On the Connection between

# No-Regret Learning, Fictitious Play, & Nash Equilibrium

Amy Greenwald

Brown University

Gunes Ercal, David Gondek, Amir Jafari

July, 2000

# Games

A game is a tuple $\Gamma = (\mathcal{I}, (S_i, r_i)_{i \in \mathcal{I}})$ where

- $\mathcal{I}$ is a set of players ($i \in \mathcal{I}$)
- $S_i$ is a set of (pure) strategies ($s_i \in S_i$)
- $r_i : S \to \mathbb{R}$ is a reward function ($s \in S = \prod_i S_i$)

A repeated game is a sequence of tuples $\Gamma^T$

# Nash Equilibrium

A mixed strategy $q_i$ is a probability distribution over pure strategy set $S_i$.

Given opposing strategy $q_{-i}$, a best-response $q_i^*$ is a strategy that maximizes (one-step) reward:

$$q_i^* = \arg\max_{q_i} r_i(q_i, q_{-i})$$

A Nash equilibrium $q^* = (q_i^*, q_{-i}^*)$ is a strategy profile *s.t.* $q_i^*$ is a best-response to $q_{-i}^*$, for all players $i$.

All (finite) games have mixed strategy Nash equilibria.

# No-Regret

Regret is the expected difference in rewards for playing mixed strategy $q_i^t$ rather than pure strategy $s_i$:

$$\rho_i^t(s_i) = r_i(s_i, q_{-i}^t) - r_i(q_i^t, q_{-i}^t)$$

An on-line learning algorithm exhibits no-regret iff $\forall s_i$:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \rho_i^t(s_i) \leq 0$$

# Learning Algorithms

Narendra and Thathachar Learning Automata

Freund and Schapire Multiplicative Updating

Foster and Vohra The Mixing Method

Hart and Mas-Colell Correlated Equilibrium

Hannan, Banos, Megiddo, etc.

## Observation

In an infinitely repeated game, if all players learn via no-regret algorithms, and if $q_i^t \to \bar{q}_i$ for all players $i$, then $\bar{q} = q^*$: *i.e.*, $\bar{q}$ is a mixed strategy Nash equilibrium.

# Proof

By assumption,

$$r_i(s_i, q_{-i}^t) \to r_i(s_i, \bar{q}_{-i})$$

$$r_i(q_i^t, q_{-i}^t) \to r_i(q_i^*, \bar{q}_{-i})$$

Therefore,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} r_i(s_i, q_{-i}^t) = r_i(s_i, \bar{q}_{-i})$$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} r_i(q_i^t, q_{-i}^t) = r_i(\bar{q}_i, \bar{q}_{-i})$$

By the no-regret property

$$\begin{aligned} & r_i(s_i, \bar{q}_{-i}) \le r_i(\bar{q}_i, \bar{q}_{-i}), \text{ for all } s_i \\ \Rightarrow \quad & r_i(s_i^*, \bar{q}_{-i}) \le r_i(\bar{q}_i, \bar{q}_{-i}) \end{aligned}$$

where $s_i^* \in \arg\max_{s_i \in S_i} r_i(s_i, \bar{q}_{-i})$: *i.e.*, $\bar{q} = q^*$.

# Simulations

## Informed Settings

$$\rho_{s_i}^T = \sum_{t=1}^{T} r_i(s_i, q_{-i}^t)$$

$$q_{s_i}^{T+1} = \frac{(1+\alpha)^{\rho_{s_i}^T}}{\sum_{\bar{s}_i \in S_i} (1+\alpha)^{\rho_{\bar{s}_i}^T}}$$

[Freund and Schapire 1995]

- $\alpha > 0$ is the learning rate

## Naive Settings

$$\begin{aligned} \widehat{r}(s_i, q^t_{-i}) &= 1^t_{s_i} \frac{r(s_i, q^t_{-i})}{\widehat{q}^t_{s_i}} \\ \widehat{q}^{T+1}_i &= (1-\epsilon) q^{T+1}_i + \frac{\epsilon}{|S_i|} \end{aligned}$$

[Auer, Cesa-Bianchi, Freund, and Schapire 1996]

- $\epsilon > 0$ is the exploration rate

# Pure Strategy Nash Equilibria

## Prisoners' Dilemma

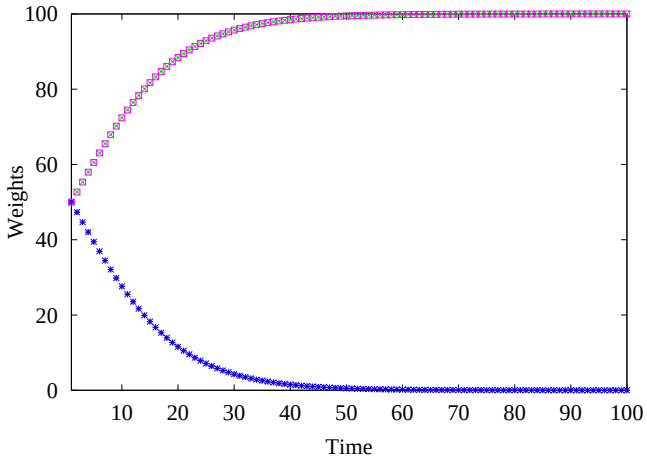
| 1 \ 2 | $C$ | $D$ |
|---|---|---|
| $C$ | 4,4 | 0,5 |
| $D$ | 5,0 | 1,1 |

## Battle of the Sexes

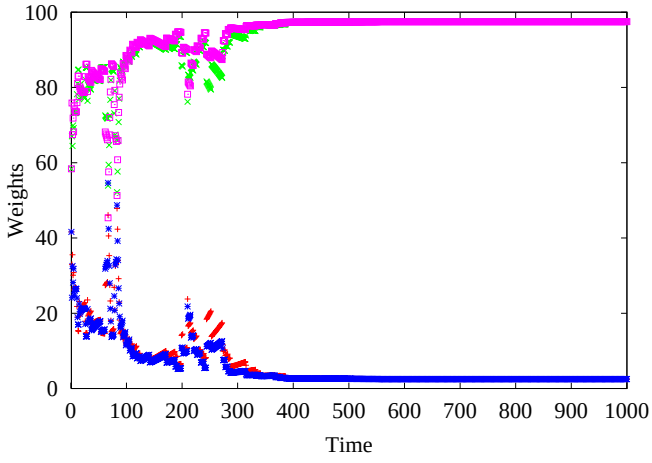
| 1 \ 2 | $B$ | $F$ |
|---|---|---|
| $B$ | 2,1 | 0,0 |
| $F$ | 0,0 | 1,2 |

# Prisoners' Dilemma

## Informed Setting

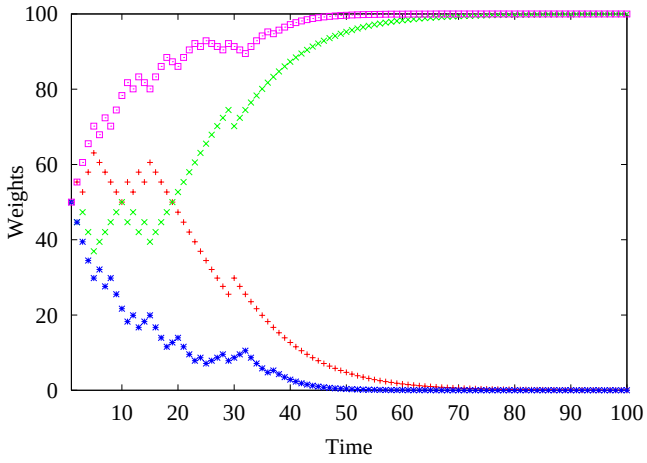



## Naive Setting

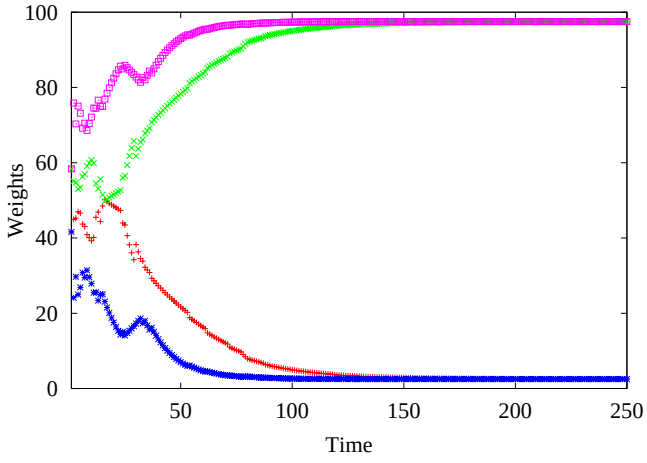


# Battle of the Sexes

## Informed Setting




## Naive Setting



# Mixed Strategy Nash Equilibria

## Matching Pennies

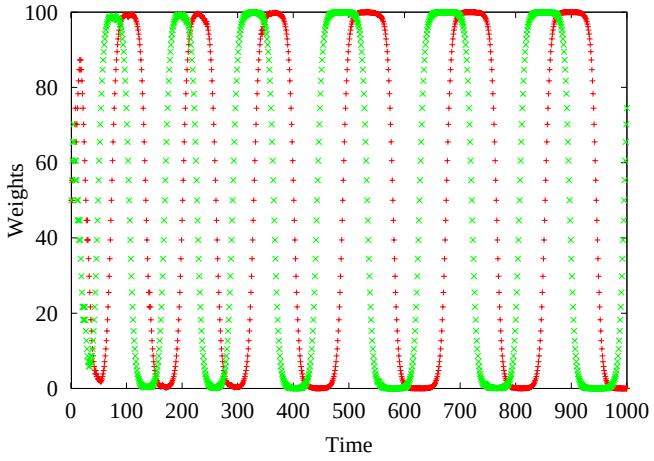
| 1 \ 2 | $H$ | $T$ |
|---|---|---|
| $H$ | 1,0 | 0,1 |
| $T$ | 0,1 | 1,0 |

## Shapley Game

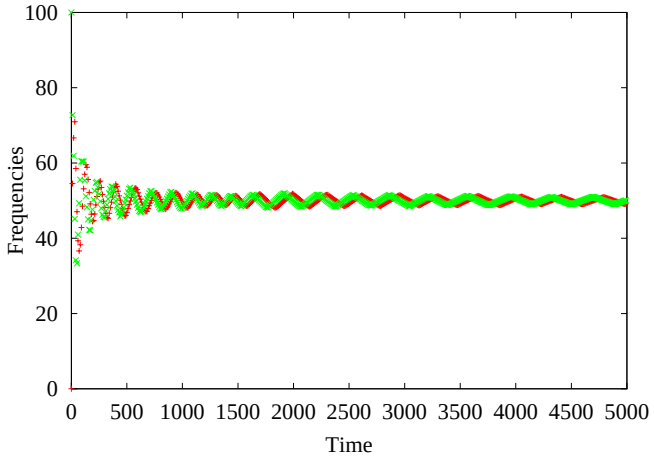
| 1 \ 2 | $L$ | $C$ | $R$ |
|---|---|---|---|
| $T$ | 1,0 | 0,1 | 0,0 |
| $M$ | 0,0 | 1,0 | 0,1 |
| $B$ | 0,1 | 0,0 | 1,0 |

# Matching Pennies

## Mixed Strategies

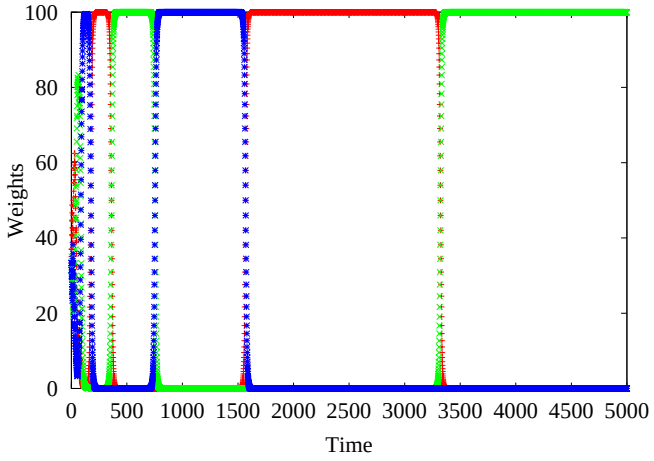



## Empirical Frequencies

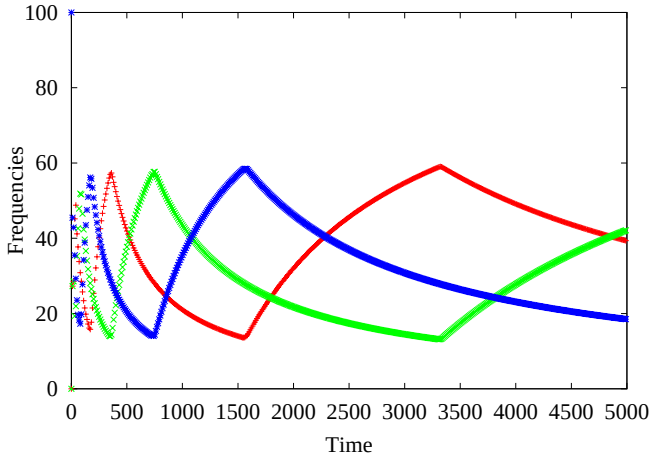


# Shapley Game

## Mixed Strategies

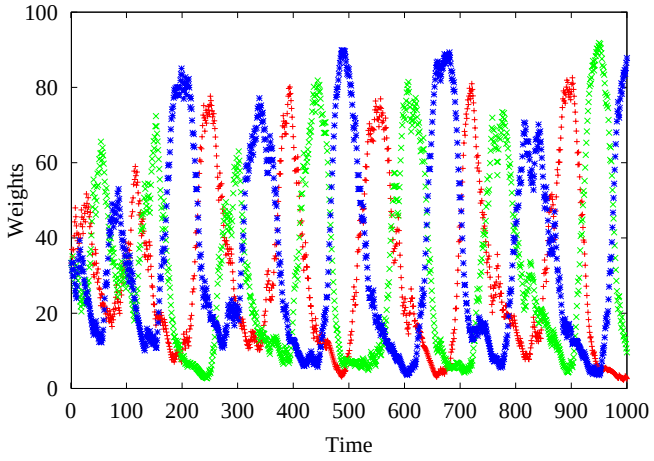



## Empirical Frequencies

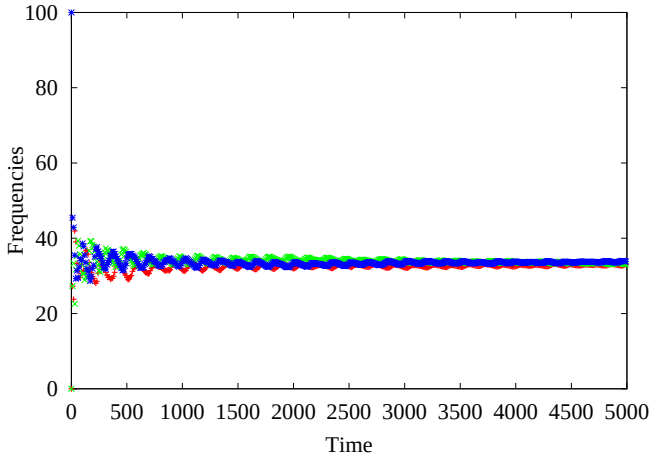


# Shapley Game

## Smoothed Mixed Strategies




## Smoothed Empirical Frequencies



# Summary of Simulations

## Theoretical

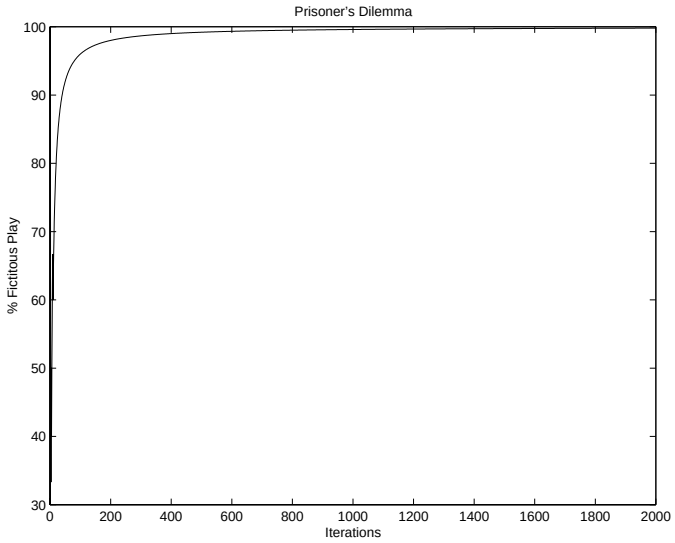
- If no-regret learning converges, it converges to Nash equilibrium.

## Experimental

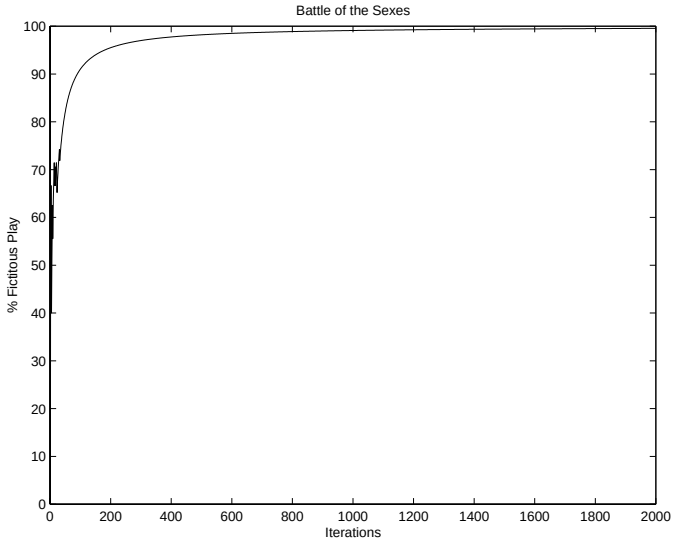
- No-regret learning converges to pure strategy Nash equilibria in games for which such equilibria exist.
- No-regret learning does not converge otherwise.
  - Empirical frequencies converge to Nash in zero-sum games.
  - Empirical frequencies converge to Nash in non-zero sum games of 2 strategies.
  - Smoothed empirical frequencies converge to Nash in games of $> 2$ strategies.

# No-Regret and Fictitious Play

## Prisoners' Dilemma

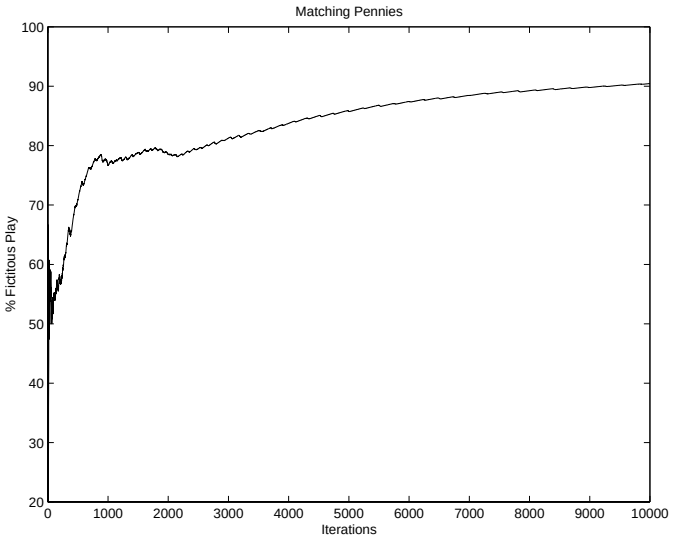


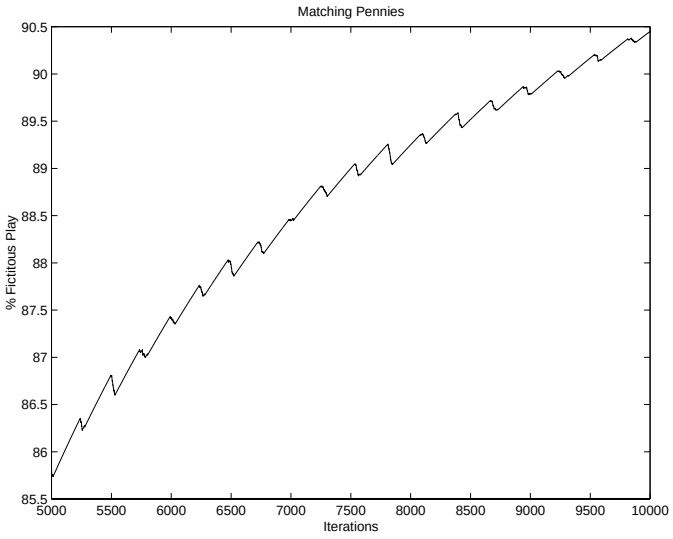

## Battle of the Sexes



# No-Regret and Fictitious Play

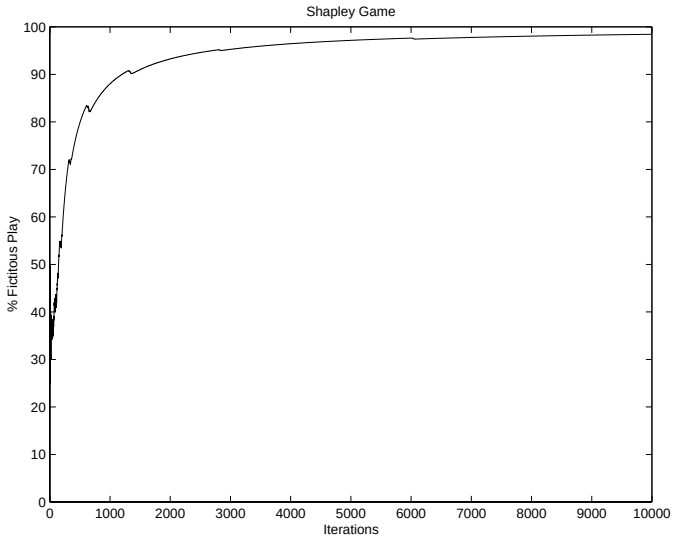
## Matching Pennies

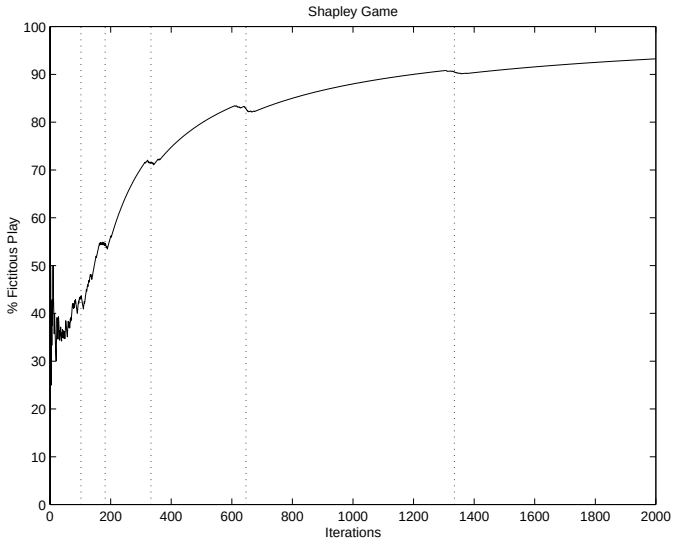


# No-Regret and Fictitious Play

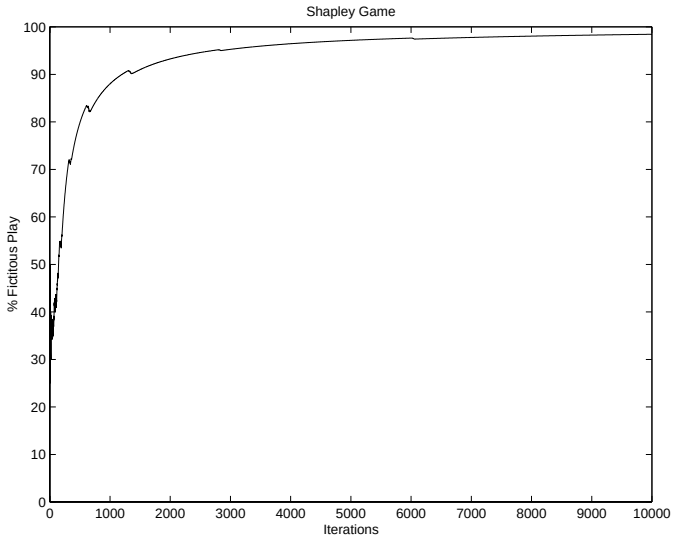
## Shapley Game

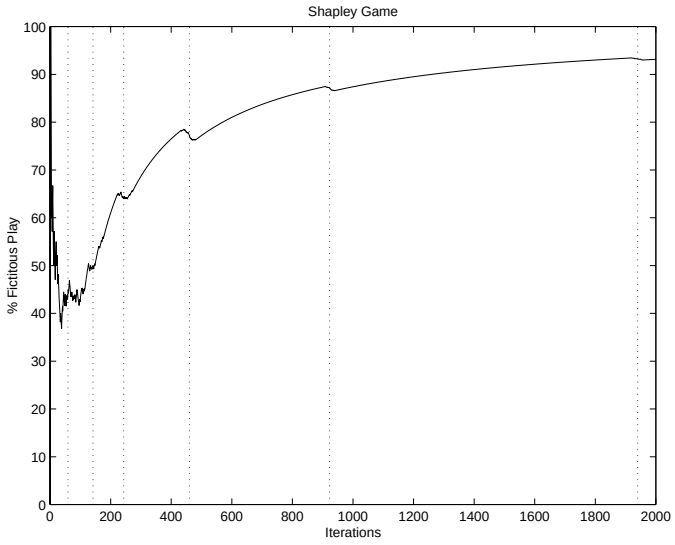


# No-Regret and Fictitious Play

## Shapley Game



## Papers

- Shopbots and Pricebots [w/ Kephart]
- Santa Fe Bar Problem [w/ Mishra & Parikh]
- Learning in Networks [w/ Friedman & Shenker]

http://www.cs.brown.edu/people/amygreen