Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory ••• (••••) •••–•••
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

Learning in games with unstable equilibria

Michel Benaïm [a], Josef Hofbauer [b], Ed Hopkins [c,*]

[a] *Institut de Mathématiques, Université de Neuchâtel, CH-2007 Neuchâtel, Switzerland*
[b] *Department of Mathematics, University of Vienna, A-1090 Vienna, Austria*
[c] *Department of Economics, University of Edinburgh, 31 Buccleuch Place, Edinburgh EH8 9JY, UK*

Received 13 July 2006; final version received 19 February 2008; accepted 7 September 2008

Abstract

We propose a new concept for the analysis of games, the TASP, which gives a precise prediction about non-equilibrium play in games whose Nash equilibria are mixed and are unstable under fictitious play-like learning. We show that, when players learn using weighted stochastic fictitious play and so place greater weight on recent experience, the time average of play often converges in these "unstable" games, even while mixed strategies and beliefs continue to cycle. This time average, the TASP, is related to the cycle identified by Shapley [L.S. Shapley, Some topics in two person games, in: M. Dresher, et al. (Eds.), Advances in Game Theory, Princeton University Press, Princeton, 1964]. The TASP can be close to or quite distinct from Nash equilibrium.

JEL classification: C72; C73; D83

Keywords: Games; Learning; Best response dynamics; Stochastic fictitious play; Mixed strategy equilibria; TASP

1. Introduction

At the basis of the theory of learning in games is the question as to whether Nash equilibria are stable or unstable. The hope is to predict play: if an equilibrium is an attractor for a plausible learning dynamic, we think that it is a possible outcome for actual play. On the other hand,

* Corresponding author.
E-mail addresses: michel.benaim@unine.ch (M. Benaïm), josef.hofbauer@univie.ac.at (J. Hofbauer), E.Hopkins@ed.ac.uk (E. Hopkins).
URLs: http://www.unine.ch/math/personnel/equipes/benaim/benaim_pers/benaim.html (M. Benaïm), http://homepage.univie.ac.at/Josef.Hofbauer/ (J. Hofbauer), http://homepages.ed.ac.uk/hopkinse/ (E. Hopkins).

0022-0531/$ – see front matter
doi:10.1016/j.jet.2008.09.003

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

if a Nash equilibrium is unstable, we would expect actual players, for example, subjects in an experiment, not to play that equilibrium or even to be close to it. Shapley [38] famously found that there are games for which learning may not approach the only Nash equilibrium but rather will continuously cycle. If we take this result seriously as an empirical prediction, then there are games in which Nash equilibrium play will never emerge. Note that Shapley's result implies that even play averaged over time should not be close to an unstable equilibrium.

In this paper, we advance the novel hypothesis that even when learning diverges from equilibrium, it is still possible to make a precise prediction about play. Surprisingly, in games with a unique unstable mixed equilibrium the time average of play may converge even when players' mixed strategies do not. If an equilibrium is unstable under stochastic fictitious play (SFP) with the classical assumption that players place an equal weight on all past experience, then both mixed strategies and time averages must diverge from equilibrium. But we find that if greater weight is placed on more recent experience, as it is in "weighted" stochastic fictitious play, then although the players' mixed strategies will approach the cycle of the type found by Shapley, the time average will converge. We show that, as the level of noise and the level of forgetting approach zero, the time average of play approaches the TASP (Time Average of the Shapley Polygon), that is, the time average of the Shapley cycle under the continuous time best response dynamics. We find that in many cases the TASP is close to the Nash equilibrium. Since the time average is much easier to observe than mixed strategies, it may well appear that play has converged to the equilibrium. We can also identify games where the TASP and Nash equilibrium are quite distinct, and so offer the possibility of a clearer empirical test between the two.

Specifically, we look at monocyclic games, a class of games that generalises Rock–Paper–Scissors and that has only mixed equilibria. We provide a sufficient condition for the instability of equilibrium in such games under both best response and perturbed best response dynamics in continuous time and show that in this case there is a unique Shapley cycle which is an attractor for the best response dynamics. We then show that this implies that the time average of play in the discrete time weighted fictitious play process will approach the TASP as its step size approaches zero. Furthermore, the time average of the smooth dynamics associated with stochastic fictitious play will also approach the TASP when one simultaneously takes the step size and the level of noise to zero. This is in contrast to the behaviour of the classical fictitious play process, under which there is no convergence for these games even in time average.

These results are not of purely theoretical interest. They, in fact, arise in direct response to recent experimental work on the economically important phenomenon of price dispersion. Cason and Friedman [12] and Morgan, Orzen, and Sefton [37] report on experimental investigations of the price dispersion models of Burdett and Judd [10] and Varian [40] respectively. Both studies report aggregate data that is remarkably close to the price distribution that would be generated if the subjects had been playing the mixed Nash equilibrium. This is surprising if one takes learning theory seriously, as earlier results by Hopkins and Seymour [34] indicate that the mixed equilibria of these models are unstable under most common learning processes. Cason, Friedman and Wagener [13] reexamine the data from Cason and Friedman [12] and indeed find that play is highly non-stationary and there are clear cycles present. They therefore reject the hypothesis that subjects were in fact playing Nash equilibrium. This is also consistent with the earlier results of [9]. They find, in an experimental study of a Bertrand–Edgeworth oligopoly market with no pure equilibrium, that prices cycle but prices averaged across the whole session still approximate the mixed equilibrium distribution. Our results explain the apparent empirical paradox. When mixed equilibria are unstable under learning, we predict persistent cycles in play. Nonetheless, if

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

players learn placing more weight on recent experience, the time average of play should converge to the TASP, which in these games is close to the Nash equilibrium.

It is true that there are existing results in learning theory that show convergence of time averages without convergence to equilibrium. For example, the evolutionary replicator dynamics cycle around mixed strategy equilibria of zero sum games, but the time average of the dynamics nonetheless converge (see, for example, [30, pp. 79, 121, 130]). That is, convergence must be to a Nash equilibrium and then only in a relatively small class of games. In contrast, we obtain convergence in a wide class of games where there is no convergence of any sort under traditional assumptions. Furthermore, we show convergence to the TASP which is distinct from both Nash equilibrium and perturbed equilibrium concepts such as quantal response or logit equilibrium. Alternatively, Hart and Mas-Colell [24] propose a learning model where the time average of play always converges to the set of correlated equilibria. However, this set can be very large, whereas the TASP is a single point.

Fictitious play was introduced many years ago with the underlying principle that players play a best response to their beliefs about opponents, beliefs that are constructed from the average past play of opponents. This we refer to as players having "classical" beliefs. It was in this framework that Shapley [38] obtained his famous result. However, even when fictitious play converges to a mixed strategy equilibrium, it does so only in time average not in marginal frequencies. This problem motivated the introduction of smooth or stochastic fictitious play (see [19] for a survey), which permits convergence in actual mixed strategies. This more recent work still employs classical beliefs. However, experimental work has found greater success with generalisations of fictitious play that allow for players constructing beliefs by placing greater weight on more recent events (see [11,14] amongst many others). This is called forgetting or recency or weighted fictitious play. Despite their empirical success, models with recency have not received much theoretical analysis, largely because they are more difficult to analyse than equivalent models with classical beliefs. This paper represents one of the few attempts to do so.[1]

Many years ago, Edgeworth [15] predicted persistent cycles in a competitive situation where the only Nash equilibrium is in mixed strategies. This view was for a long while superseded by faith that rational agents would play Nash equilibrium, no matter how complicated the model or market. In the case of mixed strategies, learning theory provides some support for Edgeworth, persistent cycles are a possibility even when agents have memory of more than the one period Edgeworth assumed (though in other games, learning will converge even to a mixed equilibrium). Furthermore, recent learning models that allow for stochastic choices do not imply the naive, predictable cycles described by Edgeworth. Cycles may only be detectable by statistical tests for non-stationarity (see [13]). In the absence of such sophisticated analysis, these perturbed Edgeworth–Shapley cycles may to an outside observer look indistinguishable from mixed equilibrium.

However, tests for cyclical play may not be sufficient to identify the TASP. Following [8], strictly one should reject the hypothesis of Nash equilibrium play by experimental subjects if one finds that their play is non-stationary. But the TASP is not the only alternative hypothesis. For example, suppose subjects were learning and this was converging to a Nash equilibrium, only asymptotically would play approach stationarity. In practice, to identify the TASP from experimental data, one would have to make a detailed econometric investigation of the dynamics

[1] Fictitious play with finite memory has been considered (see [42, Ch. 6]). Other learning models not based on fictitious play where the speed of learning does not decrease over time include [6] and [29].

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

to determine whether play was convergent. So, it would be convenient to have a simpler way of separating the TASP and equilibrium play. We therefore give an example where the TASP and Nash equilibrium are quite distinct. These should make possible a simple test simply based on average play. We are therefore optimistic that the theoretical results of this paper can and will be tested.

2. Shapley polygons and Edgeworth cycles

We start with a generalisation of the well-known Rock–Paper–Scissors game and two specific examples,[2]

$$RPS = \begin{array}{|c|c|c|}\hline 0 & -a_2 & b_3 \\ \hline b_1 & 0 & -a_3 \\ \hline -a_1 & b_2 & 0 \\ \hline \end{array}, \qquad A = \begin{array}{|c|c|c|}\hline 0 & -1 & 3 \\ \hline 2 & 0 & -1 \\ \hline -1 & 3 & 0 \\ \hline \end{array}, \qquad B = \begin{array}{|c|c|c|}\hline 0 & -3 & 1 \\ \hline 1 & 0 & -2 \\ \hline -3 & 1 & 0 \\ \hline \end{array}. \tag{1}$$

Game A and game B both have a unique Nash equilibrium in mixed strategies, for A, $x^* = (13, 10, 9)/32 = (0.40625, 0.3125, 0.28125)$ and, for B, $x^* = (9, 10, 13)/32 = (0.28125, 0.3125, 0.40625)$. They appear to be very similar. Learning theory, however, says that they are quite different. Specifically, if a single large population of players are repeatedly randomly matched to play one of these games, most learning and/or evolutionary dynamics, such as fictitious play, the replicator dynamics, reinforcement learning or stochastic fictitious play, should converge to (close to) the Nash equilibrium in game A, but should diverge from equilibrium in game B.

Shapley [38] was the first to show that there are games in which a learning process does not converge to a Nash equilibrium. Instead, the fictitious play process that he examined converged to a cycle of increasing length. We can recreate Shapley's result in the context of a single large population who are repeatedly randomly matched in pairs to play a normal form game such as A or B above. Fictitious play assumes that agents play a best response given their beliefs. The vector x_t represents the belief at time t, with x_{it} the probability given to an opponent playing his ith strategy. That is, $x_t \in S^N$ the simplex $S^N = \{x = (x_1, \ldots, x_N) \in \mathbf{R}^N\colon \sum x_i = 1, x_i \geqslant 0, \text{ for } i = 1, \ldots, N\}$. An agent then chooses a pure strategy that is in the set of best responses to her current beliefs, or $b(x_t)$.[3] The dynamic equation for the fictitious play process in a single population will be

$$x_{t+1} - x_t \in \gamma_t\big(b(x_t) - x_t\big) \tag{2}$$

with γ_t being the step size. Classically, beliefs are assumed to be based on the average of past play by their opponents, which implies that the step size will be equal to $1/(t+1)$. An alternative, that is explored in this paper, is that players place a weight of one on last period's observation, a weight δ on the previous period, and δ^{n-1} on their experience n periods ago, for some $\delta \in [0, 1)$. Then the step size γ_t will be $1 - \delta$, a constant.

Suppose that δ takes the extreme value of 0, "Cournot beliefs," so that players play a best response to the last choice of their opponent. In RPS, as Rock is the best response to Scissors which

[2] We use the evolutionary game theory convention for symmetric games and only give the payoffs for the row player. That is, for a payoff matrix with typical element a_{ij}, if the row player chooses strategy i, and the column player j, row gets a_{ij} and column a_{ji}.

[3] As $b(\cdot)$ is not in general single valued, the dynamics arising from fictitious play present certain mathematical difficulties. See [4] for a full treatment.

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

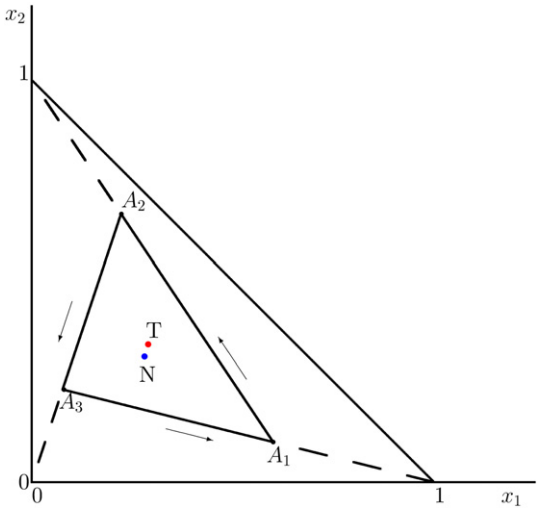

Fig. 1. The Shapley triangle for game B with the TASP (T) and the Nash equilibrium (N).

is the best response to Paper, we would see a cycle of the form $P, S, R, P, S, R, P, S, R, \ldots$. This is a very simple example of an "Edgeworth cycle" of best responses. Clearly, if players follow this cycle the time average of their play will converge to $(1/3, 1/3, 1/3)$. Of course, for some RPS games, this will be equal to or be close to the mixed Nash equilibrium. However, one would not describe this type of behaviour as equilibrium play, as it involves predictable cycles rather than randomisation. Or, more formally, there is only convergence of the time average, but not marginal frequencies.

Under classical beliefs, change will be more gradual. For example, in the case of game B if beliefs are at a point to the right of A_1 in Fig. 1, where x_1 is relatively high, the best response will be the second strategy, or $b(x_t) = e_2 = (0, 1, 0)$. Agents in the population play the second strategy and beliefs about the likelihood of seeing strategy 2 increase. Beliefs move in the direction of the vertex where $x_2 = 1$, until they approach near A_2, and strategy 3 becomes a best response. Then, beliefs will move toward the vertex $e_3 = (0, 0, 1)$ until strategy 1 becomes the best response. That is, there will be cyclical motion about the Nash equilibrium. In game A, it can be shown that over time the cycles converge on the Nash equilibrium, but in game B beliefs converge to the triangle $A_1A_2A_3$ illustrated in Fig. 1 and the cycles are persistent.

The easiest way to prove such convergence results is to use the continuous time best response (BR) dynamics, defined as

$$\dot{x} \in b(x) - x. \tag{3}$$

For a class of games including the game B given in (1), Gaunersdorfer and Hofbauer [22] show that the best response dynamics converge to the "Shapley polygon" (Gilboa and Matsui [23] use the term "cyclically stable set"). In game B this is the triangle $A_1A_2A_3$ illustrated in Fig. 1, but we can give a more general definition.

Definition 1. A *Shapley polygon* is a polygon in S^N with M vertices $A_1, \ldots, A_M$ which is a closed orbit for the best response dynamics (3).

We can then define the TASP as follows.

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

Definition 2. The *TASP* (time average of the Shapley Polygon) is

$$\tilde{x} = \frac{1}{T}\int_0^T b\big(x(t)\big)\,dt = \frac{1}{T}\int_0^T x(t)\,dt \tag{4}$$

where $x(0) = x(T)$.[4] That is, it is the time average of the best response dynamics (3) over one complete circuit of a Shapley polygon.

In the standard case where the best replies along the cycle are pure strategies, it is possible to be more specific. We label an edge the ith edge if on that edge the ith strategy is being played. That is, on that edge, $b(x) = e_i$, that is the vector with 1 at position i and zero elsewhere. Suppose that at some time t_0, the dynamics (3) are at vertex A_{i-1}. Denote the coordinates of the ith vertex as x^{A_i}. Then, because between A_{i-1} and A_i the best response $b(x)$ is e_i, the BR dynamics imply the linear differential equation $\dot{x}_i = 1 - x_i$ with initial condition $x_i(t_0) = x_i^{A_{i-1}}$. Thus, we have on that edge $x_i(t_0 + t) = 1 + \exp(-t)(x_i^{A_{i-1}} - 1)$. Let T_i be the total time spent by the continuous time BR dynamics on the ith edge. Or, let T_i solve $x_i^{A_i} = 1 + \exp(-T_i)(x_i^{A_{i-1}} - 1)$. Then, over one complete circuit of the Shapley polygon, $\tilde{x}_i$ is the proportion of time spent on side i, or,

$$\tilde{x}_i = \frac{T_i}{\sum_{j=1}^M T_j}. \tag{5}$$

Now, Shapley polygons do not exist for every game. For example, in game A in (1) the Nash equilibrium is a global attractor for the best response dynamics and there is no Shapley polygon. But for the game B, there is a Shapley triangle (which is unique and asymptotically stable) and, following [22], we can calculate that $A_1 = (6, 1, 3)/10$, $A_2 = (2, 6, 1)/9$, and $A_3 = (1, 3, 9)/13$ as shown in Fig. 1. The TASP can be computed numerically as $\tilde{x} \approx (0.29, 0.34, 0.37)$, marked as "T" in Fig. 1.

Benaïm, Hofbauer and Sorin [4] recently have extended the theory of stochastic approximation to set valued dynamics. Their results imply that for the game B under classical fictitious play beliefs the discrete time dynamic (2) will approach the Shapley polygon. That is, there will be persistent cycles in beliefs, not convergence to equilibrium. Now, under such classical beliefs, the speed of learning declines each period with accumulated experience. So, movement around the cycle is slower and slower. Observed play might look like this $P, S, R, P, P, S, S, R, R, P, P, P, S, S, S, R, R, R, \ldots$. Consequently, the time average of play does not converge, see Monderer and Shapley [36, Lemma 1] for a general proof.

But what if players place greater weight on more recent experience, with δ not at the extreme value of 0? We show in the current paper that, like for classical fictitious play, beliefs will cycle around the Shapley polygon (or close to it), but at constant speed. Consequently, we can show that, like for the simple Edgeworth cycles, average play will converge, and for δ close to one this time average will be close to the TASP.

Now, as we see in Fig. 1, the TASP is close to the Nash equilibrium of the game B. So, if the population of players do in fact learn according to weighted fictitious play, then average play will be close to the Nash equilibrium because average play will be close to the TASP. However, beliefs

[4] The equality of these two time averages follows by integrating Eq. (3) along the periodic solution $x(t)$ over one period $[0, T]$ such that $x(0) = x(T)$.

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

will continue to cycle. In contrast, in game A both beliefs and average play will converge to the Nash equilibrium. The problem is that beliefs are not directly observable, whereas average play which can be seen, can be misleading. It would be very easy for an experimenter to conclude in the case of game B that play had converged to the Nash equilibrium, when in reality only average play had converged, and to the TASP and not to the Nash equilibrium.

Talk of convergence to point close to but not identical to Nash may well remind readers of quantal response (QRE) or logit equilibria. The literature on these perturbed equilibria is now extensive and there has been considerable success in explaining empirical phenomena. See, for example, [35] or [1]. While they are certainly a competing explanation for non-Nash play, there are important differences. The most important is that QRE is an *equilibrium* concept and assumes stable play. It is, therefore, not consistent with the cycles described above or the non-stationary behaviour present in much experimental data.

Furthermore, there are games in which the TASP and any Nash or quantal response equilibrium are quite different. For example, consider a RPS game with the addition of another strategy D (for "Dumb" as for $c > 0$ and $d < 1$ it is not a best response to any pure strategy).

$$RPSD = \begin{array}{|cccc|}\hline 0 & -3 & 1 & c \\ 1 & 0 & -3 & c \\ -3 & 1 & 0 & c \\ d & d & d & 0 \\ \hline \end{array}. \tag{6}$$

When $c > 0$, then this game has no pure strategy equilibrium. For example if $c = 1/10$ and $d = -1/10$, the unique Nash equilibrium is fully mixed and equal to $(1, 1, 1, 17)/20$. It is possible to calculate that, under the BR dynamics, the Nash equilibrium is a saddle with the stable manifold being the line satisfying $x_1 = x_2 = x_3$. Thus for almost all initial conditions, the BR dynamics diverge. When the weights on the first three strategies are no longer equal, the fourth strategy is not a best reply, so that any weight on x_4 tends to die out as play diverges from equilibrium. But on the face where $x_4 = 0$, we have the original RPS game, and with the above parameter values, there will be a Shapley polygon on the face. Indeed, it is easy to calculate the TASP in this case as $(1/3, 1/3, 1/3, 0)$. That is, the Nash equilibrium places a weight of $17/20$ on the fourth strategy and the TASP places no weight on it whatsoever. For this game, the Nash equilibrium and the TASP are quite distinct.

3. The model

Stochastic fictitious play (SFP) was introduced by Fudenberg and Kreps [18] and is further analysed in [2,16,26–28,32,33]. Models of this kind have been applied to experimental data by Cheung and Friedman [14], Camerer and Ho [11] among others.

Stochastic fictitious play embodies the idea that players play, with high probability, a best response to their beliefs about opponents' actions. With classical fictitious play beliefs, beliefs are constructed from opponents' past play with every observation is given an equal weight. However, the experimental studies cited above all find that players seem to place greater weight on more recent events than is suggested by the classical model. We will consider both cases.

Here, we concentrate on the case where a large population of players are repeatedly randomly matched in pairs to play a two player matrix game with N strategies and payoff matrix A. That is, for those familiar with evolutionary game theory, we analyse a single population learning model, rather than the two population asymmetric case (see [3] for some discussion of the asymmetric

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

case). Time is discrete and indexed by $t = 1, 2, \ldots$. We write the beliefs of a player as $x_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})$, where in this context x_{1t} is the subjective probability in period t that the next opponent will play his first strategy in that period. That is, $x_t \in S^N$. This implies that the vector of expected payoffs of the different strategies for any player, given her beliefs, will be Ax_t. We write the interior of the simplex, that is where all strategies have positive representation, as $\operatorname{int} S^N$ and its complement, the boundary of the simplex as ∂S^N. We also make use of the tangent space of S^N, which we denote $\mathbf{R}_0^N = \{\xi \in \mathbf{R}^N : \sum \xi_i = 0\}$.

Given fictitious play beliefs, if a player were to adopt a strategy $p \in S^N$, she would expect payoffs of $p \cdot Ax$. Following Fudenberg and Levine [19, p. 118 ff], we suppose payoffs are perturbed such that payoffs are in fact given by

$$\pi(p, x) = p \cdot Ax + \lambda v(p) \tag{7}$$

where $\lambda > 0$ scales the size of the perturbation. One possible interpretation is that the player has a control cost to implementing a mixed strategy with the cost becoming larger nearer the boundary. In any case, given appropriate conditions on the perturbation function $v(\cdot)$ (again see Fudenberg and Levine), for each fixed $x \in S^N$ there is a unique $p = p(x) \in \operatorname{int} S^N$ which maximises the perturbed payoff $\pi(p, x)$ for the player. Note that the original formulation of SFP due to Fudenberg and Kreps [18], see also [19, p. 105 ff], involved a truly stochastic perturbation of payoffs. As Hofbauer and Sandholm [28] show, the truly stochastic formulation is a special case of the deterministic approach. In either case, the solution to the perturbed maximisation problem is a smooth function $p(x)$ which approximates the best reply correspondence. The best-known special case is the exponential or logit rule.

We now turn to the dynamic process by which beliefs are updated. We look at a large population: each period the whole population is randomly matched in pairs to play. After each round the vector $X_t \in S^N$ of actions chosen by those who play is publicly announced. The law of large numbers ensures that, given current beliefs x_t, realised play is $X_t = p(x_t)$. For example, in [14], a finite of number of subjects were repeatedly randomly matched in pairs. In the "history" treatment, after each choice they are then informed of the play of all subjects. This treatment, in which all agents play every period and all see the same information is similar to the formal model described above.[5]

In either case, each individual then updates her belief according to the rule,

$$x_{t+1} = (1 - \gamma_t)x_t + \gamma_t X_t. \tag{8}$$

The step-size γ_t will play an important role in our analysis. Under classical fictitious play one sets $\gamma_t = 1/(t + 1)$. That is

$$x_{t+1} = \frac{X_t + X_{t-1} + \cdots + X_1 + x_1}{t + 1},$$

or all observations and initial beliefs x_1 are given equal weight.[6] Here, we explore the implications if players place an exponentially declining weight on past experience with δ being the forgetting factor. This implies that $\gamma_t = 1 - \delta$, a constant, as

$$x_{t+1} = \delta x_t + (1 - \delta)X_t = (1 - \delta)\big(X_t + \delta X_{t-1} + \cdots + \delta^{t-1} X_1\big) + \delta^t x_1.$$

[5] Note that a full treatment of actual experimental protocols would allow for finite numbers, subject heterogeneity in initial beliefs and each player making different observations. The only papers that tackle these problems analytically rather than by simulation are [31] and, more recently, [20].

[6] One can give a different weight to initial beliefs and more generally still one can simply say the step size is of order $1/t$.

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

Setting $\delta = 0$ induces "Cournot" beliefs, only the last period matters, while as δ approaches 1, the updating of beliefs approaches that of classical fictitious play.

If we assume that all agents have the same initial belief and use the same updating rule then, in the large population case, the beliefs in the population will evolve according to

$$x_{t+1} - x_t = \gamma_t\big(p(x_t) - x_t\big) \tag{9}$$

where γ_t is the step size. We will also need the continuous time equivalent to the above discrete dynamic. We have already seen the BR dynamics (3) which corresponds to (2). For the perturbed process (9), we clearly have

$$\dot{x} = p(x) - x, \tag{10}$$

which we can call the perturbed best response (PBR) dynamics.

As is now well known, the steady states of SFP and, equally, the PBR dynamics are not Nash equilibria. Rather, they are perturbed equilibria known as quantal response equilibria (QRE) or logit equilibria. Specifically, a perturbed equilibrium $\hat{x}_\lambda$ satisfies

$$\hat{x}_\lambda = p(\hat{x}_\lambda). \tag{11}$$

Of course, what this equilibrium relationship implies is that beliefs must be accurate or equilibrium beliefs $\hat{x}_\lambda$ are equal to the equilibrium mixed strategy $p(\hat{x}_\lambda)$. However, how close a resulting equilibrium will be to Nash depends on the parameter λ, with the set of perturbed equilibria approaching the set of Nash equilibria as λ approaches zero. See [35] or [1] for further details.

4. Results

In this section we analyse the behaviour of SFP in games with unstable equilibria. We first examine the behaviour of weighted SFP and then contrast our results with the very different behaviour that occurs under classical beliefs. The learning processes that we analyse unfold in discrete time. However, to understand their asymptotic behaviour, it will be crucial to look at some associated continuous time dynamics, the BR (3) and PBR (10) dynamics. Clearly, these are the continuous time analogues of (2) and (9) respectively.

We consider a class of games that Hofbauer [25] calls *monocyclic* (see also, [30, Chapter 14.5]) that generalises the RPS game given in (1). They are two player normal form games with a payoff matrix A that has the following properties:

1. $a_{ii} = 0$.
2. $a_{ij} > 0$ for $i \equiv j + 1 \pmod N$ and $a_{ij} < 0$ else.

The first condition is only a convenient normalisation. Clearly, the strategic properties of these games would not be altered by the addition of a constant to a column. The second condition is much stronger and it ensures that, as the name suggests, monocyclic games have a unique cycle of best responses. Monocylic games do not have equilibria in pure strategies, only mixed equilibria. However, the equilibria of monocyclic games are not necessarily unique and do not have to be fully mixed.

Equilibria of monocyclic games can be stable or unstable under learning. For example, under the continuous time BR dynamics, there is a knife-edge. In particular, if x^* is a completely mixed Nash equilibrium, so that $x^* \cdot Ax^*$ is the equilibrium payoff, then if $x^* \cdot Ax^* < 0$, the equilibrium is unstable, but if $x^* \cdot Ax^* \geqslant 0$, then the equilibrium x^* is globally asymptotically

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

stable (see [25]). For the particular case of 3×3 monocyclic games with an unstable mixed equilibrium, Gaunersdorfer and Hofbauer [22] show that the best response dynamics converge to the "Shapley triangle" introduced in Section 2. The essence of the proof is that it establishes that the best response dynamics in monocyclic games move toward the set defined by $\max(Ax)_i = 0$. That is, the set where the best payoff against the current population state is zero. In games where equilibrium payoffs are negative, this set is distinct from the Nash equilibrium and so the dynamics must diverge from equilibrium. In contrast, the Shapley polygon is contained in this set.[7] In fact, in the 3×3 case the Shapley triangle and the set $\max(Ax)_i = 0$ are identical. Proofs are in Appendix A.

Proposition 1. *Suppose the game A is monocyclic, has a fully mixed Nash equilibrium x^* and $x^* \cdot Ax^* < 0$. Then the mixed Nash equilibrium x^* is unstable under the best response dynamics* (3)*. Furthermore, there is a Shapley polygon, and from an open, dense and full measure set of initial conditions, the best response dynamics converge to this Shapley polygon. The time average from these initial conditions converge to the TASP $\tilde{x}$. That is,*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)\,dt = \tilde{x}.$$

For more than 3 strategies, there are games that have cycles of best responses but which are not monocyclic. The problem is that in such games there may be multiple Shapley polygons. Even worse, the best response dynamics may not converge to either to a Nash equilibrium or a Shapley polygon and instead follow a chaotic orbit. This is why results for $N \times N$ games on convergence to equilibria are rare and to periodic orbits rarer still. Thus, what is remarkable about Proposition 1 is not that it employs restrictive assumptions, it is that there are any such results at all. Note that the above proposition does not claim that there is convergence to the Shapley polygon from all initial conditions. For example, there may be mixed strategy equilibria that are saddle points, and thus attract some initial conditions.[8]

We now consider what the above results on continuous time systems imply for the underlying discrete time learning process. Consider a monocyclic game, with a mixed equilibrium unstable under the best response dynamics. Clearly, we would expect beliefs for the discrete time best response dynamics (2) to diverge as well. However, what happens to the time average of play and of beliefs? Remember that under fictitious play x_t the state variable represents beliefs. The pure strategy that is actually played is given by $b(x_t)$. Let w_t be the time average of play, and $\hat{w}_t$ the time average of beliefs, under this process. That is,

$$w_t = \frac{1}{t}\sum_{s=1}^{t} b(x_s), \qquad \hat{w}_t = \frac{1}{t}\sum_{s=1}^{t} x_s.$$

For the perturbed process (9) corresponding to SFP, we can examine similar averages. We can write them as, respectively,

[7] This relies on the assumption that A is normalised so that $a_{ii} = 0$ for all i.

[8] See [3] for some examples of monocyclic games with mixed equilibria that are saddle points and of games that are not monocyclic.

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

$$z_t = \frac{1}{t}\sum_{s=1}^{t} p(x_s), \qquad \hat{z}_t = \frac{1}{t}\sum_{s=1}^{t} x_s.$$

Remember that in weighted (stochastic) fictitious play the step size of learning γ_t is equal to a constant, $1-\delta$, where δ is the recency parameter (in contrast to the classical case where γ_t is decreasing). We examine what happens to the time averages of play as δ approaches 1 and thus γ approaches zero.

Proposition 2. *Suppose the game A is monocyclic, has a fully mixed Nash equilibrium x^* and $x^* \cdot Ax^* < 0$. Assume the step size $\gamma_t = \gamma$, a constant. Then, for the discrete time best response dynamics* (2)*, for almost all initial conditions x*

$$\lim_{\gamma\to 0}\lim_{t\to\infty} w_t = \lim_{\gamma\to 0}\lim_{t\to\infty} \hat{w}_t = \tilde{x}.$$

Now the upper-semicontinuity result in the proof covers also the discretisation (9) since all limit points of $p(y)$ as $y \to x$ and $\lambda \to 0$ are contained in $b(x)$. Therefore, we obtain a similar result for SFP.

Proposition 3. *Suppose the game A is monocyclic, has a fully mixed Nash equilibrium x^* and $x^* \cdot Ax^* < 0$. Assume the step size $\gamma_t = \gamma$, a constant. Then, for the discrete time perturbed best response dynamics* (9) *from almost all initial conditions x*

$$\lim_{\lambda\to 0,\gamma\to 0}\lim_{t\to\infty} z_t = \lim_{\lambda\to 0,\gamma\to 0}\lim_{t\to\infty} \hat{z}_t = \tilde{x}.$$

The importance of this result is that the time average of play in the large population model of weighted SFP converges to the TASP.

Corollary 1. *Suppose the game A is monocyclic, has a fully mixed Nash equilibrium x^* and $x^* \cdot Ax^* < 0$. Then, in the large population model of* weighted *SFP, for any $\varepsilon > 0$, for all values of λ sufficiently small, t sufficiently large and δ sufficiently close to one, the time average of play z_t and the TASP $\tilde{x}$ satisfy $\|z_t - \tilde{x}\| < \varepsilon$.*

Furthermore, cyclic play actually leads to higher payoffs than playing the Nash equilibrium. Specifically, in monocyclic games, the condition for the Nash equilibrium to be unstable is that the equilibrium payoff is strictly negative. Under weighted SFP, for λ small, beliefs x_t will be close to the Shapley polygon. Thus, the average payoff $p(x_t) \cdot Ap(x_t)$ in the unstable case will much of the time be close to $b(x_t) \cdot Ab(x_t)$ (which is zero due to the normalisation of diagonal payoffs to zero) and hence higher than in equilibrium.

We can compare the result of fictitious play under recency with what happens to fictitious play under classical beliefs, where every observation is given an equal weight and so the step size γ_t is not constant but decreases. Under classical beliefs, the time average of play w_t and beliefs $\hat{w}_t$ are asymptotically identical. When a mixed equilibrium is unstable, typically neither will converge. That is, the limits, rather than being equal to the Nash equilibrium or to the TASP, simply do not exist.

This follows as Proposition 1 establishes that in a class of monocyclic games mixed equilibria are unstable under the BR dynamics (3), and by the stochastic approximation results of [4], beliefs under fictitious play should also diverge from these equilibria. Since by definition classical

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

beliefs are formed from the time average of play, the time average, as for the BR dynamics, for most initial conditions should approach the Shapley polygon. That is, there will be persistent cycles in the time average of play and not convergence.

In contrast, in stable games, like example A in (1), there is relatively little difference in behaviour under classical and weighted SFP. The game A is included in the class of games for which Hofbauer and Sandholm [28] show that the (perturbed) mixed equilibrium will be a global attractor for the perturbed dynamics (10). Thus, play will converge to equilibrium under classical SFP. While this is not the case for weighted SFP for all values of the recency parameter δ, standard results from the theory of stochastic approximation (for example, Theorem 3, p. 44, [7]) imply that asymptotic play will approach the perturbed equilibrium, as one takes the limit $\delta \to 1$.

5. Conclusions

Much of the recent work on learning in games has been concerned with selection between different Nash equilibria, or with providing an adaptive basis for equilibrium play. In this paper, we take a completely different approach. We found that in some games learning under stochastic fictitious play has a non-equilibrium outcome, which nevertheless gives a precise prediction about play. We introduced the TASP (time average of the Shapley polygon), building on earlier results by Shapley [38] and Gaunersdorfer and Hofbauer [22], as an outcome for the time average of play. This we suggest could be useful in understanding behaviour in a number of economically interesting models, including the Varian [40] model of price dispersion and Bertrand–Edgeworth competition.

This also represents one of the few attempts at analysis of learning in games when players place greater weight on more recent experience. Most previous work on stochastic fictitious play and reinforcement learning has examined models with learning that slows over time. This is despite the fact that most empirical work fitting learning models to experimental data has found that weighting recent experience more highly gives a better fit. The two types of models do give similar predictions when considering games that have Nash equilibria that are stable under learning. The finding here, however, is that they give radically different results when considering equilibria that are unstable.

However, there are other learning models besides fictitious play (see [42] for a recent survey) which do not predict divergence. One is due to Hart and Mas-Colell [24]. In their model, the time average of play converges to the set of correlated equilibria of the game in question. In the RPS games the only correlated equilibrium is the Nash equilibrium (see [41]) and so the Hart–Mas-Colell model predicts learning should always converge in this class of games, something that is in distinct contrast with the learning models considered here.[9] Equally, Foster and Young [17] introduce a learning model where each player forms hypotheses about the strategies of her opponents and plays (almost always) a best response given her beliefs. When her observations of her opponents' play are sufficient to reject her current hypothesis, she forms a new hypothesis. Foster and Young show that there are parameter values of the model such that players' mixed strategies will be close to some Nash equilibrium for most of the time in any game. It thus offers

[9] In contrast, Shapley's original game is an example of a game with a unique Nash equilibrium but where the set of correlated equilibria is much larger. In such games, divergence from Nash equilibrium under weighted SFP might be difficult to distinguish empirically from learning according to the model of Hart and Mas-Colell. But this only emphasises that the TASP, being a single point, offers greater precision as a prediction.

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

a different prediction from stochastic fictitious play, which predicts that players' mixed strategies should diverge from equilibrium in some games.

In this paper, we have obtained a series of theoretical results on learning. These are asymptotic results that also depend on taking limiting values of two key parameters that determine the level of noise and recency respectively. This may generate some skepticism about the results' empirical relevance, firstly because real phenomena occur in finite time, and second, because estimates of these parameters from experimental data are not close to these limit values. However, if the TASP is to be dismissed on this basis, so should Nash equilibrium. If one takes stochastic fictitious play or its variants such as EWA learning [11] seriously as models of human behaviour, Nash equilibrium play only occurs as the asymptotic limit of learning behaviour, and then only if the appropriate parameters are at their limit values. Indeed, recent research has found that perturbed equilibria, such as quantal response equilibria [35], that allow the noise parameter not to be at its limit, often fit experimental data better.

The point is that the TASP, like Nash equilibrium, offers a qualitative prediction about behaviour in games that can be made without any parameter estimation. Thus, these concepts can still be empirically useful as an initial hypothesis. One can then go on to make their predictions more precise by using richer models that employ more parameters. In the case of the TASP, it can be generalised by looking at the time average of stochastic fictitious play for which there are two parameters, noise and recency, that can affect the long run outcome. However, these parameters have been jointly estimated in existing attempts to fit stochastic fictitious play to experimental data (see [11,14] among others). Thus, there is no fundamental barrier to taking the TASP to the data.

Appendix A

Proof of Proposition 1. Let B^i be set of points $x \in S^N$ with i being the unique best reply, and B^{ij} be set of points $x \in S^N$ with precisely two pure best replies i and j. The union of all B^i is open, dense and has full $(N-1)$-dimensional Lebesgue measure in S^N. Let $B = \bigcup_{i=1}^N B^i \cup \bigcup_{i=1}^N B^{i-1,i}$. We will show that B is strongly forward invariant under the best response dynamics and all orbits there approach a unique Shapley polygon contained in B.

Suppose $x \in B^1$, i.e., $(Ax)_1 > (Ax)_j$ for all $j \neq 1$. Then $x(t) = e^{-t}x + (1-e^{-t})e_1$ and $(Ax(t))_1 = e^{-t}(Ax)_1$ and for $j \neq 1, 2$,

$$\big(Ax(t)\big)_j = e^{-t}(Ax)_j + \big(1-e^{-t}\big)a_{j1} < e^{-t}(Ax)_j < e^{-t}(Ax)_1 = \big(Ax(t)\big)_1. \tag{1}$$

So along the ray from x to e_1, the best response can only switch from 1 to 2 which indeed must happen for some $t > 0$, since $a_{21} > 0$.

Hence the orbit hits B^{12}. The only way to continue is towards e_2. Repeating the above argument shows that orbits in B^{12} move into B^{23}, etc., and finally from B^{N1} back into B^{12}. This defines a continuous return map $f : B^{N1} \to B^{N1}$. f is single-valued as solutions starting in B are unique. f is a composition of projective maps and hence a projective map itself. Being uniformly continuous, it can be extended to the closure $\overline{B^{N1}}$ of the convex polyhedron B^{N1}. A fixed point of f in B^{N1} generates a closed orbit under the best response dynamics, an invariant N-gon, i.e., a Shapley polygon. However, since $\overline{B^{N1}}$ contains the interior equilibrium x^* we cannot directly apply a fixed point theorem to prove the existence of the Shapley polygon.

Define $V(x) = \max_i (Ax)_i$. As shown above for $x \in B^1$, along any solution $x(t) \in B$, $V(x(t)) = e^{-t}V(x)$. Hence $V(x(t)) \to 0$, as $t \to \infty$.

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

The set $B_0 = B \cap \{x \in S^N\colon\ V(x) = 0\}$ is forward invariant and its closure contains no equilibrium, since $V(\hat{x}) = \hat{x} \cdot A\hat{x} < 0$ holds for each equilibrium $\hat{x}$ (by assumption for the interior equilibrium x^*, and automatically for each boundary equilibrium of a monocyclic game). Since $V(x^*) < 0$ and $V(e_i) = a_{i+1,i} > 0$, each ray from x^* to a point x near e_i hits the set $\{V = 0\}$ in a unique point which is thus contained in $B_0^{i+1} = B^{i+1} \cap \{x \in S^N\colon\ (Ax)_{i+1} = 0\}$, a convex $(N-2)$-dimensional set. The sets $B_0^{i,i+1}$ are therefore $(N-3)$-dimensional. The closure $\overline{B_0^{N1}}$ is a closed and convex polyhedron, mapped by f into itself. So by Brouwer's fixed point theorem, it contains a fixed point (which cannot be an equilibrium). Its orbit is a Shapley polygon Γ.

To prove uniqueness and stability of this Shapley polygon, we use the projective metric d, as in [22]. The distance between two points $x, y \in \operatorname{int} B_0^{N1}$ (the relative interior[10] of B_0^{N1}) is given by the logarithm of the double ratio

$$d(x, y) = \left|\log\left(\frac{xp}{xq} : \frac{yp}{yq}\right)\right|$$

with p, q being the intersection points of the line through x, y with the relative boundary of B_0^{N1}. Since $f(\overline{B_0^{N1}}) \subseteq \overline{B_0^{N1}}$, we have $d(f(x), f(y)) \leqslant d(x, y)$ for $x, y \in \operatorname{int} B_0^{N1}$. Now (1) holds for $j \neq 1, 2$ with a strict inequality even under the weaker assumption $(Ax)_1 \geqslant (Ax)_j$ for all j and $(Ax)_1 > (Ax)_2$. This shows that for $x \in \operatorname{bd} B_0^{N1}$ (with at least a third best reply j besides N and 1), $f(x) \in \operatorname{int} B_0^{N1} = B_0^{N1} \cap \operatorname{int} S^N$. Hence $f(\overline{B_0^{N1}}) \subseteq \operatorname{int} B_0^{N1}$, and hence $d(f(x), f(y)) < d(x, y)$ for $x, y \in \operatorname{int} B_0^{N1}$ with $x \neq y$. Hence, by a variant of Banach's fixed point theorem, the fixed point of f is unique and attracts all orbits in $\overline{B_0^{N1}}$.

Hence all orbits in B approach the Shapley polygon Γ, and Γ is Lyapunov stable. □

Remark. The complement of B consists of all points with at least two non-successive pure best replies (or more than two best replies). The behaviour of orbits starting outside B depends in an intricate way on the payoff matrix. Typically, solutions starting in $x \notin B$ are not unique. From every $x \notin B$ (except possibly x^*) there exists at least one solution that enters B and hence converges to Γ. The solutions staying in $S^N \setminus B$ can converge to a Nash equilibrium or, for $N \geqslant 5$, to an unstable Shapley polygon contained in $S^N \setminus B$.

Proof of Proposition 2. The Shapley polygon Γ with corners $A_1, \ldots, A_N$ is an attractor (= asymptotically invariant set) for (3) whose basin of attraction B is open and dense in S^N, and the complement $S^N \setminus B$ has zero Lebesgue measure. For small $\gamma > 0$, the map (2) has an attractor nearby with basin of attraction exhausting B as $\gamma \to 0$.[11] The time average $\hat{w}_t$ converges to a space average over the attractor of the map (2) with respect to some invariant measure, which tends to the unique measure invariant under the BR dynamics concentrated on the Shapley polygon in the limit as γ goes to zero.[12] The space average with respect to this unique invariant

[10] The relative interior $\operatorname{int} C$ of a convex set $C \subseteq \mathbf{R}^N$ is the interior of C within the affine space spanned by it. The relative boundary of C is then given by $\operatorname{bd} C = \bar{C} \setminus \operatorname{int} C$.

[11] This is well known for discretisations of differential equations, see e.g. [39] or [21]. The corresponding result for differential inclusions needed here for the BR dynamics follows readily by combining their results and methods of proof with those in [4]. See [5] for details.

[12] Again, this is well known for differential equations. For differential inclusions, the corresponding result is shown in [5].

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

measure equals the time average given by the expression (5). The other limit follows from the relation

$$w_t - \hat{w}_t = \frac{1}{t}\sum_{s=1}^{t}\big(b(x_s) - x_s\big) = \frac{1}{t}\frac{1}{\gamma}(x_{t+1} - x_1) \to 0,$$

as t approaches infinity. □

Acknowledgments

We thank Tilman Börgers, Tim Cason, Dan Friedman, Martin Hahn, Larry Samuelson, Bill Sandholm, Sylvain Sorin and Jörgen Weibull for helpful comments. Michel Benaïm thanks the Swiss National Science Foundation for support, Grant 200021-1036251/1. Josef Hofbauer thanks ELSE for support. Ed Hopkins thanks the Economic and Social Research Council for support, award reference RES-000-27-0065.

References

[1] S. Anderson, J. Goeree, C. Holt, The logit equilibrium: A perspective on intuitive behavioral anomalies, Southern Econ. J. 69 (2002) 21–47.
[2] M. Benaïm, M.W. Hirsch, Mixed equilibria and dynamical systems arising from fictitious play in perturbed games, Games Econ. Behav. 29 (1999) 36–72
[3] M. Benaïm, J. Hofbauer, E. Hopkins, Learning in games with unstable equilibria, Working paper, University of Edinburgh, 2005.
[4] M. Benaïm, J. Hofbauer, S. Sorin, Stochastic approximation and differential inclusions, SIAM J. Control Optim. 44 (2005) 328–348.
[5] M. Benaïm, J. Hofbauer, S. Sorin, Stochastic approximation and differential inclusions, Part 3: Extensions, 2008.
[6] M. Benaïm, J. Weibull, Deterministic approximation of stochastic evolution in games, Econometrica 71 (2003) 873–903.
[7] A. Benveniste, M. Métivier, P. Priouret, Adaptive Algorithms and Stochastic Approximations, Springer-Verlag, Berlin, 1990.
[8] J. Brown, R. Rosenthal, Testing the minimax hypothesis: A reexamination of O'Neill's game experiment, Econometrica 58 (1990) 1065–1081.
[9] J. Brown Kruse, S. Rassenti, S.S. Reynolds, V.L. Smith, Bertrand–Edgeworth competition in experimental markets, Econometrica 62 (1994) 343–371.
[10] K. Burdett, K. Judd, Equilibrium price dispersion, Econometrica 51 (1983) 955–969.
[11] C. Camerer, T.-H. Ho, Experience-weighted attraction learning in normal form games, Econometrica 67 (1999) 827–874.
[12] T. Cason, D. Friedman, Buyer search and price dispersion: A laboratory study, J. Econ. Theory 112 (2003) 232–260.
[13] T. Cason, D. Friedman, F. Wagener, The dynamics of price dispersion or Edgeworth variations, J. Econ. Dynam. Control 29 (2005) 801–822.
[14] Y.-W. Cheung, D. Friedman, Individual learning in normal form games: Some laboratory results, Games Econ. Behav. 19 (1997) 46–76.
[15] F.Y. Edgeworth, The pure theory of monopoly, in: Papers Relating to Political Economy, vol. 1, Burt Franklin, New York, 1925.
[16] G. Ellison, D. Fudenberg, Learning purified mixed equilibria, J. Econ. Theory 90 (2000) 84–115.
[17] D.P. Foster, H.P. Young, Learning, hypothesis testing, and Nash equilibrium, Games Econ. Behav. 45 (2003) 73–96.
[18] D. Fudenberg, D. Kreps, Learning mixed equilibria, Games Econ. Behav. 5 (1993) 320–367.
[19] D. Fudenberg, D. Levine, The Theory of Learning in Games, MIT Press, Cambridge, MA, 1998.
[20] D. Fudenberg, S. Takahashi, Heterogeneous beliefs and local information in stochastic fictitious play, Working paper, 2007.
[21] B. Garay, J. Hofbauer, Chain recurrence and discretization, Bull. Austral. Math. Soc. 55 (1997) 63–71.
[22] A. Gaunersdorfer, J. Hofbauer, Fictitious play, Shapley polygons, and the replicator equation, Games Econ. Behav. 11 (1995) 279–303.

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003

[23] I. Gilboa, A. Matsui, Social stability and equilibrium, Econometrica 59 (1991) 859–867.
[24] S. Hart, A. Mas-Colell, A simple adaptive procedure leading to correlated equilibrium, Econometrica 68 (2000) 1127–1150.
[25] J. Hofbauer, Stability for the best response dynamics, Working paper, University of Vienna, 1995.
[26] J. Hofbauer, From Nash and Brown to Maynard Smith: Equilibria, dynamics and ESS, Selection 1 (2000) 81–88.
[27] J. Hofbauer, E. Hopkins, Learning in perturbed asymmetric games, Games Econ. Behav. 52 (2005) 133–152.
[28] J. Hofbauer, W.H. Sandholm, On the global convergence of stochastic fictitious play, Econometrica 70 (2002) 2265–2294.
[29] J. Hofbauer, W.H. Sandholm, Evolution in games with randomly disturbed payoffs, J. Econ. Theory 132 (2005) 47–69.
[30] J. Hofbauer, K. Sigmund, Evolutionary Games and Population Dynamics, Cambridge University Press, Cambridge, UK, 1998.
[31] E. Hopkins, Learning, matching and aggregation, Games Econ. Behav. 26 (1999) 79–110.
[32] E. Hopkins, A note on best response dynamics, Games Econ. Behav. 29 (1999) 138–150.
[33] E. Hopkins, Two competing models of how people learn in games, Econometrica 70 (2002) 2141–2166.
[34] E. Hopkins, R. Seymour, The stability of price dispersion under seller and consumer learning, Int. Econ. Rev. 43 (2002) 1157–1190.
[35] R.D. McKelvey, T.R. Palfrey, Quantal response equilibria for normal form games, Games Econ. Behav. 10 (1995) 6–38.
[36] D. Monderer, L.S. Shapley, Fictitious play property for games with identical interests, J. Econ. Theory 68 (1996) 258–265.
[37] J. Morgan, H. Orzen, M. Sefton, An experimental study of price dispersion, Games Econ. Behav. 54 (2006) 134–158.
[38] L.S. Shapley, Some topics in two person games, in: M. Dresher, et al. (Eds.), Advances in Game Theory, Princeton University Press, Princeton, 1964.
[39] A.M. Stuart, A.R. Humphries, Dynamical Systems and Numerical Analysis, Cambridge University Press, Cambridge, 1996.
[40] H.R. Varian, A model of sales, Amer. Econ. Rev. 70 (1980) 651–659.
[41] Y. Viossat, The replicator dynamics does not lead to correlated equilibria, Games Econ. Behav. 59 (2007) 397–407.
[42] H.P. Young, Strategic Learning and Its Limits, Oxford University Press, 2004.

Please cite this article in press as: M. Benaïm et al., Learning in games with unstable equilibria, J. Econ. Theory (2009), doi:10.1016/j.jet.2008.09.003