Learning, Information and Sorting in Market Entry Games: Theory and Evidence*

John Duffy
Department of Economics
University of Pittsburgh
Pittsburgh PA 15260, USA
jduffy@pitt.edu

Ed Hopkins
Department of Economics
University of Edinburgh
Edinburgh EH8 9JY, UK
E.Hopkins@ed.ac.uk

September 2001

Abstract

Previous data from experiments on market entry games, N–player games where each player faces a choice between entering a market and staying out, appear inconsistent with either mixed or pure Nash equilibria. Here we show that, in this class of game, learning theory predicts sorting, that is, in the long run, agents play a pure strategy equilibrium with some agents permanently in the market, and some permanently out. We conduct experiments with a larger number of repetitions than in previous work in order to test this prediction. We find that when subjects are given minimal information, only after close to 100 periods do subjects begin to approach equilibrium. In contrast, with full information, subjects learn to play a pure strategy equilibrium relatively quickly. However, the information which permits rapid convergence, revelation of the individual play of all opponents, is not predicted to have any effect by existing models of learning.

Journal of Economic Literature classification numbers: C72, D83.

Keywords: Market Entry Games, Reinforcement Learning, Fictitious Play, Experiments.

*We thank David Cooper, Jack Ochs and Amnon Rapoport for helpful comments and suggestions and the U.S. National Science Foundation for research support. Errors remain our own.

1 Introduction

Market entry games are stylized representations of a very common economic problem: a number of agents have to choose independently whether or not to undertake some activity, such as enter a market, go to a bar, drive on a road, or surf the web, the utility from which is decreasing in the number of participants. Those choosing not to undertake the activity can be thought of as staying at home, staying out of the market, or simply not participating. Market entry games typically admit a large number of Nash equilibria. Pure equilibria involve considerable coordination on asymmetric outcomes where some agents enter and some stay out. The only symmetric outcome is mixed, requiring randomization over the entry decision. There may even exist asymmetric mixed equilibria, where some agents adopt pure strategies while others play mixed strategies. Given this multiplicity of equilibrium outcomes, an obvious question is: which type of equilibrium are agents likely to coordinate upon? Many previous experiments have been conducted in an effort to address this and related questions. See, for example, Rapoport et al. (2000, forthcoming(a),(b)), Seale and Rapoport (2000), Camerer and Lovallo (1999), Sundali et al. (1995), Rapoport et al. (1998) and Erev and Rapoport (1998). However, up to now, none of these studies has yielded evidence to suggest that repeated play leads to coordination on any type of Nash equilibrium, and so the question of the relevant equilibrium outcome for this class of games remains open.[1]

In this paper, we investigate the hypothesis that, given sufficient repeated play and adequate feedback, agents should learn equilibrium behavior, but only in the long–run. This assertion leads naturally to further questions: what in practice is "sufficient" and what is "adequate"? How long should we expect to wait before coordination on an equilibrium obtains? What information is necessary? How do these factors interact, for example, does better information lead to faster convergence? In this paper, we attempt to answer these questions in two ways. First, we provide formal results on long–run behavior in market entry games under a number of different models of learning, each of which differs in terms of sophistication and use of information. Second, we report the results of a new series of experiments designed to test these predictions.

These formal results from learning theory are able to discriminate between the many equilibria in market entry games. It is shown that learning must converge, but only to a subset of possible equilibria. There are two models of learning which have attracted particular interest in explaining behavior in laboratory experiments, reinforcement learning and stochastic fictitious play.[2] They differ considerably in terms of sophistication. However, we show that under both, play must converge to an asymmetric pure equilibrium that involves what could be called "sorting", where some players always enter and the remaining players always stay out. In contrast, a model of learning by imitation, which is somewhat intermediate in terms of sophistication. is shown to lead to the symmetric mixed equilibrium. However, these are all asymptotic results. Thus, even if one of these learning models accurately describes human behavior, there is no guarantee that we would see the predicted outcome in the time available for laboratory experiments. What we seek to examine is whether such results are relevant in the timeframe of experiments, and by implication whether they are relevant outside the laboratory.

[1] As Ochs (1998, p. 169) notes in a recent survey of experimental market entry game research, "...a common feature of all market game experiments... is that the aggregate distributions [of entry rates] are not produced by Nash equilibrium profiles, that is, the *individual behavior* observed in all of these experiments is at variance with that implied by the best response conditions of a Nash equilibrium." (emphasis added).

[2] There is now a very large literature. See for example, Fudenberg and Levine (1998), Erev and Roth (1998), Camerer and Ho (1999).

Previous experimental investigations of market entry games have concentrated on testing whether the symmetric mixed equilibrium or an asymmetric pure Nash equilibrium characterize the behavior of experimental subjects. In fact, the data seems to suggest a much more heterogeneous outcome, with some subjects apparently mixing between the two choices and some playing pure. However, the average number of entries per period is in rough accordance with equilibrium. Erev and Rapoport (1998) report two things of interest. First, speed of convergence of the average number of entries toward Nash equilibrium levels is faster when more information is provided. Second, distance from the symmetric mixed equilibrium is decreasing over time.

The conclusion that subjects are converging toward that mixed equilibrium should be treated with caution. We show that under both reinforcement learning and stochastic fictitious play the mixed equilibrium is a saddlepoint, and hence movement toward this equilibrium in the short run is not inconsistent with convergence to a pure strategy equilibrium in the long run. In addition, Erev and Rapoport report a decrease in "alternation" over time, that is, the frequency that an agent plays the strategy which she did not play the previous period, which suggests individuals are getting closer to pure strategies. However, these experiments have not provided ideal data sets to test the predictions of learning theory. For example, Rapoport et al. (1998) had sessions lasting 100 periods, but within that time, the parameter which determined the number of entrants in equilibrium was constantly altered. Erev and Rapoport (1998) kept the parameters constant in each session, but each session lasted 20 periods, which is probably not sufficient for long-term behavior to emerge. As the capacity parameter c changes, the profile of strategies necessary to support a Nash equilibrium also changes, making coordination on a Nash equilibrium extremely challenging.

The new experiments on which we report here have several new features. First, each session involved 100 periods of play of an unchanging market entry game to give some chance for long run behavior to be observed. Second, three different information treatments were employed. In the first "limited information" treatment, subjects were given no initial information about the game being played and each round were only told the payoff they earned. In the second, "aggregate information" treatment subjects were told the payoff function, and then were told after each round the number of subjects who had entered, the number who had stayed out, and the payoff each had earned. In the final "full information" treatment subjects were given the same information as in the aggregate information treatment, but in addition after each round the choice and payoff of each individual subject was revealed.

Our results are somewhat surprising. In the limited information treatment, there is some tendency for groups of subjects to converge upon a pure equilibrium, but only toward the very end of the 100 period session. The aggregate information treatment, despite the additional information provided, produced results very similar to those in the limited information treatment. In the full information treatment, the tendency toward sorting was much greater than in the other two treatments. This is despite the fact that all of the learning models considered predict no effect from the additional information provided in the full information treatment

2 The Game

One very simple formulation of the market entry game, found for example in Erev and Rapoport (1998), is where payoffs are linear in the number of entrants or participants. For example, if player

i's strategy is $\delta^i = 0$ stay out, or $\delta^i = 1$ go in, then her payoff is

$$\pi_i(\delta) = \begin{cases} v, & \text{if } \delta^i = 0, \\ v + r(c - m), & \text{if } \delta^i = 1. \end{cases} \tag{1}$$

Here, v, r, c are positive constants and $0 \leq m \leq N$ is the number of agents that choose entry. The constant c, therefore, has the interpretation as the capacity of the market or road or bar. In particular, the return to entry exceeds the return to staying out, if and only if $m < c$. We can assume $1 \leq c < N$.

2.1 Nash Equilibria

There are many pure strategy Nash equilibria for this class of games. If c is an integer, any profile of pure strategies which is consistent with either c or $c - 1$ entrants is a Nash equilibrium. If c is not an integer, a pure strategy Nash equilibrium involves exactly $\bar{c}$ entrants where $\bar{c}$ is the largest integer smaller than c. Moreover if c is not an integer the number of Nash equilibria is finite, while if c is an integer there is a continuum of equilibria. The latter have the form, $c - 1$ players enter, $N - c$ stay out, and one player enters with any probability. Furthermore, this implies that only when c is not an integer are the pure equilibria strict.

Additionally, for $c > 1$, there is a symmetric mixed Nash equilibrium. This has the form

$$\bar{y}^i = \frac{c-1}{N-1} \text{ for } i = 1, ..., N$$

where $\bar{y}^i$ is the probability of entry by the ith player. Note that the expected number of entrants in the symmetric mixed equilibrium is $c > N(c-1)/(N-1) > c - 1$. There are additional asymmetric mixed equilibria[3] of the form $j < c - 1$ players enter with probability one, $k < N - c$ players stay out with probability one, and the remaining $N - j - k$ players enter with probability $(c-1-j)/(N-j-k-1)$. In one of these asymmetric mixed Nash equilibria, the expected number of entrants is $j + (c-1-j)(N-j-k)/(N-j-k-1)$ which again is between c and $c - 1$. Note though that as k approaches $N - c$, the expected number of entrants approaches c.

The common feature of all these Nash equilibria is that the expected number of entrants is between c and $c - 1$. This basic fact is corroborated by the experimental evidence. However, given the multiplicity of equilibria, it would be preferable if there were some way to select among the different equilibrium possibilities.

The market entry games that we examine here can be considered as one member of the large class of coordination games, characterized by having large numbers of Nash equilibria. However, unlike games of pure coordination, where players have an incentive for all to take the same action, here successful coordination involves different players taking different actions: some enter and some stay out. Some game theorists have suggested that the only reasonable outcome in one-shot play of such games is the symmetric equilibrium, which is mixed. In contrast, the insight from the literature on learning and evolution is that in repeated interaction, individuals will learn to condition their behavior on the behavior of others and hence converge to an asymmetric equilibrium. For example, when $N = 2$ and $2 > c \geq 1$, one would expect the outcome where exactly one agent enters with probability one, rather than have both randomize over entry. One player stays out because she can identify that her rival will definitely enter.

[3] These asymmetric equilibria have not received much attention in previous experimental studies.

On the other hand, in very large and anonymous populations, one might guess that a mixed strategy outcome might be more likely, given the inability to identify the strategies of individual opponents. More recently Binmore and Samuelson (2001) have suggested that in some circumstances the identification of different players might be subject to noise and this might allow the symmetric mixed equilibrium to persist. In the model of imitation presented here, the parameter ϕ to some extent takes the same role as noise in the model of Binmore and Samuelson. As we will see, if ϕ is sufficiently large, we have a symmetric outcome. If ϕ is small, there will be an asymmetric pure outcome.

2.2 Perturbed Equilibrium

There is ample evidence from experiments that Nash equilibrium, and in particular mixed equilibrium, can perform badly in predicting the play of subjects. McKelvey and Palfrey (1995) introduce the concept of a "quantal response equilibrium" or QRE, which, though an equilibrium concept, allows for systematic errors on the part of players. As, for example Hopkins (2001) notes, QRE are equivalent to the steady states of certain stochastic learning processes. One particular example of this is the exponential version of stochastic fictitious play (see Section 3.2). In the case of the market entry games considered here, any symmetric QRE of the most common logit form would imply an expected number of entrants closer to $N/2$ than in the symmetric Nash equilibrium.

It is possible to identify one other potential form of error. For example, assume agents use some form of naive learning by imitation.[4] That is, if the returns to the two strategies are public information, agents will tend to switch to the one which currently has the higher payoff. As they fail to take into account that their own switching of strategy will alter the returns to the activity, their behavior will be biased toward entry. In Section 3.3 below, we present a model in which imitators converge to a symmetric perturbed equilibrium where the expected number of entrants is closer to N than in the symmetric Nash equilibrium.

3 Varieties of Learning

Imagine this game was played repeatedly by the same group of players at discrete time intervals $n = 1, 2, 3, ...$ How can the learning dynamics be modeled? We suppose that all players have propensities for the two possible actions. We write player i's propensities as (q^i_{1n}, q^i_{2n}). Let the probability that agent i enters in period n be y^i_n and define $y_n = (y^1_n, ..., y^N_n)$. The probability of entry is determined by one of several possible mappings from propensities, for example, a reinforcement learning rule

$$y^i_n = \frac{q^i_{1n}}{q^i_{1n} + q^i_{2n}}, \tag{2}$$

or the exponential rule

$$y^i_n = \frac{\exp \beta q^i_{1n}/n}{\exp \beta q^i_{1n}/n + \exp \beta q^i_{2n}/n}. \tag{3}$$

The principal focus of interest, however, is what information agents use in modifying their actions.

[4] There is some experimental evidence that subjects do imitate others when given the opportunity. See, for example, Duffy and Feltovich (1999), Huck et al. (1999).

3.1 Simple Reinforcement

Simple reinforcement is what is assumed in standard reinforcement learning models. That is, changes in propensities are a function only of payoffs actually received. In this case, the change in propensity for player i in period n would be,

$$q^i_{1n+1} = q^i_{1n} + \delta^i_n(v + r(c - m_n)),\ q^i_{2n+1} = q^i_{2n} + (1 - \delta^i_n)v, \tag{4}$$

where m_n is the actual number of entrants in period n. This updating rule together with the choice rule (2) is what Erev and Roth (1998) call their basic reinforcement learning model. Note that given the choice rule (2), all propensities must remain strictly positive for the probability y^i to be defined. This can be assured, given the updating rule (4), if all payoffs are strictly positive. This last assumption is not usually problematic in an experimental setting, as experiments are usually designed, as were the ones reported here, so as not to give subjects negative payoffs.

3.2 Hypothetical Reinforcement

In hypothetical reinforcement, in addition to undergoing simple reinforcement, an agent hypothesizes what she would have received if she had played strategies other than the one she actually chose. The payoff she would have received is then added to the corresponding propensity. In this context, this implies

$$q^i_{1n+1} = q^i_{1n} + v + r(c - m_n - (1 - \delta^i_n)),\ q^i_{2n+1} = q^i_{2n} + v. \tag{5}$$

Of course, use of this rule generally requires more information than simple reinforcement. Without knowledge of the payoff structure and the actions taken by opponents, it is very difficult to know what one would have received if one had acted differently.

This updating rule together with the exponential choice rule (3) is an example of stochastic fictitious play (see for example, Fudenberg and Levine, 1998, Ch4). Fictitious play is often modeled in terms of an agent having beliefs over the actions of opponents rather in terms of propensities for his own actions. The two methods are equivalent when there are two players (see for example, Camerer and Ho, 1999; Erev and Roth, 1998). In situations with more than two players, such as this one, the models potentially diverge because of the possibility of correlation between the play of opponents (see Fudenberg and Levine, 1998, Ch2). However, we show in the Appendix (Proposition 5) that for this class of market entry game, the different formulations have the same expected motion and asymptotic outcome.

3.3 Learning by Imitation

Learning by imitation has meant various things in the literature. Nonetheless, it is possible to break down the different models into two components. First, for imitation to be possible, players must observe the actions and payoffs of some other players. That is, there must be some assumption on what information is observed. Second, given that the payoffs of some other player or players are observed, there has to be some form of behavior rule which determines whether a player will drop her current strategy in favor of the strategy of some other player.

The most common informational structure in the literature is that each period every agent can observe the action and payoff of one other player. In the context of the market entry games we consider, it may be more natural to think of players observing the return to the two possible strategies. Turning to behavior rules, the literature on learning by imitation has mostly concentrated on pure strategy rules with no memory. Two particular forms have been suggested: switch to the action of the other player if she had a higher payoff; switch to the action of the other player with a probability which is increasing in the difference in payoffs.

An alternative, more in line with the previous models, and allowing for agents to place some weight on previous experience, would be for agents to have propensities and reinforce them according to the following rule,

$$q^i_{1n+1} = q^i_{1n} + (\delta^i_n + \phi(1-\delta^i_n))(v + r(c - m_n)),\ q^i_{2n+1} = q^i_{2n} + (1 - \delta^i_n + \phi\delta^i_n)v, \tag{6}$$

where $0 \leq \phi \leq 1$ is a parameter which determines the relative weight placed on others' experience. This specification reverts to simple reinforcement when ϕ is zero. For $\phi = 1$, the agent places as much weight on others' experience as his own. In this case, updating of propensities is entirely independent of the strategy actually chosen. The latter case may not be unreasonable. After all, all agents face exactly the same decision problem. One way to interpret this model of learning is as a dynamic version of the model of Binmore and Samuelson (2001).

In the case of pure imitation, by which we mean the parameter ϕ is equal to one, it is possible to make the following observations. First, there will be a tendency to excess entry because imitators do not take into account the fact that changing from staying out to going in will depress payoffs to going in. Second, if the behavior rule is identical across agents, the outcome will be symmetric. Over a period of time, every agent will enter with the same frequency as every other agent. This is simply because all agents use the same rule to react to the same aggregate statistic: the difference in payoff between the two strategies. Thus, in summary if agents adopt such a rule, it will look as though play converges to close to the symmetric mixed Nash equilibrium, though, as noted, average entry will be higher.

3.4 Learning an Entry Threshold

In the work of Rapoport, Seale and Winter (1998), Rapoport, Seale and Parco (2000), a different type of learning model, specific to the market entry games is introduced. Rather than possessing propensities for actions, agents are assumed to each have an individual entry threshold, c^i_n. They then employ the deterministic decision rule:

$$\text{enter in period } n \text{ if and and only if } c \geq c^i_n. \tag{7}$$

This rule has been used to explain the behavior of subjects in market entry experiments where a different value of c was used in each period. However, it can also be used in the present context where c is kept constant.

The updating of the entry threshold levels is as follows. Players who stayed out in period n compare their payoff to what they would have received if they had entered and c^i_n is adjusted according to

$$c^i_{n+1} = \min[c^i_n, c^i_n - wr(c - m_n - 1)], \tag{8}$$

where $w > 0$ is a parameter measuring speed of adjustment. This rule implies that the threshold level is lowered if the payoff to entry exceeds that from staying out.[5] This, in effect, assumes that players use some form of hypothetical reinforcement. For players who entered, the updating is

$$c^i_{n+1} = \max[c^i_n, c^i_n + wr(c - m_n)]. \tag{9}$$

That is, the threshold is raised if the payoff to staying out is higher than to entry. Rapoport, Seale and Winter (1998), Rapoport, Seale and Parco (2000) also consider the possibility that w takes different values for upward and downward adjustments of c^i, that w decreases over time and also the possibility that players may make stochastic deviations from the above rule. However, we go on to show that this basic model gives the same prediction as that of fictitious play and reinforcement learning: asymptotically there will be a sorting outcome.

4 Learning Dynamics

We now investigate the dynamics of the various types of learning introduced in the previous section. Each of these dynamics has differing requirements concerning information. Reinforcement learning requires only information on one own's payoff. Learning by imitation requires in addition evidence about the actions and payoffs of others. Hypothetical reinforcement learning models such as stochastic fictitious play or threshold learning, require information about both the actions of others and the structure of payoffs. Thus there is a rough ordering of the three processes in terms of information requirements, which are reflected in the three information treatments in our experiments. However, as we now show, the asymptotic behavior of these three types of learning dynamics are not ordered in terms of their informational inputs. In fact, the most and least sophisticated, reinforcement learning and fictitious play, are predicted to lead to sorting, while learning by imitation, the intermediate case in terms of information, predicts a symmetric outcome even in the long run.

To obtain some analytic results on the learning processes we consider, we make use of results from the theory of stochastic approximation. Simply put, (see the Appendix for details), this allows investigation of the behavior of a stochastic learning model by evaluating its expected motion. In the case of the classic reinforcement learning process defined by the updating rule (4) and the choice rule (2), the expected motion of the ith player's strategy adjustment can be written as

$$E[y^i_{n+1}|y_n] - y^i_n = \frac{1}{Q^i_n} y^i_n (1 - y^i_n) r(c - 1 - \sum_{j \neq i} y^j_n) + O(\frac{1}{n^2}), \tag{10}$$

where $Q^i_n = \sum_j q^i_{nj} > 0$ is a player specific scaling factor. Note that the right-hand side of the system of equations (10) is very close to the evolutionary replicator dynamics, which for this game would be the following system of differential equations:

$$\dot{y}^i = y^i (1 - y^i) r(c - 1 - \sum_{j \neq i} y^j) \tag{11}$$

[5] There is some ambiguity in the original papers, Rapoport, Seale and Winter (1998), Rapoport, Seale and Parco (2000). There it is written that players consider the payoff they would have received if they had entered. But this would be $v + r(c - m_n)$ in present notation. That is, the agents, just as in the imitation model considered in the previous subsection, do not take into account that the payoff to entry would have been lower if they had in fact entered. If the updating process (8) was reformulated in this fashion, the threshold learning model would yield results similar to the imitation learning model.

for $i = 1, ..., N$.

This class of entry game is a rescaled partnership game (Hofbauer and Sigmund, 1998, pp. 127-130), also known as a game of identical interest (Monderer and Shapley, 1996). As these names suggest, what these games have in common is that it is possible to increase the payoffs of all players simultaneously. For example, in market entry games movement toward Nash frequencies of entry makes no player worse off and some better off. This in turn leads to particular results in terms of learning. It is shown in the Appendix that in these market entry games the learning dynamics must converge to an equilibrium for the replicator dynamics (11).

But we can go further. Note that if under the replicator dynamics all players start with the same mixed strategy, that is, if $0 < y^1 = y^2 = ... = y^N < 1$, then (11) becomes

$$\dot{y}^i = y^i(1 - y^i)r(c - 1 - y^i(N - 1)).$$

It is relatively easy to see that each y^i would converge to the value given by the symmetric mixed strategy equilibrium. However, it can be shown that this is the only way for such a deterministic system to converge to this equilibrium as it is a saddle, the stable manifold being one dimensional and defined by the equation $y^1 = y^2 = ... = y^N$. This deterministic result allows us to show that the actual stochastic learning process will have a zero probability of remaining in this manifold and hence a zero probability of converging to a mixed equilibrium. And if the learning process converges to an equilibrium but not to a mixed equilibrium, it must converge to a pure equilibrium. This is the intuition behind the following proposition, the proof of which is in the Appendix.

Proposition 1 *If agents use the reinforcement learning updating rule (4) and choice rule (2), for generic values of c, with probability one the learning process converges to a pure Nash equilibrium of the game. That is,* $\Pr\{\lim_{n\to\infty} y_n \in \bar{Y}\} = 1$*, where* $\bar{Y}$ *is the set of pure Nash equilibrium profiles.*

The reference to generic values of c refers to a difficulty mentioned earlier if c is an integer. In this case, there are an infinite number of Nash equilibria where $c - 1$ agents enter with probability one, and $N - c$ agents stay out and with the remaining agent completely indifferent. Our intuition here about what a reasonable outcome constitutes and the analytic results available are both considerably weaker.

We now turn to hypothetical reinforcement and fictitious play. From Hopkins (2001), under the hypothetical updating rule (5) and the exponential choice rule (3), the expected motion of strategies can be written as

$$E[y^i_{n+1}|y_n] - y^i_n = \frac{\beta}{n+1}\left(y^i_n(1 - y^i_n)r(c - 1 - \sum_{j\neq i} y^j_n) + \frac{1}{\beta}\sigma(y^i_n)\right) + O(\frac{1}{n^2}), \qquad (12)$$

where $\sigma(y^i_n)$ is a noise term equal to

$$\sigma(y^i) = y^i_n(1 - y^i_n)(\log(1 - y^i_n) - \log y^i_n).$$

That is, the expected motion is close but not identical to the replicator dynamics. First, there is the additional noise term σ which ensures that each action will always be taken with a positive probability. Second, the expected motion is multiplied by the factor β. This has the effect that learning under stochastic fictitious play is much faster than under reinforcement learning.

The equilibrium points of such dynamics are not in general identical to Nash.[6] Define $\hat{y}$ as a perturbed equilibrium which satisfies for $i = 1, ..., N$,

$$r(c - 1 - \sum_{j \neq i} y_n^j) + \frac{1}{\beta}(\log(1 - y^i) - \log y^i) = 0. \tag{13}$$

Note that for β sufficiently large and in generic games there will be a perturbed equilibrium for each Nash equilibrium. Second, as $\beta \to \infty$, the set of perturbed equilibria $\hat{y}$ approaches the set of Nash equilibria. Furthermore, by similar methods to those used in Proposition 1, we are able to establish the following result.[7]

Proposition 2 *If all players use hypothetical reinforcement together with the exponential choice rule (3), for generic values of* c *and for* β *sufficiently large, then* $\Pr\{\lim_{n\to\infty} y_n \in \hat{Y}\} = 1$, *where* $\hat{Y}$ *is the set of perturbed equilibrium each corresponding to one of the pure Nash equilibria of the game.*

Under learning by imitation, the outcome can be quite different. First, all agents using this learning rule update their propensities in exactly the same way, which leads inevitably to a symmetric outcome. Second, they do not take into account the impact on payoffs of their change in strategy, which means that this symmetric outcome is not the same as the symmetric mixed Nash equilibrium. Define $\hat{y}$ here as a perturbed equilibrium which satisfies

$$r(c - \sum_{j=1}^{N} y^j) + \frac{1}{\beta}(\log(1 - y^i) - \log y^i) = 0. \tag{14}$$

Note that if β is large, then the expected number of entrants in this perturbed equilibrium approaches c, whereas in the symmetric mixed Nash equilibrium it is less, in fact, $N(c-1)/(N-1)$. So this form of perturbed equilibrium involves excess entry.[8]

Proposition 3 *If all players use a combination of imitation updating (6), with* $\phi = 1$, *and the exponential choice rule (3), then* $\Pr\{\lim_{n\to\infty} y_n^i = \hat{y}^i\} = 1$ *for all* i, *where* $\hat{y}$ *is a symmetric solution to (14).*

The intuition for this result is that, as everyone receives the same reinforcements, entirely independent of individual experience, the only possible limiting outcome is symmetric. Behavior may differ in the short–run until the influence of differing initial propensities is eliminated. The result in Proposition 3 however, is in conflict with the evidence from previous experiments, in which considerable heterogeneity in behavior is observed.

Proposition 4 *If all players use the threshold rule (7) and the updating rules (8) and (9), then for* w *sufficiently small and for generic initial conditions, choices converge to a pure Nash equilibrium.*

That is, asymptotically at least, there will be sorting. However in previous experiments, there has been little evidence for the perfect individual sorting predicted by fictitious play and reinforcement learning as well as this threshold model. Our new experiments give somewhat different results as we will see.

[6] See for example, Fudenberg and Levine (1998, Ch4).

[7] A similar result for two player games is proved in Hofbauer and Hopkins (2000). Monderer and Shapley (1996) prove the convergence of fictitious play in this class of game.

[8] This corresponds to the discovery by Vega Redondo (1997) that learning by imitation in Cournot oligopoly leads to a competitive outcome, that is, output in excess of Nash behavior.

5 Experimental Design

The experimental design involves repeated play of the market entry game by a group of $N = 6$ inexperienced subjects under one of three different information conditions. We begin by discussing the parameterization of the payoff function and the three information conditions. We then explain the procedures followed.

We chose to set $v = 8$, $r = 2$ and $c = 2.1$ resulting in the following payoff function (in dollars)

$$\pi_i(\delta) = \begin{cases} \$8, & \text{if } \delta^i = X, \\ \$8 + 2(2.1 - m), & \text{if } \delta^i = Y. \end{cases}$$

where $0 \leq m \leq 6$ is the number of subjects (including i) choosing Y. We chose c to be non-integer so that, as noted, the number of Nash equilibria of the game would be finite and the pure equilibria would be strict. The number of players, 6, is significantly smaller than in the previous experiments on market entry games. Our choice of $N = 6$ was based on the following considerations. We wanted a parameterization for the payoff function, in particular, a choice for the parameter r, that was similar to previous studies, and we wanted to provide subjects with reasonable compensation for their active participation. At the same time, we wanted to avoid any possibility that subjects earned *negative payoffs* that might result in ill-defined entry probabilities under the various learning models we examine.[9] These considerations favored our choice of a smaller number of subjects.

In the first "limited information" treatment, subjects were repeatedly asked to choose between two actions X or Y, (which corresponded to "stay out" or "enter" respectively) without knowing the payoff function, π_i. Indeed, subjects did not even know that they were playing a game with other subjects. In this limited information treatment, each subject was informed only of the payoff from his own choice of action. Each subject's history of action choices and payoffs was reported on their computer screens, and subjects also recorded this information on record sheets. Thus, in the limited information treatment, subjects had all the information necessary to play according to the reinforcement learning dynamic, but did not possess the information necessary for imitation of other subjects' actions or for playing according to fictitious play.

In the second "aggregate information" treatment, subjects received feedback on the payoff from their action choice as in the limited information treatment, but were fully informed of the payoff function. In particular, subjects were told the payoff function, and to insure that their payoffs from choosing Y were as transparent as possible, the instructions also included the following table revealing all possible payoff values from choosing action Y. This table was also drawn on a chalkboard for all to see.

Fraction of players who choose action Y:	1/6	2/6	3/6	4/6	5/6	6/6
Payoff each earns from choosing action Y:	\$10.20	\$8.20	\$6.20	\$4.20	\$2.20	\$0.20

The instructions also clearly stated that the payoff each subject earned from choosing action X was always \$8, and this fact was also written on the chalkboard. Following the play of each round in the aggregate information treatment, subjects were further informed of the fraction of the six

[9] Erev and Rapoport (1998) use a parameterization of the payoff function that can result in significant negative payoffs given the larger number of subjects they consider (12). However, they adjust the propensity updating process of their learning model in the event that propensities become negative. It is less clear that the human subjects would make a similar adjustment.

players who had chosen X and the fraction of the six players who had chosen Y, as well as the payoff received by all those choosing X and all those choosing Y. The past history (last 10 rounds) of the fractions choosing X and Y, along with the payoffs from each choice was always present on subjects' computer screens, and subjects were asked to record this information on record sheets as well. Hence, in the aggregate information treatment, subjects had all the information necessary to imitate the actions of others and/or play according to fictitious play.

In a final "full information" treatment, subjects were given all the information provided to subjects in the aggregate information treatment, and in addition, subjects were informed of the individual actions chosen by each of the other 5 players in the session, who were identified by their player ID numbers; this last piece of information was not available in the aggregate (or in the limited) information treatments. For example, as noted in the full information treatment instructions, the subject with ID number 3 might see that in the just completed round, the other 5 subjects' choices were:

$$1X \quad 2Y \quad 4X \quad 5X \quad 6Y$$

indicating that subject number 1 chose X, subject number 2 chose Y, subject numbers 4 and 5 both chose X and subject number 6 chose Y. The immediate past history (the last 10 rounds) of this individual action information was always present on subjects' screens, thus enabling them to assess the extent to which the other 5 subjects were consistent or inconsistent in their choice of action. Since subjects in the full information all knew the payoffs earned each round by those choosing X and those choosing Y, they were provided a complete record of the actions chosen and payoffs earned by each individual subject in every round of the session.

We conducted nine one 1–hour sessions: three sessions for each of the three different information treatments. Each session involved exactly 6 subjects who had no prior experience playing the market entry game under any treatment (54 subjects total). Subjects were recruited from the undergraduate population at the University of Pittsburgh. In each session, the group of 6 subjects were seated at computer workstations, and were given written instructions which were also read aloud. Subjects were isolated from one another and no communication among subjects was allowed.

Subjects played the market entry game by entering their choice of action in each round, X or Y, using the computer keyboard when prompted by their monitor. Once all subjects had made their action choices, the computer program determined each subject's own payoff according to the parameterization of π_i given above, and reported this payoff back to each subject. Whether additional information was provided depended on the treatment as discussed above.

The six subjects played 100 rounds of the market entry game in an experimental session lasting one hour. Because the predictions that follow from propositions 1–4 are all asymptotic, we wanted a sufficient number of repetitions to allow the predicted behavior to develop. Simulations of the various learning models (available on request) indicated that the 100 rounds allowed should be adequate at least for a pronounced movement toward equilibrium, if not actual convergence. Second, these simulations also indicated that, as learning slows over time, increasing the number of repetitions to 150, for example, would not produce radically different behavior.

The 100 rounds were broken up into four, 25–round sets. Subjects were informed that at the end of each 25–round set, an integer from 1 to 25 would be randomly drawn from a uniform distribution with replacement. The chosen integer corresponded to one of the round numbers in the just completed 25–round set. Each subject's dollar payoff in that round was added to their total cash earnings for the session. This design was chosen to prevent subjects from becoming bored during the 100 repetitions of the market entry game. In addition to the 4 cash payments, subjects received \$5 for showing up on time and participating in the experiment. The average total amount earned by each subject was \$37.87 in the limited information treatment, \$36.53 in the aggregate information treatment, and \$35.33 in the full information treatment.

6 Equilibrium Predictions and Hypotheses

Given our parameterization of the market entry game, pure strategy Nash equilibria have 2 players always entering, each earning \$8.20, and 4 players always staying out, each earning \$8.00. The unique symmetric mixed strategy Nash equilibrium prediction is that each player enters with probability .22 and earns an expected payoff of \$8.00. In this equilibrium, the expected number of entrants is 1.32. Finally, as noted in Section 3, there are many asymmetric mixed equilibria. However, play in some of the sessions seems to approach one of these in particular. In this asymmetric mixed equilibrium, 2 players always stay out and the remaining 4 players enter with probability .367, earning an expected payoff of \$8.00 each. The expected number of entrants in this asymmetric mixed equilibrium is 1.467. As noted, if subjects were to use a perturbed choice rule such as the exponential rule (3), the steady states of the learning process would not be Nash equilibria, but perturbed equilibria (also known as QRE). We report also the QRE equilibria (for a typical value of the parameter β) that correspond to the three Nash equilibria of interest. Lastly, if subjects all adopted learning by imitation, we would expect to see a symmetric perturbed equilibrium characterized by equation (14). These equilibrium predictions are summarized in Table 1.

Table 1: Equilibrium Predictions

	Number of Entrants	
Equilibrium	Mean	Standard Deviation
Pure	2	0
Symmetric Mixed	1.32	1.015
Asymmetric Mixed	1.467	0.964
Pure QRE ($\beta = 5$)	1.781	0.512
Symmetric QRE ($\beta = 5$)	1.457	1.050
Asymmetric QRE ($\beta = 5$)	1.525	0.968
Symmetric Imitation ($\beta = 5$)	2.158	1.175

Together with the theoretical results of the previous section, we can make the following hypotheses.

Hypothesis 1 *If subjects are reinforcement learners, then: (a) play should evolve over time toward a pure strategy Nash equilibrium and (b) there should be no change in the speed with which play evolves toward this equilibrium in the limited information treatment as compared with the aggregate information or full information treatments.*

Hypothesis 2 *If subjects learn through imitation (in particular, they update using (6) and choose to enter according to the probability given in (3)), then (a) play should evolve over time toward a symmetric mixed strategy (perturbed) equilibrium in the aggregate and full information treatments in which there is excess entry; (b) there should be no change in the speed with which play evolves toward this mixed strategy equilibrium in the aggregate information treatment as compared with the full information treatment.*

Hypothesis 3 *If subjects are hypothetical reinforcement learners, playing according to stochastic fictitious play or a threshold learning rule then (a) play should evolve over time toward a (perturbed) pure strategy Nash equilibrium in the aggregate and full information treatments ; (b) there should be no change in the speed with which play evolves toward a pure strategy equilibrium in the aggregate information treatment as compared with the full information treatment.*

Note that the fictitious play, learning by imitation and threshold learning models require information that was not made available to subjects in our limited information treatment. It is therefore unclear what these models predict in such circumstances. There has been more than one attempt (for example, Fudenberg and Levine, 1998, Ch4; Sarin and Vahid, 1999) to specify a fictitious play-like learning process for environments where opponents' play is not observable. However, the properties of these learning processes, and in particular, their speed with respect to fictitious play, are not well known. Therefore, we treat the fictitious play, learning by imitation and threshold learning models as making no predictions in the limited information treatment.

A related issue is the exact definition of fictitious play in games with more than two players. One might wish to claim that if fictitious play is based on beliefs about the actions of opponents, updating those beliefs is only really possible when the actions of all opponents are observable. However, the rationale behind fictitious play is that agents monitor past behavior in order to predict returns to different strategies. And in the market entry games we consider, to predict her payoff an agent would only have to form a belief about the expected total number of entrants. Finally as Proposition 5 (in the Appendix) shows, in this class of market entry games, the two approaches generate the same results.

7 Experimental Findings

7.1 Main results

The main findings are summarized in Tables 2 and 3 and Figure 1. Table 2 reports session-level means and standard deviations of the number of entrants. Figure 1 reports the round-by-round mean number of entrants across the three sessions of each treatment, along with a one standard deviation bound. Finally, Table 3 reports individual subject entry frequencies and standard deviations.

Table 2 and Figure 1 reveal that in all three treatments, the mean number of entrants generally lies between c and $c-1$, or between 2.1 and 1.1, though there are some exceptions. In particular, over the last 50 rounds of two of the three aggregate information sessions and one of the three full information sessions, the average number of entrants exceeded 2.1 by small amounts. Perhaps the most interesting finding in Table 2 and Figure 1 is that in each of the 3 full information treatments, there appears to be perfect coordination on a pure Nash equilibrium for at least one 10-round period, i.e. the standard deviation for that 10–round entry frequency was zero (see, in particular Figure 1).

Of course, to assess whether a pure Nash equilibrium was actually achieved requires further disaggregation of the data, which is done in Table 3. This table reports the mean and standard deviation of the entry frequencies for each of the six subjects in each of the nine sessions. Looking at the full information treatment results, we see that in two of the three sessions (full information sessions numbers 1 and 2), subjects did indeed achieve perfect coordination on the pure strategy

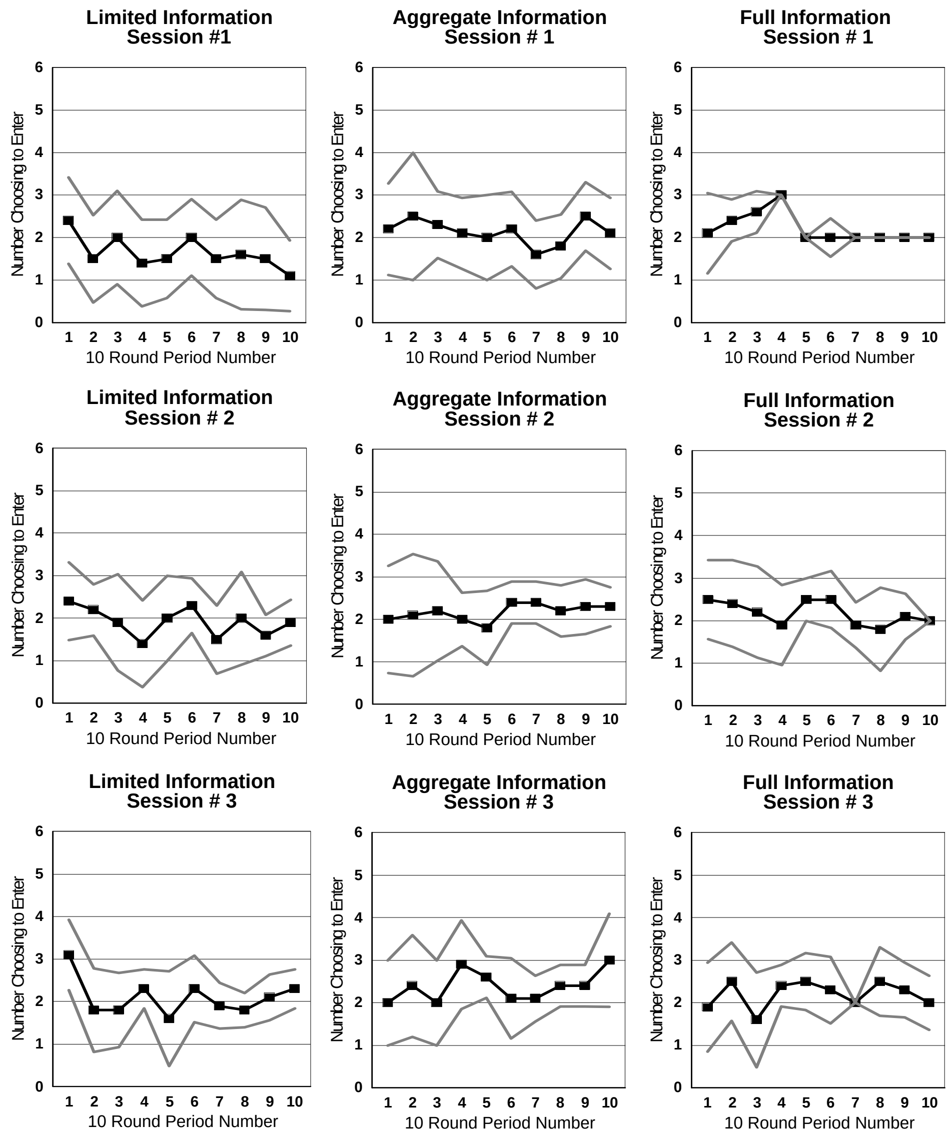

Figure 1: 10-Round Mean Number of Entrants, One-Standard Deviation Bound

Table 2
Mean Number of Entrants and Standard Deviations Over All 100 Rounds, the Last 50 Rounds and Last 10 Rounds of Each Limited Information Session

	Session #1			Session #2			Session #3		
	All 100	Last 50	Last 10	All 100	Last 50	Last 10	All 100	Last 50	Last 10
Mean	1.65	1.54	1.10	1.92	1.86	1.90	2.10	2.08	2.30
St. Dev.	1.09	1.08	0.83	0.91	0.80	0.54	0.84	0.60	0.46

Mean Number of Entrants and Standard Deviations Over All 100 Rounds, the Last 50 Rounds and Last 10 Rounds of Each Aggregate Information Session

	Session #1			Session #2			Session #3		
	All 100	Last 50	Last 10	All 100	Last 50	Last 10	All 100	Last 50	Last 10
Mean	2.13	2.04	2.10	2.17	2.32	2.30	2.39	2.40	3.00
St. Dev.	0.99	0.87	0.83	0.90	0.55	0.46	0.94	0.83	1.10

Mean Number of Entrants and Standard Deviations Over All 100 Rounds, the Last 50 Rounds and Last 10 Rounds of Each Full Information Session

	Session #1			Session #2			Session #3		
	All 100	Last 50	Last 10	All 100	Last 50	Last 10	All 100	Last 50	Last 10
Mean	2.21	2.00	2.00	2.18	2.06	2.00	2.20	2.22	2.00
St. Dev.	0.52	0.20	0.00	0.83	0.68	0.00	0.82	0.67	0.63

Nash equilibrium where 2 players always enter and 4 always stay out over the last 10 rounds of these sessions, as the standard deviation of the entry frequencies are zero for each subject.

We note further that in full information session 1, subjects actually achieved a pure strategy Nash equilibrium much earlier, from rounds 41-51 and another pure strategy equilibrium beginning in round 54; they remained in the latter pure strategy equilibrium for the last 46 rounds of the experiment (see Figure 1). In full information session 2, subjects achieved a pure strategy equilibrium in round 85 and remained in that equilibrium for the last 15 rounds of the experiment. In full information session 3, a pure strategy equilibrium was achieved only briefly from rounds 63-69 (7 rounds).[10] However, we note that by the last 10 rounds of full information session 3 four of the six players were adhering to pure strategies; one always in and three always out.

Table 3 reveals that there is some support for Hypothesis 1a: as the reinforcement learning model predicts, subjects in the limited information sessions are quite close to coordinating on a pure equilibrium by the end of each of the three limited information sessions. Note in particular, that by the final 10 rounds of each session, three or four players choose not to enter at least 90% of the time, and one or two players choose to enter more than 50% of the time. Moreover we see that the standard deviations for the individual entry frequencies are almost always lower in the last 10 rounds as compared with the last 50 rounds. On the other hand, there does not appear to be much support for Hypothesis 1b as there are some differences in both the speed of convergence as subjects are given more information in the aggregate information and full information treatments. In particular, convergence toward the pure strategy equilibrium appears to be much faster in the full information treatment as compared with the limited or aggregate information treatments.

[10]Figure 1 may give the mistaken impression that a pure strategy Nash equilibrium was obtained over rounds 61-70 of full information session 3. In fact, there were just two entrants in each round of this 10-round interval, but in rounds 63 and 70, one subject who had been an entrant in previous rounds chose not to enter while another subject who had been staying out simultaneously chose to enter. Hence, the standard deviation in the number of entrants was indeed 0 over rounds 61-70, as reported in Figure 1, but a pure equilibrium was only obtained over the shorter interval consisting of rounds 63–69.

Table 3
Individual Entry Frequencies: Means and (Standard Deviations) Over All 100 Rounds, the Last 50 Rounds and Last 10 Rounds of Each Limited Information Session

Player	Session #1			Session #2			Session #3		
Number	All 100	Last 50	Last 10	All 100	Last 50	Last 10	All 100	Last 50	Last 10
1	0.07	0.08	0.00	0.76	0.84	1.00	0.12	0.02	0.00
	(0.26)	(0.27)	(0.00)	(0.43)	(0.37)	(0.00)	(0.32)	(0.14)	(0.00)
2	0.26	0.36	0.10	0.58	0.64	0.80	0.04	0.00	0.00
	(0.44)	(0.48)	(0.30)	(0.49)	(0.48)	(0.40)	(0.20)	(0.00)	(0.00)
3	0.34	0.32	0.30	0.04	0.02	0.00	0.67	0.76	0.80
	(0.47)	(0.47)	(0.46)	(0.20)	(0.14)	(0.00)	(0.47)	(0.43)	(0.40)
4	0.56	0.58	0.60	0.00	0.00	0.00	0.24	0.30	0.50
	(0.50)	(0.49)	(0.49)	(0.00)	(0.00)	(0.00)	(0.43)	(0.46)	(0.50)
5	0.03	0.00	0.00	0.25	0.02	0.00	0.13	0.04	0.00
	(0.17)	(0.00)	(0.00)	(0.43)	(0.14)	(0.00)	(0.34)	(0.20)	(0.00)
6	0.39	0.20	0.10	0.29	0.34	0.10	0.90	0.96	1.00
	(0.49)	(0.40)	(0.30)	(0.45)	(0.47)	(0.30)	(0.30)	(0.20)	(0.00)

Individual Entry Frequencies: Means and (Standard Deviations) Over All 100 Rounds, the Last 50 Rounds and Last 10 Rounds of Each Aggregate Information Session

Player	Session #1			Session #2			Session #3		
Number	All 100	Last 50	Last 10	All 100	Last 50	Last 10	All 100	Last 50	Last 10
1	0.01	0.00	0.00	0.22	0.14	0.10	0.76	0.76	0.60
	(0.10)	(0.00)	(0.00)	(0.42)	(0.35)	(0.30)	(0.43)	(0.43)	(0.49)
2	0.53	0.46	0.50	0.49	0.74	0.20	0.10	0.00	0.00
	(0.50)	(0.50)	(0.50)	(0.50)	(0.44)	(0.40)	(0.30)	(0.00)	(0.00)
3	0.60	0.54	0.70	0.15	0.00	0.00	0.04	0.04	0.00
	(0.49)	(0.50)	(0.46)	(0.36)	(0.00)	(0.00)	(0.20)	(0.20)	(0.00)
4	0.61	0.60	0.20	0.00	0.00	0.00	0.29	0.30	0.50
	(0.49)	(0.49)	(0.40)	(0.00)	(0.00)	(0.00)	(0.46)	(0.46)	(0.50)
5	0.00	0.00	0.00	0.36	0.44	1.00	0.70	0.76	0.90
	(0.00)	(0.00)	(0.00)	(0.48)	(0.50)	(0.00)	(0.46)	(0.43)	(0.30)
6	0.38	0.44	0.70	0.95	1.00	1.00	0.50	0.54	1.00
	(0.49)	(0.50)	(0.46)	(0.22)	(0.00)	(0.00)	(0.50)	(0.50)	(0.00)

Individual Entry Frequencies: Means and (Standard Deviations) Over All 100 Rounds, the Last 50 Rounds and Last 10 Rounds of Each Full Information Session

Player	Session #1			Session #2			Session #3		
Number	All 100	Last 50	Last 10	All 100	Last 50	Last 10	All 100	Last 50	Last 10
1	0.35	0.02	0.00	0.02	0.00	0.00	0.73	0.96	1.00
	(0.48)	(0.14)	(0.00)	(0.14)	(0.00)	(0.00)	(0.44)	(0.20)	(0.00)
2	0.05	0.00	0.00	0.46	0.74	1.00	0.00	0.00	0.00
	(0.22)	(0.00)	(0.00)	(0.50)	(0.44)	(0.00)	(0.00)	(0.00)	(0.00)
3	0.81	1.00	1.00	0.25	0.08	0.00	0.00	0.00	0.00
	(0.39)	(0.00)	(0.00)	(0.43)	(0.27)	(0.00)	(0.00)	(0.00)	(0.00)
4	0.01	0.00	0.00	0.51	0.16	0.00	0.70	0.64	0.70
	(0.10)	(0.00)	(0.00)	(0.50)	(0.37)	(0.00)	(0.46)	(0.48)	(0.46)
5	0.01	0.00	0.00	0.67	0.96	1.00	0.51	0.52	0.30
	(0.10)	(0.00)	(0.00)	(0.47)	(0.20)	(0.00)	(0.50)	(0.50)	(0.46)
6	0.98	0.98	1.00	0.27	0.12	0.00	0.26	0.10	0.00
	(0.14)	(0.14)	(0.00)	(0.44)	(0.32)	(0.00)	(0.44)	(0.30)	(0.00)

In the last 20 rounds of the three aggregate information sessions, subjects appear to be somewhere between the asymmetric mixed equilibrium and the pure equilibrium. That neither equilibrium has been reached is supported by the fact that there is excessive entry relative to that predicted in either equilibrium (compare the mean number of entrants in Table 2 with the predictions in Table 1). Notice also in Table 3 that in the last 50 (and last 10 rounds) of each of the three aggregate information treatments there remain four players who are still choosing to enter with some positive frequency, and exactly two players who (almost) purely stay out.

The difference in findings between the limited and full information treatments appear to lie in the speed of convergence and not the type of equilibrium selected. In particular, it appears that additional information may affect the speed of convergence to a pure strategy equilibrium in violation of the notion that subjects are strictly reinforcement learners. On the other hand, there is not much difference in subject behavior between the final 10 rounds of limited information session number 3 and full information session number 3.

Note also that learning by imitation (Hypothesis 2) does not fare well as an explanation of the findings from the aggregate and full information treatments. The higher average entry in the aggregate information treatment might indicate that some subjects failed to take into account their own impact on the return to entry just as in naive imitation. However, outcomes in all sessions of the aggregate information treatment were far too asymmetric to support the hypothesis that imitation learning was widespread.

There is much more support for Hypothesis 3a than for 3b. Whereas play does seem to approach a pure strategy equilibrium in the aggregate and full information treatments, it also appears that the additional information provided in the full information treatment relative to the aggregate information treatment has a substantial effect on subject behavior; subjects in the full information treatment are much closer to the pure strategy equilibrium by the end of the session than are subjects in the aggregate information treatments; indeed, as noted earlier, in two of the three full information sessions subjects had achieved and sustained perfect coordination on a pure equilibrium by the end of the session.

7.2 Convergence to equilibrium

To determine how close subjects were to convergence on a particular equilibrium, we first calculated each subject's entry frequency over 10-period, non-overlapping samples, $s = 1,2,...,10$. Denote the entry frequency of subject i over sample s by y^i_s. We then calculated the mean squared deviation (msd) from a predicted equilibrium entry frequency $\hat{y}^i$, over each 10-period sample in session j, $\text{msd}^j_s = 1/6\sum_{i=1}^{6}(y^i_s - \hat{y}^i)^2$. This calculation is straight–forward for the unique symmetric mixed equilibrium, where $\hat{y}^i = .22$ for all i. Since there are many pure and asymmetric mixed equilibria, we chose to select *one* equilibrium of each type for each session. Each pure equilibrium was selected by determining the two players who were closest to playing the pure strategy of always entering over the last 10 rounds of each session. The other four players were regarded as being closest to playing the pure strategy of always staying out. In all sessions, the assignment of pure strategies to players based on final 10-round entry frequencies was readily apparent. (See the data in Table 3 for the last 10 rounds). Depending on this categorization, the predicted entry frequency, $\hat{y}^i$, would be either 1 or 0, and using these predictions, we calculated the msd from "the" pure strategy for each s in each session. Similarly for the asymmetric mixed equilibrium, we used the final 10-round entry frequencies to determine the two players in each session who were closest to playing the pure strategy of always staying out, $\hat{y}^i = 0$. The other four players were regarded as being closest to

Table 4: Estimates from OLS Regression: msd $= \gamma_0 + \gamma_1 s + \epsilon$
(Standard Errors in Parentheses)

Treatment	msd from:	constant	s	R^2
Limited Info.	Pure	0.187***	–0.014***	.30
		(0.025)	(0.004)	
	Asym. Mixed	0.062***	0.001	.02
		(0.017)	(0.003)	
	Sym. Mixed	0.074***	0.004	.05
		(0.019)	(0.003)	
Aggregate Info.	Pure	0.244***	–0.009	.08
		(0.034)	(0.005)	
	Asym. Mixed	0.053***	0.005*	.12
		(0.016)	(0.002)	
	Sym. Mixed	0.085***	0.010***	.27
		(0.018)	(0.003)	
Full Info.	Pure	0.292***	–0.028***	.31
		(0.048)	(0.008)	
	Asym. Mixed	0.098***	0.006	.08
		(0.023)	(0.004)	
	Sym. Mixed	0.119***	0.011***	.24
		(0.022)	(0.004)	

* significantly different from zero at the 10-percent level, ** at the 5-percent level, *** at the 1-percent level.

playing the mixed strategy which has a predicted entry probability of $\hat{y}^i = .367$.[11] Again, these assignments were readily apparent.

Figure 2 shows the sequence of 10-period, mean squared deviations averaged over the three sessions of each information treatment, $(1/3 \sum_{j=1}^{3} \text{msd}_s^j)$. In all three information treatments, the (average) msd from "the" pure equilibrium is initially much higher than the msd from the other two types of equilibria, but by the final 10 rounds, the msd from the pure equilibrium is less than the msd from these other equilibrium types. In the case of the full information treatment, the msd from the pure equilibrium falls below the msd from the other equilibrium types between periods 50-60, and remains there for the duration of the full information sessions. Notice also that in the aggregate and full information treatments, the msd from the asymmetric and the symmetric mixed equilibria appears to be rising over time. Indeed, ordinary least squares regressions of the session level data msd_s^j on a constant and "time", s (30 observations per regression) confirm that there is a slight negative time trend in the msd from the pure strategy equilibrium and a slight positive time trend in the msd from the asymmetric and symmetric mixed equilibria in all three treatments, though not all of these slope coefficient estimates are significantly different from zero. See Table 4 below.

For comparison purposes, we have also conducted a simulation exercise in which six artificial players play according to the stochastic reinforcement or the stochastic fictitious play dynamic for

[11] We recognize that msd can be an imperfect measure of convergence to a mixed strategy equilibrium as it cannot detect sequential dependencies in player's entry choices. However, since we do not find that players converge to a mixed strategy or asymmetric mixed equilibrium using our msd convergence criterion, it seems unlikely that alternative convergence criteria that were capable of detecting sequential dependencies in entry choices would alter our findings.

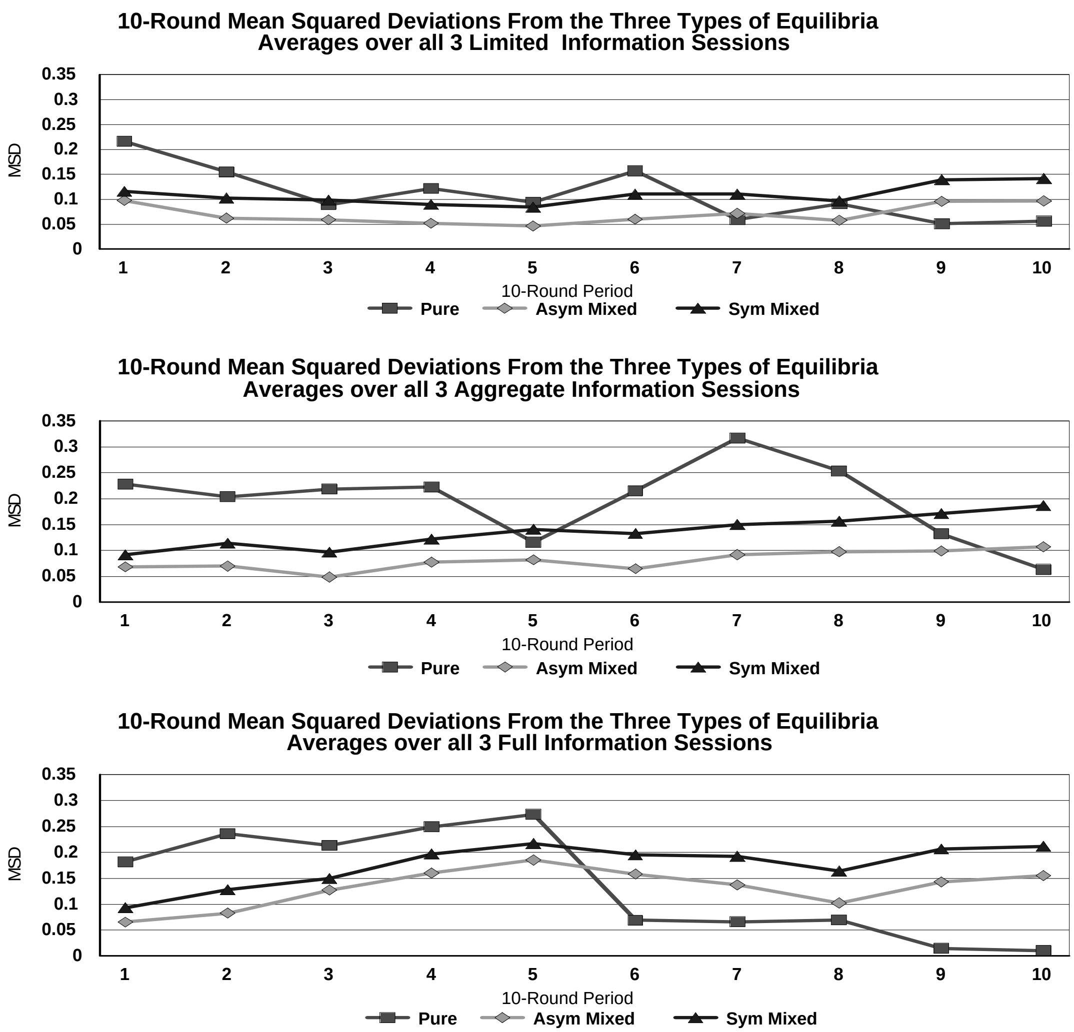

Figure 2: 10-Round Mean Squared Deviations from the Three Types of Equilibria: Averages over all 3 sessions of a treatment

100 rounds. We conducted 100 runs of these two simulation exercises and calculated the mean squared deviations from each of the three equilibrium types in the same manner that was used in the construction of Figure 2. The results are reported in Figure 3. We see that reinforcement learning shows very slow movement toward pure strategy equilibrium. In fact, within the 100 rounds considered, one can see that there is some motion toward the symmetric mixed equilibrium, even though from our asymptotic results we know that eventually reinforcement learning must converge to a pure equilibrium. This initial move toward the mixed equilibrium reflects the fact that this equilibrium is a saddlepoint under the replicator dynamics. In contrast, learning under stochastic fictitious play is much faster even with a parameter value of $\beta = 5$. Movement is definitely away from the mixed equilibrium, and by the end of 100 periods, play is close to a pure equilibrium. Though as noted in Section 4, for finite β, exact convergence to pure equilibrium is not possible.[12] Figure 3 also reveals that the rate at which the learning dynamics are converging toward equilibrium is quite slow by the end of 100 periods. This suggests that extending the experiments a few periods further would yield little additional insight. A comparison of Figures 2 and 3 reveals that actual play is quite similar to the simulated play, though actual play is somewhat more variable. The extra smoothness of the simulations can be attributed to the larger number of sessions (100 as opposed to 3) used to construct the plots.

7.3 The role of information

While it appears that the amount of information that subjects are given affects their behavior, we have yet to provide any direct evidence that subjects are reacting differently to the different types of information they are given across the three treatments. In particular, we want to examine the extent to which subjects can be viewed as reinforcement learners across the three treatments, and whether they switch to behavior that is more consistent with hypothetical reinforcement when additional information is provided. In an effort to address this issue we have conducted a number of logit regressions where the dependent variable is simply the action chosen by subject i in period t, $a_i(t) \in \{0, 1\}$, where 1 denotes entry. The logit regressions estimate the probability of entry, $\Pr[a_i(t) = 1]$. In addition to a constant term and session dummy variables, we considered two explanatory variables. The first of these is each player's *own relative payoff* from entry, defined as the average payoff from entering over all rounds played (through round $t - 1$), less the average payoff from non–entry over all rounds played, or \$8.[13] The second explanatory variable is the hypothetical payoff each player would expect to earn from entering, given his own past entry decision and the historical average frequency with which the other five players (excluding himself) have chosen to enter over all rounds played (through round $t - 1$), less the payoff for non–entry, \$8. We refer to this explanatory variable as the *hypothetical relative payoff* from entry. If players are strict reinforcement learners, the own relative payoff variable should be the only one of the two payoff variables that matters in explaining the probability of entry. On the other hand, if players are strict hypothetical reinforcement learners, then in the aggregate and full information treatments, the hypothetical relative payoff variable should be the only payoff variable that matters

[12] Table 1 above gives details of perturbed equilibria for $\beta = 5$. One can also calculate that if play was at such a perturbed pure equilibrium the mean squared deviation from the corresponding pure equilibrium would be 0.004.

[13] For the logit regressions involving the limited information sessions, the average payoff from entry or non–entry was initially set equal to zero, as subjects were not made aware of the payoff function. Once they had played either strategy, the average payoff from that strategy would become nonzero, and in particular, the average payoff from not entering would become \$8.

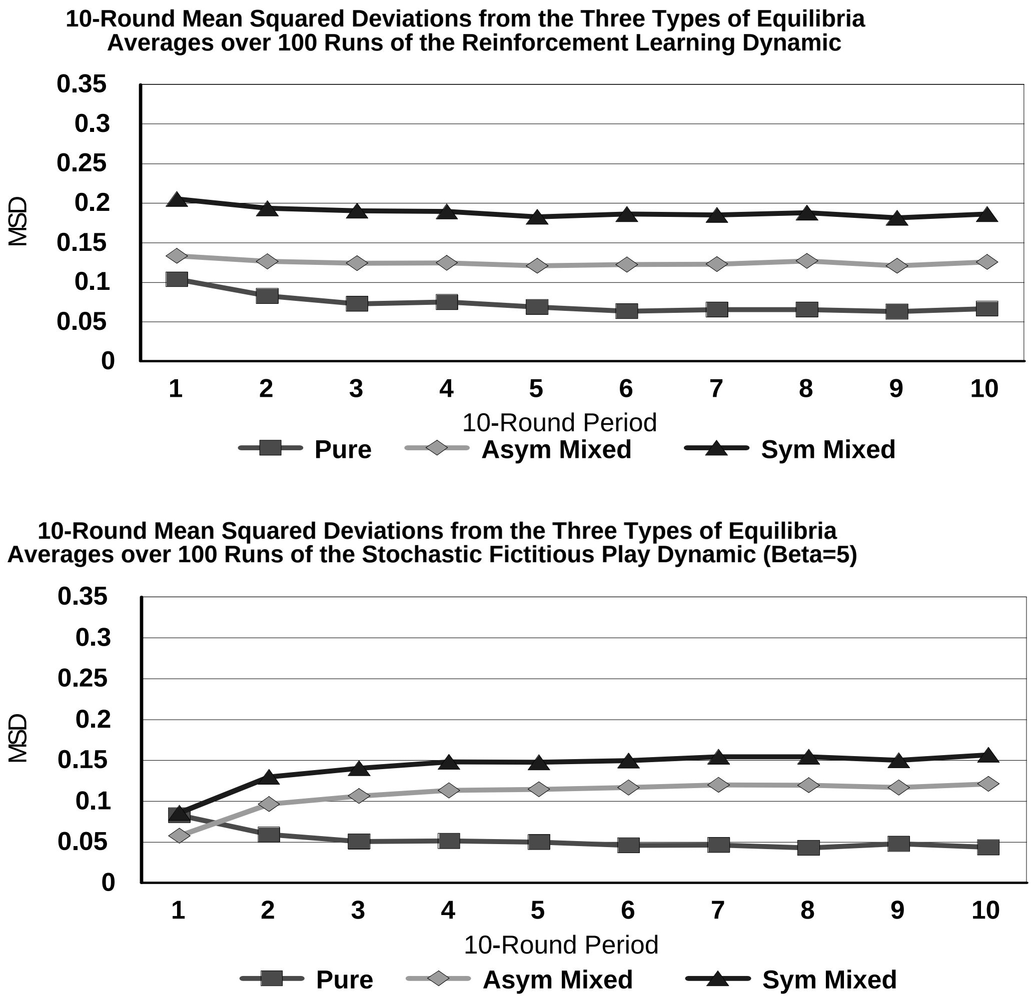

Figure 3: 10-Round Mean Squared Deviations from the Three Types of Equilibria: Simulated Results.

in explaining the probability of entry.

The logit regressions were conducted using pooled individual-level data from each of the three sessions of a treatment - a total of 1800 observations per regression; in addition to a constant term, two session dummy variables were included in every regression specification to capture any fixed differences across sessions within a treatment. Three regression specifications were used. The first included a constant, the two dummy variables and both payoff variables. The second and third specifications purged one of the two payoff variables from the first specification. Table 5 reports the estimated coefficients from all logit regressions, and reports results from a likelihood ratio test of the null hypothesis that the second or third specification is not significantly different from the first specification.

Table 5: Estimates from Logit Regressions of Probability of Entry
(Std. Errors in Parentheses) Two session dummy coefficients suppressed.

Treatment:	Limited Information			Aggregate Information			Full Information		
Specification:	1	2	3	1	2	3	1	2	3
Constant	0.184	-0.143	0.240**	2.417***	0.819***	2.293***	2.577***	0.188*	2.651***
	(0.124)	(0.114)	(0.124)	(0.177)	(0.128)	(0.175)	(0.190)	(0.104)	(0.193)
Own Relative Payoff	0.449***	0.737***	—	0.326***	0.559***	—	0.267***	0.514***	—
	(0.059)	(0.060)		(0.041)	(0.044)		(0.041)	(0.044)	
Hypothetical Relative Payoff	0.748***	—	1.389***	1.679***	—	2.103***	1.866***	—	2.178***
	(0.114)		(0.098)	(0.119)		(0.115)	(0.116)		(0.114)
-ln L	931.8	952.3	975.6	889.6	1003.2	929.1	863.2	1036.1	887.1
l.r. test χ^2		40.85	87.44		227.2	79.03		345.6	47.6
$p > \chi^2$		0.00	0.00		0.00	0.00		0.00	0.00

* significantly different from zero at the 10-percent level, ** at the 5-percent level, *** at the 1-percent level.

The regression results reveal that subjects are neither strict reinforcement learners nor strict hypothetical reinforcement learners. In particular, the coefficient estimate on the own relative payoff variable is positive and significantly different from zero in all three treatments, as is the coefficient estimate on the hypothetical relative payoff variable. Furthermore, likelihood ratio tests allow us to reject the null hypothesis that there is no difference between specifications 1 and 2 or 1 and 3 across all three treatments, in favor of the alternative that specification 1 is always preferred.

The significance of hypothetical relative payoffs in explaining entry decisions in the *limited* information case may appear anomalous, as subjects in this treatment did not have access to the information necessary to construct this hypothetical payoff variable. On the other hand, the hypothetical payoff variable may be positively correlated with past entry decisions, and if there is serial correlation in such decisions, the regression may simply reflect this correlation. For example, in the pure equilibrium, where two players are always in and 4 players are always out, those who always enter would have a hypothetical relative payoff that was always positive and those who always stayed out would have a hypothetical relative payoff that was always negative. One could also hypothesize that in such a simple game as this, players might be able to perform hypothetical reasoning without the information usually required. In particular, some subjects might recognize that the payoff to one action was constant and in effect reinforce the action of staying out with a payoff of $8 even when entering.

Notice, however, that the *magnitude* of the coefficient estimates on the own relative payoff variable decrease as more information is provided (that is as we move from the limited to the full information case), while the magnitude of the coefficient estimates on the hypothetical relative payoff variable increase with the amount of information provided. This finding is consistent with the notion that players are using the additional information to become more sophisticated learners. However, the fact that the coefficient on the relative hypothetical payoff increases further in the full information treatment relative to the aggregate information treatment suggests that the additional information provided in the full information treatment allows subjects to behave in a more sophisticated manner, in contrast to the predictions of hypothetical reinforcement learning models, which posit that the additional information in the full information treatment should not affect learning behavior.

In summary, subjects appear to be neither reinforcement learners, hypothetical reinforcement learners nor imitation learners. While they do seem to respond differently and non-monotonically, to different information partitions, their behavior across treatments is not easily predicted by any one of the variety of learning rules considered in this paper which represent most of the rules currently studied in the learning literature.

8 Conclusions

We have derived new results on learning behavior in market entry games and have carried out an experiment to test our predictions. The theoretical predictions appear to have some support. In most sessions, toward the end of 100 rounds, play was at or close to the pure equilibrium outcome predicted by the reinforcement, fictitious play and threshold learning models. However, the symmetric mixed strategy equilibrium predicted by the imitation learning model was not observed. These findings call into question the notion that imitation is an important means by which individuals learn in market entry–type games. More importantly, these findings suggest that it may take a substantial number of repetitions before the play of experimental subjects in market entry games (and possibly other games as well) approaches the asymptotic predictions of learning models. Consequently, caution appears called for in using asymptotic results for learning models to predict or characterize behavior in economic decision–making experiments, which are typically conducted for relatively shorter lengths of time.

Our experimental design also enabled us to investigate subjects' use of information. Our main conclusion here is that individuals are adaptable in ways that are not captured by current learning models. When individuals possess the minimal amount of information assumed by reinforcement learning models, as in our limited information treatment, such that they do not even know that they are playing a game, they are still capable of learning equilibrium behavior. However, reinforcement learning does not capture the change in behavior that occurs when more information is provided. Similarly, belief based learning models, such as fictitious play, do not capture the qualitative difference in play between our aggregate and full information treatments.

No one learning model appears to capture the behavior observed across our three experimental treatments. Alternatively, if one were to employ a hybrid learning model such as the experience weighted attraction (EWA) approach of Camerer and Ho (1999) which encompasses the various learning models examined here, one would obtain quite different parameter estimates for our three different treatments. That there seems to be no one model that performs robustly across different treatments of the same game, must be disappointing for those who might hope for a learning model that was applicable in all situations. Such a model is only likely to be forthcoming if and when there is a better understanding of how information is used to play games in practice. We hope that this paper sheds light on some of the shortcomings of existing learning models, and spurs other researchers to provide further improvements.

Appendix

This appendix gives the proofs behind the results in the text. We analyze stochastic processes of the form

$$x_{n+1} - x_n = \gamma_n f(x_n) + \eta_n(x_n) + O(\gamma_n^2) \tag{15}$$

for $x_n \in I\!R^n$. We can think of η as the random component of the process with $E[\eta_n | x_n] = 0$. γ_n is the step size of the process. For all the learning models we consider γ_n is a strictly decreasing sequence, with $\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$.

To obtain results on the asymptotic behavior of these stochastic learning processes, we examine the behavior of the mean or averaged ordinary differential equations (ODE's) derived from the stochastic process above as follows,

$$\dot{x} = f(x). \tag{16}$$

We show that in fact the averaged ODE's arising from both reinforcement learning and stochastic fictitious play are both closely related to the evolutionary replicator dynamics (11).

In particular, we apply two classic results from the theory of stochastic approximation. First, Corollary 6.6 of Benaïm (1999) states that if the dynamic (16) admits a strict Liapunov function and possesses a finite number of equilibrium points, then with probability one the stochastic process (15) must converge to one of these points. We show below that suitable Liapunov functions exist for this class of games for all learning models we consider. Second, Theorem 1 of Pemantle (1990) establishes that the stochastic process (15) will converge to an unstable equilibrium point of (16) with probability zero. This is important in that we can show that all mixed strategy equilibria in this class of market entry game are unstable under the replicator dynamics (Lemma 1 below). This combined with the application of Corollary 6.6 of Benaïm (1999) implies that for both reinforcement learning and stochastic fictitious play, convergence must be to a pure strategy equilibrium.

First we examine reinforcement learning. Using the results of Hopkins (2001) it is possible to establish that the mean ODE associated with the model of reinforcement learning given by choice rule (2) and updating rule (4) will be given by the following equations on $[0,1]^N$,

$$\dot{y}^i = \mu^i y^i (1 - y^i) r (c - 1 - \sum_{j \neq i} y^j). \tag{17}$$

If each μ^i were exactly one then we would have the standard replicator dynamics. The additional factor μ^i arises because in the original stochastic learning process there is a different step size, equal to $1/Q_n^i$, for each player. We take the step size of the first player $1/Q_n^1$ to be the step size γ_n of the whole system, and introduce $\mu^i = Q^1/Q^i > 0$ to keep track of the relative speed of learning of the different players. Because each μ^i is not constant over time, strictly, we also require a further set of equations

$$\dot{\mu}^i = \mu^i \left(v + y^i r(c - 1 - \sum_{j \neq i} y^j) - \mu^i (v + y^1 r(c - 1 - \sum_{j \neq 1} y^j)) \right), \tag{18}$$

for $i = 2, 3, ..., N$. There is no equation for μ^1 because from the definition μ^1 is exactly one and hence $\dot{\mu}^1$ is identically zero.

Lemma 1 *For market entry games with generic values of c, the only equilibria of the replicator dynamics (17) together with (18) which are asymptotically stable are pure Nash equilibria. All other equilibria are unstable.*

Proof: For generic, that is, non integer values of c, this class of market entry games has only a finite number of Nash equilibria each of which is isolated. The fixed points of the replicator dynamics consist of these equilibria and in addition all pure strategy profiles. It can be verified that any equilibrium point of the standard replicator dynamics is an equilibrium point for the joint system (17), (18).[14]

We first show that the local stability of any such equilibrium is entirely determined by the replicator dynamics and not by the additional equations (18). The linearization at any fixed point will be of the form

$$\begin{pmatrix} J & 0 \\ d\dot{\mu}/dy & d\dot{\mu}/d\mu \end{pmatrix}, \tag{19}$$

where J is the Jacobian of the linearized replicator dynamics. Because of the block of zeros to the upper right, it can be shown that every eigenvalue of a matrix of the above form is an eigenvalue for either J or $d\dot{\mu}/d\mu$. The latter matrix is diagonal and has only negative elements. Hence, if J has one or more positive eigenvalues, the equilibrium point is unstable for the joint dynamics, if it has N negative eigenvalues, the equilibrium point is asymptotically stable for the joint dynamics.

We now investigate the structure of J. At any fully mixed equilibrium where all players enter with probability $\bar{y}$, the Jacobian J of the linearized replicator dynamics has the form $J_{ii} = \mu^i(1 - 2y^i)r(c - 1 - \sum_{j \neq i} y^j)$ which equals zero if $y^i = \bar{y}$ for $i = 1, ..., N$, and $J_{ij} = -\mu^i \bar{y}(1 - \bar{y})r$. $-J$ has therefore only nonnegative entries and one can see that $(-J)^2$ has only strictly positive entries. So by the theorem of Perron-Frobenius, $-J$ has a strictly positive eigenvalue. Now, since the trace of $-J$ is zero, it must have a negative eigenvalue. Clearly, J also has at least one positive and one negative eigenvalue. Hence, we have a saddlepoint.

At any asymmetric mixed equilibrium let the first $N - j - k$ players randomize over entry and the remaining $j + k$ players play pure. Then one can calculate that in this case that the Jacobian evaluated at this equilibrium has the form

$$J = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix},$$

[14] In fact, for each equilibrium point of the standard replicator dynamics, there are two for the joint system, one with μ positive and the other with μ equal to zero. However, the latter is always unstable and is never an asymptotic outcome for the reinforcement learning process.

where A is a $N-j-k \times N-j-k$ block of the form found at a symmetric mixed equilibrium as described above, and C is a diagonal matrix. It is easy to show that the eigenvalues of J consist of the eigenvalues of C, which are negative, and of A, which by the above argument are a mixture of positive and negative.

At any pure profile, one can calculate that the Jacobian is diagonal. Furthermore, if this profile is not a Nash equilibrium then at least one diagonal element is positive. In contrast, at a pure Nash equilibrium which must be strict as c is non-integer, all elements are negative. ∎

Proof of Proposition 1: As outlined above, the proof is in two steps. We identify a suitable Liapunov function which ensures convergence of the stochastic process. Then, we show that the stochastic process cannot converge to a mixed equilibrium. For technical reasons that will become apparent, this proof is actually for the following perturbed system which is arbitrarily close to the original. We replace the original simple reinforcement updating rule with

$$q^i_{1n+1} = q^i_{1n} + \delta^i_n(v + r(c - m_n) - \lambda \log y^i_n),\ q^i_{2n+1} = q^i_{2n} + (1 - \delta^i_n)(v - \lambda \log(1 - y^i_n)), \tag{20}$$

for a very small $\lambda > 0$. Hopkins (2001) shows the associated ODE in this case is

$$\dot{y}^i = \mu^i y^i (1 - y^i) \left(r(c - 1 - \sum_{j \neq i} y^j) + \lambda(\log(1 - y^i_n) - \log y^i_n) \right). \tag{21}$$

For λ sufficiently close to zero, the fixed points of the modified dynamic system will be arbitrarily close to those of the original but on the interior of $[0,1]^N$. Then define

$$V_0(y) = -r \left(y^1 y^2 + y^1 y^3 + ... + y^2 y^3 + ... + y^{N-1} y^N \right) + r(c-1) \sum_{i=1}^N y^i. \tag{22}$$

This function has a local maximum at each pure Nash equilibrium and a local minimum at each pure state which is not Nash. Now consider, the modified Liapunov function

$$V_1(y) = V_0(y) - \lambda \sum_{i=1}^N \left(y^i \log y^i + (1 - y^i) \log(1 - y^i) \right).$$

Note that

$$\frac{\partial V_1(y)}{\partial y^i} = r(c - 1 - \sum_{j \neq i} y^j) + \lambda(\log(1 - y^i) - \log y^i).$$

This implies that, first, the critical points of V_1 correspond to perturbed equilibria of the dynamics (21), and second, given the dynamic system (21) and (18),

$$\dot{V}_1(y) = \frac{dV_1(y)}{dy} \cdot \dot{y} = \sum_{i=1}^N \mu^i y^i_n (1 - y^i_n) \left(r(c - 1 - \sum_{j \neq i} y^j_n) + \lambda(\log(1 - y^i) - \log y^i) \right)^2 \geq 0$$

with equality only where $\dot{y} = 0$. Hence, $V_1(y)$ is a strict Liapunov function in the sense of Corollary 6.6 of Benaïm (1999). Second, for generic values of c, this class of game possesses a finite number of equilibria. Hence, by that Corollary, the stochastic process must converge to an equilibrium point.

However, if we take the limit $\lambda \to 0$, the linearization of the dynamics (18) and (21) approaches the linearization of the replicator dynamics (19). Hence, the above Lemma demonstrates that all equilibria of these dynamics except pure Nash equilibria are unstable under the deterministic dynamics (18) and (21) for λ sufficiently small.

The next step is to apply Theorem 1 of Pemantle (1990) which states that if an equilibrium point of the associated ODE is unstable, there is a zero probability of convergence of the stochastic process to that point, subject to four technical conditions. The first is simply that the linearization of the ODE at the equilibrium point has at least one positive eigenvalue. The second and fourth require that the step size of the stochastic process be of order $1/n$. Here, the step size is $1/Q_n^1$. The third condition is that

$$E[\max(0, \theta \cdot \eta_n)|y_n \in M] \geq C/n$$

where M is a neighborhood of the equilibrium, θ is any unit vector and C is a positive constant. This condition demands that each η_n^i must have positive realizations, something which is only true in the system we consider in the interior of $[0,1]^N$.Hence, our assumption that $\lambda > 0$.

So the learning process converges to the fixed points of the replicator dynamics apart from those which are pure Nash equilibria with probability zero. It follows that the learning process must converge to a pure Nash equilibrium with probability one. ∎

Proof of Proposition 2: In the case of the exponential version of stochastic fictitious play, given the expected motion (12), (see Hopkins (2001) for details), the associated ODE will be

$$\dot{y}^i = \beta \left(y^i(1-y^i)r(c-1-\sum_{j\neq i} y^j) + \frac{1}{\beta} y^i(1-y^i)(\log(1-y^i) - \log y^i) \right). \tag{23}$$

It is relatively easy to verify that $V_1(y)$ can be used as a Liapunov function if one sets $\lambda = 1/\beta$. Again because of the existence of a suitable Liapunov function, the learning process will converge to the set of perturbed equilibria. With the exponential dynamics (23), as β becomes large, the dynamics approach a positive scalar transformation of the replicator dynamics (11). So for β large enough the results of Lemma 1 will hold. Again by Theorem 1 of Pemantle (1990) convergence to any equilibrium other than a pure Nash equilibrium is impossible. Condition 3 here is easily verified as all perturbed equilibria are strictly interior. ∎

Proof of Proposition 3: One can calculate that in this case the associated ODE will be

$$\dot{y}^i = \beta \left(y^i(1-y^i)r(c-\sum_{j=1}^{N} y^j) + \frac{1}{\beta} y^i(1-y^i)(\log(1-y^i) - \log y^i) \right). \tag{24}$$

Again we can find a Liapunov function, in this case

$$W(y) = \frac{r}{2}(c - \sum_{j=1}^{N} y^j)^2 - \frac{1}{\beta} \sum_{i=1}^{N} \left(y_n^i \log y_n^i + (1-y_n^i)\log(1-y_n^i) \right).$$

Again the stochastic process must converge to an equilibrium point. Clearly, as under this learning rule updating is identical for all agents, the limiting outcome is symmetric. Hence, in the limit, the learning process must converge to the unique symmetric perturbed equilibrium which satisfies (14). ∎

Proof of Proposition 4: Note that each c^i changes if and only if the payoff to the action not taken is higher than to the action taken. This immediately implies that a vector $c_n = (c_n^1, ..., c_n^N)$ is a fixed point of the process if and only if it induces an action profile which is a pure Nash equilibrium. Note further that if $m_n > (<)\bar{c}$ where $\bar{c}$ is the largest integer smaller than c, each entrant will raise her c^i (each player staying out will lower his c^i). For generic initial conditions (that is, no two players have the same initial threshold level) and if the rate of adjustment w is sufficiently small, the number of entrants will converge monotonically to $\bar{c}$. ∎

Finally, we establish an identical result to Proposition 2 for stochastic fictitious play in terms of beliefs. Let z_{in}^j be the probability that i assigns to the possibility that j will enter in period n. In fictitious play, the beliefs are equal to the empirical frequency of j's actions, which implies the following updating rule

$$z_{in+1}^j = z_{in}^j + \frac{\delta_n^j - z_{in}^j}{n+1}. \tag{25}$$

It is also possible to write the exponential choice rule (3) in terms of expected payoffs given beliefs

$$y_n^i = \frac{\exp \beta\pi_{1n}^i}{\exp \beta\pi_{1n}^i + \exp \beta\pi_{2n}^i}, \tag{26}$$

where $\pi_{1n}^i = v + r(c - 1 - \sum_{j \neq i} z_{in}^j)$ and $\pi_{2n}^i = v$. In this case, the associated ODE's (see for example, Benaim and Hirsch, 1999) are the system of equations, for $i = 1, ..., N$,

$$\dot{z}^i = y^i(z) - z^i \tag{27}$$

where $y^i(z)$ is the exponential choice rule given in (26).

Proposition 5 *If all players learn according to stochastic fictitious play in terms of the belief updating rule (25) and associated exponential choice rule (26), for generic values of c and for* β *sufficiently large, then* $\Pr\{\lim_{n\to\infty} y_n \in \hat{Y}\} = 1$, *where* $\hat{Y}$ *is the set of perturbed equilibrium each corresponding to one of the pure Nash equilibria of the game.*

Proof: One can show that $V_1(z)$ works as a Liapunov function for these dynamics in a similar way as $V_1(y)$ does for the dynamics (23) (see also, Theorem 3.3 of Hofbauer and Hopkins, 2000). Hopkins (2001) shows the dynamics (27) have the same linearization as the dynamics (23) generated by fictitious play in terms of propensities, and so all but pure equilibria are asymptotically unstable. Thus again, the stochastic process must converge with probability one to a pure equilibrium. ∎

References

Benaïm, M. (1999) "Dynamics of stochastic algorithms," in *Séminaire de Probabilités XXXIII*, J. Azéma et al. Eds, Berlin: Springer-Verlag.

Benaïm, M., Hirsch, M.W. (1999) "Mixed equilibria and dynamical systems arising from fictitious play in perturbed games," *Games Econ. Behav.*, **29**, 36-72.

Benveniste, A., Métivier, M., Priouret, P. (1990) *Adaptive Algorithms and Stochastic Approximations*. Berlin: Springer-Verlag.

Binmore, K., Samuelson, L. (2001) "Evolution and mixed strategies," *Games Econ. Behav.* **34**, 200-226.

Camerer, C., Ho, T-H. (1999). "Experience-weighted attraction learning in normal form games," *Econometrica*, **67**, 827-874.

Camerer, C., Lovallo, D. (1999). "Overconfidence and excess entry: an experimental approach," *Amer. Econ. Rev.*, **89**, 306-318.

Duffy J., Feltovich, N. (1999). "Does observation of others affect learning in strategic environments? An experimental study," *Int. J. Game Theory*, **28**, 131-152.

Erev, I., Rapoport, A. (1998). "Coordination, "magic", and reinforcement learning in a market entry game," *Games Econ. Behav.*, **23**, 146-175.

Erev, I., Roth, A.E. (1998). "Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria," *Amer. Econ. Rev.*, **88**, 848-881.

Fudenberg, D., Levine D. (1998). *The Theory of Learning in Games*. Cambridge, MA: MIT Press.

Hofbauer, J., Hopkins, E. (2000). "Learning in perturbed asymmetric games," working paper, http://www.ed.ac.uk/~ehk/

Hofbauer, J., Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge, UK: Cambridge University Press.

Hopkins, E. (2001). "Two competing models of how people learn in games," forthcoming in *Econometrica*, http://www.ed.ac.uk/~ehk/

Huck S., Norman, H., Oechssler, J. (1999). "Learning in Cournot oligopoly–an experiment," *Econ. J.*, **109**, pp. C80-95.

Monderer, D., Shapley, L. (1996). "Fictitious play property for games with identical interests," *J. Econ. Theory*, **68**, 258-65.

Ochs, J (1998), "Coordination in market entry games," in D.V. Budescu, I. Erev and R. Zwick, (eds.) *Games and Human Behavior* 143–172.

Pemantle, R. (1990). "Nonconvergence to unstable points in urn models and stochastic approximations," *Ann. Probability*, **18**, 698-712.

Rapoport, A., Seale, D.A., Erev, I., Sundali, J.A. (1998), "Equilibrium play in large group market entry games," *Manag. Sci.*, **44**, 119-141.

Rapoport, A., Seale, D.A. and Parco, J.E. (2000), "Coordination in the aggregate without common knowledge or outcome information," to appear in R. Zwick and A. Rapoport (eds.) *Experimental Business Research* Norwell, MA: Kluwer academic press.

Rapoport, A., Seale, D.A. and Winter, E. (1998), "Coordination and learning behavior in large groups with asymmetric players," *Games Econ. Behav.*, in press.

Rapoport, A., Seale, D.A. and Winter, E. (2000), "An experimental study of coordination and learning in iterated two–market entry games," *Econ. Theory* 16, 661–687.

Sarin, R., Vahid, F. (1999). "Payoff assessments without probabilities: a simple dynamic model of choice," *Games Econ. Behav.*, **28**, 294-309.

Seale, D.A. and Rapoport, A. (2000), "Elicitation of strategy profiles in large group coordination games," *Exp. Econ.* 3, 153–179.

Sundali, J.A., Rapoport, A., Seale, D.A. (1995). "Coordination in market entry games with symmetric players," *Organ. Behav. Human Decision Processes*, **64**, 203-218.

Vega-Redondo, F. (1997). "The evolution of Walrasian behavior," *Econometrica*, **65**, 375-84.