap H \neq \emptyset$ .

Again, by Lemma 2,  $H$  is an attracting set for  $\Phi$ , thus  $L \cap H$  is also an attracting set for  $\Phi|L$ . Since  $\Phi|L$  admits no proper attracting set, we have  $L \subset H$ .

Since the definition of internal chain-transitivity depends only on the semi-flow restricted to  $L$ ,  $L$  is internally chain-transitive for  $\Phi|H$ .  $\square$

Theorem 1 shows that in the long run all agents have the same beliefs, and moreover that the system converges to one of the internally chain-transitive sets of (Unitary), just as it does in the case of a single agent in each population, and in the ‘‘global information’’ case where there are several agents in each population, and each agent observes the outcomes in all matches. The next step is to rule out convergence to linearly unstable equilibria. Let  $\langle x, y \rangle$  be the inner product of  $x$  and  $y$ . If  $y$  is a unit vector, then  $|\langle x, y \rangle|$  is equal to the length of the orthogonal projection of  $x$  on the  $y$  direction. The following is a simple extension of Theorem 5.1 in Benaïm and Hirsch [1999].

**Lemma 3.**  $\liminf_{n \rightarrow \infty} E(|\langle U_{pi,n+1}, e \rangle| | \mathcal{F}_n) > 0$  almost surely for every unit vector  $e \in \mathbb{R}^S$  with  $\sum_{s \in S} e(s) = 0$  under (FP1).

*Proof.* Since  $E(|\langle U_{pi,n+1}, e \rangle| | (\theta_{q1,n}, \dots, \theta_{qM,n}) = (\theta_{q1}, \dots, \theta_{qM}))$  is independent of  $n$  and continuous in  $(\theta_{q1}, \dots, \theta_{qM}) \in \Sigma^M$ , it is sufficient to show

$$E(|\langle U_{pi,n+1}, e \rangle| | (\theta_{q1,n}, \dots, \theta_{qM,n}) = (\theta_{q1}, \dots, \theta_{qM})) > 0$$

for every  $(\theta_{q1}, \dots, \theta_{qM})$ .

Since  $(1/M) \sum_j \overline{BR}(\theta_{qj})$  is in the interior of  $\Sigma$ , there exists  $s \in S$  such that

$$\langle s, e \rangle > \left\langle \frac{1}{M} \sum_j \overline{BR}(\theta_{qj}), e \right\rangle.$$

Thus

$$\begin{aligned} & E(|\langle U_{pi,n+1}, e \rangle| | (\theta_{q1,n}, \dots, \theta_{qM,n}) = (\theta_{q1}, \dots, \theta_{qM})) \\ & \geq \min_j \overline{BR}(\theta_{qj,n+1})(s) \left( \langle s, e \rangle - \left\langle \frac{1}{M} \sum_j \overline{BR}(\theta_{qj}), e \right\rangle \right) > 0 \end{aligned}$$

since  $\overline{BR}(\theta_{qj,n+1})(s) > 0$  for every  $j$ .  $\square$

**Theorem 2.** *If  $(\sigma_1^*, \sigma_2^*)$  is a linearly unstable rest point for (Unitary), then*

$$\begin{aligned} \lim_{n \rightarrow \infty} \theta_{11,n} &= \cdots = \lim_{n \rightarrow \infty} \theta_{1M,n} = \sigma_1^*, \\ \lim_{n \rightarrow \infty} \theta_{21,n} &= \cdots = \lim_{n \rightarrow \infty} \theta_{2M,n} = \sigma_2^* \end{aligned}$$

*with probability 0.*

*Proof.* Let  $(\sigma_1^*, \sigma_2^*)$  be a linearly unstable equilibrium for (Unitary). Every eigenvector of the Jacobian of (Unitary) at  $(\sigma_1^*, \sigma_2^*)$  is an eigenvector of the Jacobian for (Hetero) at  $(\sigma_1^*, \dots, \sigma_1^*, \sigma_2^*, \dots, \sigma_2^*)$ . Since (Unitary) has an eigenvalue with a strictly positive real part, so does (Hetero). Hence  $(\sigma_1^*, \dots, \sigma_1^*, \sigma_2^*, \dots, \sigma_2^*)$  is a linearly unstable rest point for (Hetero). The theorem now follows from Theorem B and Lemmas 1 and 3.  $\square$

Theorems 1 and 2 let us immediately extend the past results on SFP that were mentioned at the end of Section 2.

**Heterogeneous Perturbations** Now we relax the assumption that all agents use the same smoothed best response function. This will not make any difference for Lemmas 1–3, so Theorem A applies and the subspace  $H$  of homogeneous beliefs is a globally attracting set; the semi-flow  $\Phi|_H$  is now

$$\begin{aligned} \dot{\theta}_1(t) &= \frac{1}{M} \sum_{i=1}^M \overline{BR}_{1i}(\theta_2(t)) - \theta_1(t), \\ \dot{\theta}_2(t) &= \frac{1}{M} \sum_{j=1}^M \overline{BR}_{2j}(\theta_1(t)) - \theta_2(t). \end{aligned} \tag{HP}$$

The question now is to relate the asymptotic behavior of this flow to that of (Unitary) and thus to past results. The case where all agents in population 1 use  $\overline{BR}_1$ , while agents in population 2 use  $\overline{BR}_2$  is trivial because the literature on fictitious play with two populations has allowed asymmetric payoff functions. The following lemma is useful for the more general case:

**Lemma 4.** *Suppose that, for every  $p$  and  $i$ ,  $\overline{BR}_{pi}$  is derived from maximizing perturbed utility functions of the form  $u(\sigma, \sigma') + v_{pi}(\sigma)$ , where  $v_{pi}$  satisfies the assumptions of (DP). Then there exists  $v_{pA}$  that satisfies the assumptions*

of (DP) and generates  $\overline{BR}_{pA}$  such that  $\overline{BR}_{pA}(\sigma') = (1/M) \sum_i \overline{BR}_{pi}(\sigma')$  for every  $\sigma' \in \Sigma$ . Thus the system (HP) is equivalent to the system

$$\begin{aligned}\dot{\theta}_1(t) &= \overline{BR}_{1A}(\theta_2(t)) - \theta_1(t), & (\text{Aggregate}) \\ \dot{\theta}_2(t) &= \overline{BR}_{2A}(\theta_1(t)) - \theta_2(t).\end{aligned}$$

*Proof.* This follows from the proof of Theorem 2.1 in Hofbauer and Sandholm [2002], which shows that a choice function  $C: \mathbb{R}^S \rightarrow \text{int}(\Sigma)$  can be derived from (DP):

$$C(\pi) = \arg \max_{\sigma} (\langle \sigma, \pi \rangle + v(\sigma))$$

if and only if its derivative  $DC(\pi)$  is symmetric, negative definite, and satisfies  $\sum_s DC(\pi)(s) = 0$ : these properties are inherited by convex combinations of such functions.  $\square$

Several results in the literature on SFP in two-player games hold for all smoothed best response functions, so they apply to the aggregate best response function  $\overline{BR}_{pA}$ . This is the case for example for Hofbauer and Sandholm's [2002] global convergence results in two-player zero-sum games and in potential games with where all of the rest points are hyperbolic, so in these cases we can conclude that the aggregate system converges to a rest point of (Aggregate). Moreover, such a rest point is close to Nash equilibria of the unperturbed stage game if each agent's stochastic or deterministic payoff perturbations are small enough.

**Theorem 3.** *If (i) every  $\overline{BR}_{pi}$  is derived from additive random utility model (SP) with vanishing random components,  $\varepsilon_{pi} \rightarrow 0$  in probability, or (ii) every  $\overline{BR}_{pi}$  is derived from deterministic perturbation model (DP) with vanishing deterministic perturbation,  $\sup_{\sigma, \sigma'} |v_{pi}(\sigma) - v_{pi}(\sigma')| \rightarrow 0$ , then a fixed point of  $(\overline{BR}_{1A}, \overline{BR}_{2A})$  converges to Nash equilibria of the unperturbed stage game.*

*Proof.* See Appendix.  $\square$

In contrast, in some  $3 \times 3$  games, the stability of an equilibrium can depend on the details of the perturbation used to generate the best response; at this point it is an open question whether any of the various sufficient conditions for stability and instability are preserved by aggregation.

### 3.2 Synchronous Matching in One Population

Now we consider SFP in a single population with  $M$  agents, and revert to the assumption that each agent uses the same smoothed best response  $\overline{BR}$ . Here each agent sees a slightly different sample since they do not play against themselves. Intuitively, we expect that this would not make much difference to play in sufficiently large populations, since each agent is only  $1/M$  of the population; we will see that the right measure of “sufficiently large” depends on the sensitivity of the smoothed best response function to small changes in beliefs.

Specifically, in every period  $n$ , each agent  $i$  is matched with another agent  $\mu_n(i)$ ; for each  $j \neq i$ ,  $\mu_n(i) = j$  with probability  $1/(M-1)$ , and  $\mu_n$  is independent across periods. Agent  $i$ 's strategy at period  $n$  is denoted by  $s_{i,n}$ .

Agent  $i$  with initial weights  $\kappa_{i,0}$  updates his belief  $\theta_{i,n}$  about his opponent's play in the next period according to

$$\begin{aligned}\theta_{i,n+1} - \theta_{i,n} &= \frac{1}{n + \bar{\kappa}_{i,0} + 1} (s_{\mu_{n+1}(i),n+1} - \theta_{i,n}) \\ &= \frac{1}{n+1} (s_{\mu_{n+1}(i),n+1} - \theta_{i,n} + b_{i,n+1}),\end{aligned}$$

where  $\bar{\kappa}_{i,0} = \sum_{s \in S} \kappa_{i,0}(s)$  and

$$b_{i,n+1} = -\frac{\bar{\kappa}_{i,0}}{n + \bar{\kappa}_{i,0} + 1} (s_{\mu_{n+1}(i),n+1} - \theta_{i,n}).$$

Since  $E(s_{j,n+1} | \mathcal{F}_n) = \overline{BR}(\theta_{j,n})$  and  $\mu_{n+1}(i)$  is uniformly drawn from  $\{1, \dots, i-1, i+1, \dots, M\}$ , we have

$$\theta_{i,n+1} - \theta_{i,n} = \frac{1}{n+1} \left( \frac{1}{M-1} \sum_{j \neq i} \overline{BR}(\theta_{j,n}) - \theta_{i,n} + U_{i,n+1} + b_{i,n+1} \right), \quad (\text{SFP2})$$

where

$$U_{i,n+1} = s_{\mu_{n+1}(i),n+1} - \frac{1}{M-1} \sum_{j \neq i} \overline{BR}(\theta_{j,n}).$$

As in Lemma 1,  $\{U_n\}$  and  $\{b_n\}$  are adapted to  $\{\mathcal{F}_n\}$ ,  $E(U_{i,n+1} | \mathcal{F}_n) = 0$ ,  $|U_{i,n}| \leq 1$ , and  $|b_{i,n}| \leq \bar{\kappa}_{i,0}/n$ .

Stochastic approximation yields the system of ODEs:

$$\dot{\theta}_i(t) = \frac{1}{M-1} \sum_{j \neq i} \overline{BR}(\theta_j(t)) - \theta_i(t). \quad (\text{Hetero-Sym})$$

This system induces semi-flow  $\Phi'$  on  $\Sigma^M$ . The corresponding system with unitary beliefs is

$$\dot{\theta}(t) = \overline{BR}(\theta(t)) - \theta(t). \quad (\text{Unitary-Sym})$$

Let  $H' = \{(\theta_1, \dots, \theta_M) \in \Sigma^M \mid \theta_1 = \dots = \theta_M\}$  be the subspace where all agents have the same beliefs, and recall that  $K$  is a Lipschitz constant for  $\overline{BR}$ .

**Lemma 5.** *If  $M > K + 1$ , then  $H'$  is a globally attracting set for the semi-flow  $\Phi'$  induced by (Hetero-Sym).*

*Remark 6.* This result follows from a more general result (Lemma 7) in the next subsection. Intuitively, regardless of the history of play, the maximum difference  $|\theta_i - \theta_j|$  in the beliefs of agents  $i$  and  $j$  is  $1/(M-1)$ . This is magnified by the slope  $K$  of  $\overline{BR}$ , so that the difference decreases if  $K/(M-1) < 1$ .

*Remark 7.* The lemma only gives a sufficient condition, but as we discuss in the next section, simulations suggest that the bound is fairly tight and that for  $M < K + 1$  heterogeneous beliefs can persist in an anti-coordination game.

Hofbauer and Sandholm [2002] show that (Unitary-Sym) characterizes the long-run behavior of a system where two agents are drawn from a single population to play each period, all agents observe the outcome each period, and agents update on the basis of past outcomes, treating observations of their own past behavior the same as observations of others; this last assumption explains why their results are independent of the population size. This is not the case here, as we can see by considering  $M = 2$ , where (Hetero-Sym) is equivalent to the system (Unitary) that describes the usual two-population model, and the symmetric game with payoffs.

$$\begin{array}{cc} (1, 1) & (2, 3) \\ (3, 2) & (1, 1) \end{array}$$

In this case the system (Unitary-Sym) converges to the mixed equilibrium, while the standard two-population SFP converges to a rest point that approximates one of the asymmetric pure equilibria. Thus not only do the

beliefs of the two agents remain different, aggregate play in the long run is different from that in the unitary symmetric model.

Let  $L$  be an  $\omega$ -limit for (SFP2). Replacing Lemma 2 by Lemma 5 and modifying Lemmas 1 and 3, we obtain the following theorems as in the previous subsection.

**Theorem 4.** *If  $M > K + 1$ , then  $L$  is a subset of  $H'$  and internally chain-transitive for  $\Phi'|H'$  with probability 1.*

**Theorem 5.** *If  $\sigma^*$  is a linearly unstable equilibrium for (Unitary-Sym), then*

$$\lim_{n \rightarrow \infty} \theta_{1,n} = \dots = \lim_{n \rightarrow \infty} \theta_{M,n} = \sigma^*$$

*with probability 0.*

Theorems 4 and 5 show that the results mentioned at the end of Section 2 apply here provided that  $M$  is sufficiently large.<sup>13</sup> As the smoothed best responses become close to the exact best responses,  $M$  needs to grow to keep  $M > K + 1$ .

Based on Theorems 4 and 5, we can extend some results that are specific to one-population models: Hofbauer and Sandholm [2002] show that play converges to the unique rest point of (Unitary-Sym) in games with an interior ESS. On the other hand, Hopkins [1999b] shows that if the payoff matrix corresponding to payoff function  $u$  is positive definite and the smoothed best response takes the logit form with parameter  $\beta$  sufficiently large, then the equilibrium corresponding to a fully mixed symmetric Nash equilibrium of the stage game is linearly unstable. Theorems 4 and 5 show that these results carry over to (SFP2).

---

<sup>13</sup>As we have seen, the aggregate outcome of (Hetero-Sym) can differ from that of (Unitary-Sym) if  $M < K + 1$ . However, we conjecture that the two models have the same implications when the unitary model has a globally attracting equilibrium whenever both  $M$  and  $K$  are large. In particular, suppose that under (Unitary-Sym) with logit best response, for all  $\beta > \bar{\beta}$ , there is  $\sigma_\beta^*$  that is globally stable, and that  $\sigma_\beta^* \rightarrow \sigma^*$  as  $\beta \rightarrow \infty$ . Then we conjecture that there is  $\bar{M}$  such that for all  $M > \bar{M}$ , and all  $\beta > \bar{\beta}$ , there is  $\sigma_{\beta,M}^*$  such that the submanifold where average play corresponds to  $\sigma_{\beta,M}^*$  is globally stable under (Hetero-Sym) with  $d(\sigma_{\beta,M}^*, \sigma_\beta^*) < \varepsilon$ . In this case, with large populations the unitary model is a good approximation to aggregate play regardless of the relative magnitudes of  $M$  and  $K$ .

**Heterogeneous Perturbations** Once we introduce heterogeneous smoothed best responses functions  $\{\overline{BR}_i\}$ , the asymptotic behavior of SFP is approximated by

$$\dot{\theta}_i(t) = \frac{1}{M-1} \sum_{j \neq i} \overline{BR}_j(\theta_j(t)) - \theta_i(t).$$

In this model, the set of homogeneous beliefs,  $H'$ , may not be globally attracting, as each agent is observing and reacting to a different distribution of opponents' strategies. For this reason, we cannot extend Theorems 4 or 5 to heterogeneous perturbations. However, since the set of chain-recurrent points is upper semi-continuous in  $\beta$ ,<sup>14</sup> if the heterogeneity among agents is sufficiently small, the agents' beliefs converge near an internally chain-transitive set of (Hetero-Sym) with some smooth best response function  $\overline{BR}$ ; our previous results imply that this is near an internally chain-transitive set of (Unitary-Sym) with the same  $\overline{BR}$ .

**Theorem 6.** *Fix a smoothed best response function  $\overline{BR}$  with Lipschitz constant  $K$ . Assume that  $M > K + 1$ . For any  $\varepsilon > 0$ , there exists  $\delta > 0$  such that, if  $|\overline{BR}_i(\sigma) - \overline{BR}(\sigma)| < \delta$  for every  $i = 1, \dots, M$  and  $\sigma \in \Sigma$ , then  $(\theta_1, \dots, \theta_M)$  converges to an  $\varepsilon$ -neighborhood of an internally chain-transitive set for  $\Phi|H'$  with probability 1.*

### 3.3 One Population—Asynchronous Clocks

The models we have considered so far suppose that all agents play every period, so that every agent knows that his current opponent has played the game exactly as often as he has. This is typically the case in game theory experiments, but in many field settings, agents play at varying frequencies, and do not know how many times their current opponents have played the game. To model such situations, we now we consider a model with “asynchronous clocks.”

Specifically, at each moment, two agents are drawn randomly from a single population of size  $M$  agents; we denote the chosen agents at moment  $n$  by the set  $C_n$ . When an agent is chosen, he chooses a smoothed best response in game  $G$ , given his assessment of the opponent's actions. After the play, he updates his assessment based on the realized strategy of his partner. If a

---

<sup>14</sup>This will be proved in a forthcoming paper by Benaïm, Hofbauer, and Sorin.

player is not chosen, on the other hand, then he keeps the same assessment as before.

We can describe this situation more formally. As in SFP with synchronous clocks, at the beginning of period  $n$ , each agent  $i$  has weights  $\kappa_{i,n}$ , and forms the probability  $\theta_{i,n}$  that agent  $i$  assigns to his opponent's strategy at the next period. If agent  $i$  is drawn from the population as one of the two active players at period  $n$  (i.e.,  $i \in C_n$ ), then he chooses a strategy  $s_{it}$  according to  $\overline{BR}(\gamma_{i,n})$ .

The difference between this system and the usual SFP is that each agent only updates his assessment when he is drawn to play:

$$\kappa_{i,n+1}(s') = \begin{cases} \kappa_{i,n}(s') + 1 & \text{if } C_n = \{i, j\} \text{ and } s_{j,n-1} = s' \text{ for some } j, \\ \kappa_{i,n}(s') & \text{otherwise.} \end{cases}$$

Thus,  $\kappa_{i,n}(s')$  is equal to the initial weight on  $s'$ ,  $\kappa_1^i(s')$ , plus the number of periods in which player  $i$  was drawn and his opponent chose  $s'$  before period  $n$ . Set  $\bar{\kappa}_{i,n} = \sum_{s \in S} \kappa_{i,n}(s)$ ; then  $1/\bar{\kappa}_{i,n}$  is agent  $i$ 's *step size* at period  $n$ , which governs the size of influence of a new observation. In order to control different step sizes for different agents, we introduce variables that count the step size for each agent, and then apply stochastic approximation to this extended state space; these step size variables cannot be “run backwards” to time negative infinity, which is why we need to consider semi-flows and not flows.

Let

$$y_{i,n} = \frac{\bar{\kappa}_{i,n} + 1}{n + 1}$$

be the ratio between the step sizes of synchronous and asynchronous clock models. Using this variable, we write the recursion for  $\{\theta_{i,n}\}$  as follows:

$$\begin{aligned} \theta_{i,n+1} - \theta_{i,n} &= I(i \in C_{n+1}) \frac{1}{n+1} \frac{1}{y_{i,n}} (s_{j,n+1} - \theta_{i,n}) \\ &= \frac{1}{n+1} \left[ \frac{2}{M} \frac{1}{y_{i,n}} \left( \frac{1}{M-1} \sum_{j \neq i} \overline{BR}(\theta_{j,n}) - \theta_{i,n} \right) + U_{i,n+1} \right], \end{aligned} \tag{SFP3}$$

where  $I$  is the indicator function, and

$$U_{i,n+1} = \frac{1}{y_{i,n}} \left[ I(i \in C_{n+1}) (s_{j,n+1} - \theta_{i,n}) - \frac{2}{M} \left( \frac{1}{M-1} \sum_{j \neq i} \overline{BR}(\theta_{j,n}) - \theta_{i,n} \right) \right].$$

Note that  $\{y_{i,n}\}$  satisfies the following recursion:

$$\begin{aligned} y_{i,n+1} - y_{i,n} &= \frac{1}{n+2} (I(i \in C_{n+1}) - y_{i,n}) \\ &= \frac{1}{n+1} \left( \frac{2}{M} - y_{i,n} + U'_{i,n+1} + b'_{i,n+1} \right), \end{aligned}$$

where

$$\begin{aligned} U'_{i,n+1} &= I(i \in C_{n+1}) - \frac{2}{M}, \\ b'_{i,n+1} &= -\frac{1}{n+2} (I(i \in C_{n+1}) - y_{i,n}). \end{aligned}$$

We will apply Theorems A and B to the process  $\{(\theta_{1,n}, \dots, \theta_{M,n}, y_{1,n}, \dots, y_{M,n})\}$  and approximate the process by the following system of ODEs:

$$\begin{aligned} \dot{\theta}_i(t) &= \frac{2}{M} \frac{1}{y_i(t)} \left( \frac{1}{M-1} \sum_{j \neq i} \overline{BR}(\theta_j(t)) - \theta_i(t) \right), & (\text{Uniform-Sym}) \\ \dot{y}_i(t) &= \frac{2}{M} - y_i(t), \end{aligned}$$

which induces a semiflow  $\Phi''$  on  $\Sigma^M \times \mathbb{R}_{++}^M$ .

**Lemma 6.** *The process (SFP3) defined above satisfies the following properties:*

1. *The system of ODEs (Uniform-Sym) is globally integrable,*
2.  *$\{(U_{i,n}, U'_{i,n})\}$  and  $\{b'_{i,n}\}$  are adapted to  $\{\mathcal{F}_n\}$ ,*
3.  *$E(U_{i,n+1} | \mathcal{F}_n) = E(U'_{i,n+1} | \mathcal{F}_n) = 0$ ,  $\limsup_n E(|U_{i,n+1}|^2 | \mathcal{F}_n) \leq (M/2)^2$  almost surely,  $|U'_{i,n}| \leq 1$ ,  $\liminf_n E(|\langle U_{i,n+1}, e \rangle| | \mathcal{F}_n) > 0$  almost surely for every  $e \in \mathbb{R}^S$  with  $\sum_s e(s) = 0$ ,*
4.  *$|b'_{i,n}| \leq (\bar{\kappa}_{i,0} + 1)/(n+1)$ , the right-hand side is square-summable, and*
5.  *$\{(\theta_{1,n}, \dots, \theta_{M,n}, y_{1,n}, \dots, y_{M,n})\}$  is precompact almost surely.*

*Proof.* 1. Since  $\min(y_i(0), 2/M) \leq y_i(t) \leq \max(y_i(0), 2/M)$  for  $t \geq 0$ , (Uniform-Sym) is globally integrable in the positive direction, i.e., the semiflow  $\Phi''$  is well defined.

2. The measurability is obvious.
3. The martingale difference condition follows from the fact that  $E(I(i \in C_{n+1})|\mathcal{F}_n) = 2/M$  and  $E(s_{j,n+1}|\mathcal{F}_n, C_{n+1} = \{i, j\}) = \overline{BR}(\theta_{j,n})$ .  
It is easy to see that  $|U_{i,n+1}| \leq 1/y_{i,n}$  and  $|U'_{i,n}| \leq 1$ . Since  $y_{i,n} \rightarrow 2/M$  almost surely by the law of large number,  $\limsup_{n \rightarrow \infty} E(|U_{i,n+1}|^2|\mathcal{F}_n) \leq (M/2)^2$  almost surely.  
Similarly to Lemma 3, we have  $E(|\langle U_{i,n+1}, e \rangle||\mathcal{F}_n) \geq c/y_{i,n}$  with some  $c > 0$ . Since  $y_{i,n} \rightarrow 2/M$  almost surely,  $\liminf_n E(|\langle U_{i,n+1}, e \rangle||\mathcal{F}_n) \geq 2c/M > 0$  almost surely.
4. Since  $I(i \in C_n) \leq 1$  and  $y_{i,n-1} \leq y_{i,0} = \bar{\kappa}_{i,0} + 1$ , we have  $|b_{i,n}| \leq (\bar{\kappa}_{i,0} + 1)/(n + 1)$ .
5. By the law of large number,  $y_{i,n} \rightarrow 2/M$  is bounded away from 0 and  $\infty$  almost surely. □

Let  $L$  be an  $\omega$ -limit for (SFP3). Using Lemma 6, we can apply Theorems A and B to (SFP3) to obtain the following results. Recall that  $\Phi'$  is the semi-flow induced by (Hetero-Sym), and  $H' = \{(\theta_1, \dots, \theta_M) | \theta_{1,n} = \dots = \theta_{M,n}\}$ .

**Theorem 7.**  $L = L_0 \times \{(2/M, \dots, 2/M)\}$  for some  $L_0 \subset \Sigma^M$ , and, if  $M > K + 1$ , then  $L_0$  is a subset of  $H'$  and internally chain-transitive for  $\Phi'|H'$  with probability 1.

*Proof.* By Theorem A and Lemma 6,  $L$  is internally chain-transitive for  $\Phi''$  with probability 1. Since  $X_0 = \Sigma^M \times \{(2/M, \dots, 2/M)\}$  is a globally attracting set for  $\Phi''$ ,  $L = L_0 \times \{(2/M, \dots, 2/M)\}$  for some  $L_0 \subset \Sigma^M$ . Since  $\Phi''|X_0$  is the same as  $\Phi'$  except for the last  $M$  components, by Lemma 5, if  $M > K + 1$ , then  $L_0 \subset H'$ . Thus  $L_0$  is internally chain-transitive for  $\Phi'|H'$ . □

**Theorem 8.** If  $\sigma^*$  is a linearly unstable equilibrium for (Unitary-Sym), then

$$\lim_{n \rightarrow \infty} \theta_{1,n} = \dots = \lim_{n \rightarrow \infty} \theta_{M,n} = \sigma^*$$

with probability 0.

*Proof.* The theorem follows from Theorem B and Lemma 6 since the eigenvalues of the Jacobian matrix for (Unitary-Sym) at  $\sigma^*$  are a part of the eigenvalues of the Jacobian matrix for (Uniform-Sym) at  $(\sigma_1^*, \dots, \sigma_M^*, 2/M, \dots, 2/M)$ . □

**Non-uniform Matching** Now we extend the previous analysis to i.i.d. but non-uniform matching. Assume that, for each period  $n$ ,  $C_n = \{i, j\}$  with probability  $p_{ij}$ . Note that  $p_{ii} = 0$ ,  $p_{ij} = p_{ji} \geq 0$ , and  $\sum_{1 \leq i < j \leq M} p_{ij} = 1$ . Also assume that, for each agent  $i$ , there exists  $j \neq i$  such that  $p_{ij} > 0$ .<sup>15</sup> As in the related literature on social learning about a common payoff parameter (cf. Ellison and Fudenberg [1993], Bala and Goyal [1998], DeMarzo, Vayanos and Zweibel [2003], Golub and Jackson [2007]), interaction probabilities  $\{p_{ij}\}$  can be interpreted as a network; for example, the circle model of Ellison [1993] is roughly analogous to  $p_{ij} = 1/2$  for  $i - j = \pm 1 \pmod{M}$  and  $p_{ij} = 0$  otherwise.<sup>16</sup>

Since each agent plays the game with a positive probability, his step size at period  $n$  is approximately proportional to  $1/n$ , which allows us to use stochastic approximation and obtain

$$\begin{aligned}\dot{\theta}_i(t) &= \frac{1}{y_i(t)} \sum_{j \neq i} p_{ij} \overline{BR}(\theta_j(t)) - \theta_i(t), \\ \dot{y}_i(t) &= \sum_{j \neq i} p_{ij} - y_i(t).\end{aligned}$$

Since  $y_i(t) \rightarrow \sum_{j \neq i} p_{ij}$  as  $t \rightarrow \infty$ , it is enough to analyze the following system of ODEs:

$$\dot{\theta}_i(t) = \sum_{j \neq i} q_{ij} \overline{BR}(\theta_j(t)) - \theta_i(t), \quad (\text{Nonuniform-Sym})$$

where  $q_{ij} = p_{ij} / \sum_{k \neq i} p_{ik}$  is the probability that agent  $i$ 's opponent is agent  $j$  conditional on agent  $i$  being drawn from the population to play the game. This system induces semi-flow  $\Phi^q$  on  $\Sigma^M$ . We extend Lemma 5 to non-uniform matching, showing that  $H'$  is a globally attracting set for  $\Phi^q$  if the population is large enough and matching is close enough to uniform.

---

<sup>15</sup>This assumption does not lose generality. If, for some  $i$ ,  $p_{ij} = 0$  for all  $j \neq i$ , then we can exclude player  $i$  from our analysis.

<sup>16</sup>Of course, the interpretation of the network structure is different in the two types of models. Models of myopic best responses on networks, such as Ellison's, assume that each agent knows the current actions of all of her neighbors. Each period, each agent plays a best response to the mixed strategy corresponding to the current actions of the neighbors, where the action of each agent  $j$  gets weight  $p_{ij}$ , thus the interaction process is treated as deterministic, perhaps as the result of a round-robin tournament within each period. In contrast, random selection of opponents would make our model stochastic even if each agent played a myopic best response to either her last observation or to the current opponent's play in the previous period.

Let

$$\Delta = \max_{1 \leq i < j \leq M} \sum_{k=1}^M \max(q_{ik} - q_{jk}, 0)$$

be the maximum difference between the distributions of two agents' opponents. For example,  $\Delta = 1/(M - 1)$  under uniform matching. This is a measure of the maximum difference of the beliefs of any two agents, so we can conclude that the beliefs and play must converge when  $K\Delta < 1$  (Lemma 5).

**Lemma 7.** *If  $K\Delta < 1$ , then  $H'$  is a globally attracting set for the semi-flow  $\Phi^q$  induced by (Nonuniform-Sym).*

*Proof.* See Appendix. □

## 4 Convergence to a Cycle

In some games, the usual SFP converges to a cycle, and never converges to any of the Nash equilibria. Our results show that the same is true for SFP with heterogeneous beliefs, even in the case of asynchronous clocks. Specifically, consider the rock-scissors-paper “game B” from Benaïm, Hofbauer, and Hopkins [2006]:

$$\begin{array}{ccc} (0, 0) & (-3, 1) & (1, -3) \\ (1, -3) & (0, 0) & (-2, 1) \\ (-3, 1) & (1, -2) & (0, 0) \end{array} .$$

This game has a unique Nash equilibrium  $\sigma^* = (9, 10, 13)/32$ ; Benaïm, Hofbauer, and Hopkins show that under the exact best response dynamics, the mixed equilibrium  $\sigma^*$  is unstable, and that there is a “Shapley polygon”  $C$  (i.e., a cycle) that attracts all orbits that do not start at the equilibrium.

Now consider logistic best response dynamics with parameter  $\beta \rightarrow \infty$  so that the smoothed best response puts weight tending to 1 on the exact best responses, and let  $\sigma^\beta$  be the Nash distribution (the fixed point of the best response functions) as a function of  $\beta$ . Because the set of chain-recurrent points is upper semi-continuous in  $\beta$ , the only chain-recurrent points of (Unitary-Sym) for large  $\beta$  are all close to either  $C$  or  $\sigma^*$ . Hopkins [2002] shows that the Nash distribution  $\sigma^\beta$  is linearly unstable for large  $\beta$ , so Theorem B shows that SFP does not converge to  $\sigma^\beta$ . Benaïm, Hofbauer, and

Hopkins use a Lyapunov function to show that  $\sigma^\beta$  is a repeller for large  $\beta$ ; a small modification of their argument shows that the repelling neighborhood does not vanish as  $\beta \rightarrow \infty$ . This shows that SFP cannot remain even in a neighborhood of  $\sigma^\beta$  and so (from the upper semi-continuity result) must converge to a neighborhood of  $C$ .

**Lemma 8.** *There are  $\delta > 0$  and  $\bar{\beta}$  such that, for all  $\beta > \bar{\beta}$ , the only chain-recurrent set of (Unitary-Sym) in the  $\delta$ -neighborhood  $N_\delta$  of  $\sigma^*$  is the Nash distribution  $\sigma^\beta$ .*

*Proof.* See Appendix. □

**Theorem 9.** *For every  $\delta > 0$ , there is  $\bar{\beta}$  such that for all  $\beta > \bar{\beta}$ , the  $\omega$ -limit of SFP is in the  $\delta$ -neighborhood of  $C$  with probability 1.*

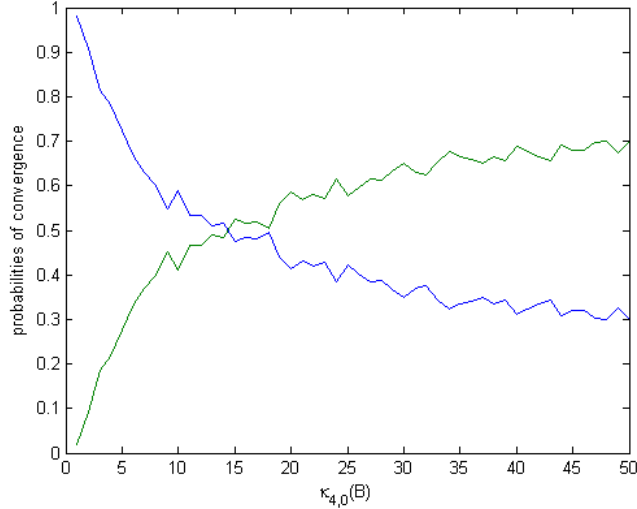
*Proof.* This follows from the lemma, the upper semi-continuity of chain-recurrent points, and the fact that, if an internally chain-transitive set intersects the basin of an attractor, then it is contained in that attractor (Benaïm [1999], Corollary 5.4). □

## 5 Simulations: Initial Condition Dependence and the Role of “Prior Confidence”

We have shown that in the long run the system must converge to an internally chain-transitive set of the unitary system, and that it cannot converge to a linearly unstable equilibrium. We now use some simulations to explore how the relative probabilities of various attractors depend on the parameters of the system, and in particular on its initial conditions. Both sets of simulations are for the model with asynchronous clocks and uniform matching, and suppose that the agents use the logit best response function with parameter  $\beta = 7$ ; with the payoff matrices we use, this implies that the maximum slope of the smooth best response is  $K = 3.5$ . The figures all report the distribution of the state of the system after 10,000 periods, where the distribution comes from averaging over 100 runs of simulation for each parameter configuration.

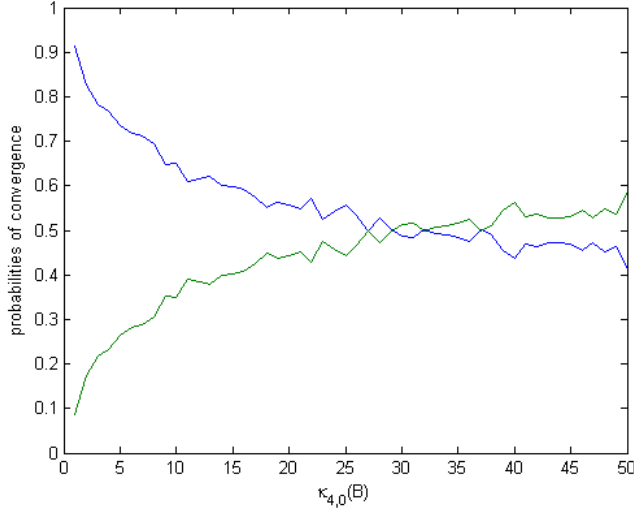
Our first set of simulations concerns a pure coordination game with payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} (1, 1) & (0, 0) \\ (0, 0) & (1, 1) \end{pmatrix}, \end{array}$$



where both actions are equally efficient. The population size is set at  $M = 4$ , and we suppose that three of the four agents have prior beliefs  $\kappa_{i,0} = (1, 0)$ , so they initially believe that  $A$  is very likely, but they are not very certain of that belief. The fourth agent has prior  $\kappa_{4,0} = (0, \kappa_{4,0}(B))$ , where we vary the parameter  $\kappa_{4,0}(B)$  to investigate the impact of the agent’s confidence in his prior. Thus, in the initial period, 3 of the 4 agents play  $A$  with high probability, while, in the second round 2, agents play  $A$  with high probability, one agent is indifferent and gives each action equal weight, and one agent plays  $B$  with high probability (if  $\kappa_{4,0}(B)$  is much larger than 1).

The further development of the system depends on the extent of agent 4’s prior conviction, and the long-run result is plotted in Figure 1. The  $y$ -axis in this figure plots two curves: the blue line is the fraction of runs where at the terminal date  $N = 10,000$ ,  $\theta_{i,N}(A) > .9$  for all agents  $i$ ; the black line plots the fraction of runs where  $\theta_{i,N}(B) > .9$ . The sum of the two plots ranges from .996 to 1, indicating that the system almost always converges to one of the equilibrium distributions, as predicted by the theoretical results. From the figure we see that if  $\kappa_{4,0}(B)$  is only 3 or 4, the system converges to a state where all agents expect action  $A$ ; this is because agent 4 is likely to start playing  $A$  before other agents become “convinced” of  $B$ . When  $\kappa_{4,0}(B)$  is about 15, the system is equally likely to converge to either equilibrium, while for large values of  $\kappa_{4,0}(B)$ , play tends to converge to  $B$ .



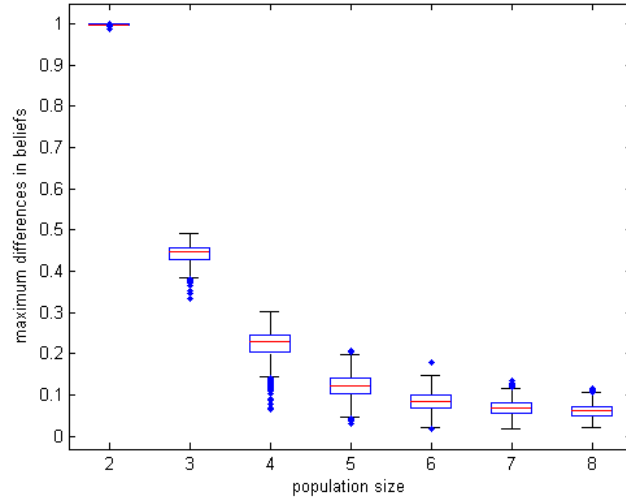
The next figure plots the same data for the “global information” model where agents have the same priors as above, but all agents observe the outcome of every match. Here all 4 agents shift their priors towards  $A$  after the first period outcome, so we expect that the one “very convinced”  $B$ -agent will have less impact and that play should be less likely to converge to  $B$ . The figure shows that this is the case, though the effect is not pronounced: when  $\kappa_{4,0}(B) = 15$ ,  $A$  has empirical probability of about .6 instead of .5 and  $\kappa_{4,0}(B) = 25$  is required for the two equilibria to be equally likely.

We also use simulations to test the tightness of the condition  $M > K + 1$  for conformity in the one-population model with uniform matching. This plot is for the game with payoffs

$$\begin{array}{cc} 0, 0 & 1, 1 \\ 1, 1 & 0, 0 \end{array}$$

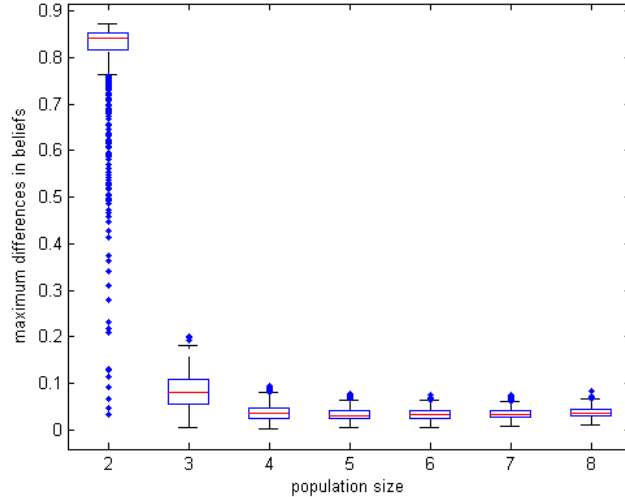
The initial weights for all agents are  $\kappa_{i,0}(A) = \kappa_{i,0}(B) = 1$ .

The  $x$ -axis is the population size,  $M = 2, 3, \dots, 8$ ; the  $y$ -axis plots the distribution of  $\max_{i,j} |\theta_{i,N}(A) - \theta_{j,N}(A)|$  at the end of  $N = 10,000$  periods. The simulation was run 100 times for each value of  $M$ , and the diagram shows the .25, .50, and .75 fractiles of the resulting distribution, along with the outliers. The theorem predicts that the difference should be 0 once



$M > K + 1 = 4.5$ , i.e.,  $M \geq 5$ ; the plot shows a marked decrease by  $M = 5$ , but even with  $M = 8$  the dispersion is non-trivial, which suggests that the convergence to uniformity can be very slow.

The plot for  $\beta = 3$  show a sharp decrease at  $M = 2$ , which conforms with the theoretical prediction; the dispersion falls to a lower level than with  $\beta = 7$  but it is still not 0.



## 6 Concluding Remarks

### 6.1 Related Literature

Hopkins [1999a] considers a deterministic model of “population fictitious play” with a continuum of agents.<sup>17</sup> He supposes that, over a small interval of time  $\Delta t$ , agents are matched an “arbitrarily large” number of times, but that they adjust their beliefs by  $\Delta t$  as much as they would in a period of length 1; this corresponds to an (unmodelled) limit on the frequency of interaction within each period. He then adds a form of once-off perturbation to the initial attractiveness of each strategy,<sup>18</sup> and takes the continuous-time limit, which is the same whether players observe the outcomes of all matches or only the outcomes of their own. He shows that the resulting process is a positive definite dynamic, as is SFP; this means that the eigenvalues of the

<sup>17</sup>Fudenberg and Levine [1993b] also used a deterministic, continuum-population model to study non-equilibrium Bayesian learning. They allowed for heterogeneous beliefs, which can persist because the agents’ observation functions are endogenous, but they only consider the steady states of the system and do not analyze its dynamics.

<sup>18</sup>These perturbations do not correspond either to a permanent payoff shock (because their effect asymptotically vanishes) or to a variation in initial beliefs (because they can make the agents play a strictly dominated strategy). Instead, the perturbations correspond to a payoff shock whose magnitude decreases over time at a given rate.

linearized versions of both processes are purely imaginary at any fully mixed equilibrium of a constant-sum game, and that an interior ESS profile, if it exists, is stable under both processes.

Benaïm and Weibull [2003] assume that a pair of agents is selected in each period, and that all agents have the same observations and beliefs, as in the one-population model of Hofbauer and Sandholm [2002]. They consider a different sort of limit, sending the population size to infinity to obtain a deterministic approximation. Benaïm et al. [2006] consider a variant of one-population, unitary-belief SFP where agents give less weight to older observations, and use stochastic approximation techniques to analyze the iterated limit where first the “discounting” of past observations vanishes and then the parameter  $\beta$  in the logit best response goes to infinity. They show that, in this case, the time average of play always converges, even though this need not be the case under SFP (Benaïm and Hirsch [1999].) It would be very interesting to extend their analysis to models with personal histories considered here.

Hopkins [2002] applies stochastic approximation to a model of perturbed reinforcement learning. In the first formulation of this model, the step sizes are stochastic and different for each agent, as in our asynchronous-clock model; to obtain convergence to a perturbed equilibrium, he then considers a version where the step size are normalized to be proportional to  $1/n$ .

Hart and Mas-Colell [2000, 2001] study systems of “regret learners” with a fixed population of  $m$  agents playing an  $m$ -player game without trying to manipulate the future play of their opponents. Just as with SFP, this strategic myopia can be justified by appeal to a large population of agents and random matching, so it would be interesting to know when the Hart and Mas-Colell convergence results extend to large populations with heterogeneous histories of play and hence heterogeneous regrets. Foster and Young [2006] study an aspiration-based learning rule in a two-player game with two agents in a setting where the agents see only their realized payoffs, and do not realize that they are playing a game. When agents are unaware that they are involved in a specific game, the payoffs they receive may in fact come from interactions with several different other agents, so once again it would be interesting to know when the convergence results extend to large populations and heterogeneous “beliefs.”

The work of Acemoglu et al. [2007], and Cripps et al. [2006], like ours, is based on Bayesian learning; the papers differ in that what is being observed is exogenous. In Acemoglu et al., the agents have a common set of observations,

but each one uses a different likelihood function to interpret the data, and the question is when all agents come to have the same beliefs about the distribution of an underlying parameter  $\theta$ . In Cripps et al., agents receive private signals with a known joint distribution. The signals are assumed to identify  $\theta$ , so in the long run all agents assign probability converging to 1 to the true value; the focus is on the higher order beliefs of the agents, and when the true value of  $\theta$  becomes almost common knowledge.

## 6.2 Conclusion

In addition to the results in the line of Theorems A and B, the literature also has results of the form “every attractor of the continuous time system has a positive probability of containing the  $\omega$ -limit of the system,” as in Benaïm [1999, Theorem 7.3]. We have proved similar results for our synchronous-clock models with one or two populations, but not for the asynchronous-clock model; the complication there is that the system is not well behaved in the neighborhood of  $y_i = 0$ . Instead, we have used simulations to get more precise estimates of the relative probabilities of various attractors.

Since our conclusions are about the long-run behavior of the learning system, they should be insensitive to the details of play in the initial periods. For example, the result that beliefs converge to a unitary state should extend to learning rules that are “asymptotically empirical” and “asymptotically myopic” in the sense of Fudenberg and Kreps [1993].

Heterogeneous beliefs and private histories seem natural in a number of other learning models in addition to SFP, for example in models where agents have a short memory and play a best response to their most recent observation, or to a subsample of the last few outcomes, as in Young [1993]; we hope to explore this in future work.

## Appendix

*Proof of Theorem 3.* If every  $\overline{BR}_{pi}$  is derived from (SP), then  $\overline{BR}_{pA}$  is derived from (SP) with random component  $\varepsilon_{pA}$  whose cumulative distribution function is the average of the cumulative distribution functions of  $\varepsilon_{pi}$ . Thus, if  $\varepsilon_{pi} \rightarrow 0$  in probability for every  $i$ , then  $\varepsilon_{pA} \rightarrow 0$  in probability as well. Therefore, the statement follows from Hofbauer and Sandholm [2002, Proposition 3.1].

Suppose that  $\overline{BR}_{pi}^k$  is derived from (DP) with deterministic perturbation  $v_{pi}^k$  such that  $\sup_{\sigma, \sigma'} |v_{pi}^k(\sigma) - v_{pi}^k(\sigma')| \rightarrow 0$  as  $k \rightarrow \infty$ . Let  $(\sigma_1^k, \sigma_2^k)$  be a fixed point of  $(\overline{BR}_{1A}^k, \overline{BR}_{2A}^k)$  that converges to  $(\sigma_1^*, \sigma_2^*)$ . For every  $k$ , every  $p \neq q$ , and every  $\sigma_p \in \Sigma$ , we have

$$u(\overline{BR}_{pi}(\sigma_q^k), \sigma_q^k) + v_{pi}^k(\overline{BR}_{pi}(\sigma_q^k)) \geq u(\sigma_p, \sigma_q^k) + v_{pi}^k(\sigma_p)$$

by the definition of  $\overline{BR}_{pi}$ . Since  $\sigma_p^k = \overline{BR}_{pA}(\sigma_q^k)$ , we have

$$\begin{aligned} u(\sigma_p^k, \sigma_q^k) + \frac{1}{M} \sum_i v_{pi}^k(\overline{BR}_{pi}(\sigma_q^k)) &= u(\overline{BR}_{pA}(\sigma_q^k), \sigma_q^k) + \frac{1}{M} \sum_i v_{pi}^k(\overline{BR}_{pi}(\sigma_q^k)) \\ &= \frac{1}{M} \sum_i (u(\overline{BR}_{pi}(\sigma_q^k), \sigma_q^k) + v_{pi}^k(\overline{BR}_{pi}(\sigma_q^k))) \\ &\geq u(\sigma_p, \sigma_q^k) + \frac{1}{M} \sum_i v_{pi}^k(\sigma_p). \end{aligned}$$

As  $k \rightarrow \infty$ , we have  $u(\sigma_p^*, \sigma_q^*) \geq u(\sigma_p, \sigma_q^*)$ . Since this holds for any  $p \neq q$  and any  $\sigma_p$ ,  $(\sigma_1^*, \sigma_2^*)$  is an exact Nash equilibrium.  $\square$

We use the following lemma to prove Lemma 7.

**Lemma 9.** *Every linear combination of vectors  $\{x_i\}$  with zero sum of coefficients,  $\sum_i a_i x_i$ , is equal to a linear combination of  $\{x_i - x_j\}$ ,  $\sum_{i,j} b_{ij}(x_i - x_j)$ , with  $b_{ij} \geq 0$  and  $\sum_{i,j} b_{ij} = \sum_{i < j} \max(a_i - a_j, 0)$ .*

*Proof.* This lemma is easily proved by induction on the number of vectors.  $\square$

*Proof of Lemma 7.* By (Nonuniform-Sym), we have

$$e^t \theta_i(t) = \theta_i(0) + \int_0^t e^s \sum_{k \neq i} q_{ik} \overline{BR}(\theta_k(s)) ds.$$

Let  $d(t) = e^t \max_{1 \leq i < j \leq M} |\theta_i(t) - \theta_j(t)|$ . Then, for each pair  $(i, j)$ , by Lemma 9, we have

$$\begin{aligned}
e^t |\theta_i(t) - \theta_j(t)| &= \left| \theta_i(0) - \theta_j(0) - \int_0^t e^s \sum_k (q_{ik} - q_{jk}) \overline{BR}(\theta_k(s)) ds \right| \\
&= \left| \theta_i(0) - \theta_j(0) - \int_0^t e^s \sum_{k,l} w_{k,l}^{i,j} (\overline{BR}(\theta_k(s)) - \overline{BR}(\theta_l(s))) ds \right| \\
&\leq |\theta_i(0) - \theta_j(0)| + \int_0^t e^s \sum_{k,l} w_{k,l}^{i,j} |\overline{BR}(\theta_k(s)) - \overline{BR}(\theta_l(s))| ds \\
&\leq 1 + K \int_0^t e^s \sum_{k,l} w_{k,l}^{i,j} |\theta_k(s) - \theta_l(s)| ds \\
&\leq 1 + K \Delta \int_0^t d(s) ds,
\end{aligned}$$

where  $w_{k,l}^{i,j} \geq 0$  and  $\sum_{k,l} w_{k,l}^{i,j} = \sum_k \max(q_{ik} - q_{jk}, 0) \leq \Delta$ . Thus

$$d(t) \leq 1 + K \Delta \int_0^t d(s) ds.$$

By the Gronwall inequality, we have  $d(t) \leq \exp(K \Delta t)$ . Since  $K \Delta < 1$ ,

$$|\theta_i(t) - \theta_j(t)| \leq \exp[(K \Delta - 1)t] \rightarrow 0$$

as  $t \rightarrow \infty$  uniformly in the initial state.  $\square$

*Proof of Lemma 8.* We use the same Lyapunov function as in Benaïm, Hofbauer, and Hopkins [2006]:

$$L(\theta) = (\overline{BR}(\theta) - \theta)^T A \theta + \beta^{-1} (v(\overline{BR}(\theta)) - v(\theta)),$$

where  $A$  is the payoff matrix corresponding to utility function  $u$ , and  $v(\sigma) = -\sum_s \sigma(s) \ln(\sigma(s))$  is the deterministic payoff perturbation that induces the logistic best response.

Then

$$\dot{L}(\theta) = (\overline{BR}(\theta) - \theta)^T A (\overline{BR}(\theta) - \theta) + \beta^{-1} (\overline{BR}(\theta) - \theta)^T (v'(\overline{BR}(\theta)) - v'(\theta)),$$

and because  $v''$  exists, we can use the mean value theorem to conclude that in a neighborhood  $N_\delta$  of  $\sigma^*$ , there is  $\theta'(\theta)$  such that

$$\dot{L}(\theta) = (\overline{BR}(\theta) - \theta)^T (A + \beta^{-1}v''(\theta'(\theta))) (\overline{BR}(\theta) - \theta).$$

Because  $v''$  is bounded on  $N_\delta$  and independent of  $\beta$ , we see that  $\dot{L}$  is positive for sufficiently large  $\beta$ ,  $\dot{L}$  is positive on the set  $N_\delta/\{\sigma^\beta\}$ , so the repelling neighborhood does not vanish as  $\beta \rightarrow \infty$ .

□

## References

- [1] Acemoglu, D., A. Chernozukov, and M. Yildiz [2007] “Learning and Disagreement in an Uncertain World,” mimeo.
- [2] Bala, V. and S. Goyal [1998] “Learning from Neighbours,” *Review of Economic Studies* 65: 595-621.
- [3] Benaïm, M. [1999] “Dynamics of Stochastic Approximation Algorithms,” In: Azéma, J. et al. (Eds.), *Séminaire de Probabilités 33, Lecture Notes in Mathematics 1709*. Springer-Verlag, Berlin/Heidelberg, 1-68.
- [4] Benaïm, M. and M. Hirsch [1999] “Learning Processes, Mixed Equilibria and Dynamical Systems Arising from Fictitious Play in Perturbed Games,” *Games and Economic Behavior*, 29: 36-72.
- [5] Benaïm, M., J. Hofbauer, and E. Hopkins [2006] “Learning in Games with Unstable Equilibria,” mimeo.
- [6] Benaïm, M., J. Hofbauer, and S. Sorin [2005] “Stochastic Approximations and Differential Inclusions,” *SIAM Journal on Control and Optimization*, 44: 328-348.
- [7] Benaïm, M. and J. Weibull [2003] “Deterministic Approximation of Stochastic Evolution in Games,” *Econometrica*, 71: 873-903.
- [8] Brandière, O [1998] “Some Pathological Traps for Stochastic Approximation,” *SIAM Journal on Control and Optimization*, 36: 1293-1314.
- [9] Brandière, O. and M. Dufflo [1996] “Les algorithmes stochastiques contournent-ils les pièges ?,” *Annales de l’institut Henri Poincaré (B) Probabilités et Statistiques*, 32: 395-427.
- [10] Brown, G. W. [1951] “Iterative Solutions of Games by Fictitious Play,” in *Activity Analysis of Production and Allocation*, ed, by T. Koopmans, New York: Wiley, 374-376.
- [11] Cripps, M., J. Ely, G. Mailath, and L. Samuelson [2006] “Common Learning,” mimeo.

- [12] DeMarzo, P., D. Vayanos, and J. Zweibel [2003] “Persuasion Bias, Social Influence, and Unidimensional Opinions,” *Quarterly Journal of Economics* 118: 909-968.
- [13] Diaconis, P. and D. Freedman [1990] “On the Uniform Consistency of Bayes Estimates for Multinomial Probabilities,” *The Annals of Statistics*, 18: 1317-1327.
- [14] Ellison, G. [1993] “Learning, Local Interaction, and Coordination,” *Econometrica* 61: 1047-1071.
- [15] Ellison, G. [1997] “Learning from Personal Experience: One Rational Guy and the Justification of Myopia,” *Games and Economic Behavior*, 19: 180-210.
- [16] Ellison, G. and D. Fudenberg [1993] “Rules of Thumb for Social Learning,” *Journal of Political Economy* 101: 612-643.
- [17] Ellison, G. and D. Fudenberg [2000] “Learning Purified Equilibria,” *Journal of Economic Theory*, 90: 84-115.
- [18] Ellison, G., D. Fudenberg, and L. Imhof [2007] “Random Matching in Adaptive Dynamics,” mimeo.
- [19] Foster, D. and P. Young [2006] “Regret Testing: A Simple Payoff-Based Procedure for Learning Nash Equilibrium,” *Theoretical Economics*, 1: 341-367.
- [20] Fudenberg, D. and D. Kreps [1993] “Learning Mixed Equilibria,” *Games and Economic Behavior*, 5: 320-367.
- [21] Fudenberg, D. and D. K. Levine [1993a] “Self-Confirming Equilibrium” *Econometrica*, 61: 523-546.
- [22] Fudenberg, D. and D. K. Levine [1993b] “Steady State Learning and Nash Equilibrium,” *Econometrica*, 61: 547-574.
- [23] Fudenberg, D. and D. K. Levine [1995] “Consistency and Cautious Fictitious Play,” *The Journal of Economic Dynamics and Control*, 19: 1065-1089.

- [24] Fudenberg, D. and D. K. Levine [1997] “Measuring Players’ Losses in Experimental Games,” *Quarterly Journal of Economics*, 112: 479-506.
- [25] Golub, B. and M. Jackson [2007] “Naive Learning in Social Networks: Convergence, Influence, and the Wisdom of Crowds,” mimeo.
- [26] Harsanyi, J. [1973] “Games with Randomly Disturbed Payoffs,” *International Journal of Game Theory*, 2: 1-32.
- [27] Hart, S. and A. Mas-Colell [2000] “A Simple Adaptive Procedure Leading to Correlated Equilibrium,” *Econometrica* 68: 1127-1150.
- [28] Hart, S. and A. Mas-Colell [2001] “A General Class of Adaptive Strategies,” *Journal of Economic Theory*, 98: 26-54.
- [29] Hofbauer, J. and E. Hopkins [2005] “Learning in Perturbed Asymmetric Games,” *Games and Economic Behavior* 52: 133-152.
- [30] Hofbauer, J. and W. Sandholm [2002] “On the Global Convergence of Stochastic Fictitious Play,” *Econometrica* 70: 2265-2294.
- [31] Hofbauer, J. and W. Sandholm [2007] “Evolution in Games with Randomly Disturbed Payoffs,” *Journal of Economic Theory* 132: 47-69.
- [32] Hopkins, E. [1999a] “Learning, Matching and Aggregation,” *Games and Economic Behavior*, 26: 79-110.
- [33] Hopkins, E. [1999b] “A Note on Best Response Dynamics,” *Games and Economic Behavior*, 29: 138-150.
- [34] Hopkins, E. [2002] “Two Competing Models of How People Learn in Games,” *Econometrica*, 70: 2141-2166.
- [35] Hopkins, E., and M. Posch [2005] “Attainability of Boundary Points under Reinforcement Learning,” *Games and Economic Behavior*, 53: 110-125.
- [36] Jordan, J. [1993] “Three Problems in Learning Mixed Strategy Nash Equilibria,” *Games and Economic Behavior*, 5: 368-396.
- [37] Kaniovski, Y., and P. Young [1995] “Learning Dynamics in Games with Stochastic Perturbations,” *Games and Economic Behavior*, 11: 330-363.

- [38] Kiefer, J., and J. Wolfowitz [1952] “Stochastic Estimation of the Maximum of a Regression Function,” *Annals of Mathematical Statistics*, 23: 462-466.
- [39] Pemantle, R. [1990] “Nonconvergence to Unstable Points in Urn Models and Stochastic Approximations,” *Annals of Probability*, 18: 698-712.
- [40] Robbins, H., and S. Monro [1951] “A Stochastic Approximation Method,” *Annals of Mathematical Statistics*, 22: 400-407.
- [41] Shapley, L. [1964] “Some Topics in Two Person Games,” in eds. M. Dresher et al., *Advances in Game Theory*. Princeton: Princeton University Press.
- [42] Tarrès, P. [2000] “Pieges Repulsifs,” *Comptes rendus de l’académie des sciences*, t.330, Série 1, 125-130.
- [43] Young, H. P. [1993] “The Evolution of Conventions,” *Econometrica*, 61, 57-84.