UNIVERSITY OF LIVERPOOL

COMP 532
Machine Learning and BioInspired Optimization

Lecture 17: Multi-Agent Learning

Dr. Shan Luo

Department of Computer Science

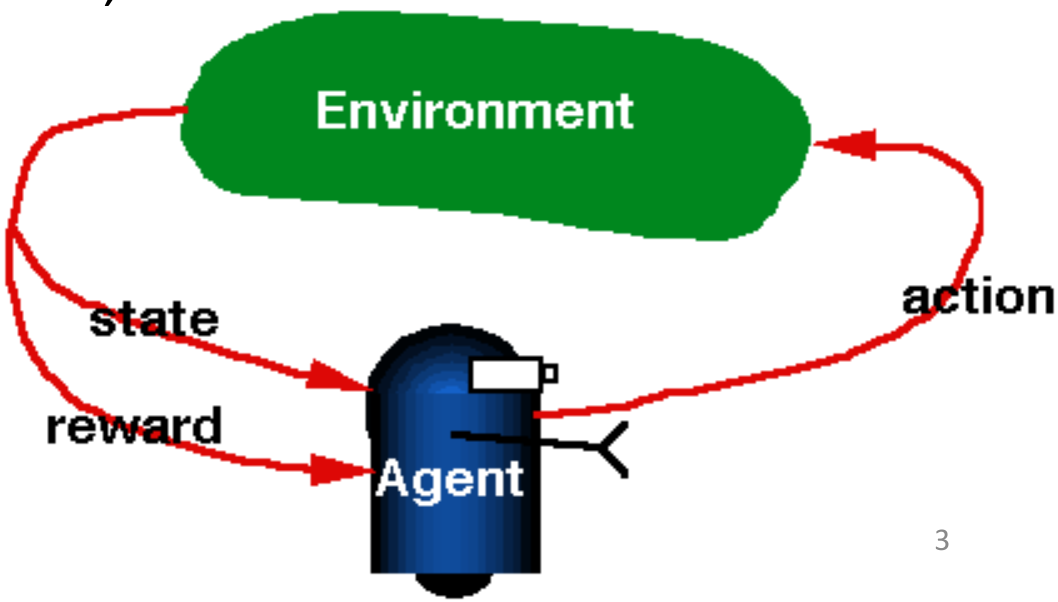
shan.luo@liverpool.ac.uk

Outline (two lectures)

- **Problem revisited**
- **Game Theory**
  - Repeated Games
  - Stochastic or Markov Games
- **“Naïve” approaches to multi-agent learning**
  - Fictitious play
- **More on stochastic games**
- **Spectrum of approaches**
- **Basic/state-of-the-art approaches**
  - minimax-Q, Nash-Q
  - tinkering with learning rates: WoLF (Bowling)
- **Challenges and Opportunities**

Problem revisited
Normal single-agent learning

- Assume that environment has observable states, characterizable expected rewards and state transitions, and all of the above is stationary (MDP-ish)
- Non-learning, theoretical solution to fully specified problem: Dynamic Programming formalism
- Learning: solve by trial and error without a full specification: RL + exploration, Monte Carlo, ...



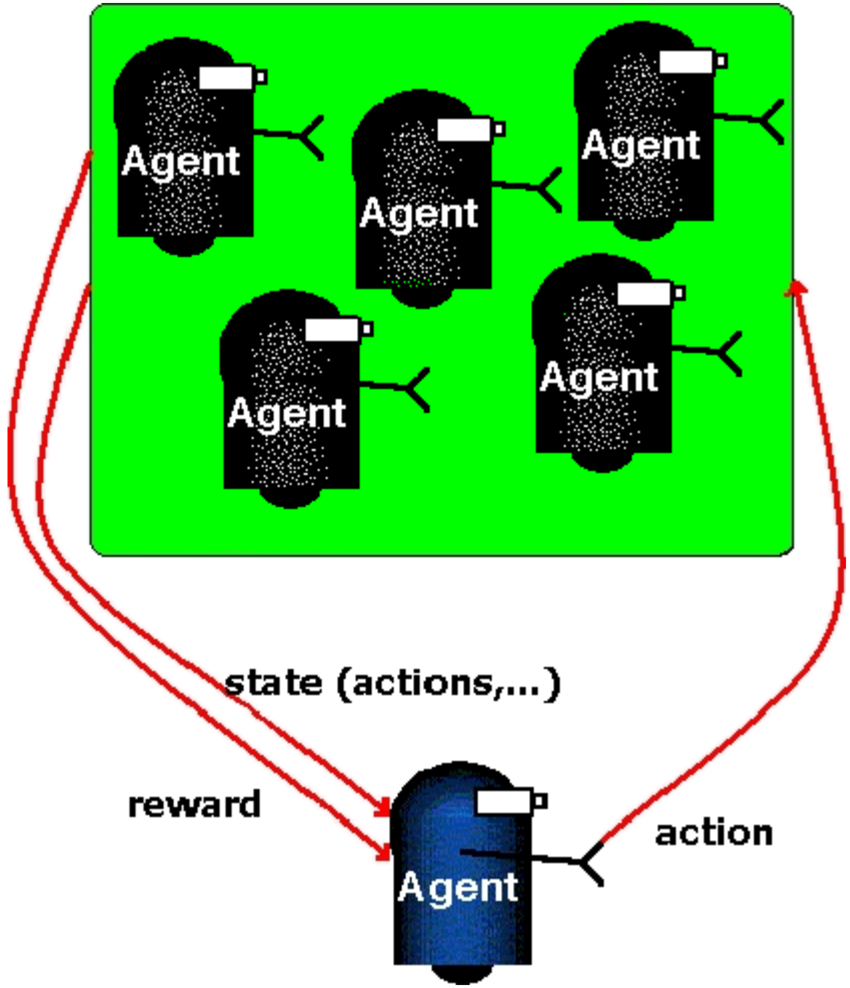
Problem revisited
Multi-Agent Learning Problem:

- Agent tries to solve its learning problem, while other agents in the environment also are trying to solve their own learning problems. challenging non-stationarity.

- Main scenarios: (1) cooperative; (2) self-interested

- Agent may know very little about other agents:
  - payoffs may be unknown
  - learning algorithms unknown

- Traditional method of solution: **game theory**



Problem revisited
Multi-Agent Learning Problem:

- A precise problem formulation is still lacking!

- Some MAL objectives:
  - Learning should converge to a stationary strategy
  - In “self-play” learning (all agents use the same learning algorithm), learners should jointly converge to an equilibrium strategy
  - Learning should achieve payoffs as good as a best-response to other agents’ strategies
  - (Worst case bound) Learning should guarantee a minimum payoff

Game Theory

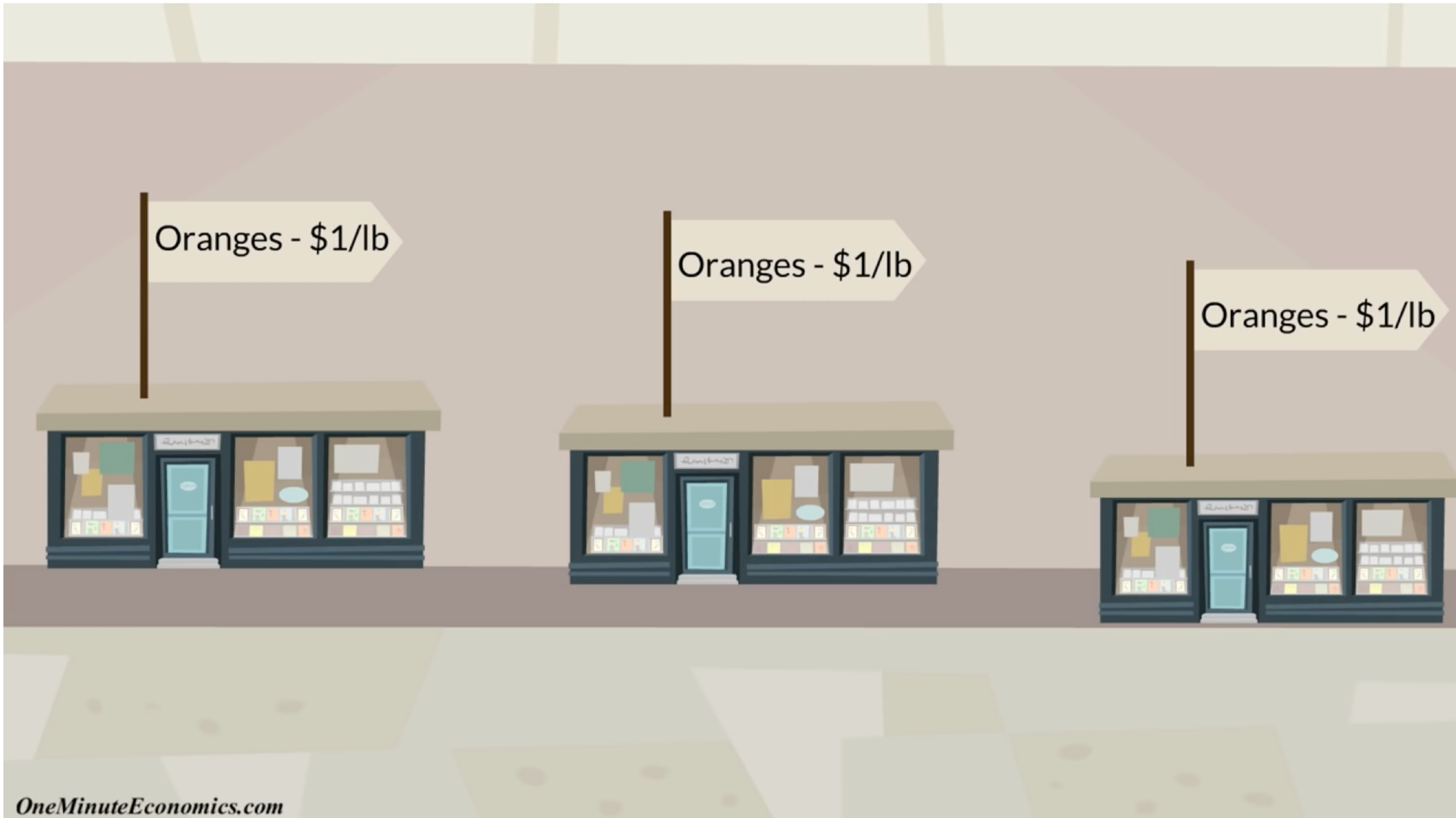
- Essential theoretical / conceptual tools for studying multi-agent learning

Wikipedia definition:

> Game theory is most often described as a branch of applied mathematics and economics that studies situations where players choose different actions in an attempt to maximize their returns. The essential feature, however, is that it provides a formal modelling approach to social situations in which decision makers interact with other minds.

- Today, widely used in politics, business, economics, biology, psychology, computer science etc.

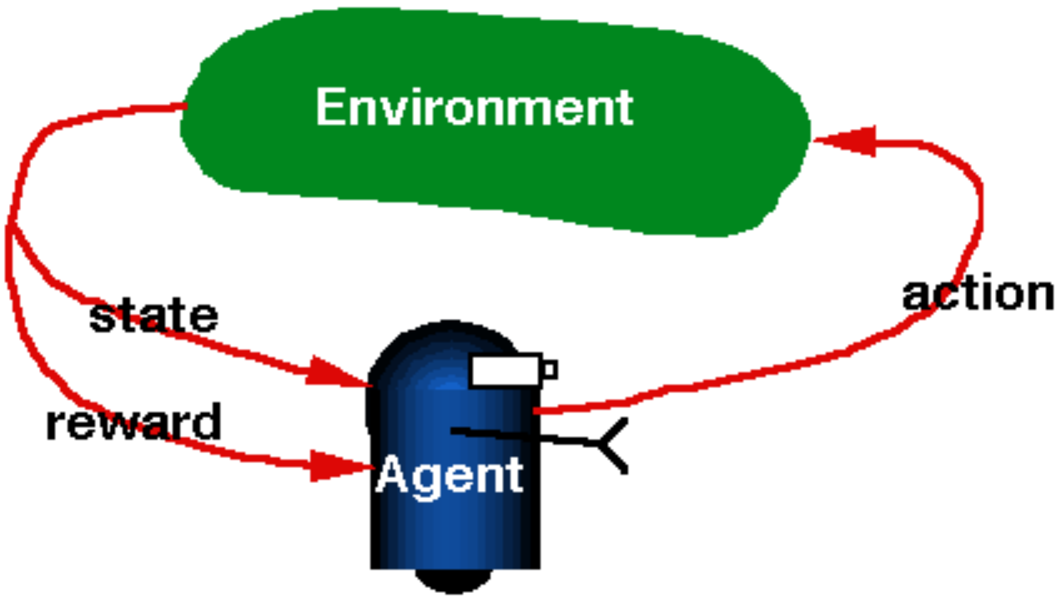
Game Theory



https://www.youtube.com/watch?v=YueJukoFBMU

Traditional Game Theory

- Rational or perfectly logical players
- A rational **player/agent** will make decisions that maximize her individual expected utility (= expected payoff for simplicity) given her understanding/beliefs about the problem. Also, perfectly indifferent to payoffs received by other players.



Game Theory

- A game is specified by: players (1…n), actions, and (expected) payoff matrices (functions of joint actions)

B's action

$$
\begin{array}{cccc}
 & R & P & S \\
R & 0 & -1 & +1 \\
P & +1 & 0 & -1 \\
S & -1 & +1 & 0
\end{array}
\qquad
\begin{array}{cccc}
 & R & P & S \\
R & 0 & +1 & -1 \\
P & -1 & 0 & +1 \\
S & +1 & -1 & 0
\end{array}
$$

A's action

A's payoff B's payoff

A Rock, Paper, Scissors game

- If payoff matrices are "identical", A and B are cooperative, else non-cooperative (zero-sum = purely competitive)
- AKA games in Normal Form.

Normal-Form Games

- n players, A_i action set for player i
- A is joint action space: A_1 x … x A_n.
- R_i reward to player i as a function of the joint action A.

$$R_1 \quad \begin{array}{cc} & a_2 \\ a_1 & \begin{bmatrix} & R_1(a) & \end{bmatrix} \end{array} \qquad R_2 \quad \begin{array}{cc} & a_2 \\ a_1 & \begin{bmatrix} & R_2(a) & \end{bmatrix} \end{array}$$

Game Theory: Basic lingo

- Games with no states: matrix games
- Games with states: stochastic games, Markov games; (state transitions are functions of joint actions)

- Games with simultaneous moves: normal form
- Games with alternating turns: extensive form

- No. of rounds = 1: one-shot game
- No. of rounds > 1: repeated game

- deterministic action policy: pure strategy
- non-deterministic action policy: mixed strategy
  e.g. Prob(R,P,S) = (½,¼,¼)

Stochastic vs. Matrix Games

- A stochastic game (a.k.a. "Markov game") generalizes MDPs to multiple agents
  - set of **n** agents / players — $\mathbf{i \in I}$
  - finite state space — $\mathbf{s \in S}$
  - action set — $\mathbf{<a_1(s),\ldots,a_n(s)>}$
  - stationary reward distribution — $\mathbf{r_i(s, <a_1,\ldots,a_n>)}$
  - stationary transition probabilities — $\mathbf{P(s'|s, <a_1,\ldots,a_n>)}$

- A matrix game has no state information, only actions and payoffs (**|S| = 1**)

Basic Analysis

$\boxed{x = \pi}$

- Agent i's mixed strategy $\boldsymbol{x}_i$ is a best-response to others' $\boldsymbol{x}_{-i}$ if it maximizes payoff given $\boldsymbol{x}_{-i}$
- $\boldsymbol{x}_i$ is a dominant strategy if it maximizes payoff regardless of what others do
- A joint strategy $\boldsymbol{x}$ is an equilibrium if each agent's strategy is simultaneously a best-response to everyone else's strategy, i.e., no incentive to deviate. Nash equilibrium is the main one, but there are others (e.g. correlated equilibrium)
- A Nash equilibrium always exists, but there may be exponentially many of them, and very hard to compute

equilibrium selection is a big problem
(players need to agree on which eqm to choose)

Basic Analysis

https://www.youtube.com/watch?v=ZoZe2tPd3Hk

Real-Life vs. Game Theory games

- NFL playoffs
- World Series of Poker
- World of Warcraft
- Buying a house
- Salary negotiations
- Competitive pricing:
  - Best Buy vs. Circuit City
  - Airline fare wars
- OPEC production cuts
- NASDAQ, NYSE, …
- FCC spectrum auctions

- Matching Pennies
- Rock-Paper-Scissors
- Prisoners' Dilemma
- Battle-of-the-Sexes
- Chicken
- Ultimatum

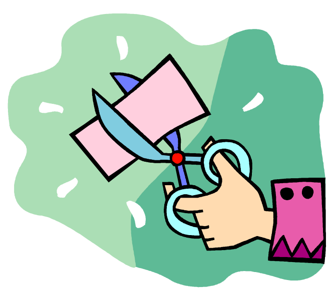
Assumptions in Normal-Form Games

- Game specification is fully known; actions and payoffs are fully observable by all players (not scalable!)
- Players act “simultaneously”, i.e. without observing actions of others
- Assume no communication between players, or it doesn’t affect play (communication is “cheap talk”)
- Basic analysis assumes the game is only played once (called one-shot)

Presentation of Rock Paper Scissors Payoffs in a Bimatrix

Column player

Row player		R	P	S
	R	0, 0	-1, +1	+1, -1
	P	+1, -1	0, 0	-1, +1
	S	-1, +1	+1, -1	0, 0



- This is a **zero-sum game** since for each pair of joint actions, the players’ payoffs add up to zero.
- This is a **symmetric game:** invariant under swapping of player labels
- This game has a unique mixed strategy Nash equilibrium: both players play uniform random strategies:
  Prob(R, P, S) = (1/3, 1/3, 1/3)

Prisoners’ Dilemma Game

Prisoner 2

Prisoner 1		Confess (Defect)	Hold out (Cooperate)
	Confess (Defect)	1 1	5 0
	Hold out (Cooperate)	0 5	3 3

Prisoners' Dilemma Game

Prisoner 2

Prisoner 1		Confess (Defect)	Hold out (Cooperate)
	Confess (Defect)	1 / 1	5 / 0
	Hold out (Cooperate)	0 / 5	3 / 3

Whatever Prisoner 2 does,
the best that Prisoner 1 can do is Confess

Prisoners' Dilemma Game

		Prisoner 2	
		Confess (Defect)	Hold out (Cooperate)
Prisoner 1	Confess (Defect)	1 1	5 0
	Hold out (Cooperate)	0 5	3 3

Whatever Prisoner 1 does,
the best that Prisoner 2 can do is Confess

Prisoners' Dilemma Game

Prisoner 2

Prisoner 1		Confess (Defect)	Hold out (Cooperate)
	Confess (Defect)	1 1	5 0
	Hold out (Cooperate)	0 5	3 3

Each player has a dominant strategy to Confess.

The dominant strategy equilibrium is (Confess, Confess)

A strategy is a dominant strategy if it is a player's strictly best response to any strategies the other players might pick.

A dominant strategy equilibrium is a strategy combination consisting of each players dominant strategy.

Prisoners’ Dilemma Game

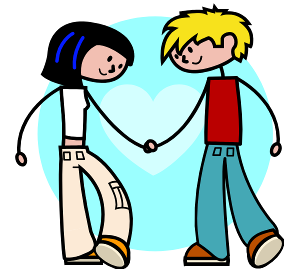
Prisoner 2

		Confess (Defect)	Hold out (Cooperate)
Prisoner 1	Confess (Defect)	1 1	5 0
	Hold out (Cooperate)	0 5	3 3

The payoff in the dominant strategy equilibrium (1,1) is worse for both players than (3,3), the payoff in the case that both players hold out. Thus, the Prisoners’ Dilemma Game is a game of social conflict.

Opportunity for multi-agent learning: by learning during repeated play, the Pareto optimal solution (3,3) can emerge as a result of learning (also can arise in evolutionary game theory).

Battle of the Sexes

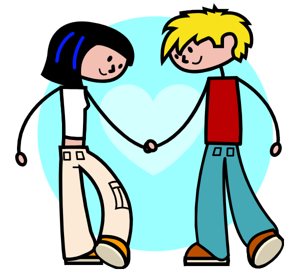


Bob

Alice	Ballet	Football
Ballet	2 1	0 0
Football	0 0	1 2

https://en.wikipedia.org/wiki/Battle_of_the_sexes_(game_theory)

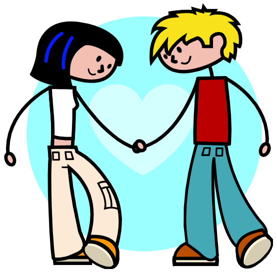
Battle of the Sexes



Bob

Alice		Ballet	Football
	Ballet	2 1	0 0
	Football	0 0	1 2

Battle of the Sexes



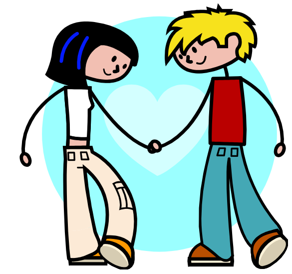
Bob

Alice	Ballet	Football
Ballet	2, 1	0, 0
Football	0, 0	1, 2

This game has

- no (iterated) dominant strategy equilibrium

Battle of the Sexes

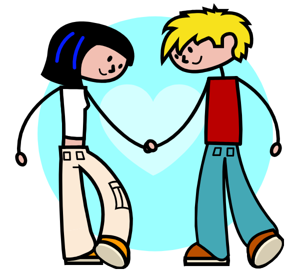


Alice \ Bob	Ballet	Football
Ballet	2, 1	0, 0
Football	0, 0	1, 2

This game has

- no (iterated) dominant strategy equilibrium
- two Nash equilibria (Football, Football) and (Ballet, Ballet)

Battle of the Sexes



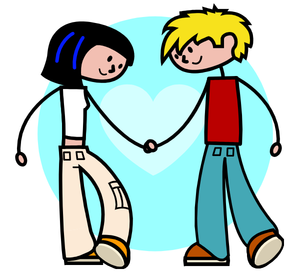
Bob

Alice	Ballet	Football
Ballet	2, 1	0, 0
Football	0, 0	1, 2

This game has two pure Nash equilibria

Can you see what the third equilibrium is?

Battle of the Sexes



		Bob		
		Ballet	Football	
Alice	Ballet	2 / 1	0 / 0	x_B
	Football	0 / 0	1 / 2	x_F
		y_B	y_F	

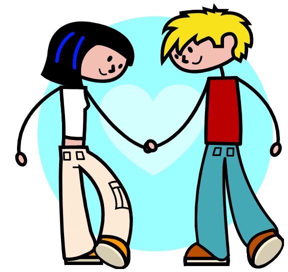
Players will only mix their strategy x when indifferent w.r.t. payoff, given the other player's strategy y. E.g. for Alice:

$$E[r|B] = E[r|F]$$
$$2y_B = y_F$$
$$2y_B = 1 - y_B$$
$$y_B = {}^{1}/_{3}$$
$$y_F = 1 - y_B = {}^{2}/_{3}$$

Similar reasoning for Bob given Alice's mixed strategy

$$x_F = {}^{1}/_{3} \qquad x_B = {}^{2}/_{3}$$

Battle of the Sexes



Bob

		Ballet	Football	
Alice	Ballet	2, 1	0, 0	x_F
	Football	0, 0	1, 2	x_B
		y_F	y_B	

Which Nash equilibrium is best?

Let’s look at the payoffs:

(B, B): Alice gets 2, Bob gets 1
(F, F): Alice gets 1, Bob gets 2

Unfair… what about the mixed NE?

$(2/3, 1/3)$: Both only get 2/3!

Worse payoff due to chance of miscoordination.

How can these two players coordinate ?

Simple Approaches to Multi-Agent Learning

- Fictitious Play
- Independent Q-learning (one end of the spectre)
- Joint Action Learning (other end of the spectre)

Simple Approaches to Multi-Agent Learning

Basic idea: agent adapts, ignoring non-stationarity of other agents' strategies

- **Fictitious play:** Agent observes time-average frequency of other players' action choices, and models:

$$prob(action\, k) = \frac{\#times\ k\ observed}{total\ \#\,observations}$$

  agent then plays best-response to this model

- Variants of fictitious play: exponential recency weighting, "smoothed" best response (~softmax), small adjustment toward best response, ...

What if all agents use fictitious play?

- Strict Nash equilibria are absorbing points for fictitious play
- Typical result is limit-cycle behavior of strategies
- In certain cases, product of empirical distributions converges to Nash even though actual play cycles (penny matching example)

Wrapping up

- Problem revisited
- Game Theory
  - Repeated Games
  - Stochastic or Markov Games
  - Example games
- **“Naïve” approaches to multi-agent learning**
  - Fictitious play