Plans for Remainder of the Course

- **Course Topics**
  1. Multi-agent Learning
  2. Reinforcement learning from human feedback (RLHF)
- **Logistics**: one more HW (likely out by this weekend) and one more quiz (next Thur)

DATA 37200: Learning, Decisions, and Limits
(Winter 2025)

Basics of Game Theory and Multi-agent Learning

Instructor: Haifeng Xu

Outline

- Game Theory and Nash Equilibrium
- Zero-Sum Games: theory and learning

What is Game Theory About?

➢ Consequences of our decisions are often affected by others

- Buying a house
- Choose stocks to invest
- Applying to universities
- …

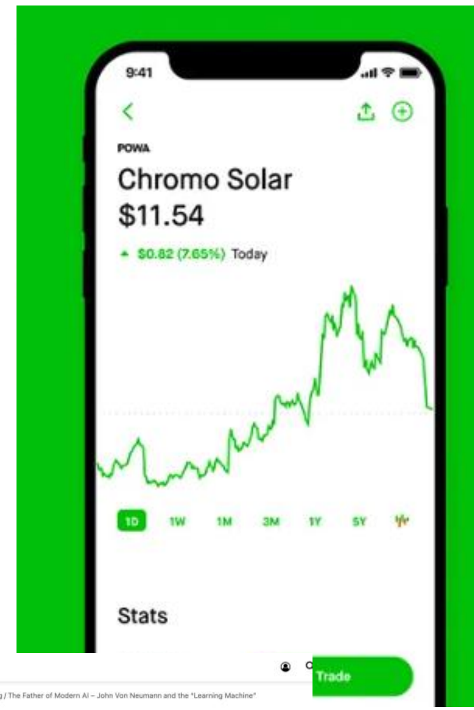
➢ Game theory offers a framework to reason about decisions in such multi-agent situations

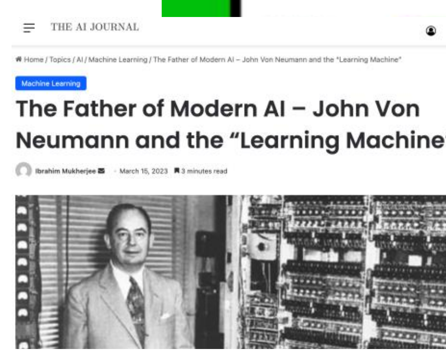
What is Game Theory About?

➢ Consequences of our decisions are often affected by others
- Buying a house
- Choose stocks to invest
- Applying to universities
- …

➢ Game theory offers a framework to reason about decisions in such multi-agent situations

➢ It has deep connections to AI
- For a long period, RL research was primarily motivated by solving games (e.g. checkers, chess, go, poker, etc.)
- In fact, a key founder of AI John von Neumann is a founder of game theory





Example I: Prisoner's Dilemma

- Two members A,B of a criminal gang are arrested
- They are questioned in two separate rooms
  - No communications between them

A \ B	B stays silent	B betrays
A stays silent	-1 -1	0 -3
A betrays	-3 0	-2 -2

Q: How should each prisoner act?

- Betray is always the best action

Example I: Prisoner's Dilemma

- Two members A,B of a criminal gang are arrested
- They are questioned in two separate rooms
  - No communications between them

B A	B stays silent	B betrays
A stays silent	-1 -1	0 -3
A betrays	-3 0	-2 -2

Q: How should each prisoner act?

- Betray is always the best action

Example I: Prisoner's Dilemma

- Two members A,B of a criminal gang are arrested
- They are questioned in two separate rooms
  - No communications between them

A \ B	B stays silent	B betrays
A stays silent	-1, -1	-3, 0
A betrays	0, -3	-2, -2

equilibrium

Q: How should each prisoner act?

- Betray is always the best action
- But, (-1,-1) is a better outcome for both
- Why? What goes wrong?
  - Selfish behaviors lead to inefficient outcome

Example 2: Rock-Paper-Scissor

Player 2

Player 1

	Rock	Paper	Scissor
Rock	(0, 0)	(-1, 1)	(1, -1)
Paper	(1, -1)	(0, 0)	(-1, 1)
Scissor	(-1, 1)	(1, -1)	(0, 0)

Q: what is an equilibrium?

- Need to randomize – any deterministic action pair cannot make both players happy
- Common sense suggests (1/3,1/3,1/3)

Main Components of a Game

- Players: participants of the game, each may be an individual, organization, a machine or an algorithm, etc.
- Strategies: actions available to each player
- Outcome: the profile of player strategies
- Payoffs: a function mapping an outcome to a utility for each player

Normal-Form Representation

- n players, denoted by set $[n] = \{1, \cdots, n\}$
- Player i takes action $a_i \in A_i$
- An outcome is the action profile $a = (a_1, \cdots, a_n)$
  - As a convention, $a_{-i} = (a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_n)$ denotes all actions excluding a_i
- Player i receives payoff $u_i(a)$ for any outcome $a \in \Pi_{i=1}^{n} A_i$
  - $u_i(a) = u_i(a_i, a_{-i})$ depends on other players' actions
- $\{A_i, u_i\}_{i \in [n]}$ are public knowledge

This is the most basic game model

- There are game models with richer and more intricate structures

Illustration: Prisoner's Dilemma

- 2 players: 1 and 2
- $A_i = \{\text{silent}, \ \text{betray}\}$ for $i = 1,2$
- An outcome can be, e.g., $a = (\text{silent}, \ \text{silent})$
- $u_1(a), u_2(a)$ are pre-defined, e.g., $u_1(\text{silent}, \ \text{silent}) = -1$
- The whole game is public knowledge; players take actions simultaneously
  - Equivalently, acts without knowing the others' actions

Equilibrium

- An outcome a^* is an *equilibrium* if no player has incentive to deviate unilaterally. More formally,

$$u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*), \qquad \forall a_i \in A_i$$

  - It is a special case of Nash equilibrium, called *pure strategy NE*

A \ B	B stays silent	B betrays
A stays silent	-1, -1	-3, 0
A betrays	0, -3	-2, -2

Prisoner's Dilemma

Equilibrium

- An outcome a^* is an *equilibrium* if no player has incentive to deviate unilaterally. More formally,

$$u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*), \qquad \forall a_i \in A_i$$

  - It is a special case of Nash equilibrium, called *pure strategy NE*

Quiz: find equilibrium

	L	M	R
U	4,3	5,1	6,2
M	2,1	8,4	3,6
D	3,0	9,6	2, 5

Equilibrium

- An outcome a^* is an *equilibrium* if no player has incentive to deviate unilaterally. More formally,

$$u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*), \qquad \forall a_i \in A_i$$

  - It is a special case of Nash equilibrium, called *pure strategy NE*

Quiz: find equilibrium

	L	M	R
U	4,3	5,1	6,2
M	2,1	8,4	3,6
D	3,0	9,6	2, 5

Equilibrium

➢ An outcome a^* is an *equilibrium* if no player has incentive to deviate unilaterally. More formally,

$$u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*), \qquad \forall a_i \in A_i$$

- It is a special case of Nash equilibrium, called *pure strategy NE*

What about this?

	Rock	Paper	Scissor
Rock	(0, 0)	(-1, 1)	(1, -1)
Paper	(1, -1)	(0, 0)	(-1, 1)
Scissor	(-1, 1)	(1, -1)	(0, 0)

Pure strategy NE does not always exist…

Pure vs Mixed Strategy

- Pure strategy: take an action deterministically
- Mixed strategy: can randomize over actions
  - Described by a distribution x_i where $x_i(a_i)$ = prob. of taking action a_i
  - $|A_i|$-dimensional simplex $\Delta_{A_i} := \{x_i: \sum_{a_i \in A_i} x_i(a_i) = 1, x_i(a_i) \geq 0\}$ contains all possible mixed strategies for player i
  - Players draw their own actions *independently*
- Given strategy profile $x = (x_1, \cdots, x_n)$, expected utility of i is

$$\sum_{a \in A} u_i(a) \cdot \Pi_{i \in [n]} x_i(a_i)$$

  - Often denoted as $u_i(x)$ or $u_i(x_i, x_{-i})$ or $u_i(x_1, \cdots, x_n)$
  - When x_i corresponds to some pure strategy a_i, we also write $u_i(a_i, x_{-i})$

Best Responses

Fix any x_{-i}, x_i^* is called a best response to x_{-i} if

$$u_i(x_i^*, x_{-i}) \geq u_i(x_i, x_{-i}), \quad \forall\, x_i \in \Delta_{A_i}.$$

Claim. There always exists a pure best response

Proof: if randomization over a few actions are optimal, then each of these actions must be equally good and all optimal (i.e. pure best response)

Nash Equilibrium (NE)

A mixed strategy profile $x^* = (x_1^*, \cdots, x_n^*)$ is a **Nash equilibrium** if

$$u_i(x_i^*, x_{-i}^*) \geq u_i(x_i, x_{-i}^*), \qquad \forall\, x_i \in \Delta_{A_i}, \forall i \in [n].$$

That is, for any i, x_i^* is a best response to x_{-i}^*.

A relaxation

A mixed strategy profile $x^* = (x_1^*, \cdots, x_n^*)$ is a ϵ-approximate **Nash equilibrium** (ϵ-NE) if

$$u_i(x_i^*, x_{-i}^*) \geq u_i(x_i, x_{-i}^*) - \epsilon, \qquad \forall\, x_i \in \Delta_{A_i}, \forall i \in [n].$$

That is, for any i, x_i^* is an ϵ-approximate best response to x_{-i}^*.

More often used because we usually will not (or cannot) find exact optimal, but approximately optimal strategies

Nash Equilibrium (NE)

A mixed strategy profile $x^* = (x_1^*, \cdots, x_n^*)$ is a **Nash equilibrium** if

$$u_i(x_i^*, x_{-i}^*) \geq u_i(x_i, x_{-i}^*), \qquad \forall\, x_i \in \Delta_{A_i}, \forall i \in [n].$$

That is, for any i, x_i^* is a best response to x_{-i}^*.

Remarks

- An equivalent condition: $u_i(x_i^*, x_{-i}^*) \geq u_i(a_i, x_{-i}^*), \forall\, a_i \in A_i, \forall i \in [n]$
  - Since there always exists a pure best response
- Fundamentally, equilibrium is not about optimality, but rather about stability
  - Recall prisoner's dilemma, both are at bad situations but no one wants to deviate
  - This happens quite often in strategic interactions
- It is not clear yet that such a mixed strategy profile would exist
  - Recall that pure strategy Nash equilibrium may not exist

Nash Equilibrium (NE)

Theorem (Nash, 1951): Every finite game (i.e., finite players and actions) admits at least one mixed strategy Nash equilibrium.

- Nash proved this result during his PhD at Princeton
- Time proves that this is a foundational result in game-theory, later won Nobel Prize in Econ

That's just a fixed point theorem

Nash Equilibrium (NE)

Theorem (Nash, 1951): Every finite game (i.e., finite players and actions) admits at least one mixed strategy Nash equilibrium.

- Example: rock-paper-scissor – what is a mixed strategy NE?
  - $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is a best response to $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

		$1/3$	$1/3$	$1/3$
		Rock	Paper	Scissor
ExpU = 0	Rock	(0, 0)	(-1, 1)	(1, -1)
ExpU = 0	Paper	(1, -1)	(0, 0)	(-1, 1)
ExpU = 0	Scissor	(-1, 1)	(1, -1)	(0, 0)

Nash Equilibrium (NE)

Theorem (Nash, 1951): Every finite game (i.e., finite players and actions) admits at least one mixed strategy Nash equilibrium.

- A game can have many, even infinitely many, NEs
  - Unlike (single-agent) optimization which often has a unique optimal value
  - Which equilibrium will the game stabilize at? → the issue of equilibrium selection

	L	M	R
U	4,3	5,1	6,2
M	2,1	8,4	3,6
D	3,0	9,6	2, 5

Outline

- Game Theory and Nash Equilibrium
- Zero-Sum Games: theory and learning

Zero-Sum Games

- Two players: player 1 action $i \in [m] = \{1, \cdots, m\}$, player 2 action $j \in [n]$
- The game is **zero-sum** if $u_1(i,j) + u_2(i,j) = 0,\ \forall i \in [m],\ j \in [n]$
  - Models strictly competitive scenarios
  - "Zero-sum" almost always mean "2-player zero-sum" games
  - n-player games can also be zero-sum, but not particularly interesting
- Let $u_1(x,y) = \sum_{i \in [m], j \in [n]} u_1(i,j) x_i y_j$ for any $x \in \Delta_m, y \in \Delta_n$
- (x^*, y^*) is a NE for the zero-sum game if: (1) $u_1(x^*, y^*) \geq u_1(i, y^*)$ for any $i \in [m]$; (2) $u_1(x^*, y^*) \leq u_1(x^*, j)$ for any j $\in [m]$
  - Condition $u_1(x^*, y^*) \leq u_1(x^*, j) \Leftrightarrow u_2(x^*, y^*) \geq u_2(x^*, j)$
  - We can "forget" u_2; Instead think of player 2 as minimizing player 1's utility

Maximin and Minimax Strategy

- Previous observations motivate the following definitions

Definition. $x^* \in \Delta_m$ is a maximin strategy of player 1 if it solves

$$\max_{x \in \Delta_m} \min_{j \in [n]} u_1(x, j).$$

The corresponding utility value is called maximin value of the game.

Remarks:

- x^* is player 1's best action if he was to move first

Maximin and Minimax Strategy

➢ Previous observations motivate the following definitions

Definition. $x^* \in \Delta_m$ is a maximin strategy of player 1 if it solves

$$\max_{x \in \Delta_m} \min_{j \in [n]} u_1(x, j).$$

The corresponding utility value is called maximin value of the game.

Definition. $y^* \in \Delta_n$ is a minimax strategy of player 2 if it solves

$$\min_{y \in \Delta_n} \max_{i \in [m]} u_1(i, y).$$

The corresponding utility value is called minimax value of the game.

Remark: y^* is player 2's best action if he was to move first

Duality of Maximin and Minimax

Fact. $\max_{x\in\Delta_m} \min_{j\in[n]} u_1(x,j) \le \min_{y\in\Delta_n} \max_{i\in[m]} u_1(i,y)$.

That is, moving first is no better in zero-sum games.

- Let $y^* = \arg\min_{y\in\Delta_n} \max_{i\in[m]} u_1(i,y)$, so

$$\min_{y\in\Delta_n} \max_{i\in[m]} u_1(i,y) = \max_{i\in[m]} u_1(i,y^*)$$

- We have

$$\max_{x\in\Delta_m} \min_{j\in[n]} u_1(x,j) \le \max_{x\in\Delta_m} u_1(x,y^*) = \max_{i\in[m]} u_1(i,y^*)$$

Duality of Maximin and Minimax

Fact. $$\max_{x\in\Delta_m} \min_{j\in[n]} u_1(x,j) \leq \min_{y\in\Delta_n} \max_{i\in[m]} u_1(i,y).$$

Minimax Theorem. $$\max_{x\in\Delta_m} \min_{j\in[n]} u_1(x,j) = \min_{y\in\Delta_n} \max_{i\in[m]} u_1(i,y).$$

As far as I can see, there could be no theory of games … I thought there was nothing worth publishing until the Minimax Theorem was proved

Duality of Maximin and Minimax

Fact. $$\max_{x \in \Delta_m} \min_{j \in [n]} u_1(x, j) \le \min_{y \in \Delta_n} \max_{i \in [m]} u_1(i, y).$$

Minimax Theorem. $$\max_{x \in \Delta_m} \min_{j \in [n]} u_1(x, j) = \min_{y \in \Delta_n} \max_{i \in [m]} u_1(i, y).$$

As far as I can see, there could be no theory of games … I thought there was nothing worth publishing until the Minimax Theorem was proved

Next: a modern proof of minimax theorem using no regret concept

✓ This proof illustrates core principles underlying many modern AI/RL approaches for solving games (e.g., poker, Go)

Online learning is a natural way to play repeated games

Repeated game: the same game played repeatedly for many rounds

- Think about how you play rock-paper-scissor repeatedly
- In reality, we play like online learning
  - You try to analyze the past patterns, then decide which action to respond, possibly with some randomness
  - This is basically online learning!

Repeated Zero-Sum Games and Players' Regret

Basic Setup:

- A zero-sum game with payoff matrix $U \in \mathbb{R}^{m \times n}$
- Row player maximizes utility and has actions $[m] = \{1, \cdots, m\}$
  - Column player thus minimizes utility
- The game is played repeatedly for T rounds
- Each player uses an online learning algorithm to pick a mixed strategy at each round

Repeated Zero-Sum Games and Players' Regret

- From row player's perspective, the following occurs in order at round t
  - Picks a mixed strategy $x_t \in \Delta_m$ over actions in $[m]$
  - Her opponent, the column player, picks a mixed strategy $y_t \in \Delta_n$
  - Action $i_t \sim x_t$ is chosen and row player receives utility $U(i_t, y_t) = \sum_{j \in [n]} y_t(j) \cdot U(i_t, j)$
  - Row player observes y_t (for future use)
- Column player has a symmetric perspective, but will think of $U(i, j)$ as his cost

Difference from online learning: utility/cost vector determined by the opponent, instead of being stochastic as in MAB

Repeated Zero-Sum Games and Players' Regret

- Expected total utility of row player $\sum_{t=1}^{T} U(x_t, y_t)$
  - Note: $U(x_t, y_t) = \sum_{i,j} U(i,j) x_t(i) y_t(j) = (x_t)^T U y_t$

- Regret of row player is

$$\max_{i \in [m]} \sum_{t=1}^{T} U(i, y_t) - \sum_{t=1}^{T} U(x_t, y_t)$$

- Regret of column player is

$$\sum_{t=1}^{T} U(x_t, y_t) - \min_{j \in [n]} \sum_{t=1}^{T} U(x_t, j)$$

From No Regret to Minimax Theorem

Now we are ready to prove the minimax theorem, using the fact that no regret (i.e., $o(T)$ regret) algorithms exist

From No Regret to Minimax Theorem

- Assume both players use no-regret learning algorithms
- For row player, we have

$$R_T^{row} = \max_{i \in [m]} \sum_{t=1}^{T} U(i, y_t) - \sum_{t=1}^{T} U(x_t, y_t)$$

$$\Leftrightarrow \frac{1}{T} \sum_{t=1}^{T} U(x_t, y_t) + \frac{R_T^{row}}{T} = \frac{1}{T} \max_{i \in [m]} \sum_{t=1}^{T} U(i, y_t)$$

$$= \max_{i \in [m]} U\left(i, \frac{\sum_t y_t}{T}\right)$$

$$\geq \min_{y \in \Delta_n} \max_{i \in [m]} U(i, y)$$

From No Regret to Minimax Theorem

- Assume both players use no-regret learning algorithms
- For row player, we have

$$\frac{1}{T}\sum_{t=1}^{T} U(x_t, y_t) + \frac{R_T^{row}}{T} \geq \min_{y\in\Delta_n} \max_{i\in[m]} U(i, y)$$

- Similarly, for column player,

$$R_T^{column} = \sum_{t=1}^{T} U(x_t, y_t) - \min_{j\in[n]} \sum_{t=1}^{T} U(x_t, j)$$

implies

$$\frac{1}{T}\sum_{t=1}^{T} U(x_t, y_t) - \frac{R_T^{column}}{T} \leq \max_{x\in\Delta_m} \min_{j\in[n]} U(x, j)$$

From No Regret to Minimax Theorem

- Assume both players use no-regret learning algorithms
- For row player, we have

$$\frac{1}{T}\sum_{t=1}^{T} U(x_t, y_t) + \frac{R_T^{row}}{T} \geq \min_{y \in \Delta_n} \max_{i \in [m]} U(i, y)$$

- Similarly, for column player,

$$R_T^{column} = \sum_{t=1}^{T} U(x_t, y_t) - \min_{j \in [n]} \sum_{t=1}^{T} U(x_t, j)$$

implies

$$\frac{1}{T}\sum_{t=1}^{T} U(x_t, y_t) - \frac{R_T^{column}}{T} \leq \max_{x \in \Delta_m} \min_{j \in [n]} U(x, j)$$

- Let $T \rightarrow \infty$, no regret implies $\frac{R_T^{row}}{T}$ and $\frac{R_T^{column}}{T}$ tend to 0. We have

$$\min_{y \in \Delta_n} \max_{i \in [m]} U(i, y) \leq \max_{x \in \Delta_m} \min_{j \in [n]} U(x, j)$$

From No Regret to Minimax Theorem

- Assume both players use no-regret learning algorithms

$$\frac{1}{T}\sum_{t=1}^{T} U(x_t, y_t) + \frac{R_T^{row}}{T} \geq \min_{y\in\Delta_n} \max_{i\in[m]} U(i, y)$$

$$\frac{1}{T}\sum_{t=1}^{T} U(x_t, y_t) - \frac{R_T^{column}}{T} \leq \max_{x\in\Delta_m} \min_{j\in[n]} U(x, j)$$

$$\Rightarrow \min_{y\in\Delta_n} \max_{i\in[m]} U(i, y) \leq \max_{x\in\Delta_m} \min_{j\in[n]} U(x, j)$$

From No Regret to Minimax Theorem

➢Assume both players use no-regret learning algorithms

$$\frac{1}{T}\sum_{t=1}^{T} U(x_t, y_t) + \frac{R_T^{row}}{T} \geq \min_{y \in \Delta_n} \max_{i \in [m]} U(i, y)$$

$$\frac{1}{T}\sum_{t=1}^{T} U(x_t, y_t) - \frac{R_T^{column}}{T} \leq \max_{x \in \Delta_m} \min_{j \in [n]} U(x, j)$$

$$\Rightarrow \min_{y \in \Delta_n} \max_{i \in [m]} U(i, y) \leq \max_{x \in \Delta_m} \min_{j \in [n]} U(x, j)$$

➢Recall that min-max ≥ max-min also holds, because moving second will not be worse for the row player

✓ These conclude the proof of minimax theorem

✓ Minimax Theorem implies nice equilibrium characterization in zero-sum games

Characterizing the NE in Zero-Sum Games

Theorem. In 2-player zero-sum games, (x^*, y^*) is a NE if and only if x^* and y^* are the maximin and minimax strategy, respectively.

$\Leftarrow$: if x^* [y^*] is the maximin [minimax] strategy, then (x^*, y^*) is a NE

- Want to prove $u_1(x^*, y^*) \geq u_1(i, y^*), \forall i \in [m]$

$$
\begin{aligned}
u_1(x^*, y^*) &\geq \min_{j} u_1(x^*, j) \\
&= \max_{x \in \Delta_m} \min_{j} u_1(x, j) \\
&= \min_{y \in \Delta_n} \max_{i \in [m]} u_1(i, y) \\
&= \max_{i \in [m]} u_1(i, y^*) \\
&\geq u_1(i, y^*), \forall i
\end{aligned}
$$

- Similar argument shows $u_1(x^*, y^*) \leq u_1(x^*, j), \forall j \in [n]$
- So (x^*, y^*) is a NE

Characterizing the NE in Zero-Sum Games

Theorem. In 2-player zero-sum games, (x^*, y^*) is a NE if and only if x^* and y^* are the maximin and minimax strategy, respectively.

⇒: if (x^*, y^*) is a NE, then x^* [y^*] is the maximin [minimax] strategy

- Observe the following inequalities

$$\begin{aligned} u_1(x^*, y^*) &= \max_{i \in [m]} u_1(i, y^*) \\ &\geq \min_{y \in \Delta_n} \max_{i \in [m]} u_1(i, y) \\ &= \max_{x \in \Delta_m} \min_{j} u_1(x, j) \\ &\geq \min_{j} u_1(x^*, j) \\ &= u_1(x^*, y^*) \end{aligned}$$

- So the two "≥" must both achieve equality.
  - The first equality implies y^* is the minimax strategy
  - The second equality implies x^* is the maximin strategy

Characterizing the NE in Zero-Sum Games

Theorem. In 2-player zero-sum games, (x^*, y^*) is a NE if and only if x^* and y^* are the maximin and minimax strategy, respectively.

Corollary: In repeated zero-sum games, suppose both players use learning algorithms with regret R_T to select action sequence $\{x_t\}$ and $\{y_t\}$. Then $(\frac{\sum_{t=1}^{T} x_t}{T}, \frac{\sum_{t=1}^{T} y_t}{T})$ is an ϵ-NE of the game with $\epsilon = \frac{R_T}{T}$.

Thank You

Haifeng Xu

University of Chicago

haifengxu@uchicago.edu