Fair and Efficient Solutions to the Santa Fe Bar Problem

Julie Farago
Department of Computer Science
Harvard University
Cambridge, MA 02138
farago@fas.harvard.edu

Amy Greenwald Keith Hall
Department of Computer Science
Brown University, Box 1910
Providence, RI 02912
[amy|kh]@cs.brown.edu

Abstract

This paper asks the question: can adaptive, but not necessarily rational, agents learn Nash equilibrium behavior in the Santa Fe Bar Problem? To answer this question, three learning algorithms are simulated: fictitious play, no-regret learning, and Q-learning. Conditions under which these algorithms can converge to equilibrium behavior are isolated. But it is noted that the pure strategy Nash equilibria are unfair, while the (symmetric) mixed strategy equilibrium is inefficient. Thus, SFBP is redesigned to induce adaptive agents to learn fair and efficient equilibrium outcomes.

1 Introduction

The *Santa Fe Bar Problem* (SFBP) was introduced by Brian Arthur, an economist at the Santa Fe Institute. Arthur challenges the behavioral predictions of rational choice theory, since, in SFBP, *rationality precludes learning*.

> 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 that 60 [or, some fixed but perhaps unknown capacity c] percent of the possible 100 are present . . . a person or agent goes if she expects fewer than 60 to show up and stays home if she expects more than 60 to go, Choices are unaffected by previous visits; there is no collusion or prior communication among the agents; the only information available is the number who came in the past weeks.[1]

The bar in Santa Fe is a *congested* resource, characterized as such because the value to an agent of attending the bar depends on the number of other agents that attend the bar. In other words, agents' valuations of congested resources are not exogenously-determined, but rather are endogenous functions of one another's actions. This so-called congestion effect is apparent in many real-world situations, ranging from fishermen fishing in common waters, to farmers polluting common water supplies, to network users monopolizing bandwidth, to other versions of the tragedy of the commons [6] that are characterized by *negative externalities*.[1]

Let us analyze SFBP under the assumption of *rationality*: *i.e.*, agents maximize utility given their beliefs. Define an *uncrowded* bar as one in which attendance is less than or equal to the capacity, 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 $+1$ and the utility of going to a crowded bar be -1; the utility of staying home is 0, regardless of the state of the bar. In this setup, rational choice theory dictates the following behavior: if an agent believes that the bar will be uncrowded with a probability p, then his rational choices are to go to the bar if $p > 1/2$ and to stay home if $p < 1/2$. In the case where $p = 1/2$, rational agents are indifferent between attending the bar and staying home and may behave arbitrarily: *e.g.*, randomize. The true probability p that the bar is uncrowded is determined by the agents (possibly randomized) strategies. If all the agents learn the correct value of p, then we stumble upon the fundamental contradiction. Suppose the agents learn 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, suppose the agents learn that the bar will be uncrowded with probability $p > 1/2$; then the bar will be empty with probability 0.[2] Thus, rational agents cannot learn in SFBP: *rationality precludes learning*.

[1] An externality is a third-party effect. An example of a negative (positive) externality is pollution (standardization).

[2] Schelling [11] refers to phenomena of this kind as self-negating prophecies. Yogi Berra also observed this phenomenon. He said, "Nobody goes there anymore; it's too crowded."

In his original paper, Arthur demonstrated via simulations that certain types of *boundedly rational* agents are capable of learning to center their collective attendance around the capacity of the bar. Arthur's approach is based on complex modeling of cognitive aspects of inductive reasoning. In Greenwald [4], boundedly rational, no-regret learning agents are simulated and shown to converge to the neighborhood of the symmetric mixed strategy Nash equilibrium in which all agents attend the bar with probability $\sim c/N$. This paper continues this thread of research, asking the question: can adaptive, but not necessarily rational, agents learn Nash equilibrium behavior in SFBP? To answer this question, three learning algorithms are simulated: fictitious play [10], which is an example of a model-based learning algorithm (see Sec. 3); no-regret learning [2], which minimizes the learner's regret in the worst case (see Sec. 4); and Q-learning [12], a reinforcement learning algorithm (see Sec. 5).

2 Game-Theoretic Analysis

The Santa Fe bar problem is a repeated game. The agents, or players, are the inhabitants of Santa Fe; notation $\mathcal{N} = \{1, \ldots, N\}$, with $n \in \mathcal{N}$. For agent n, the (pure, or deterministic) 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, or randomized) strategy $q_n \in Q_n$. The expected utilities, or payoffs, obtained by agent n depend on the particular strategic choice taken by agent n, the value to agent n of attending the bar, and the negative externality.

Let s_n denote the realization of mixed strategy q_n of agent 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 $E(s) = +1$; 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 agent n of attending the bar. Now the utility function for agent n is given by:

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

2.1 Nash Equilibrium

A Nash equilibrium is a vector of strategies, one per agent, from which no agent has any incentive to deviate [8]. Since the set of Nash equilibria in SFBP is very large, we restrict our attention to equilibria that satisfy fairness and/or efficiency. A *fair* outcome requires that agents with identical utilities be equally likely to attend the bar. *Efficiency* is a measure of collective, or total agent, utility achieved relative to its maximum value.

At any pure strategy Nash equilibrium of SFBP, exactly c agents attend the bar, while $N - c$ agents stay home. Those agents n attending the bar obtain utility $\alpha_n > 0$; thus, they have no incentive to stay home, where they would obtain utility 0. On the other hand, those agents m that stay home (and obtain utility of 0) have no incentive to attend the bar; by doing so, they would obtain utility $-\alpha_m < 0$, since attendance by any one of them would suddenly cause congestion. This outcome is an efficient equilibrium, but it is obviously unfair.

Assuming $\alpha_n = \alpha$ for all agents n, the (symmetric) mixed strategy Nash equilibrium is the probability p at which the agents are indifferent between attending the bar and staying home—if the agents were not indifferent between their pure strategies, then they would not employ mixed strategies. Thus, p satisfies the following equation:

$$\alpha \sum_{i=0}^{c} \binom{N}{i} p^i (1-p)^{N-i} - \alpha \sum_{i=c+1}^{N} \binom{N}{i} p^i (1-p)^{N-i} = 0 \tag{1}$$

In other words, the expected utility of going to the bar equals 0, the expected utility of staying home. Equivalently,

$$\sum_{i=0}^{c} \binom{N}{i} p^i (1-p)^{N-i} = \sum_{i=c+1}^{N} \binom{N}{i} p^i (1-p)^{N-i} \tag{2}$$

The solution to this equation is approximately c/N. Since it is symmetric, this solution generates a fair outcome. But notice that if all the agents employ this mixed strategy Nash equilibrium, then efficiency is near zero: half the time the bar is uncrowded, yielding positive collective utility for the agents, but half the time the bar is crowded, yielding negative collective utility for the agents (see Table 1).

p	Collective Utility	p	Collective Utility	p	Collective Utility
0	0.0	.450	45.8	.58	18.8
.1	10.0	.475	47.2	.6	0.5
.2	20.0	.5	47.8	.7	-67.5
.3	30.0	.52	46.5	.8	-80.0
.4	40.0	.54	41.9	.9	-90.0
.425	42.4	.56	33.0	1.0	-100.0

Table 1: Collective utility vs. p, for $N = 100$ and $c = 60$. The mixed strategy Nash equilibrium $p = 0.6$ yields collective utility near 0, and $p = 0.5$ yields the highest efficiency.

3 On Learning an Efficient Outcome

Fictitious play is a model-based learning algorithm, where agents model their opponents' strategic behavior. In its standard formulation [10], the model is taken to be the opponents' empirical distribution of play. Given this model, fictitious play dictates that an agent play one of its utility-maximizing strategies. Fictitious play converges to Nash equilibrium in certain restricted classes of games (*e.g.*, zero-sum games).

One straightforward implementation of fictitious play in SFBP is to simply compute the empirical distribution over the two events $\{$*uncrowded*, *crowded*$\}$. Formally, let $a^0 = 0$ and

$$a^{t+1} = a^t + \begin{cases} 1 & \text{if } s^t \le c \\ 0 & \text{otherwise} \end{cases}$$

For all times $t > 0$, the probability that the bar is uncrowded $p^t = a^t/t$. The expected utility for agent n at time $t+1$ is computed in terms of this probability p^t:

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

Fictitious play prescribes that agent n play any strategy s_n^{t+1} at time $t+1$ that satisfies the following:

$$s_n^{t+1} \in \arg\max_{s_n \in S_n} \mathbb{E}_n^{t+1}[\pi_n(s_n, s)] \tag{3}$$

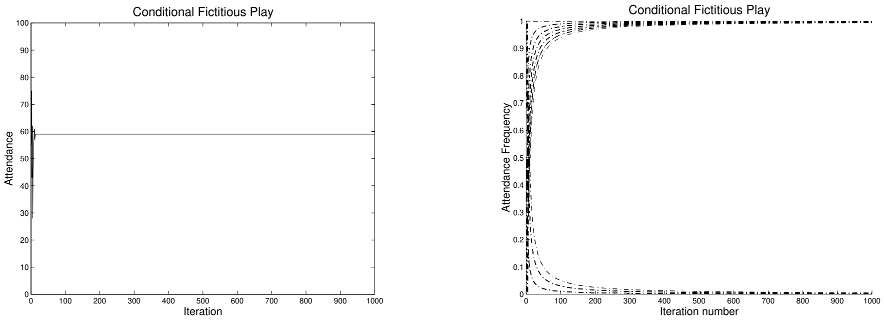
By definition, fictitious play agents maximize their utilities given their beliefs; thus, fictitious play is a rational learning algorithm. By the argument put forth in the introduction, this formulation of fictitious play leads to oscillatory behavior that does not converge to Nash equilibrium in SFBP, assuming $\alpha_n = \alpha$ for all agents n.

But now consider the following variation of fictitious play: each agent computes the empirical distribution over the two events $\{$*uncrowded*, *crowded*$\}$, *conditioned on its own action*. This algorithm is not subject to the paradoxical outcome of the previous algorithm, where beliefs were homogeneous by design, because conditioning on actions leads to *heterogeneous beliefs*. In particular, let $b_n^0 = 0$, $c_n^0 = 0$, and

$$b_n^{t+1} = b_n^t + \begin{cases} 1 & \text{if } s^t \le c \text{ and } s_n^t = 1 \\ 0 & \text{if } s^t > c \text{ and } s_n^t = 1 \end{cases} \qquad c_n^{t+1} = c_n^t + \begin{cases} 1 & \text{if } s_n^t = 1 \\ 0 & \text{otherwise} \end{cases}$$

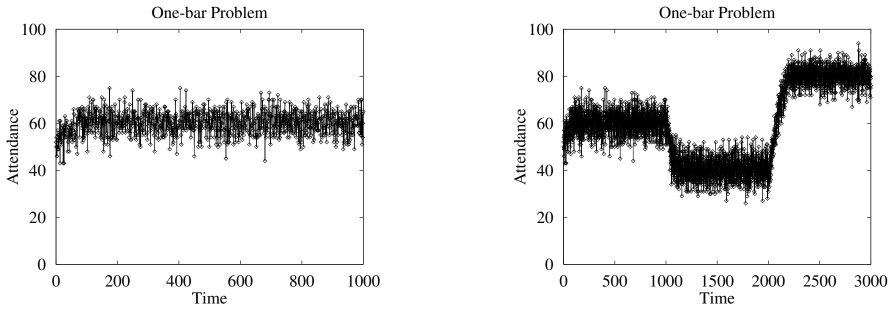
Now for all times $t > 0$, the conditional probability that the bar is uncrowded and agent n attends the bar is $p_n^t = b_n^t/c_n^t$. The expected utility for agent n at time $t+1$ is computed exactly as before, but in terms of the conditional probability p_n^t. And as above, this version of fictitious play based on conditional probabilities prescribes play that is utility-maximizing with respect to one's beliefs.

Interestingly, this conditional fictitious play algorithm is not subject to the paradox of rational learning in SFBP. Fig. 1 depicts the results of simulations of this algorithm, for $N = 100$, $c = 60$, and $\alpha_n = 1$ for all agents n. Total attendance converges to exactly 60, and the likelihood of each individual agent attending the bar converges to either 1 or 0. In other words, conditional fictitious play converges to a pure strategy Nash equilibrium in SFBP. As stated above, this outcome is efficient, but it is not fair. In the next section, we study no-regret learning in SFBP, which obtains an outcome that is fair, but is not efficient.



(a) Total Attendance. (b) Individual Frequencies.

Figure 1: Conditional Fictitious Play.



(a) $N = 100$ and $c = 60$. (b) $c = 60$, $c = 40$, and $c = 80$.

Figure 2: No-regret learning.

4 On Learning a Fair Outcome

We now study learning among agents that are not rational. On the contrary, we consider boundedly-rational agents that exhibit *no-regret*. No-regret learning algorithms do not maintain models over the space of opponents' strategies or utility functions. Instead, they specify that agents *explore* their own strategy space by playing all pure strategies with some non-zero probability, and *exploit* successful strategies by increasing the probability of employing those strategies that generate high utilities. Unlike model-based learning, simple techniques of this nature do not rely on any complex modeling of prior probabilities over possible states of the world. Also, unlike Arthur's original approach, they are are not based on inherently complex models of human cognition.

Freund and Schapire [2] study a no-regret learning algorithm based on an exponential updating scheme. Let $P_n^t(s_n)$ denote the cumulative utility obtained by agent n through time t by employing 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:

$$q_n^{t+1}(s_n) = \frac{(1+\beta)^{P_n^t(s_n)}}{\sum_{s'_n \in S_n} (1+\beta)^{P_n^t(s'_n)}} \tag{4}$$

Fig. 2(a) plots attendance at the bar over time, assuming 100 agents employ this algorithm with $\beta = 0.01$, $c = 60$ and as above, $\alpha_n = 1$ for all agents n. Attendance centers around 60, although it does not converge to exactly 60 as in Fig. 1(a). Specifically, the mean attendance is 60.04 and the variance is 5.11. Moreover, these results are robust in the sense that the agents readily adapt if ever the capacity of the bar changes (see Fig. 2(b)) ($\beta = 0.05$).

Learning via this and similar no-regret algorithms [3], which do not necessitate perfectly rational behavior, yield *mixed strategy* Nash equilibrium outcomes in SFBP. As argued previously, such outcomes are fair (since on average all agents attend the bar 60% of the time) but inefficient, since collective utility is near zero. Thus, both fictitious play and no-regret learning algorithms converge to (different) Nash equilibria in SFBP. But these algorithms have excessive informational requirements: an agent somehow knows whether or not the bar is crowded even if s/he does not attend. In the next section, we simulate Q-learning, a reinforcement learning algorithm that does not rely on any knowledge other than the utility associated with the strategy that the agent actually employs.

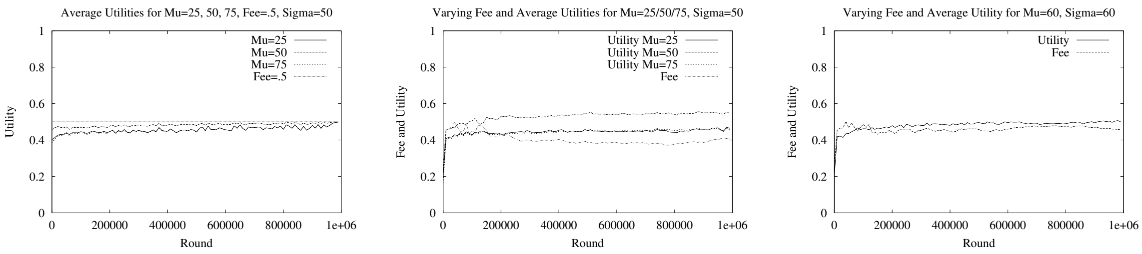


Figure 3: (a) A heterogeneous population of Q-learners learn the equilibrium of the mechanism. (b) A heterogeneous population of Q-learners learn the equilibrium of the mechanism while a derivative follower simultaneously implements an efficient outcome. (c) A homogeneous population of Q-learners learn the equilibrium of the mechanism while a derivative follower simultaneously implements a fair and efficient outcome.

5 A Fair and Efficient Mechanism

In Arthur's original formulation of SFBP, it is assumed that all agents have identical utilities, and moreover that these utilities are known. Now suppose that the value to each agent of going to the bar is unknown to a central planner, say the mayor, whose goal is to design a mechanism, the equilibria of which maximize the total utility of the townspeople. Assume each agent i's utility u_i is an endogenous function of the other agents' actions: let u_i be a concave function of the attendance at the bar, say λ, and let μ_i be the point at which agent i's utility peaks: *i.e.*, $u_i(\lambda) = \max\{1 - (\lambda - \mu_i)^2/\sigma_i^2, 0\}$ for $\sigma_i \geq 0$. In this setting, total utility is maximized at the median value of the μ_i's, say μ_i^*, with those agents whose valuations are closest to μ_i^* attending the bar. Thus, if 25 agents' utilities peak at 25, 50 agents' utilities peak at 50, and 25 agents' utilities peak at 75, with $\sigma_i = 50$ for all agents i, total utility is maximized if the 50 agents for which $\mu_i = 50$ go to the bar. Can the mayor elicit this efficient outcome?

Let us begin our analysis by giving the mayor access to an oracle that informs him of all the agents' peak valuations. In this case, he can compute the median, which is 50 in our example. But can he induce precisely those 50 agents whose peak utilities occur at 50 to attend the bar so as to maximize total utility? The following taxation scheme is designed to achieve this outcome [4]: charge agents that attend the bar an entrance 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 utilities to those λ agents are $u_i(\lambda) - x$, if the bar is uncrowded, and $-(u_i(\lambda) - x)$, if the bar is crowded. The utilities to those $N - \lambda$ agents that do not go to the bar are $\lambda x/(N - \lambda)$. An entrance fee $x = 0.5$ yields an equilibrium in which the 50 agents whose peak valuations are 50 attend the bar, and the remaining 50 agents stay home. All agents achieve utility 0.5 in equilibrium. Allowing Q-learning [12] agents to adapt their behavior to this mechanism, average utility per agent approaches this equilibrium value (see Fig. 5(a)).

But in fact the mayor does not have access to an oracle that informs him of all the agents' peak valuations. Thus, he cannot compute the median, and he cannot compute the equilibrating fee. His knowledge is constrained; likewise, his behavior is taken to be boundedly rational. Assume the mayor sets a fee according to his beliefs about the population of Santa Fe, and updates the fee weekly according to the trend in utility. One simple algorithm for adjusting the fee is *derivative-following* [5], which experiments with incremental changes, continuing to move in the same direction until average utility decreases, at which point the direction of movement is reversed: given increment $\gamma > 0$, the fee $f_{t+1} = f_t + \gamma[\text{sign}(f_t - f_{t-1})\text{sign}(u_t - u_{t-1})]$, where u_t is the agents' average utility at time t. Fig. 5(b) depicts simulations not only of boundedly rational agents adapting their behavior to the mechanism, but the mayor adjusts the mechanism in response to the agents' collective choices as well. Both the fee and average utilities per agent converge near the equilibrium value of 0.5.

Now suppose the agents' valuation functions are such that $\mu_i = \sigma_i = 60$ for all i. In this setting, it is appropriate to seek not only as an efficient outcome, but among all efficient outcomes, an outcome should be *fair*: *i.e.*, agents with identical utilities should be equally likely to attend the bar. Q-learning agents together with a mayor that abides by the derivative-following algorithm generate fair outcomes in this scenario that approximate efficiency (see Fig. 5(c).) Thus, in this paper we have achieved the following: (i) defined conditional fictitious play, a rational learning algorithm that learns pure strategy Nash equilibria in SFBP, thereby avoiding the rational learning paradox; (ii) established empirically that no-regret learning converges to the symmetric mixed strategy Nash equilibrium in SFBP; and (iii) implemented a taxation mechanism that elicits fair and efficient collective behavior among Q-learning agents.

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. Greenwald, A. Jafari, G. Ercal, and D. Gondek. On no-regret learning, Nash equilibrium, and fictitious play. In *Proceedings of Eighteenth International Conference on Machine Learning*, pages 226–233, June 2001.

[4] A. Greenwald, B. Mishra, and R. Parikh. The Santa Fe bar problem revisited: Theoretical and practical implications. Presented at *Stonybrook Festival on Game Theory: Interactive Dynamics and Learning*, July 1998.

[5] A.R. Greenwald and J.O. Kephart. Shopbots and pricebots. In *Proceedings of Sixteenth International Joint Conference on Artificial Intelligence*, volume 1, pages 506–511, August 1999.

[6] G. Hardin. The tragedy of the commons. *Science*, 162:1243–1248, 1968.

[7] R.B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, 1991.

[8] J. Nash. Non-cooperative games. *Annals of Mathematics*, 54:286–295, 1951.

[9] M. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, Cambridge, 1994.

[10] J. Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54:298–301, 1951.

[11] T. Schelling. *Micromotives and Macrobehavior*. W.W. Norton and Company, New York, 1978.

[12] C. Watkins. *Learning from Delayed Rewards*. PhD thesis, Cambridge University, Cambridge, 1989.