The El Farol Bar Problem Revisited: Reinforcement Learning in a Potential Game

Duncan Whitehead
Edinburgh University

Date
September 2008

THE UNIVERSITY *of* EDINBURGH

The El Farol Bar Problem Revisited: Reinforcement Learning in a Potential Game

Duncan Whitehead
University of Edinburgh

September 17, 2008

Abstract

We revisit the El Farol bar problem developed by Brian W. Arthur (1994) to investigate how one might best model bounded rationality in economics. We begin by modelling the El Farol bar problem as a market entry game and describing its Nash equilibria. Then, assuming agents are boundedly rational in accordance with a reinforcement learning model, we analyse long-run behaviour in the repeated game. We then state our main result. In a single population of individuals playing the El Farol game, learning theory predicts that the population is eventually subdivided into two distinct groups: those who invariably go to the bar and those who almost never do. In doing so we demonstrate that learning theory predicts sorting in the El Farol bar problem.

1 Introduction

The El Farol bar problem was introduced by Brian W. Arthur (1994) as a framework to investigate how one models bounded rationality in economics. It was inspired by the El Farol bar in Santa Fe, New Mexico, which offered Irish music on Thursday nights. The original problem was constructed as follows:

> "N people decide independently each week whether to go to a bar that offers entertainment on a certain night. For correctness, let us set N at 100. Space is limited, and the evening is enjoyable if things are not too crowded – specifically, if fewer than 60 percent of the possible 100 are present. There is no sure way to tell the numbers coming in advance; therefore a person or an agent goes (deems it worth going) if he expects fewer than 60 to show up or stays home if he expects more than 60 to go."[1]

Arthur's (1994) preliminary results from the field of computational economics show that the number of people attending the bar converges quickly and then hovers around the capacity level of the resource.

Our contribution to the literature on the El Farol bar problem and theory of learning in games is fourfold. First, we apply the Erev and Roth (1998) model of reinforcement learning to the El Farol framework. We believe the

---

[1] Arthur (1994), pp 409.

Erev and Roth (1998) model of reinforcement learning is the most appropriate individual learning model to apply in this instance, because in general people who stay at home do not know what payoff they would have received if they had gone to the bar. We then prove analytically that long-run behaviour will converge asymptotically to the set of pure strategy Nash equilibria of the El Farol stage game.[2] In other words the number of people attending the bar converges and then hovers around the capacity level of the resource. Furthermore, learning theory predicts sorting in the El Farol bar problem; that is, in a single population of individuals playing the El Farol game, learning theory predicts that the population is eventually subdivided into two distinct groups: those who invariably go to the bar and those who almost never do.

Second, we demonstrate that the El Farol bar problem may be modelled as a market entry game with boundedly rational reinforcement learners. We build upon the work of Duffy and Hopkins (2005), who have proved that in market entry games, where payoffs are decreasing in a continuous manner with respect to the number of other market entrants, the only asymptotically stable Nash equilibria are those corresponding to pure Nash profiles. Our main result also proves asymptotic convergence to those equilibria corresponding to pure Nash profiles in the market entry game. In addition our result also proves that this is the case when payoffs are decreasing in a discontinuous way with respect to the number of other market entrants.

Third, Sandholm (2001) has proved that, under a broad class of evolutionary dynamics, behaviour convergences to Nash equilibrium from all initial conditions in potential games with continuous player sets. Sandholm's (2001) convergence results assume that individual behaviour adjustments should satisfy what was termed *positive correlation*; meaning any myopic adjustment dynamic that exhibits a positive relationship between growth rates and payoffs in each population. Our result contributes to this literature by proving that, for the evolutionary dynamics associated with Erev and Roth's (1998) model of reinforcement learning, long-run behaviour converges in potential games with finite sets of players.

Finally, there is a contribution to be made to the extensive literature on the El Farol bar problem and its associated problem, the Minority Game in the field of complex systems.[3] Currently, it would appear that the opportunity to apply convergence results from models of individual learning to situations like those represented by the El Farol bar problem has been overlooked.

We will begin by using the tools of game theory to model the El Farol bar problem as a non-cooperative coordination game in which payoffs are determined by negative externalities. We then model the El Farol bar problem as a repeated market-entry game with boundedly rational agents. Analysis of the stage game will show that there are a large number of Nash equilibria. Therefore, equilibria refinement/coordination becomes problematic. In order to refine the equilibria set, we allow players to learn from experience. The analytical tools developed in Duffy and Hopkins (2005), Hopkins and Posch (2005) and Monderer and Shapley (1996) will be employed to study the predicted outcome of play under the Erev and Roth (1998) model of reinforcement learning.

---

[2] This is in contrast with Franke's (2003) use of numerical simulations of reinforcement learning applied to the El Farol bar problem.

[3] See http://www.unifr.ch/econophysics/minority/ for research on the Minority Game.

Reinforcement learning assumes that individuals only have access to the attendance figures of the bar for each week that they attend.[4] The long-run behaviour of agents under this adaptive rule will then be considered, and it will be shown that under this learning process, play will converge to the set of pure strategy Nash equilibria with probability one.

The intuition behind our main result is that in the El Farol bar problem reinforcement learners who do not regularly attend are more often than not disappointed when they do choose to do so. Similarly, those who regularly attend always seem to have a good time, and thus are more likely to attend in the future.

A good way to think about this outcome is to imagine that all players in one week play a mixed strategy. It is quite likely that the bar actually turns out to be busy. Therefore, all agents who attended will be reinforced with the lower payoff. This will reduce their propensity to attend in the future. The following week the probability of the bar being overcrowded will be diminished. Those who do attend will most likely receive high payoff reinforcement from attending and their propensity to attend in the future will increase again while that of the players who stayed away will be reduced. Therefore, we have two positive feedback loops. One causes those who attend regularly to do so more often. The other leads those who stay at home to be more likely to do so in the future. We can therefore see that any mixed strategy Nash equilibrium is asymptotically unstable under the dynamics of Erev and Roth (1998) reinforcement learning.

In Section 2 we review the El Farol bar problem as introduced by Arthur (1994). We set out his modelling approach to bounded rationality in the El Farol bar problem and summarise the initial results from his computational experiments. We discuss the use of the inductive thinking approach to modelling bounded rationality, both in the El Farol bar problem and its closely related problem, the Minority Game. We then outline our motivation for the application of the individual learning approach to capturing the bounded rationality of decision makers and suggest a reinforcement learning model for the El Farol framework. In Section 3 we introduce our model of the El Farol bar problem, define the El Farol stage game and characterise the set of Nash equilibria, set out in detail the Erev and Roth (1998) model of reinforcement learning within the El Farol framework, and write down an expression for player's expected strategy adjustment. In Section 4 we state and prove our main result; that in the El Farol bar problem a population of boundedly rational agents who behave in accordance with the Erev and Roth (1998) reinforcement learning model are sorted into those who always attend the El Farol bar and those who always stay at home. Finally, we provide some concluding remarks in Section 5.

2   The El Farol Bar Problem

The El Farol bar problem was created by Arthur (1994) as a device to investigate how one might best model bounded rationality in economics. It was inspired by the El Farol bar in Santa Fe, New Mexico, which offered Irish music on Thursday nights. The problem is set out as follows: there is a finite population

---

[4]However, the results presented here within are easily extended to allow for the more generic set-up where all individuals learn attendance figures whether they attend or not. This is often referred to as hypothetical reinforcement or fictitious play learning.

of people and every Thursday night all of the them want to go to the El Farol bar. However, the El Farol bar is quite small, and it is not enjoyable to go there if it is too crowded. So much so, in fact, that the following rules are in place:

- If less than 60% of the population go to the bar, those who go have a more enjoyable evening at the bar than they would have had had they stayed at home.

- If 60% or more of the population go to the bar, those who go have a worse evening at the bar than they would have had had they stayed at home.

Unfortunately, it is necessary for everyone to decide at the same time whether they will go to the bar or not. They cannot wait and see how many others go on a particular Thursday before deciding to go themselves on that Thursday.

The important characteristic of the El Farol bar problem is that if there was an obvious method that all individuals could use to base their decisions on, then it would be possible to find a deductive solution to the problem. However, no matter what method each individual uses to decide if they will go to the bar or not, if everyone uses the same method it is guaranteed to fail. Therefore, from the point of view of the individual, the problem is ill-defined and no deductive rational solution exists.

Situations like those represented by the El Farol bar problem highlight two specific reasons why perfect deductive reasoning might fail to provide clear solutions to some theoretical problems. The first is simply a question of the cognitive limitations of the mind. Beyond a certain level of complexity, logical capacity fails to cope. The second is that in complex strategic situations individuals cannot always rely on persons they are interacting with to behave under assumptions of perfect rationality. In situations like the El Farol bar problem, individuals are forced into a world where they must choose their strategies based on guesses of their opponents' likely behaviour. Without objective, well-defined, shared assumptions, these types of problems become ill-defined and cannot be solved rationally.

The question that arises is how does one best model bounded rationality in economics when perfect rationality fails? Given the defining characteristic of the El Farol bar problem, namely that finding a deductive rational solution is impossible, it follows that the problem itself could provide a useful framework to explore models of bounded rationality in general.

2.1 Inductive Reasoning in the El Farol Framework

Arthur (1994) notes that there is a consensus among psychologists that in situations that are either complicated and/or ill-defined, humans tend to look for patterns in order to develop internal models on which they can base their decisions. These methods are inherently inductive. In the El Farol bar problem, Arthur (1994) follows this line of thought and postulates that individuals decide whether they will go to the bar or not by employing mental models to predict expected future attendance. In other words they create forecasting models. If an individual using a specific forecasting model predicts attendance to be low then, based on that model, that individual would attend and vice-versa if attendance is predicted to be high.

4

As previously discussed, and deriving from the ill-defined nature of the El Farol bar problem itself, we can conclude that no forecasting model can be employed by all individuals and be accurate at the same time. We can easily demonstrate this fact by assuming that a forecasting model exists that predicts that the attendance in the coming week, given attendance in past weeks, is going to be high. If all individuals use this forecasting model to base their decisions on, then nobody will go to the bar.[5] This then renders the forecast invalid and implies that there exists no single forecasting model that all individuals can use upon which to base their attendance decisions. No deductive solution exists to this problem.

2.1.1 The Inductive Thinking Approach

Arthur's (1994) approach to modelling bounded rationality in the El Farol bar problem is to assume that each individual has access to a number of forecasting models which they use to make their decisions. Furthermore, they score and rank these models at the end of each week according to their accuracy in order to determine which particular model they should base their decision on.

Formally, Arthur (1994) imagines that each individual utilises a number of forecasting models, denoted s^k, to predict attendance in the coming week. Each model forecasts attendance for the coming week given the history of attendance over the last d weeks, denoted $d(h_{t-1}) \in D$, where D is the set of all possible attendance profiles for the last d weeks and d is an exogenously fixed parameter. Then, following the disclosure of the number of individuals who attended the El Farol bar on the most recent Thursday night, a score is associated with each forecasting model. Specifically, the score, denoted $U_t(s^k)$, is calculated by computing the weighted average of the score of the same model in the previous week and the absolute difference between the forecasting model's last prediction, denoted $s^k(d(h_{t-1}))$, and the most recent realised turnout, denoted y_t. Equation (1) formulises this calculation.[6]

$$U_t(s^k) = \lambda U_{t-1}(s^k) + (1-\lambda) |s^k(d(h_{t-1}) - y_t)| \tag{1}$$

In each week the forecasting model with the highest score is referred to as the active predictor. On each Thursday individuals undertake the action of either attending the El Farol bar or not in accordance with their active predictor. If an individual's active predictor forecasts the attendance on the coming evening to be high, then that individual will choose not to go to the bar. Conversely, if the active predictor forecasts attendance to be low, then that individual will deem it worthwhile going to the bar and they will anticipate an enjoyable evening of Irish music. Once all individuals have made their decisions, i.e. whether to attend the El Farol bar or not, they are then informed of the actual turnout at the bar. This information is made know publicly to all individuals. Each individual then realises their payoffs, updates the score for all their available forecasting models, and confirms their active predictor for next Thursday's decision.

---

[5] This is reminiscent of Yogi Berra's famous comment, "Oh, that place. It's so crowded nobody goes there anymore."

[6] It should be noted that I have taken specific care to outline the El Farol bar problem and Arthur's proposed model of the problem as he originally formulated it. This has been possible due to the work of Zambrano (2004) who re-analysed Arthur's original code.

2.1.2 Agent-Based Computer Simulations

Arthur (1994) investigated this model of the El Farol bar problem through the use of computational experiments. He designed artificial agents and simulated their dynamic interaction over time.

In Arthur's (1994) computer simulations, as in the original formulation of the problem, the size of the population, N, is set to 100 and the enjoyable capacity of the El Farol bar, C, is set to 60. Arthur (1994) then creates a finite set of diverse forecasting models, or predictors, which map attendance histories to a predicted bar attendance for the coming week. These models were doled out uniformly and randomly, such that each agent was endowed with a non-transferable set of K forecasting models.[7] Each simulation experiment was then run for 100 periods with the combined runs totalling to $10,000$ periods.
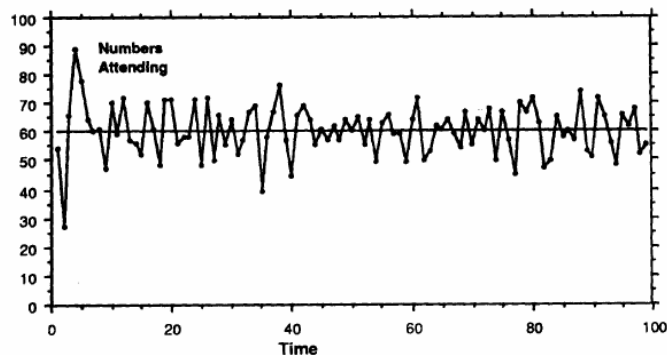

Figure 1: Attendance According to Arthur's (1994) Simulations.

The first thing to note about the results of these computer experiments is that, given the starting conditions and the fixed set of predictors available to each simulated agent, the dynamics are completely deterministic. Nevertheless, the simulations produce some interesting results. Two observations become immediately apparent.

First, mean attendance always converges to the capacity of the bar. Second, on average 40% of the active predictors forecasted attendance to be higher than the capacity level and 60% below. Arthur (1994) expands on these observations by noting that, "the predictors self organise into an equilibrium pattern or 'ecology'."[8] An example of the attendance rates from a typical run of 100 periods can be seen in Figure 1.

2.1.3 The Minority Game

There has been much interest in the El Farol bar problem as a system to study agents in market-like interactions. This has led to the definition of a similar problem called the Minority Game which embodies some basic market mechanisms, while keeping mathematical complexity to a minimum.

---

[7] This did not preclude the possibility that the agents' predictor sets might overlap.
[8] Arthur (1994), pp.409.

The Minority Game is a repeated game where N agents have to decide between two actions, such as buy or sell or attend or not. With N odd this procedure identifies a minority action as that chosen by the minority. Agents who take the minority action are rewarded with one payoff unit. Agents cannot communicate with one another and they have access to publicly available information on the history of past outcomes for a fixed number of periods. As in the El Farol bar problem, the set up requires a prohibitive computational task and, from a strategic point of view, the problem is ill-defined. Again it is postulated that in such complex strategic interactions, agents may prefer to simplify their decision tasks by seeking out behaviour rules, or heuristics, that allocate an action for each possible observed history of outcomes.

The literature on the Minority Game concludes, through both agent-based and analytical models, that there exists a cooperative phase of play when the ratio of the number of unique possible histories to the number of agents, N, is large enough. That is, with respect to the so-called 'random agent' state, in which each agent chooses their action by flipping a coin, agents are better off because the system moves to a sort of 'coordinated' state. The analytical research on the Minority Game employs techniques borrowed from statistical physics in order to describe the game as a spin system, thus enabling the system's properties to be outlined. It should be noted that this avenue of investigation does not enable the study of individual behaviour, but only the system as a whole.

One aspect of this approach, and indeed Arthur's (1994) original investigations, to the El Farol bar problem and bounded rationality is that the theory does not explicitly detail the predictors that should/would be available to each individual/agent. In reality there most likely exists an evolutionary process that regulates the set of predictors as a whole and their availability to each individual agent. Arthur (1994) draws on the following metaphor to make the point: "Just as species, to survive and reproduce, must prove themselves by competing and being adapted within the environment created by other species, in this world hypothesis, to be accurate and therefore acted upon, must prove themselves by competing and being adapted within and environment created by other agents' hypothesis."[9]

2.2 Individual Learning in the El Farol Framework

The El Farol bar problem represents a complex strategic environment where rational deductive thinking fails to provide any clear solutions. The question we wish to address is what we should put in place of perfect rationality. In the previous section, we reviewed the literature reporting work that has been directed at achieving this goal within the El Farol framework through the use of inductive reasoning. Suppose instead that individuals in the El Farol bar problem can find their way to an optimal solution by trial and error, i.e. learning.[10] In effect we propose that this is the role that, loosely speaking, the predictors fulfil in Arthur's (1994) original paper on inductive reasoning and bounded rationality in the El Farol bar problem. Recall that if a predictor correctly forecasts attendance, it is more likely to be used as an active predictor.

---

[9] Arthur, (1994), pp. 408.

[10] A player cannot adapt to situations that are only encountered once. With this in mind, we must consider players learning equilibria in an identically repeated game environment.

If not, it will not be used. Following this argument it seems reasonable to consider the El Farol bar problem as one with boundedly rational agents who gradually adjust their behaviour over time, until there is no longer any room for improvement in their payoffs.

In game theory the techniques for modelling this type of adaptation process are closely related to replicator dynamics. The idea of replicator dynamics was introduced by Smith (1974) to model dynamic processes in the biological sciences. Essentially, replicator dynamics says that if an individual of a certain type earns an above average payoff, then that individual type's frequency in the population rises. When modelling an individual learning process in a repeated game, we modify this interpretation of replicator dynamics to the following: if an individual who has a propensity to use a particular strategy earns an above average payoff from that strategy, then the propensity to use that strategy in the future increases.

The El Farol bar problem will now be modelled as a repeated market entry game where players adhere to a pre-specified learning process. The manner in which individual learning is modelled in repeated games is simple and quite intuitive. Essentially, individual learning is an algorithm that each player follows in each period of play. Imagine that each individual in the El Farol bar problem, whether they go to the bar or not, keeps an urn by their side. In the urn there are a number of balls coloured either green or red. We can consider these balls to be replacing the function of Arthur's (1994) predictors in the El Farol bar problem.[11] Instead of each individual making their action choice dependent on the forecast of their active predictor, players will choose a ball from their urn and obey its colour coding. In other words if a green ball is selected that individual will go to the bar and if a red ball is chosen they will stay at home. Once a ball is drawn and the corresponding action is taken, the ball is then placed back into the urn.

The learning model is then specified by an updating rule. This is the set of instructions that dictates how many balls and of what colours should be added to the urn after each round of play. Using this framework we can describe each player as having propensities for each action. The propensity to undertake a certain action is a function of the number of correspondingly coloured balls in the urn.[12] The probability that a ball of a certain colour will be chosen from a particular individual's urn is determined by the choice rule, which is a mapping from propensities to a number in the unit interval. To find the equilibrium, we calculate in the limit, as the number of repetitions of the game tends to infinity, the probability that each action will be taken.

Let us now recall in detail the motivation for employing an individual learning model of bounded rationality in the El Farol bar problem. As previously stated, the complexity of the problem makes it reasonable to assume that individuals suffer from cognitive limitations. Furthermore, we have already demonstrated that the complexity of beliefs means that, from a strategic viewpoint, individuals are unable to employ deductive reasoning to identify optimal/coordinated strategies. Given these constraints we suppose that individuals find their optimal strategies in the El Farol bar problem through

---

[11] This is not to be taken literally, but they will provide the same decision function as the predictors do in Arthur's formulation.

[12] It is also dependent on the choice rule specified in the learning model which shall be expanded on later in the paper.

8

repeated interaction and the application of an adaptive algorithm. It will be assumed that any adaptive algorithm will adhere to some basic principles of individual learning.

First, the law of effect: choices that have led to good outcomes in the past are more likely to be repeated in the future. Second, the power law of practice: learning curves should initially be steep and then later they should be flatter. This is paramount to assuming that in any adaptive process the adjustments become smaller over time. Finally, choice behavior should be probabilistic. This is a basic assumption in most mathematical learning theories proposed in psychology. Erev and Roth (1998) have developed a robust model of reinforcement learning which incorporates all these principles that shall be applied to our model of the El Farol bar problem.

3   A Model of the El Farol Bar Problem

The El Farol bar problem is essentially a repeated simultaneous move game. There are N players with identical preferences who attempt to coordinate their actions of either going to the bar or staying at home in such a way as to maximise their individual payoffs, subject to the crowding externality from going to the bar. Players need to coordinate their actions, independently and without prior communication, such that:

- when a player decides to go to the bar, i.e. deems it worthy of going to the bar, they can look forward to a payoff that is greater than what they would have received had they stayed at home and

- when a player decides to stay at home, i.e. deems it not worthy of going to the bar, they can look forward to a payoff that is greater than what they would have received had they not stayed at home.

The El Farol bar problem can be interpreted as a market entry game (Franke 2003). In general market entry games are interpreted as truncated two-stage games (Selten and Güth 1982). In the first stage, players simultaneously choose either to enter or stay out of the market. Then, in the second stage, the payoffs of the entrants are determined from their market actions. Usually these payoffs are negatively related to the number of market entrants in a continuous way. However, in the El Farol bar problem, payoffs to players entering the bar are related to the number of bar attendants in a discontinuous manner.

Alternatively, the El Farol bar problem may be viewed as a congestion model and thus can be modelled, a la Rosenthal (1973), as a congestion game.[13] It is a congestion game, because each player's payoff depends on the number of other players who choose to utilise the same resource, namely the El Farol bar. This interpretation has been referred to in many studies of the El Farol bar problem in the literature (e.g. Greenwald, Mishra, and Parikh 1998, Bell and Sethares 1999, Bell and Sethares 2001, Bell, Sethares, and Bucklew 2003, Farago, Greenwald, and Hall 2002, Zambrano 2004), but has rarely been developed.

In this paper we shall initially interpret the El Farol bar problem as a market entry game. Later on in our discussions we shall return to the idea of congestion

---

[13]Clearly market entry games are a subset of the larger class of congestion games.

games, because they have important properties that are useful in understanding the long-run behaviour of boundedly rational agents learning in accordance with a reinforcement model in the El Farol bar problem.

3.1 The El Farol Stage Game

Let C, a positive no-zero integer, represent the capacity of the bar. If less than C players choose to go to the bar, then the payoff they receive is allied with the notion that *ex post* those players deemed it worthwhile going. They receive a payoff strictly greater than the payoff they would have received had they stayed at home. On the other hand, if C or more players choose to go to the bar, then the payoff the bar entrants receive is allied with the notion that, *ex post,* those players did not deem it worthwhile going to the bar. In other words they receive a payoff strictly less than the payoff they would have received had they stayed at home.

State

		Uncrowded	Crowded
Player *i*	Go to the Bar	*G*	*B*
	Stay at Home	*S*	*S*

where ***G* > *S* > *B***

Figure 2: State Dependent Payoff for Player i in the El Farol stage game.

The payoff function for each player i consists of an unconditional payoff for staying at home, denoted by S, and a conditional payoff, denoted by G or B, dependent on the state of the bar. There are two states of the bar, crowded or not crowded, and the state is determined by the remaining $N-1$ players. To ensure the strategic form of the game, the payoffs must be strictly ordered such that $G > S > B$. The payoff structure for representative player i for an isolated Thursday in the El Farol bar problem can be represented by the following payoff matrix (see Figure 2).

Given the above preliminaries, we can now define the El Farol stage game as a single-stage market entry game with discontinuous, but weakly monotonic, payoffs in other players' actions.

Definition 1 *Define the El Farol stage game as the one shot strategic game* $\Gamma =< N, \Delta, \pi^i >$ *consisting of,*

- *N players indexed by $i \in \{1, 2, , ..., N\}$,*

- *a finite set of actions $\Delta = \{0, 1\}$ indexed by δ, where $\delta^i = 1$ denotes player i's action 'go to the bar' and $\delta^i = 0$ denotes player i's action 'stay at home' and*[14]

- *a payoff function $\pi_i : \delta^i \times \delta^{-i} \to \mathbb{R} = \{S, B, G\}$, such that $G > S > B$, where $\delta^{-i} = \prod_{i \neq j} \delta^{-j}$ defines the state of the bar.*

Formally we can write the payoff function as,

$$\pi^i \left(\delta^i \right) = \begin{cases} G & \text{if } \delta^i = 1 \quad \text{and } \sum_{j \neq i} \delta^j < C \\ B & \text{if } \delta^i = 1 \quad \text{and } \sum_{j \neq i} \delta^j \geq C \\ S & \text{if } \delta^i = 0 \end{cases}$$

where $C \in \mathbb{Z}$.

3.1.1 Nash Equilibria in the El Farol Stage Game

Let us now characterise the equilibria of the El Farol stage game. The first thing to note is that the number of Nash equilibria in the El Farol stage game is large and rises quickly as N increases. Furthermore, the number of Nash equilibria is maximised for any given N when $C \approx N/2$. There are essentially three types of Nash equilibria, namely:

- **Pure Strategy Nash Equilibria**

  Nash equilibria where all players play a pure strategy.

- **Symmetric Mixed Strategy Nash Equilibria**

  Nash equilibria where all players play a mixed strategy.

- **Asymmetric Mixed Strategy Nash Equilibria**

  Nash equilibria where some players play a pure strategy and the remaining play a mixed strategy.

Let $\bar{Y}$ denote the set of Nash equilibria of the El Farol stage game. It can be shown that $\bar{Y}$ contains a finite number of elements. In Proposition 1 we state the number of pure strategy Nash equilibria, denoted $\bar{Y}_P$. Next, we show via Propositions 2 and 3 that there exists a unique symmetric mixed strategy Nash equilibrium, denoted $\bar{Y}_S$. And finally in Proposition 4, we show that the number of asymmetric mixed strategy Nash equilibria, denoted $\bar{Y}_A$, is countable. Therefore, the number of Nash equilibria in the El Farol stage game is finite.[15]

Proposition 1 *The number of pure strategy Nash equilibria in the El Farol stage game with $N \in \mathbb{N}$ players and a capacity of $C \in \mathbb{N}$ is,*

$$\begin{pmatrix} N \\ C \end{pmatrix} = \frac{N!}{C! \, (N - C)!} \tag{2}$$

---

[14]It should be noted that although we employ the notation Δ to denote the set of only two actions available to each player, we do so only to indicate how the reinforcement learning model would be extended to games with more than two distinct actions.

[15]Note that $\bar{Y} = \bar{Y}_P \cup \bar{Y}_S \cup \bar{Y}_A$.

Proof See Section A.1 in Appendix A. ∎

The following two propositions together demonstrate that a symmetric mixed strategy Nash equilibrium exists and is unique. In Proposition 2 we prove that there is a symmetric mixed strategy Nash equilibrium where all players play the same mixed strategy and that it is unique. In Proposition 3 we then prove that if all players are playing a mixed strategy they must be playing the same mixed strategy. Therefore, we have a unique symmetric mixed strategy Nash equilibrium in the El Farol stage game.[16]

Proposition 2 *In the El Farol stage game there is a symmetric mixed strategy equilibrium where all players play the same mixed strategy defined by the strategy tuple $(\alpha, [1 - \alpha])$, where α denotes the probability of going to the bar and $[1 - \alpha]$ denotes the probability of staying at home. Furthermore, α is uniquely defined by the following relationship:*

$$\left(\frac{S - B}{G - B} \right) = \sum_{m=0}^{C-1} \left(\begin{array}{c} N - 1 \\ m \end{array} \right) \alpha^m [1 - \alpha]^{N-1-m} \tag{3}$$

Proof See Section A.2 in Appendix A. ∎

Proposition 3 *In a Nash equilibria in the El Farol stage game where all players employ a mixed strategy, all agents must play the same mixed strategy.*

Proof See section A.3 in Appendix A. ∎

Let us now consider the asymmetric mixed strategy Nash equilibria. Given that we can calculate the number of pure strategy Nash equilibria from (2) and that there is a unique symmetric mixed strategy Nash equilibrium, an approach can be tabled to demonstrate that the number of asymmetric mixed strategy Nash equilibria is finite.

Proposition 4 *The number of asymmetric mixed strategy Nash equilibria in the El Farol stage game is countable.*

Proof See Section A.4 in Appendix A. ∎

We have now characterised the Nash equilibria of the El Farol stage game. Furthermore, we have shown that the number of Nash equilibria is finite. This finding will be employed later in proving our main result.

3.2 The El Farol Game

For completeness we define the El Farol bar problem as the repeated El Farol stage game with boundedly rational agents who learn in accordance with the Erev and Roth (1998) reinforcement learning model. Let us begin by defining the El Farol game.

Definition 2 *The El Farol game is the infinitely repeated El Farol stage game.*

---

[16] A similar result has been proved by Cheng (1997).

3.3 Erev and Roth (1998) Reinforcement Learning

We now set out the procedure for the Erev and Roth (1998) reinforcement learning model in detail. In this learning model, each player i has a propensity to undertake each action in each period, denoted $q_t^i(\delta)$. The timeline of the learning procedure is that in each period t each player i chooses to undertake one of their available actions $\delta \in \Delta = \{0, 1\}$ in accordance with a mapping from the propensities to the unit interval $[0, 1]$. This mapping is defined by the choice rule. The player i then undertakes the action dictated by the choice rule and receives a payoff in that period associated with that action. Player i then updates his propensities. The updating procedure is determined by the updating rule. In the Erev and Roth (1998) reinforcement learning model, the only propensities to be updated are those corresponding to the actual action taken. We can now define the model formally. The learning procedure comprises of three components: the initial conditions, a choice rule and an updating rule.

3.3.1 Initial Conditions

Let $q_t^i(\delta)$ be player i's propensity to play action $\delta \in \Delta$ in period t. In the initial period, $t = 0$, we assume that all players have positive propensities for all possible actions. That is,

$$q_t^i(\delta) > 0 \text{ for } t = 0 \text{ and for all } i \in N \text{ and } \delta \in \Delta \tag{4}$$

This assumption, along with positive payoffs, will also ensure that $q_t^i(\delta) > 0$ for all t and $\delta \in \Delta$.

3.3.2 Choice Rule

Each player i has positive a propensity, $q_t^i(\delta)$, to take action $\delta \in \Delta = \{0, 1\}$ in period t. In models of reinforcement learning, the choice rule provides a mapping from propensities to strategies. Let $\left(y_t^i, \left[1 - y_t^i\right]\right)$ represent player i's mixed strategy in period t with two possible actions $\delta \in \Delta = \{0, 1\}$, where y_t^i is the probability placed by agent i on action $\delta = 1$ in period t and $\left[1 - y_t^i\right]$ is the probability placed by agent i on action $\delta = 0$ in period t. The choice rule employed in the Erev and Roth (1998) reinforcement learning model is often referred to as the simple choice rule. It is a straightforward probability mapping from propensities to the unit interval $[0, 1]$. That is,

$$\Pr(\delta = 1) = y_t^i = \frac{q_t^i(1)}{\sum_{\delta \in \Delta} q_t^i(\delta)} = \frac{q_t^i(1)}{Q_t^i} \tag{5}$$

where $Q_t^i = \sum_{\delta \in \Delta} q_t^i(\delta)$.[17]

3.3.3 Updating Rule

Let $\bar{\sigma}^i\left(\delta_t^i, m_t^{-i}\right)$ denote the realised increment to player i's propensity in period t from taking action $\delta \in \Delta = \{0, 1\}$ given the aggregate actions taken by the

---

[17]Note that since there are only two possible actions for each player i we can write

$$\Pr(\delta = 0) = \left(1 - y_t^i\right) = \frac{q_t^i(0)}{\sum_{\delta \in \Delta} q_t^i(\delta)} = \frac{q_t^i(0)}{Q_t^i}$$

remaining $N-1$, denoted by m_t^{-i} where $m_t^{-i} = \sum_{j\neq i} \delta_t^j$. To complete, and most crucial to, our reinforcement learning model, we must state the means by which players update their propensities. Specifically, in the Erev and Roth (1998) reinforcement learning model, it takes the form that if agent i takes action δ in period t, then the agent's δth propensity is increased by an increment equal to agent i's realised payoff in that period. All other propensities remain unchanged. In other words only realised payoffs act as reinforcers. We thus have the following updating rule,[18]

$$q_{t+1}^i\left(\delta\right) = q_t^i\left(\delta\right) + \bar{\sigma}^i\left(\delta_t^i, m_t^{-i}\right) \text{ for all } \delta \in \Delta = \{0,1\} \tag{6}$$

3.4 Reinforcement Learning in the El Farol Game

We will now model the El Farol bar problem as the El Farol game with boundedly rational agents who learn according to the Erev and Roth (1998) reinforcement learning model. To study the long-run dynamics of the El Farol game with bounded rational agents learning in accordance with the Erev and Roth (1998) reinforcement model, we need to first write the expected motion of the ith player's $\delta = 1$ strategy adjustment. In order to accomplish this task, we must first define player i's expected payoff increment.

Let $\hat{\sigma}^i\left(\delta_t^i, y_t^{-i}\right)$ denote the expected increment to player i's propensity in period t from taking action δ given the aggregate actions taken by the remaining $N-1$ players, denoted by y_t^{-i}, where y_t^{-i} is a vector strategy profile. Note that the updating rule in the Erev and Roth (1998) reinforcement learning model is a function of realised payoffs. However, the expected motion of the ith player's $\delta = 1$ strategy adjustment will be a function of expected payoff increments. This is quantitatively and qualitatively different from realised payoff increments.

3.4.1 Expected Strategy Adjustment in the El Farol Game

To obtain analytical results from the application of Erev and Roth (1998) reinforcement learning model to the El Farol game, we make use of results from the theory of stochastic approximation. In essence we investigate the behaviour of the stochastic learning model by evaluating its expected motion as $t \longrightarrow \infty$. In the case of the Erev and Roth (1998) learning model defined by the choice rule (5) and updating rule (6), we can write down the expected motion of the ith player's $\delta = 1$ strategy adjustment through the following proposition:

Proposition 5 *Given the choice rule (5) and the updating rule (6), the expected motion of the ith player's $\delta = 1$ strategy adjustment in the repeated El Farol game is:*

$$E\left[y_{t+1}^i | y_t^i\right] - y_t^i = \frac{1}{Q_t^i} y_t^i \left[1 - y_t^i\right] \left[\hat{\sigma}^i\left(1, y_t^{-i}\right) - \hat{\sigma}^i\left(0, y_t^{-i}\right)\right]$$
$$+ O\left(\frac{1}{\left[Q_t^i\right]^2}\right) \tag{7}$$

Proof See Section B.1 in Appendix B. ∎

---

[18]Note that this updating rule reveals why in this model of reinforcement learning all payoffs must be positive. Otherwise, there would be a possibility of propensities becoming negative and thus leading to choice probabilities that are undefined.

4   Long-run Behaviour in the El Farol Game

We now arrive at our main result. We consider the behaviour of the expected motion of the players' $\delta = 1$ strategy adjustment as $t \longrightarrow \infty$. We begin by stating the main result.

Theorem 1 (Main Result) *If agents in the repeated El Farol game as defined employ the choice rule (5) and reinforcement updating rule (6) for all of $N \in \mathbb{N}$ and $C \in \mathbb{N}$ such that $C \leq N - 1$ and payoffs such that $G > S > B \geq 0$, with probability one the Erev and Roth (1998) reinforcement learning process converges to a pure Nash equilibrium of the one-shot El Farol game. That is,*

$$\Pr\left\{\lim_{t \longrightarrow \infty} y_t \in \bar{Y}_P\right\} = 1,$$

where $y_t = \left\{y_t^1, y_t^2, ..., y_t^N\right\}$, $y_t \in Y$, is a strategy profile for the N agents and $\bar{Y}_P$ is the set of pure Nash equilibrium profiles.

Now we prove the main result, Theorem 1. In the El Farol game with identical boundedly rational agents, learning according to the Erev and Roth (1998) reinforcement learning model, long-run behaviour converges asymptotically to the set of pure strategy Nash equilibria of the El Farol stage game. This result is established by studying the convergent behaviour of the discrete time stochastic process (7) describing the expected strategy adjustment of player i's action of going to the bar. In essence we wish to investigate the limit of this process as $t \to \infty$.

We accomplish this task in two main stages: a positive convergence statement and a negative one. Drawing these two results together we prove our main result. Each stage employs results from the literature on stochastic approximation. First, a result of Benaïm (1999, Corollary 6.6) is employed to demonstrate that the stochastic process will, in the limit as $t \to \infty$, converge asymptotically to one of the fixed points of the adjusted replicator dynamics. Second, two results of Hopkins and Posch (2005, Proposition 2 and 3) are utilised to demonstrate that the stochastic process describing the expected strategy adjustment of player i's action of going to the bar will *not* converge asymptotically to any fixed points that do *not* correspond to a Nash equilibria of the El Farol stage game or to any corresponding Nash equilibria that are unstable under the adjusted replicator dynamics. These two stages combined will imply that the discrete time stochastic process describing the expected strategy adjustment of player i's action of going to the bar converges asymptotically to the set of pure strategy equilibria of the El Farol stage game.

4.1   Proof of Main Result: First Stage

In the first stage of the proof, we show that the discrete time stochastic process (7) converges with probability one to one of the fixed points of the standard replicator dynamics.

Consider for a moment the behaviour of the following stochastic process (Benveniste, Métivier, and Priouret 1990):

$$x_{t+1} - x_t = \gamma_t f(x_t) + \gamma_t \eta_t(x_t) + O\left([\gamma_t]^2\right) \tag{8}$$

where x_t lies in $\mathbb{R}^N$, $E\left[\eta_t\left(x_t\right)|x_t\right] = 0$ and γ_t defines the nature of the gain in this adaptive process. For our purposes γ_t is interpreted as the step size of the learning algorithm. In our analysis we wish to study the generic convergence properties of stochastic processes of this form as $t \longrightarrow \infty$.

It turns out that the nature of the gain is important in determining what inferences can be made about the behaviour of (8) in the limit. In fact the stronger results from the theory of stochastic approximation apply to adaptive algorithms with decreasing gain, that is stochastic processes with decreasing step size.

Definition 3 *The stochastic process (8) is said to have decreasing gain if*

$$\sum_t \left(\gamma_t\right)^\alpha < \infty \text{ for some } \alpha > 1 \text{ where } \sum_t \gamma_t = +\infty$$

For example a common step size of $\gamma_t = 1/t$ would ensure that (8) has decreasing gain. It emerges that as $t \longrightarrow \infty$ there is a close relationship between the behaviour of stochastic processes (8) with a decreasing gain and the mean or averaged ordinary differential equation of the stochastic process.

$$\dot{x} = f\left(x\right) \tag{9}$$

In particular it can be shown via Benaïm (1999, Corollary 6.6) that if (9) meets certain criteria, the stochastic process (8) must converge with probability one to one of the fixed points of the mean or averaged ordinary differential equation (9).

Theorem 2 (Benaïm (1999, Corollary 6.6)) *If the dynamic process (9) admits a strict Lyapunov function and processes a finite number of fixed points, then with probability one the stochastic process (8) converges to one of these fixed points.*

We now have a method of illustrating that the long-run behaviour of boundedly rational agents, adjusting their strategies according to the Erev and Roth (1998) reinforcement learning model, in the El Farol game converges to one of the fixed points of mean or averaged differential equation (9) associated with the vector of player's expected strategy adjustments.

In order to apply this general result, we must first identify the mean or averaged differential system associated with players' expected strategy adjustment. Furthermore, it must be shown that the mean or averaged differential system admits a strict Lyapunov function. And finally, we must establish that the mean or averaged differential system possesses a finite number of isolated fixed points. In the next three subsections we purport to demonstrate just that.

4.1.1 The Joint Dynamic System

One might hope that the standard replicator dynamics represent the mean or averaged differential system derived from the discrete time stochastic process (7).

$$\dot{y}^i = y^i\left[1 - y^i\right]\left[\hat{\sigma}^i\left(1, y^{-i}\right) - \hat{\sigma}^i\left(0, y^{-i}\right)\right] \tag{10}$$

Unfortunately, the standard replicator dynamics (10) do not for two simple reasons. First, in the Erev and Roth (1998) model, the step size is endogenous; that is, it is determined by the accumulation of payoffs, and thus is not exogenously fixed. Second, the step size is not a scalar.

In order to account for these discrepancies, let us introduce a common step size of $\gamma_t = 1/t$ and N new variables μ_t^i, such that:

$$\mu_t^i = \frac{t}{Q_t^i}$$

We can now substitute $\gamma_t \mu_t^i$ for $1/Q_t^i$ in our discrete time stochastic process (7) and arrive at the following *corrected* expected motion of the ith player's strategy adjustment of going to the bar:

$$E\left[y_{t+1}^i | y_t^i\right] - y_t^i = \gamma_t \mu_t^i y_t^i \left[1 - y_t^i\right] \left[\hat{\sigma}^i \left(1, y_t^{-i}\right) - \hat{\sigma}^i \left(0, y_t^{-i}\right)\right]$$
$$+ \gamma_t \xi_t \left(y_t^i\right) + O\left(\left[\gamma_t\right]^2\right) \tag{11}$$

Since we have assumed that all payoffs in the El Farol game are positive to ensure that choice probabilities are well defined, it follows that μ_t^i is bounded away from zero. Furthermore, since $\mu_t^i = t/Q_t^i$ equals the inverse of the average payoff in the limit as $t \to \infty$, it follows that the associated mean or averaged differential equation (9) associated with the corrected discrete time stochastic process (11) is:

$$\dot{y}^i = \mu^i y^i \left[1 - y^i\right] \left[\hat{\sigma}^i \left(1, y^{-i}\right) - \hat{\sigma}^i \left(0, y^{-i}\right)\right] \tag{12}$$

In equilibrium this amounts to the standard definition of the adjusted replicator dynamics. This is extremely useful because there are many results in the literature on the equilibrium behaviour of the adjusted replicator dynamics (see Fudenberg and Levine 1998, Hopkins 2002). We shall revisit some of these findings later in this proof of the main result.

Because each μ_t^i varies over time, we require a further set of N equations describing the discrete time stochastic process of μ_t^i. Using the method we previously employed to write player i's expected strategy adjustment of going to the bar, we now find the expected change player i's step size.

Lemma 1 *Given the choice rule (5) and the updating rule (6), the expected motion of the ith player's step size in the El Farol game is:*

$$E\left[\mu_{t+1}^i | \mu_t^i\right] - \mu_t^i = \gamma_t \mu_t^i - \gamma_t \left[\mu_t^i\right]^2 \hat{\sigma}^i \left(0, y_t^{-i}\right)$$
$$+ \gamma_t \left[\mu_t^i\right]^2 y_t^i \left[\hat{\sigma}^i \left(0, y_t^{-i}\right) - \hat{\sigma}^i \left(1, y_t^{-i}\right)\right]$$
$$+ \gamma_t \xi_t \left(y_t^i\right) + O\left(\left[\gamma_t\right]^2\right) \tag{13}$$

Proof First, imagine that player i chooses to attend the bar in period t. The expected change in the player step size can be written as:

$$\mu_{t+1}^i - \mu_t^i(t) = \frac{t+1}{Q_t^i + \hat{\sigma}^i \left(1, y_t^{-i}\right)} - \frac{t}{Q_t^i}$$
$$= 1 - \mu_t^i \hat{\sigma}^i \left(1, y_t^{-i}\right) + O\left(\gamma_t\right)$$

17

Now consider the expected change in step size if player i stays at home.

$$\mu_{t+1}^i - \mu_t^i(t) = \frac{t+1}{Q_t^i + \hat{\sigma}^i\left(0, y_t^{-i}\right)} - \frac{t}{Q_t^i}$$
$$= 1 - \mu_t^i \hat{\sigma}^i\left(0, y_t^{-i}\right) + O\left(\gamma_t\right)$$

The expected motion of each player i's step size given μ_t^i can now be written as the expected motion in the step size given y_t^i times the step size in period t.

$$E\left[\mu_{t+1}^i | \mu_t^i(t)\right] - \mu_t^i(t) = \gamma_t \mu_t^i y_t^i \left[1 - \mu_t^i \hat{\sigma}^i\left(1, y_t^{-i}\right) + O\left(\gamma_t\right)\right]$$
$$+ \gamma_t \mu_t^i \left[1 - y_t^i\right]\left[1 - \mu_t^i \hat{\sigma}^i\left(0, y_t^{-i}\right) + O\left(\gamma_t\right)\right]$$

and after some more algebraic manipulation we arrive at (13). ∎

The mean or averaged differential equation derived from the discrete time stochastic process (7) has now been corrected for the endogenous and non-scalar step size. Therefore, we have the following mean or averaged differential system consisting of $2N$ differential equations with $2N$ endogenous variables:

$$\dot{y}^i = \mu^i y^i \left[1 - y^i\right]\left[\hat{\sigma}^i\left(1, y^{-i}\right) - \hat{\sigma}^i\left(0, y^{-i}\right)\right] \tag{14a}$$
$$\dot{\mu}^i = \mu^i \left[1 - \mu^i \left[\hat{\sigma}^i\left(0, y^{-i}\right) + y^i \left[\hat{\sigma}^i\left(0, y^{-i}\right) - \hat{\sigma}^i\left(1, y^{-i}\right)\right]\right]\right] \tag{14b}$$

Let us refer to this as the joint dynamic system describing the evolution of player i's strategy adjustment of going to the bar in the El Farol game.

4.1.2 Admission of a Strict Lyapunov Function

We must show that the associated mean or averaged ordinary differential system, the joint dynamic system (14), admits a strict Lyapunov function. Let us begin with some definitions.

Definition 4 *Let (9) be an ordinary differential equation defined on some subset Y of $\mathbb{R}^N$. Let $V : Y \to \mathbb{R}$ be a continuously differentiable function. Furthermore, let $\bar{y}$ be a fixed point of $V(y)$. $V(y)$ is a Lyapunov function if,*

$$\dot{V}(y) \geq 0, \qquad \forall \quad y \in Y \ and \tag{15a}$$
$$\dot{V}(\bar{y}) = 0, \qquad \forall \quad \bar{y} \in \theta \tag{15b}$$

where θ is the set of fixed points of (9).

Definition 5 *A strict Lyapunov function is a Lyapunov function $V(y)$ such that:*

$$\dot{V}(y) > 0, \qquad \forall \quad y \notin \theta \tag{16}$$

In general it can be difficult and time consuming to identify a suitable Lyapunov function for a particular system. It is often a process of trial and error. An approach to this aspect of the problem developed in the existing literature on the convergence of learning models in games (see Duffy and Hopkins 2005)

has been to explicitly derive a suitable function for $V(y)$ and then show that it admits a strict Lyapunov function. In theory, but not always in practice, this can be accomplished by first assuming that $V(y)$ indeed admits a strict Lyapunov function. If this is the case, then the partial derivative $\partial V(y)/\partial y^i$ represents the expected payoff increment to player i from going to the bar.

$$\frac{\partial V(y)}{\partial y^i} = \hat{\sigma}^i\left(1, y^{-i}\right) - \hat{\sigma}^i\left(0, y^{-i}\right) \tag{17}$$

It should then just be a question of integrating $\partial V(y)/\partial y^i$ with respect to y^i in order to find a suitable $V(y)$ and checking that both conditions (15) and (16) defining strict Lyapunov functions are met.

The difficulty with this approach is in explicitly finding a function $V(y)$. Expressing $\hat{\sigma}^i\left(1, y^{-i}\right) - \hat{\sigma}^i\left(0, y^{-i}\right)$ in a compact form is not as straightforward as one might first hope. This can be demonstrated by examining $\partial V(y)/\partial y^i$ further. Note that (17) can be expressed as:

$$\begin{aligned}
\frac{\partial V(y)}{\partial y^i} &= E\left[\pi^i | \delta = 1\right] - E\left[\pi^i | \delta = 0\right] \\
&= [B - S] + \phi[G - B]
\end{aligned}$$

where

$$\phi = \sum_{j=0}^{C-1} \Pr\left(m_t^{-i} = j\right) \tag{18}$$

ϕ is the probability that $C-1$ players or less of the remaining $N-1$ players choose to go to the El Farol bar. It is writing out this latter probability expression (18) for ϕ that is unfortunately problematic and can get cumbersome very quickly. Therefore, this turns out to be an intractable method of demonstrating that the joint dynamic system (14) admits a strict Lyapunov function.

An alternative approach is to employ a result of Monderer and Shapley (1996, Theorem 3.1) from the theory of potential games to demonstrate that the joint system (14) admits a strict Lyapunov function. The argument is as follows: the El Farol game is a congestion game therefore it is a potential game and thus admits a potential function. The properties of potential functions are similar to those of strict Lyapunov functions and therefore, it follows that the joint dynamic system (14) admits a strict Lyapunov function.

Let us now begin with some definitions and a restating of Monderer and Shapley (1996, Theorem 3.1).

Definition 6 *Let* $\Gamma\left(N, Y^i, \pi^i\right)$ *be a game in strategic form.* Γ *is called a potential game if it admits a potential function.*

Definition 7 *A function* $P : Y \rightarrow \mathbb{R}$ *is a potential function for* Γ*, if for every* $i \in N$ *and for every* $y^{-i} \in Y^{-i}$

$$\pi^i\left(x, y^{-i}\right) - \pi^i\left(x', y^{-i}\right) = P\left(x, y^{-i}\right) - P\left(x', y^{-i}\right) \qquad \forall \quad x, x' \in Y^i$$

Theorem 3 (Monderer and Shapley (1996, Theorem 3.1)) *Every congestion game is a potential game.*

19

Now we can show that the joint dynamic system (14) admits a strict Lyapunov function.

Lemma 2 *The joint dynamic system (14) admits a strict Lyapunov function.*

Proof The El Farol stage game is a congestion game and therefore by Theorem 3, Monderer and Shapley (1996, Theorem 3.1), it is a potential game. Thus, there exists a function $P : \delta^i \times \delta^{-i} \to \mathbb{R}$ for every $i \in N$ and for every $\delta^{-i} \in \Delta^{-i}$ such that:

$$\pi^i \left(1, \delta^{-i}\right) - \pi^i \left(0, \delta^{-i}\right) = P\left(1, \delta^{-i}\right) - P\left(0, \delta^{-i}\right) \qquad \forall \quad \delta \in \Delta = \{0, 1\}$$

Given that there is a continuous set of mixed strategies, we can write the potential function $P(y)$ as a smooth function with respect to the strategy space $y \in [0, 1]^N$. $P(y)$ is therefore continuously differentiable. Therefore, for every $i \in N$ and for every $x^{-i} \in [0, 1]^{N-1}$,

$$\pi^i \left(x, y^{-i}\right) - \pi^i \left(x', y^{-i}\right) = P\left(x, y^{-i}\right) - P\left(x', y^{-i}\right) \qquad \forall \quad x, x' \in [0, 1]$$

Now choose x and x' equal to 0 and 1 respectively and take expectations of both sides. It follows that for every $i \in N$ and for every $y^{-i} \in [0, 1]^{N-1}$,

$$\hat{\sigma}^i \left(1, y^{-i}\right) - \hat{\sigma}^i \left(0, y^{-i}\right) = P\left(1, y^{-i}\right) - P\left(0, y^{-i}\right)$$

Or otherwise stated,

$$\frac{\partial P(y)}{\partial y^i} = \hat{\sigma}^i \left(1, y^{-i}\right) - \hat{\sigma}^i \left(0, y^{-i}\right) \tag{19}$$

Furthermore,

$$\begin{aligned}
\dot{P}(y) &= \frac{dP(y)}{dy^i} \dot{y}^i \\
&= \hat{\sigma}^i \left(1, y^{-i}\right) - \hat{\sigma}^i \left(0, y^{-i}\right) \dot{y}^i \\
&= \mu^i y^i \left[1 - y^i\right] \left[\hat{\sigma}^i \left(1, y^{-i}\right) - \hat{\sigma}^i \left(0, y^{-i}\right)\right]^2 \geq 0
\end{aligned}$$

By assumption, $\mu^i > 0$ and $y^i \in [0, 1]$. Since $\left[\hat{\sigma}^i \left(1, y^{-i}\right) - \hat{\sigma}^i \left(0, y^{-i}\right)\right]^2 \geq 0$ we have that $\dot{P}(y)$ is non negative. Additionally, at any fixed point $\bar{y} \in \theta$ either $\bar{y}^i = 0$, $\left(1 - \bar{y}^i\right) = 0$ or $\left[\hat{\sigma}^i \left(1, y^{-i}\right) - \hat{\sigma}^i \left(0, y^{-i}\right)\right] = 0$. Thus $P(y)$ admits a Lyapunov function.
At any $y \notin \theta$, $\dot{y}^i \neq 0$. It should be obvious that:

$$\dot{P}(\bar{y}) = \mu^i y^i \left(1 - y^i\right) \left[\hat{\sigma}^i \left(1, y^{-i}\right) - \hat{\sigma}^i \left(0, y^{-i}\right)\right]^2 > 0.$$

Therefore, $P(y)$ admits a strict Lyapunov function. It follows that the joint dynamic system (14) admits a strict Lyapunov function. ∎

4.1.3 Fixed Points of the Joint Dynamic System

Definition 8 *The fixed points of the joint dynamic system (14) are defined as $\bar{x} = (\bar{y}, \bar{\mu})$ such that $\dot{y} = 0$ and $\dot{\mu} = 0$.*

Lemma 3 *The joint dynamic system (14) possesses a finite number of isolated fixed points.*

Proof Consider the joint dynamic system (14). The fixed points of the N equations describing the evolution of the step size occur when either:

$$\bar{\mu}^i = 0, \frac{1}{\left[\hat{\sigma}^i\left(0, y_t^{-i}\right) + y^i\left[\hat{\sigma}^i\left(0, y_t^{-i}\right) - \hat{\sigma}^i\left(1, y_t^{-i}\right)\right]\right]}$$

By assumption, all payoffs are positive therefore $\bar{\mu}^i$ is bounded away from zero. This means that the fixed points of the joint dynamic system (14) with $\bar{\mu}^i = 0$ are always unstable (see Hopkins 2002, Duffy and Hopkins 2005) and therefore are never asymptotic outcomes. We can thus concentrate on the latter case.

Consider the first N equations of the joint dynamic system (14). Once we substitute for $\bar{\mu}^i$ and multiply both sides by the denominator we have:

$$y^i\left[1 - y^i\right]\left[\hat{\sigma}^i\left(1, y_t^{-i}\right) - \hat{\sigma}^i\left(0, y_t^{-i}\right)\right] = 0$$

In other words the fixed points of the joint dynamic system (14) are exactly the same as those under the adjusted replicator dynamics (12) and, consequently, the standard replicator dynamics (10). The characterisation of the fixed point of the standard replicator dynamics (10) is well known (see Weibull 1995) and consists of the union of all pure states and Nash equilibria of the underlying game.

The number of pure states is obviously finite and, as proved in Propositions 2-4, the number of Nash equilibria in the underlying El Farol game is countable. Therefore, the joint dynamic system (14) possesses a finite number of fixed points points. ■

Just to be absolutely clear, the fixed points of the joint dynamic system (14) consist of the following:

- **Pure strategy Nash equilibria**

  These are the pure states of the joint dynamic system (14) that correspond to the pure strategy Nash equilibria of the underlying game.

- **Symmetric mixed strategy Nash equilibrium**

  This is the full interior state of the joint dynamic system (14) that corresponds to the symmetric mixed strategy Nash equilibria of the underlying game. That is, the Nash equilibrium where all players play a unique mixed strategy best response.

- **Asymmetric mixed strategy Nash equilibria**

  These are boundary states of the joint dynamic system (14) that correspond to asymmetric mixed strategy Nash equilibria of the underlying game. By boundary states we mean those where a subset of the N players play a unique mixed strategy best response while the remainder play a pure strategy.

- **Fixed points that are not Nash equilibria**

  Not all fixed points of the joint dynamic system (14) correspond to Nash equilibria of the underlying game. There are pure states of the

joint dynamic system (14) that do not correspond to pure strategy Nash equilibria of the underlying game. Note that it is not possible to have interior fixed points or fixed points on some boundary of the state space of the joint dynamic system (14) that do not correspond to Nash equilibria of the underlying game.

4.1.4 Positive Convergence Result

Proposition 6 *The discrete time stochastic process (7) converges with probability one to one of the fixed points of the standard replicator dynamics (10).*

Proof By Lemma 2 the joint dynamic system (14) admits a strict Lyapunov function. By Lemma 3 the joint dynamic system (14) possesses a finite number of fixed points which are identical to those of the standard replicator dynamics (10). Therefore, by Theorem 2, Benaïm (1999, Corollary 6.6), the discrete time stochastic process (7) converges to one of the fixed points of the standard replicator dynamics (10). ∎

4.2 Proof of Main Result: Second Stage

In the second part of the proof of the main result, we show that the discrete time stochastic process (7) does not converge to any equilibria corresponding to Nash equilibria of the underlying game which are unstable under the adjusted replicator dynamics (12) or equilibria that do not corresponding to a Nash of the underlying game. We tackle this in two steps.

First, we show that the stability properties of a fixed point of the joint dynamic system (14) are entirely determined by the stability properties of the corresponding fixed point under the adjusted replicator dynamics (12). We then determine the stability properties of the Nash equilibria under the adjusted replicator dynamics (12). We conclude that only the pure strategy Nash equilibria are stable under the adjusted replicator dynamics (12). Finally, we employ Hopkins and Posch (2005, Proposition 2) to show that the discrete time stochastic process (7) cannot converge to any fixed points unstable under the adjusted replicator dynamics (12).

Second, we employ Hopkins and Posch (2005, Proposition 3) to demonstrate that the discrete time stochastic process (7) cannot converge to any fixed point not corresponding to a Nash equilibria under the underlying game. Therefore, we have our negative convergence result.

4.2.1 Unstable Equilibria in the Adjusted Replicator Dynamics

Definition 9 *A fixed point $\bar{x} = (\bar{y}, \bar{\mu})$ of the joint dynamic system (14) is unstable if its linearisation evaluated at $\bar{x}$ has at least one eigenvalue with a positive real part.*

Theorem 4 (Hopkins and Posch (2005, Proposition 2)) *Let $\bar{x}$ be a Nash equilibrium that is linearly unstable under the adjusted replicator dynamics (12). Then the Erev and Roth (1998) reinforcement learning process defined by the choice rule (5) and the updating rule (6) asymptotically converges to one of these points with probability zero.*

Lemma 4 *The stability properties of the fixed points of the joint dynamic system (14) are entirely determined by the stability properties of the corresponding fixed points of the adjusted replicator dynamics (12).*

Proof The linearisation of the joint dynamic system (14) at any fixed point, $\bar{x}$, will be of the form:

$$\begin{pmatrix} \dfrac{\partial \dot{y}}{\partial y} & \dfrac{\partial \dot{y}}{\partial \mu} \\ \dfrac{\partial \dot{\mu}}{\partial y} & \dfrac{\partial \dot{\mu}}{\partial \mu} \end{pmatrix} \tag{20}$$

Consider the partitions of the above matrix (20) evaluated at a fixed point of the joint dynamic systems (14) in turn. $d\dot{y}/d\mu$ is obviously the null matrix.

$$\frac{\partial \dot{y}^i}{\partial \mu^j} = 0 \text{ for all } i, j \tag{21}$$

Given (21), every eigenvalue of the matrix (20) is an eigenvalue for either $d\dot{y}/dy$ or $d\dot{\mu}/d\mu$. The latter matrix is diagonal.

$$\frac{\partial \dot{\mu}^i}{\partial \mu^j} = 0 \text{ for } i \neq j$$

$$\frac{\partial \dot{\mu}^i}{\partial \mu^j} \neq 0 \text{ for } i = j$$

And the diagonal elements are all negative.

$$\frac{\partial \dot{\mu}^i}{\partial \mu^i} = 1 - 2\bar{\mu}^i \left[\hat{\sigma}^i \left(0, y_t^{-i}\right) + \bar{y}^i \left[\hat{\sigma}^i \left(0, y_t^{-i}\right) - \hat{\sigma}^i \left(1, y_t^{-i}\right) \right] \right] < 0$$

Therefore, all the eigenvalues of $d\dot{\mu}/d\mu$ are negative. Now consider the elements of $d\dot{y}/dy$. This is the linearisation, or otherwise referred to as the Jacobian, of the adjusted replicator dynamics (12).

$$J = \begin{pmatrix} \dfrac{\partial \dot{y}^1}{\partial y^1} & \dfrac{\partial \dot{y}^1}{\partial y^2} & \cdots & \dfrac{\partial \dot{y}^1}{\partial y^N} \\ \dfrac{\partial \dot{y}^2}{\partial y^1} & \dfrac{\partial \dot{y}^2}{\partial y^1} & \cdots & \dfrac{\partial \dot{y}^2}{\partial y^N} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial \dot{y}^N}{\partial y^1} & \dfrac{\partial \dot{y}^N}{\partial y^2} & \cdots & \dfrac{\partial \dot{y}^N}{\partial y^N} \end{pmatrix}$$

If the linearisation of the adjusted replicator dynamics (12) has one or more positive eigenvalues, then the fixed point of the joint dynamic system (14) at which the Jacobian is evaluated is unstable for the joint dynamic system (14). Otherwise, the fixed point is asymptotically stable for the joint dynamic system (14). ∎

Now consider the stability properties of the fixed points of the adjusted replicator dynamics (12) that correspond to Nash equilibria in the El Farol game.

Lemma 5 *The fixed points of the adjusted replicator dynamics (12) corresponding to the pure strategy Nash equilibria of the El Farol stage game are asymptotically stable.*

Proof Given that the pure strategy Nash equilibria are strict, they constitute an evolutionary stable strategy of the El Farol stage game. By Weibull (1995), all evolutionary stable strategies are asymptotically stable under the replicator dynamics. ∎

Lemma 6 *The fixed point of the adjusted replicator dynamics (12) corresponding to the symmetric mixed strategy Nash equilibrium of the El Farol stage game is asymptotically unstable.*

Proof The fixed point of the joint dynamic system (14) corresponding to the symmetric mixed strategy Nash equilibria of the El Farol stage game is unique and is a fully mixed equilibrium. Furthermore, at this fully mixed fixed point of the joint dynamic system (14), $\bar{y}^i = \bar{y}^j$. Let us consider the diagonal elements of J:

$$\frac{\partial \dot{y}^i}{\partial y^i} = \mu^i \left[1 - 2y^i\right] \left[\hat{\sigma}^i \left(1, y_t^{-i}\right) - \hat{\sigma}^i \left(0, y_t^{-i}\right)\right]$$
$$= 0 \text{ if } y^i = \bar{y}$$

Since all the diagonal elements of J equal zero, the trace of J is zero. Now consider the off diagonal elements:

$$\frac{\partial \dot{y}^i}{\partial y^j} = y^i \left[1 - y^i\right] \left[\begin{array}{c} \mu^i \left(\dfrac{\partial \left[\hat{\sigma}^i \left(1, y_t^{-i}\right)\right]}{\partial y^j} - \dfrac{\partial \left[\hat{\sigma}^i \left(0, y_t^{-i}\right)\right]}{\partial y^j} \right) \\ + \dfrac{\partial \mu^i}{\partial y^j} \left[\hat{\sigma}_t^i \left(1, y_t^{-i}\right) - \hat{\sigma}_t^i \left(0, y_t^{-i}\right)\right] \end{array} \right]$$

Since all players earn the same payoff in this fully mixed symmetric equilibrium, we have that $\mu^i = \mu^j$ and therefore, J is symmetric. Therefore, J has no complex eigenvalues. With a zero trace, the real eigenvalues sum to zero. Therefore, there must be at least one eigenvalue which is positive. Hence, $\bar{x}$ is linearly unstable with respect to the joint dynamic system (14). ∎

Lemma 7 *The fixed points of the adjusted replicator dynamics (12) corresponding to the asymmetric mixed strategy Nash equilibria of the El Farol stage game are asymptotically unstable.*

Proof At the fixed points of the joint dynamic system (14) corresponding to the asymmetric mixed strategy Nash equilibria, $N - j - k$ players randomise over entry while the remaining $j + k$ players play a pure strategy. One can then calculate the Jacobian, J, evaluated at this fixed point which is of the form:

$$J = \left(\begin{array}{cc} A & B \\ 0 & C \end{array} \right)$$

where A is a $(N - j - k) \times (N - j - k)$ matrix of the form found at the symmetric fixed point as described in Lemma 6.

It is easily verified that C is a diagonal matrix of negative elements. By the same argument as put forward in Lemma 6, A is a mixture of positive and negative eigenvalues. Therefore, J has at least one positive eigenvalue and it follows that the fixed point associated with the asymmetric mixed strategy Nash equilibrium is unstable under the adjusted replicator dynamics (12). ∎

4.2.2 Non-Nash Fixed Points of the Joint Dynamic System

Theorem 5 (Hopkins and Posch (2005, Proposition 3)) *Let $\bar{x}$ be a fixed point of the replicator dynamics (10) which is not a Nash equilibrium. Then the Erev and Roth (1998) reinforcement learning process defined by choice rule (5) and the updating rule (6) asymptotically converges to one of these points with probability zero.*

Therefore, the discrete time stochastic process (7) cannot converge to any fixed point not corresponding to a Nash equilibrium under the underlying game.

4.2.3 Negative Convergence Result

Proposition 7 *The discrete time stochastic process (7) converges with probability zero to equilibria corresponding to Nash equilibria of the underlying game unstable under the adjusted replicator dynamics (12) or equilibria not corresponding to a Nash equilibrium of the underlying game.*

Proof The result follows from Theorem 4, Hopkins and Posch (2005, Proposition 2), and Theorem 5, Hopkins and Posch (2005, Proposition 3). ∎

4.3 Proof of Main Result: Concluding Stage

Proposition 8 *In the El Farol game with identical bounded rational agents learning in accordance with the Erev and Roth (1998) reinforcement learning model, long-run behaviour converges asymptotically to the set of pure strategy Nash equilibria of the El Farol stage game.*

Proof The result follows directly from our positive convergence result, Proposition 6, and our negative convergence result, Proposition 7. ∎

5 Conclusion

The results obtained from modelling the El Farol bar problem as a repeated game with boundedly rational agents implies that people tend to minimise bad experiences and maximise good ones. This is exactly what is assumed by the Erev and Roth (1998) reinforcement learning model.

The application of the Erev and Roth (1998) reinforcement learning model implies that the average attendance converges to the capacity of the El Farol bar as in Arthur's (1994) 'inductive thinking' approach to modelling boundedly rational agents. The difference lies in who, in the long-run, attends the bar. The most salient aspect of this result is that in the El Farol game the population of boundedly rational agents, who behave in accordance with the Erev and Roth (1998) reinforcement learning model, are partitioned into those who always attend the El Farol bar and those who always stay at home. This differs from

the outcome of Arthur's (1994) model where agents are differentiated by the forecasting methods, not by attendance.

The main result implies sorting and is crucially dependent on the fact that the game in question is a potential game. It can be shown that the result is robust when compared to other variants of reinforcement learning. In fact Duffy and Hopkins's (2005) paper shows how this would be the case with stochastic fictitious play. It is also possible to derive some general results for an extension to this treatment where players have idiosyncratic payoff functions. Milcataich (1996) shows that any multi-player coordination game with two identical pure actions for each player admits a potential function and by definition, is a potential game. Therefore, even if one considers players in the El Farol game with heterogenous preferences, it appears that reinforcement learning will lead to sorting.

References

ARTHUR, B. W. (1994): "Complexity in Economic Theory. Inductive Reasoning and Bounded Rationality.," *American Economic Review*, 82(2), 406–411.

BELL, A. M., AND W. A. SETHARES (1999): "The El Farol Problem and the Internet: Congestion and Coordination Failure," Society for Computational Economics: Computing in Economics and Finance Woking Paper, No. 812.

———— (2001): "Avoiding Global Congestion Using Decentralised Adaptive Agents," *IEEE Transactions on Signal Processing*, 49(11), 2873–2879.

BELL, A. M., W. A. SETHARES, AND J. A. BUCKLEW (2003): "Coordination Failure as a Source of Congestion in Information Networks," *IEEE Transactions on Signal Processing*, 51(3), 875–885.

BENAÏM, M. (1999): *Dynamics of Stochastic Approximation Algorithms*, Le Séminaire de Probabilites: Lecture Notes in Mathematics, Vol 1709. Springer-Verlag, New York.

BENVENISTE, A., M. MÉTIVIER, AND P. PRIOURET (1990): *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, London.

CASTI, J. L. (1996): "Seeing the Light at El Farol," *Complexity*, 1(5), 7–10.

CHALLET, D., M. MARSILI, AND G. OTTINO (2004): "Shedding Light on El Farol," *Physica A: Statistical Mechanics and its Applications*, 332(1), 469–482.

CHENG, J. Q. (1997): "The Mixed Strategy Equilibria and Adaptive Dynamics in the Bar Problem," Santa Fe Institute: Santa Fe Working Paper, No. 96-08-063.

DUFFY, J., AND E. HOPKINS (2005): "Learning, Information and Sorting in Market Entry Games: Theory and Evidence," *Games and Economic Behavior*, 51, 31–62.

EREV, I., AND A. E. ROTH (1998): "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, 88(4), 848–881.

FARAGO, J., A. GREENWALD, AND K. HALL (2002): "Fair and Efficient Solutions to the Santa Fe Bar Problem," in *Proceedings of the Grace Hopper Celebration of Women in Computing*.

FRANKE, R. (2003): "Reinforcement Learning in the El Farol Model," *Journal of Economic Behavior and Organization*, 51, 367–388.

FUDENBERG, D., AND D. K. LEVINE (1998): *The Theory of Learning in Games*. MIT Press.

GREENWALD, A., B. MISHRA, AND R. PARIKH (1998): "The Santa Fe Problem Revisited: Theoretical and Practical Implications," in *Proceedings of the Festival on Game Theory: Interactive Dynamics and Learning*, Stony Brook. SUNY.

HOPKINS, E. (2002): "Two Competing Models of How People Learn in Games," *Econometrica*, 70(6), 2141–2166.

HOPKINS, E., AND M. POSCH (2005): "Attainability of Boundary Points under Reinforcement Learning," *Games and Economic Behavior*, 53(1), 110–125.

MILCATAICH, I. (1996): "Congestion Games with Player-Specific Payoff Functions," *Games and Economic Behavior*, 13, 111–124.

MONDERER, D., AND L. S. SHAPLEY (1996): "Potential Games," *Games and Economic Behavior*, 14(1), 124–143.

ROSENTHAL, R. W. (1973): "A Class of Games Possessing Pure-Strategy Nash Equilibria," *International Journal of Game Theory*, 2, 65–67.

SANDHOLM, W. H. (2001): "Potential Games with Continuous Player Sets," *Journal of Economic Theory*, 97, 81–108.

SELTEN, R., AND W. GÜTH (1982): "Equilibrium Point Selection in a Class of Market Entry Games," in *Games, Economic Dynamics and Time Series Analysis*, ed. by M. Deistler, E. Fürst, and G. Schwödiauer, chap. Equilibrium point selection in a class of market entry games. Physica-Verlag, Vienna.

SMITH, J. M. (1974): "The Theory of Games and Animal Conflicts," *Journal of Theoretical Biology*, 47, 209–221.

WEIBULL, J. W. (1995): *Evolutionary Game Theory*. MIT Press, London.

ZAMBRANO, E. (2004): "The Interplay Between Analytics and Computation in the Study of Congestion Externalities: The Case of the El Farol Problem.," *Journal of Public Economic Theory*, 6(2), 375–395.

A First Appendix

A.1 Proof to Proposition 1

Proof The number of pure strategy Nash equilibria in the one-shot El Farol game is the number of ways C different players can be chosen out of the set of N players at a time. ∎

A.2 Proof to Proposition 2

Proof *Existence* Define the binary state of the bar, either uncrowded or crowded, as a binomial distribution, denoted $P^i(N,C,\alpha)$, over the number of players in the game, denoted by N, the capacity of the bar, denoted by C, and the mixed strategies employed by player i, denoted by the probability α of attending and $(1-\alpha)$ of staying at home. In particular in a mixed strategy equilibria, each player i should be indifferent between going to the bar and staying at home.

$$E\left[\pi^i|\delta=1\right] = E\left[\pi^i|\delta=0\right], \qquad \forall \quad i \in \{1,2,...,N\}$$

We have that:

$$\begin{aligned} E\left[\pi^i|\delta=1\right] &= G \cdot P^i(N,C,\alpha) + B \cdot \left[1 - P^i(N,C,\alpha)\right] \\ E\left[\pi^i|\delta=0\right] &= S \end{aligned} \tag{22}$$

where $P^i(N,C,\alpha)$ denotes the probability of the bar being in the uncrowded state or otherwise stated as the probability that less than C other players out of the $(N-1)$ remaining players choose to attend the bar. Given that players are homogeneous in preferences and the El Farol game is symmetric in payoffs, we may write:

$$P^i(N,C,\alpha) = P(N,C,\alpha) \text{ for all } i \in \{1,2,...,N\}.$$

Returning to (22), we can now substitute $P(N,C,\alpha)$ for $P^i(N,C,\alpha)$ and solve for $P(N,C,\alpha)$.

$$P(N,C,\alpha) = \frac{(S-B)}{(G-B)}$$

where $P(N,C,\alpha)$ is defined by the following binomial probability:

$$P(N,C,\alpha) = \sum_{m=0}^{C-1} C_m^{N-1} \alpha^m (1-\alpha)^{(N-1)-m}.$$

Therefore, we have (3). *Uniqueness:* $P(N,C,\alpha)$ is continuous and well defined over the closed interval $[0,1]$. Given that $\lim_{\alpha \to 0} P(N,C,\alpha) = 1$ and $\lim_{\alpha \to 1} P(N,C,\alpha) = 0$, (3) has a unique solution if the partial derivative of $P(N,C,\alpha)$, denoted $P_\alpha(N,C,\alpha)$, is less than zero.

$$\begin{aligned} P_\alpha(N,C,\alpha) &= \frac{\partial}{\partial\alpha}\left(\sum_{m=0}^{C-1} C_m^{N-1}\alpha^m(1-\alpha)^{(N-1)-m}\right) \\ &= \left(\sum_{m=1}^{C-1}(N-1)C_{m-1}^{N-2}\alpha^{m-1}(1-\alpha)^{(N-1)-m}\right) \\ &\quad -\left(\sum_{m=0}^{C-1}(N-1)C_m^{N-2}\alpha^m(1-\alpha)^{((N-1)-m)-1}\right) \end{aligned}$$

Now let $k = (m-1)$.

$$
\begin{aligned}
P_\alpha(N, C, \alpha) &= (N-1)\left(\sum_{k=0}^{C-2} C_k^{N-2}\alpha^k (1-\alpha)^{N-2-k}\right) \\
&\quad - (N-1)\left(\sum_{m=0}^{C-1} C_m^{N-2}\alpha^m (1-\alpha)^{N-2-m}\right) \\
&= (N-1)(-1)C_{m-1}^{N-2}\alpha^{m-1}(1-\alpha)^{N-1-m} \\
&< 0
\end{aligned}
$$

Hence (3) has a unique solution and thus the number of mixed strategy Nash equilibria where all agents employ the same 'mixing' is one. ∎

A.3 Proof to Proposition 3

Proof Here we use the fact that there are only two pure strategies available to each player. We show via contradiction that if no players employ a pure strategy, all players must play the same mixed strategy. First, assume that here are two players, say i and j, who employ different mixed strategies in a mixed strategy equilibrium in which their probabilities of attending the bar are α_i and α_j respectively, where $\alpha_i \neq \alpha_j$. Note that the remaining $N-2$ players also play mixed strategies. Since player i uses a mixed strategy in equilibrium, it must be that the probability of less than C attending the bar is $(S-B)/(G-B)$ given α_j and the mixed strategies employed by the $(N-2)$ remaining players. Likewise if player j uses a mixed strategy in equilibrium, it must be that the probability of less than C attending the bar is $(S-B)/(G-B)$ given α_i and the mixed strategies employed by the $(N-2)$ remaining players. If $\alpha_i \neq \alpha_j$, both these statements cannot be true. Consider the case where $\alpha_i > \alpha_j$. The probability of less than C attending the bar is $(S-B)/(G-B)$ given α_j and the mixed strategies employed by the $(N-2)$ remaining players. It is then impossible to have the probability of less than C attending the bar, given α_i and the mixed strategies employed by the $(N-2)$ remaining players equal $(S-B)/(G-B)$. We have the similar argument for $\alpha_i < \alpha_j$. Recall that agents have only two possible pure strategies. This contradiction tells us that in an equilibrium where all players play a mixed strategy, they must all play the same mixed strategy. ∎

A.4 Proof to Proposition 4

Proof Recall that an asymmetric mixed strategy Nash equilibria is a Nash equilibrium where players from a subset of the population play either of the available pure strategies, and the remaining players play the symmetric mixed strategy which supports the asymmetric mixed strategy Nash equilibria. In an asymmetric mixed strategy equilibrium, we require at least two players to play a mixed strategy and at least one player to play a pure strategy. Given that all the players playing a mixed strategy are playing the same mixed strategy, we can simply count the number of asymmetric mixed strategy Nash equilibria. ∎

B   Second Appendix

B.1   Proof to Proposition 5

Proof First, suppose that in period t player i chooses to attend the bar. The ith player's strategy adjustment of going to the bar given that player i chooses to go to the bar is,

$$
\begin{aligned}
E\left[y_{t+1}^i | \delta_t^i = 1\right] - y_t^i &= \frac{q_t^i(1) + \hat{\sigma}_t^i\left(1, y_t^{-i}\right)}{Q_t^i + \hat{\sigma}_t^i\left(1, y_t^{-i}\right)} - \frac{q_t^i(1)}{Q_t^i} \\
&= \frac{\left[1 - y_t^i\right]\hat{\sigma}_t^i\left(1, y_t^{-i}\right)}{Q_t^i} \\
&\quad - \left[\left[1 - y_t^i\right]\left[\frac{\left(\hat{\sigma}_t^i\left(1, y_t^{-i}\right)\right)^2}{Q_t^i\left(Q_t^i + \hat{\sigma}_t^i\left(1, y_t^{-i}\right)\right)}\right]\right] \\
&= \frac{\left[1 - y_t^i\right]\hat{\sigma}_t^i\left(1, y_t^{-i}\right)}{Q_t^i} + O\left(\frac{1}{\left(Q_t^i\right)^2}\right)
\end{aligned}
$$

Similarly, the ith player's strategy adjustment of going to the bar given that player i chooses to stay at home is,

$$
\begin{aligned}
E\left[y_{t+1}^i | \delta_t^i = 0\right] - y_t^i &= \frac{q_t^i(1)}{Q_t^i + \hat{\sigma}_t^i\left(0, y_t^{-i}\right)} - \frac{q_t^i(1)}{Q_t^i} \\
&= \frac{-y_t^i\hat{\sigma}_t^i\left(0, y_t^{-i}\right)}{Q_t^i} - \left[y_t^i\left[\frac{\left(\hat{\sigma}_t^i\left(0, y_t^{-i}\right)\right)^2}{Q_t^i\left(Q_t^i + \bar{\sigma}_0^i\right)}\right]\right] \\
&= \frac{-y_t^i\hat{\sigma}_t^i\left(0, y_t^{-i}\right)}{Q_t^i} + O\left(\frac{1}{\left(Q_t^i\right)^2}\right)
\end{aligned}
$$

Recall that by definition player i goes to the bar with probability y_t^i and stays at home with probability $\left[1 - y_t^i\right]$. Therefore, the expected motion of the ith player's $\delta = 1$ strategy adjustment in the repeated El Farol game is,

$$
\begin{aligned}
E\left[y_{t+1}^i | y_t^i\right] - y_t^i &= y_t^i\left[\frac{\left[1 - y_t^i\right]\hat{\sigma}_t^i\left(1, y_t^{-i}\right)}{Q_t^i}\right] \\
&\quad + \left[1 - y_t^i\right]\left[\frac{-y_t^i\hat{\sigma}_t^i\left(0, y_t^{-i}\right)}{Q_t^i}\right] + O\left(\frac{1}{\left(Q_t^i\right)^2}\right)
\end{aligned}
$$

and after further algebraic simplification we arrive at (7). ∎