Learning in the Santa Fe Bar Problem

Amy Greenwald AMYGREEN@CS.BROWN.EDU
Department of Computer Science, Brown University, Providence, RI 02912

Bud Mishra MISHRA@CS.NYU.EDU
Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012

Rohit Parikh RIPBC@CUNYVM.CUNY.EDU
Department of Computer Science, City University of New York, 365 Fifth Avenue, New York, NY 10016

Abstract

This paper investigates the Santa Fe (El Farol) bar problem (SFBP) from the point of view of rational learning. It is argued that rationality together with belief-based learning (*e.g.*, Bayesian updating) yields unstable behavior in this game. More specifically, two conditions sufficient for convergence to Nash equilibrium, namely rationality and predictivity, are shown to be incompatible. Low-rationality learning algorithms, however, which are are known in theory to converge to correlated equilibrium, in fact converge to the symmetric Nash equilibrium in SFBP. Efficiency is suboptimal, however, even at Nash equilibrium. Thus, this paper also proposes a simple modification to SFBP, whereby agents that attend the bar are charged an entry fee that is divided equally among those agents who do not attend the bar. In this modified scenario, low-rationality algorithms, which learn Nash equilibrium strategies, also learn Pareto-optimal behavior: *i.e.*, individual rationality coincides with collective rationality.

1. Introduction

The *Santa Fe bar problem* (SFBP) was introduced by Brian Arthur [1], an economist at the Santa Fe Institute, in the study of bounded rationality and inductive learning. Here is the scenario:

> *N [(say, 100)] people decide independently each week whether to go to a bar that offers entertainment on a certain night ... Space is limited, and the evening is enjoyable if things are not too crowded – especially, if fewer than* 60 *[or, some fixed but perhaps unknown capacity c] percent of the possible* 100 *are present ... a person or 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. Choices are unaffected by previous visits; there is no collusion or prior communication among the agents; and the only information available is the number who came in past weeks.*

We motivate our negative results with the following intuitive analysis of SFBP under the standard economic assumption of rationality. Define an *uncrowded* bar as one in which attendance is less than or equal to c, and define a *crowded* bar as one in which attendance is strictly greater than c. Let the utility of going to an uncrowded bar be $1/2$ and the utility of going to a crowded bar be $-1/2$; in addition, the utility of staying at home is 0, regardless of the state of the bar. If an agent predicts that the bar will be uncrowded with probability p, then his rational (*i.e.*, best) reply is to go to the bar if $p > 1/2$ and to stay home if $p < 1/2$. (In the case where $p = 1/2$, the agents are indifferent between attending the bar and staying home and may behave arbitrarily.) Now, if the agents indeed learn to predict probability p accurately, then their predictions eventually come to match the actual probability that the bar is uncrowded, as it is determined by their (possibly randomized) strategic best-replies. Herein lies a contradiction. If the agents learn to predict that the bar will be uncrowded with probability $p < 1/2$, then, in fact the bar will be empty with probability 1; on the other hand, if the agents learn to predict that the bar will be uncrowded with probability $p > 1/2$, then the bar will be empty with probability 0. We conclude that rational agents cannot learn to make accurate predictions. Rationality precludes learning.

1.1 Logical Implications

This paradoxical outcome in SFBP is arrived at via a diagonalization process in the spirit of Russell's paradox [15]. Russell's set is the set of all sets that are not elements of themselves: *i.e.*, $\mathcal{R} = \{X | X \notin X\}$. Thus, $\mathcal{R} \in \mathcal{R}$ iff $\mathcal{R} \notin \mathcal{R}$. Just as the truth of being in Russell's set depends on the fact of (not) being in the set, the value of going to the bar depends on the act of going (or not going) to the bar. For the sake of argument, consider a bar of capacity 1/2 in a world of a single agent.[1] If the agent does not go to the bar, then the bar is uncrowded, in which case her best-reply is to go to the bar. But now the bar is crowded, and so her best-reply is to stay at home. Thus rationality dictates that this agent should go to the bar if and only if she should not go to the bar.

The aforementioned paradox similarly arises in the two-player game of matching pennies, where player 1 aims to *match* player 2, while player 2 aims to *mismatch* player 1. In fact, matching pennies can be viewed as a variant of SFBP: if player 1 prefers to go to the bar only when player 2 attends as well, while player 2 prefers to go to the bar only when player 1 stays at home, then player 1 is the matcher while player 2 is the mismatcher. In matching pennies, if player 1 prefers *heads*, then player 2 prefers *tails*, but then player 1 prefers *tails*, at which point player 2 actually prefers *heads*, and finally, player 1 prefers *heads* once again. It follows that player 1 prefers *heads* iff player 1 prefers *tails*. Similarly, for player 2.

The logical conflict that arises in the game of matching pennies is closely related to the fact that the game has no pure strategy Nash equilibria [13]; similarly, SFBP has no symmetric pure strategy Nash equilibria, except in degenerate cases. In order to resolve these paradoxes, game-theorists introduce mixed strategies. The unique Nash equilibrium in matching pennies is for both players to play each of *heads* and *tails* with probability 1/2; a mixed strategy Nash equilibrium in SFBP is for all agents to go to the bar with probability $p \approx c/N$ and to stay at home with probability $1 - p$.[2]

1.2 Game-Theoretic Implications

This paper presents negative results on convergence to Nash equilibrium in SFBP which formalizes the above diagonalization argument. Two sufficient conditions for convergence to Nash equilibrium are *rationality* and *predictivity*. By rationality, we mean that players play best-replies to their beliefs. Predictivity is one way in which to capture the notion of learning: a player is said to be predictive if that player's beliefs eventually coincide with (or approach) the truth of what he is predicting. If players learn to predict (*i.e.*, if beliefs indeed converge to other players' actual strategies), then best-replies to beliefs constitute a Nash equilibrium. In what follows, we argue that if the players employ predictive learning algorithms, assuming rationality, play does not converge to Nash equilibrium in SFBP. Equivalently, *if play converges to Nash equilibrium, then either play is not rational or play is not learned.*

In a seminal work by Kalai and Lehrer [10], sufficient conditions are presented for predictivity—specifically, an absolute continuity assumption—which suggests that convergence to Nash equilibrium is at least possible. Our negative results complement the work of Nachbar [12] and Foster and Young [6], who argue that the conditions sufficient for prediction are unlikely to ever hold. Nachbar shows that unless players' initial beliefs somehow magically coincide with Nash equilibrium, repeated play of strategic form games among Bayesian rational players does not converge to Nash equilibrium. Similarly, Foster and Young prove that in two-player games of incomplete information with unique mixed strategy Nash equilibria, rationality is not compatible with predictivity. Our theorems argue in a similar vein that unless certain strict regularity conditions are satisfied, no means of rational learning converges to Nash equilibrium in SFBP.

1.3 Computer Science Implications

SFBP and its natural extensions (*e.g.*, multiple bars) serve as abstractions of various congestion control problems that arise in networking. Many authors (*e.g.*, [16]) who capitalize on the potential for the theory of repeated games as a model of networking environments do so because of the difficulty to enforce cooperation in large-scale networks; instead, it is more realistic and more general to assume noncooperative networks. This generality is modeled in repeated games by assuming that agents are rational. Those same authors who study networking games assuming rationality often also assume that the network operating point is a Nash equilibrium. One might hope to justify this assumption on the grounds that Nash equilibrium is the outcome of rational learning. It is the conclusion of this study, however, that Nash equilibrium is *not* the outcome of rational learning in games that model networking environments.

---

[1] Similarly, one could consider a bar of capacity 1 and a married couple who act in unison.

[2] Technically, this symmetric Nash equilibrium is the solution p to the equation $\sum_{x=0}^{c} \binom{N}{x} p^x (1-p)^{N-x} = \sum_{x=c+1}^{N} \binom{N}{x} p^x (1-p)^{N-x}$, which is approximately c/N.

In light of the suite of negative theoretical results, the second part of this paper aims to resolve the paradoxes of the first via simulation experiments in computational learning. Apparently, it is necessary to relax one or more of the usual economic assumptions in order to approach equilibrium behavior. More specifically, we can relax either rationality or the implicit assumption that beliefs are deterministic. Foster and Young [6] take the latter approach. In this paper, it is shown that *low-rationality* learning yields equilibrium behavior. Similarly, in Arthur's original paper, he demonstrated via simulations that boundedly rational agents are capable of generating collective attendance centered around the capacity of the bar. In contrast to Arthur's approach, which is based on complex modeling of cognitive aspects of inductive reasoning, the algorithms of interest in this study are simple and straightforward, and are therefore more apt for use in networking applications. While (highly) rational learning does not validate the assumption that Nash equilibrium describes the solution of network games, low-rationality learning indeed yields Nash equilibrium behavior.

1.4 Economic Implications

Thus far, we have considered rational and low-rationality learning from the point of view of individual agents. Now let us shift our attention to collective behavioral issues. Observe that the symmetric Nash equilibrium in SFBP is not efficient. If all agents go to the bar with probability c/N, then half the time the bar is uncrowded, yielding positive utilities for the agents, but half the time the bar is crowded, yielding negative utilities for the agents. Therefore, the expected utility of the collective is zero. But, the maximum utility, achieved when exactly N agents attend the bar, is $cN/2$! A simple solution to this problem (advocated by Bell, *et al.* [3]) is to devise algorithms by which agents learn pure strategy Nash equilibrium: *i.e.*, c agents go to the bar and $N-c$ agents stay home, *forever*. This solution, albeit efficient, is unfair, since some agents learn to never attend the bar.

In this paper, we introduce a *fair* market mechanism that rectifies the inherent inefficiency in the symmetric equilibria of SFBP. We propose to charge those agents who attend the bar an entry fee, which is divided equally among those agents who remain at home. This idea is related to many in the economic literature on user fees and public goods (*e.g.*, Varian [17]). For example, like gasoline taxes which negatively impact drivers only, our bar entry fee negatively impacts only those agents that attend the bar; also like European gasoline taxes, which are converted into subsidies for public transportation systems, bar entry fees are distributed among those agents that do not attend. This market mechanism does not change the Nash equilibria of SFBP. Therefore, low-rationality learning algorithms, which learn the symmetric Nash equilibrium in SFBP, generate fair and efficient collective behavior.

2. Formalization

The Santa Fe bar problem is a repeated game of negative externalities.[3] We now formally define both the one-shot strategic form game, and the corresponding repeated game. The players are the inhabitants of Santa Fe; notation $\mathcal{N} = \{1, \ldots, N\}$, with $n \in \mathcal{N}$. For player n, the strategy set $S_n = \{0, 1\}$, where 1 corresponds to *go to the bar* while 0 corresponds to *stay home*. Let Q_n denote the set of probability distributions over S_n, with mixed strategy $q_n \in Q_n$. The expected payoffs obtained by player n depend on the particular strategic choice taken by player n, the value to player n of attending the bar, and a negative externality, which are defined as follows.

Let s_n denote the realization of mixed strategy q_n of player n; thus, $s = \sum_{n \in \mathcal{N}} s_n$ is the realized attendance at the bar. In addition, let $c \in \{0, \ldots, N\}$ denote the capacity of the bar. The externality E depends on s and c as follows: if the bar is uncrowded (*i.e.*, $s \leq c$), then $E(s) = 0$; on the other hand, if the bar is crowded (*i.e.*, $s > c$), then $E(s) = 1$. Let $0 \leq \alpha_n \leq 1$ denote the value to player n of attending the bar, and without loss of generality assume $\alpha_n \leq \alpha_{n+1}$.[4] Now the payoff function for player n is given by:

$$\begin{aligned} \pi_n(s_n, s) &= \begin{cases} \alpha_n - E(s) & \text{if } s_n = 1 \\ 0 & \text{otherwise} \end{cases} \\ &= s_n[\alpha_n - E(s)] \end{aligned}$$

As usual, the expected payoffs $\mathbb{E}_{q_n}[\pi_n(s_n, s)]$ obtained by player n via mixed strategy q_n are given by $\mathbb{E}_{q_n}[\pi_n(s_n, s)] = \sum_{s_n \in S_n} q_n(s_n)\pi_n(s_n, s)$. SFBP is a discretization of an ordered externality game in the sense of Friedman [7].

The one-shot strategic form SFBP is described by the tuple $\Gamma = (\mathcal{N}, (S_n, \pi_n)_{n \in \mathcal{N}}, c)$, and the infinitely repeated SFBP is given by Γ^∞. A history h^t of length $t \in \mathbb{N}$ is defined to be a sequence of t outcomes drawn from the set $S = \{0, 1, \ldots, N\}$; the history

[3] An externality is a third-party effect. An example of a negative externality is pollution; an example of a positive externality is standardization. Although externalities are typically external to a game, it is natural to view payoffs in terms of externalities when the numbers of players is large.

[4] The results in this paper are restricted to the *uniform* SFBP in which $\alpha_n = \alpha_m$ for all players $n, m \in \mathbb{N}$.

$h^t = (s^1, \ldots, s^t)$ indicates the number of players who attended the bar during periods 1 through t. Let h^0 denote the null history, let H^t denote the set of all histories of length t, and let $H = \bigcup_0^\infty H^t$.

A belief-based learning algorithm is a function from the set of all possible histories to the set of possible beliefs. We assume that beliefs in the repeated SFBP take the form of a subjective probability over the space of possible externality effects $\mathcal{E} = \{$*uncrowded*, *crowded*$\}$. Recall that the event *uncrowded* obtains at time t whenever $s^t \leq c$; otherwise, the event *crowded* obtains. Let $\Delta(\mathcal{E})$ be the set of probability distributions over the set $\mathcal{E}$. Formally, a belief-based learning algorithm for player n is a function $f_n : H \to \Delta(\mathcal{E})$.[5] Since the event space $\mathcal{E}$ is of cardinality 2, the sequence of beliefs $\{f_n^t(h^t)\} = \{(p_n^{t+1}, 1 - p_n^{t+1})\}$ is denoted simply $\{p_n^{t+1}\}$, where p_n^{t+1} is the probability that player n attributes to the bar being uncrowded at time $t+1$.

The expected payoff for player n at time t is computed in terms of the beliefs that player n holds at time t:

$$\mathbb{E}_n^t[\pi_n(s_n, s)] = \begin{cases} p_n^t \alpha_n - (1 - p_n^t)(1 - \alpha_n) & \text{if } s_n = 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $p_n^* \equiv 1 - \alpha_n$. Player n is indifferent between his two pure strategies whenever $p_n^t = p_n^*$, since this implies $\mathbb{E}_n^t[\pi_n(1, s)] = \mathbb{E}_n^t[\pi_n(0, s)] = 0$. The actual (*i.e.*, objective) probability that the bar is uncrowded at time t is denoted by p_0^t. The existence of such probabilities is implied by the assumption that in general players employ mixed strategies: *i.e.*, p_0^t can be computed directly from the players' strategies q_n^t.

Definition 2.1 SFBP is *uniform* iff for all $n, m \in \mathcal{N}$, $\alpha_n = \alpha_m \equiv \alpha$, and thus, $p_n^* = p_m^* \equiv p^*$.

3. An Example: Best-Reply Dynamics

We now show that best-reply dynamics, a learning algorithm for which Cournot proved convergence to pure strategy Nash equilibrium in models of duopoly [5], yield oscillatory behavior in SFBP. Note that since best-reply dynamics is one method of belief-based learning the result presented in this section follows as an immediate corollary of the more general results derived in later sections. We begin by reminding the reader of the definition of best-reply dynamics.

Definition 3.1 A strategy $q_n^t \in Q_n$ is said to be a *best-reply* for player n at time t iff $q_n^t \in \arg\max_{q_n \in Q_n} \pi_n(q_n, p_n^t)$: *i.e.*, $\pi_n(q_n^t, p_n^t) \geq \max_{q_n \in Q_n} \pi_n(q_n, p_n^t)$.

Definition 3.2 A given player n is said to employ *best-reply dynamics* in SFBP iff for all times t, player n assumes that $p_n^{t+1} = 1$ if $s^t \leq c$, and $p_n^{t+1} = 0$ otherwise, and moreover, player n plays only best-replies to these beliefs. In other words, if player n utilizes best-reply dynamics, then $q_n^t \in \arg\max_{q_n \in Q_n} \pi_n(q_n, s^t)$.

Assume that all players employ best-reply dynamics. If at time t, $s^t \leq c$, then $p_n^{t+1} = 1$ for all n, to which the best response at time $t+1$ is pure strategy $s_n^{t+1} = 1$. But then $s^{t+1} > c$, so that $p_n^{t+2} = 0$ for all n, to which the best response at time $t+2$ is pure strategy $s_n^{t+2} = 0$. Now, it follows that $s^{t+2} \leq c$ and $p_n^{t+3} = 1$. This pattern repeats itself indefinitely, generating oscillatory behavior that is far from equilibrium. A similar argument arises in the case in which $s^t > c$. The following remark captures this intuition.

Remark 3.3 In the uniform repeated SFBP, best-reply dynamics do not converge: *i.e.*, $\forall n$, $\lim_{t,t' \to \infty} | p_n^t - p_n^{t'} | \neq 0$ and $\lim_{t,t' \to \infty} | s^t - s^{t'} | \neq 0$.

4. A Preliminary Result

In what follows, we generalize our observation pertaining to best-reply dynamics. Specifically, we argue that if players are rational and if they learn according to Bayes' rule, their strategies do not converge to equilibrium behavior. In fact, this result is not contingent on the assumption of Bayesian learning and is readily applicable to any predictive belief-based learning mechanism. We arrive at this negative result by first deriving a seemingly positive result, namely that beliefs can *sometimes* converge to Nash equilibrium, assuming rationality and predictivity. Later we show that in practice sometimes amounts to *almost never*.

Definition 4.1 A learning algorithm is said to be *predictive* iff it generates a sequence of beliefs $\{p_n^t\}$ for player n *s.t.* $\lim_{t \to \infty} | p_n^t - p_0^t | = 0$.

In words, if player n utilizes a predictive learning algorithm, then the difference between player n's subjective beliefs p_n^t and the objective probabilities p_0^t converges to zero. This definition does not require that the objective probabilities themselves converge, only that player n's subjective beliefs approach the objective probabilities.

Definition 4.2 A set of players $\mathcal{N}$ is said to reach *consensus* iff for all $n, m \in \mathcal{N}$, $\lim_{t \to \infty} | p_n^t - p_m^t | = 0$.

[5] The given definition precludes any notion of correlated beliefs, in which players might attempt to correlate the behavior of an individual, such as oneself, with attendance at the bar.

Lemma 4.3 *If all players within set $\mathcal{N}$ are predictive, then they reach consensus.*

Proof 4.4 By the triangle inequality, $|p_n^t - p_m^t| \leq |p_n^t - p_0^t| + |p_0^t - p_m^t|$, for all $n, m \in \mathcal{N}$, and for all t. Taking limits and applying the definition of predictivity, it follows that $\lim_{t\to\infty} |p_n^t - p_m^t| = 0$. □

Definition 4.5 A player n is *rational* iff he plays only best-replies to his beliefs p_n^t.

The following theorem states that in the uniform version of the repeated SFBP, whenever players exhibit rationality and predictivity, beliefs converge to p^*. It follows by predictivity that the objective probabilities, which correspond to the players' joint strategies, must converge to p^* as well. Thus, rational players who play best-replies to their beliefs, ultimately play best-replies to actual strategies: *i.e.*, play converges to Nash equilibrium. This seemingly positive result is contested in later sections.

Theorem 4.6 *In the uniform repeated* SFBP*, if players are rational and predictive,* $\lim_{t\to\infty} |p_n^t - p^*| = 0$.

Proof 4.7 Suppose not. Two cases arise. First suppose $\exists \epsilon > 0$ *s.t.* $p_n^t > p^* + \epsilon$ infinitely often (i.o.). It follows by Lemma 4.3 that for all m, $p_m^t > p^* + \delta$ i.o., for all $0 < \delta < \epsilon$. Now by rationality, all players play best-replies, which for such t is to go to the bar: *i.e.*, for all n, $s_n^t = 1$ i.o.. This ensures that the bar will be crowded with probability 1, yielding $p_0^t = 0 < p^* + \epsilon < p_n^t$ i.o., which implies that $p_n^t - p_0^t > \epsilon$ i.o., contradicting predictivity. The argument in the second case is analogous. □

It follows immediately from Theorem 4.6 and the definition of predictivity that whenever players are rational and predictive, strategies (objective probabilities), as well as beliefs, must converge to p^*. The question now arises as to whether p^* can indeed be the actual objective probability, for if not, players cannot be both rational and predictive.

5. A First Negative Result

We now present our first negative result, namely that no mechanism of rational, predictive, belief-based learning (including Bayesian updating) gives rise to objective probabilities that converge to p^*, except in rare circumstances. Before making any general claims, we construct an example of one such special p^* for which convergence is possible. We assume that indifferent players (*i.e.*, those players n for which $p_n^t = p^*$) flip a fair coin; for if, on the contrary, players were to flip a biased coin favoring one strategy or another, they would not be truly indifferent between the two alternatives.

Example 5.1 Let $f(t) \to 0$ be a monotonically decreasing function of t: *e.g.*, $f(t) = 1/t$.

- Suppose G players (the optimists) hold beliefs $p^* + f(t)$. These players' beliefs converge to p^*. By rationality, these players *always* go to the bar.
- Let H players (the pessimists) hold beliefs $p^* - f(t)$. These players' beliefs also converge to p^*. By rationality, these players *never* go to the bar.
- Let I players (the realists) hold beliefs exactly p^* at all times t. These players are indifferent between going to the bar and not going. They flip a fair coin.

Given that players' beliefs converge to p^*, we now consider the conditions under which the players' strategies (*i.e.*, actual probabilities) also converge to p^*. Let the excess capacity of the bar $d = c - G$ for the I indifferent players, after accommodating the G players who go to the bar in every period. Suppose indifferent players go to the bar iff their coin flips show heads. In this scenario, the probability p that the bar is uncrowded is the probability that with I flips of a fair coin,[6] at most d heads appear:

$$p = \begin{cases} 0 & \text{if } d < 0 \\ 1 & \text{if } d \geq I \\ \frac{1}{2^I} \sum_{j=0}^{d} \binom{I}{j} & \text{otherwise} \end{cases} \quad (2)$$

Now as $t \to \infty$, $p_n^t \to p^*$ and $p_0^t \to p$ (in fact, $p_0^t = p$, for all t). Thus, if $p^* = p$, then both beliefs and strategies converge to p^*. □

Using the layout of Example 5.1, it is possible to describe all possible values of p in Equation 2. At fixed time t, let G denote the number of players who definitely go to the bar; let H denote the number of players who definitely stay at home; and let I denote the number of players who are indifferent and therefore flip a

[6]Mathematically, this result holds in the more general case when players flip a coin of bias q. In particular, Equation 2 becomes

$$p = \begin{cases} 0 & \text{if } d < 0 \\ 1 & \text{if } d \geq I \\ \sum_{j=0}^{d} \binom{I}{j} q^j (1-q)^{I-j} & \text{otherwise} \end{cases} \quad (1)$$

We do not present this case, however, since this assumption is more difficult to justify.

fair coin in deciding whether or not to attend the bar. The following set F describes all the realizable probabilities under these circumstances: $F = \{p \mid \exists G, H, I \in \{0, \ldots, N\}$ *s.t.* p is defined by Equation 2$\}$. F is a finite set since there are only finitely many possible values of G, H, and I. The next theorem states that objective probabilities cannot converge to p^*, if $p^* \notin F$.

Theorem 5.2 *In the uniform repeated* SFBP*, given rational and predictive players,* $\lim_{t\to\infty} |p_0^t - p^*| \neq 0$, *if* $p^* \notin F$*, provided indifferent players flip a fair coin.*

In SFBP, assuming a bar of capacity c, if players are rational and predictive, then strategies can only converge to p^* if p^* happens to be an element of the finite set F. Thus, it is only on rare occasions that players exhibit both rationality and predictivity, such that both beliefs and strategies converge to p^*: *i.e.*, play converges to Nash equilibrium. In general, play does not converge to Nash equilibrium in SFBP.

Example 5.3 Consider an instance of SFBP in which players are both rational and predictive. Let $N = I = 10$, and assume $c = 6$. In other words, there are 10 players, all of whom are indifferent and flip a fair coin. According to Equation 2, there exists $p^* \approx .828$ such that the players' strategies (*i.e.*, actual probabilities) converge to p^*. By Theorem 4.6, beliefs also converge to p^*. Thus, in this particular instance of SFBP, if by chance $\alpha = 1 - p^* \approx .172$, then play converges to Nash equilibrium. □

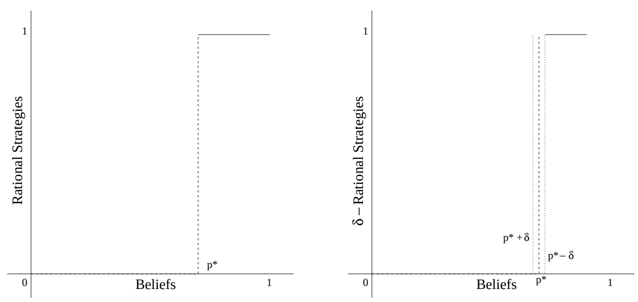
As Theorems 4.6 and 5.2 yield contradictory conclusions in all but finitely many cases, we deduce that together the assumptions of rationality and predictivity are inconsistent: *in general, there is no rational learning in* SFBP.

Corollary 5.4 *In the uniform repeated* SFBP*, players cannot be both rational and predictive, unless* $p^* \in F$.

This concludes the discussion of our first negative result. It was argued that two conditions which together are sufficient for convergence to Nash equilibrium, namely rationality and predictivity, are incompatible in SFBP. A similar analysis appeared in Greenwald [8]. In the next section, a second negative result is derived, which is based on the work of Mishra [11].

6. A Second Negative Result

The negative result presented in the previous section is contingent on the fact that in SFBP rational best-replies are in general deterministic. At fixed time t, if $p_n^t < p^*$, then player n believes that the bar is crowded, so his best-reply is to stay home, or go to the bar with probability 0; on the other hand, if $p_n^t > p^*$, then player n believes that the bar is uncrowded, so his best-reply is to go to the bar with probability 1. Although Theorem 4.6 ensures that players' beliefs converge to p^*, only when beliefs exactly coincide with p^* is player n indifferent between going to the bar and staying at home; at this point, player n may play any mixed strategy q_n^t. The step function that represents these deterministic best-replies is depicted in Figure 1(a). Beliefs p_n^t are plotted along the x-axis, and rational strategies q_n^t are plotted along the y-axis.



Rationality   δ-Rationality

Figure 1. (a) Rationality. (b) δ-Rationality.

This section describes a second negative result based on the following slightly weaker notion of rationality. Here, players' behavior near p^* is not so clear-cut, but rather it is unspecified, encouraging players to perhaps experiment with randomized strategies anywhere in the range $(p^* - \delta, p^* + \delta)$, as depicted in Figure 1(b).

Definition 6.1 Let $\delta > 0$. A player n is δ-*rational* iff he plays only best-replies to his beliefs in the range $[0, p_n^t - \delta]$ and $[p_n^t + \delta, 1]$, but plays arbitrarily in the range $(p_n^t - \delta, p_n^t + \delta)$.

The proof of Theorem 4.6 is easily extended to show that if players are δ-rational and predictive, then their beliefs converge to near p^*. But δ-rationality allows for the possibility that players experiment with nondeterministic strategies not only when beliefs precisely equal p^*, but even when beliefs are in the neighborhood of p^*. In other words, δ-rationality and predictivity imply δ-indifference.

The next theorem states that no means of δ-rational, predictive learning gives rise to objective probabilities that converge to p^*, unless the capacity of the bar fortuitously lies between $N/2 - k_1\sqrt{N}$ and $N/2 + k_2\sqrt{N}$, for certain $k_1, k_2 > 0$, provided that δ-indifferent players flip a fair coin. This result is explained intuitively as follows. By Theorem 4.6 (extended to the case of

δ-rational players), beliefs converge to p^*: *i.e.*, players are eventually δ-indifferent. By assumption, the players flip a fair coin. Thus, attendance at the bar is likely to be near $N/2$. The theorem states that unless the capacity of the bar also happens to be near $N/2$, δ-rational learning is also ineffectual in SFBP.

Theorem 6.2 *Let* $0 < \delta < \min\{(1-\alpha)-(1-\alpha)/e, \alpha - \alpha/e\}$. *In the uniform repeated* SFBP, *assuming players are δ-rational and predictive,* $\lim_{t\to\infty} |p_0^t - p^*| > \delta$, *provided that δ-indifferent players flip a fair coin,*[7] *and* $c \leq N/2 - k_1\sqrt{N}$ *where* $k_1 = 1 + \ln(1/(1-\alpha))$, *or* $c \geq N/2 + k_2\sqrt{N}$ *where* $k_2 = [3 + 3\ln(1/\alpha)]/2$.

Proof 6.3 Theorem 4.6 (in the case of δ-rational players) states that beliefs converge to within δ of p^*. Thus, all players are eventually δ-indifferent, from which point on they flip a fair coin. It follows that attendance at the bar is eventually binomially distributed $\sim S(N, 1/2)$. Two distinct cases arise, depending on the capacity of the bar. This proof utilizes the multiplicative variant of the Chernoff bound [4].

Case 6.3.1 Assume $c \leq N/2 - k_1\sqrt{N} = N/2 - \sqrt{[1+\ln(1/(1-\alpha))]N}$. In this case,

$$
\begin{aligned}
p_0^t &= \Pr[S(N,1/2) < c] \\
&\leq \Pr[S(N,1/2) < (N/2)\{1 - \\
&\quad \sqrt{[4+4\ln(1/(1-\alpha))]/N}\}] \\
&\leq e^{-([4+4\ln(1/(1-\alpha))]/2N)(N/2)} \\
&= (1-\alpha)/e \\
&< 1-\alpha-\delta, \text{ since } \delta < (1-\alpha)-(1-\alpha)/e \\
&= p^* - \delta
\end{aligned}
$$

Therefore, $p^* - p_0^t > \delta$. Contradiction.

Case 6.3.2 Assume $c \geq N/2 + k_2\sqrt{N} = N/2 + \sqrt{[3+3\ln(1/\alpha)][N/2]}$. This case is analogous. □

This theorem shows that even for δ-rational players, unless the capacity of the bar is near $N/2$, rational and predictive (*i.e.*, Nash) behavior is impossible.

7. A Positive Result

We now study learning among computational agents that are not highly rational; on the contrary, they exhibit low-rationality (*i.e.*, non-Bayesian) learning. Low-rationality learning algorithms do not maintain belief-based models over the space of opponents' strategies or payoff structures. Instead, they specify that agents *explore* their own strategy space by playing all strategies with some non-zero probability, and *exploit* successful strategies by increasing the probability of employing those strategies that generate high payoffs. Simple reinforcement techniques of this nature are advantageous because unlike Bayesian learning, they do not depend on any complex modeling of prior probabilities over possible states of the world. Also, unlike Arthur's approach, they are are not based on inherently complex models of human cognition.

Freund and Schapire [2] study a low-rationality learning algorithm based on an exponential updating scheme.[8] Let $P_n^t(s_n)$ denote the cumulative payoffs obtained by agent n through time t via strategy s_n: *i.e.*, $P_n^t(s_n) = \sum_{x=1}^{t} \pi_n(s_n, s^x)$. The weight assigned to strategy s_n at time $t+1$, for $\beta > 0$, is given by:

$$
q_n^{t+1}(s_n) = \frac{(1+\beta)^{P_n^t(s_n)}}{\sum_{s'_n \in S_n} (1+\beta)^{P_n^t(s'_n)}} \tag{3}
$$

We simulated SFBP assuming a bar of capacity 60 and 100 computational agents learning according to the algorithm of Freund and Schapire, with $\beta = 0.01$. Figure 2 plots attendance at the bar over time. Note that attendance centers around 60—the capacity of the bar. Specifically, the mean attendance is 60.04 and the variance is 5.11. Learning via this and other low-rationality algorithms, which do not necessitate perfectly rational behavior, yield equilibrium outcomes in SFBP. Moreover, these results are robust in the sense that the agents readily adapt if ever the capacity of the bar changes.

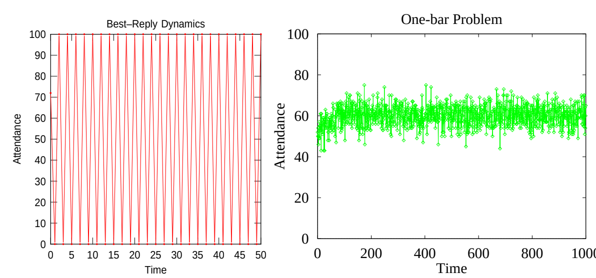


Figure 2. Rational vs. low-rationality learning.

[7] Mathematically, this result holds for arbitrary probabilities p_n, p_m *s.t.* $|p_n - p_m| < \epsilon$, for small values of $\epsilon > 0$, where p_n and p_m denote the probabilities that players n and m, respectively, go to the bar. We do not present this case since this assumption is more difficult to justify.

[8] This algorithm depends on the cumulative payoffs achieved by all strategies including the surmised payoffs of strategies which are not in fact played. Techniques exist, however, by which to convert this algorithm to an algorithm that depends only on the payoffs of those strategies that are in fact employed (see Greenwald, *et al.* [9]).

8. A Fair and Efficient Mechanism

We now consider the collective behavior of agents in SFBP. As noted earlier, the symmetric Nash equilibrium learned by low-rationality algorithms is not efficient. The collective utility of a set of low-rationality learners is zero, since half the time the bar is uncrowded, yielding positive utilities for the agents, but half the time the bar is crowded, yielding negative utilities for the agents. In this section, we introduce a market mechanism based on user fees by which low-rationality learning leads to fair and efficient outcomes.

Our proposed mechanism is straightforward: charge agents that attend the bar a user fee, say x, and distribute the proceeds evenly among those agents that do not attend the bar. Suppose λ agents go to the bar. The payoffs to those λ agents are now $\alpha - x$, if the bar is uncrowded, and $\alpha - 1 - x$, if the bar is crowded. The payoffs to those $N - \lambda$ agents that do not go to the bar is now $\lambda x/(N - \lambda)$. Note that this mechanism does not change the symmetric Nash equilibrium of SFBP.

Figure 3 plots efficiency (total utility obtained by all agents divided by the bar's capacity) as a function of the value of the fee x, for $\alpha = 1$, $N \in \{10, 100, 1000\}$, and $c = 0.6N$. Note that as $N \to \infty$, the optimal value of x approaches 0.4 and efficiency approaches 100%. The value $x = 0.4$ equates the payoffs of those agents that go $(1 - 0.4 = 0.6)$ with the payoffs of those agents that stay home (24/40), whenever precisely 60 agents go to the bar (*i.e.*, efficiency equals 100%).

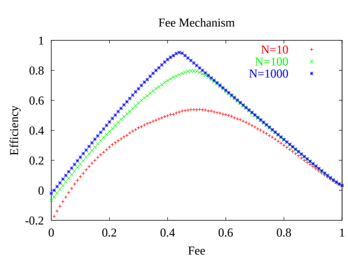


Figure 3. Efficiency vs. fee: as $N \to \infty$, optimal fee approaches 0.4 and efficiency approaches 100%.

9. Conclusion

This paper studied learning dynamics in the Santa Fe bar problem. Specifically, we investigated whether Nash equilibrium could arise the outcome of learning. We reported two negative results—neither rational nor approximately rational learning lead to Nash equilibrium behavior—and one positive result—low-rationality learning, a simple form of reinforcement learning, indeed converges to Nash equilibrium.

We also took a close look at the Nash equilibria of SFBP. The pure strategy asymmetric Nash equilibria are efficient, but unfair. The mixed strategy symmetric Nash equilibrium are fair, but inefficient. Thus, we proposed an alternative SFBP mechanism for which the symmetric Nash equilibrium is efficient. Under this mechanism, individually low-rationality learning, which generates Nash behavior, yields an outcome that collectively is both fair and efficient.

References

[1] W.B. Arthur. Inductive reasoning and bounded rationality. *Complexity in Economic Theory*, 84(2):406–411, 1994.

[2] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, pages 322–331. ACM Press, November 1995.

[3] A. M. Bell, W. A. Sethares, and J. A. Bucklew. Coordination failure as a source of congestion in information networks. May 1999.

[4] H. Chernoff. A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–509, 1952.

[5] A. Cournot. *Recherches sur les Principes Mathematics de la Theorie de la Richesse*. Hachette, 1838.

[6] D. Foster and P. Young. When rational learning fails. Working Paper, 1998.

[7] E. Friedman. Dynamics and rationality in ordered externality games. *Games and Economic Behavior*, 16:65–76, 1996.

[8] A. Greenwald. *Learning to Play Network Games*. PhD thesis, Courant Institute of Mathematical Sciences, New York University, New York, May 1999.

[9] A. Greenwald, A. Jafari, G. Ercal, and D. Gondek. On no-regret learning, Nash equilibrium, and fictitious play. Submitted to ICML 2001, January 2001.

[10] E. Kalai and E. Lehrer. Rational learning leads to Nash equilibrium. *Econometrica*, 61:1019–1045, 1993.

[11] B. Mishra. A tale of two bars: A negative result for the Santa Fe bar problem. Working Paper, 1998.

[12] J. Nachbar. Prediction, optimization, and learning in repeated games. *Econometrica*, 65:275–309, 1997.

[13] R. Parikh. A connection between Nash equilibria and propositional logic. Working Paper, 1998.

[14] R. Parikh. Social software. *Synthese*, Forthcoming, 2001.

[15] B. Russell. Letter from Russell to Frege. In *From Frege to Gödel*, pages 124–125, Cambridge, 1867. Harvard University Press.

[16] S. Shenker. Making greed work in networks: A game-theoretic analysis of switch service disciplines. *IEEE/ACM Transactions on Networking*, 3:819–831, 1995.

[17] H. Varian. *Microeconomic Analysis*. W.W. Norton & Company, Inc., New York, 1978.

[18] D. Wolpert, K. Wheeler, and K. Tumer. Collective intelligence for control of distributed dynamical systems. *Europhysics Letters*, 49(6), 2000.