Reinforcement Learning China Summer School

RLChina 2021

Game Theory Basics

Haifeng Zhang
Institute of Automation, Chinese Academy of Sciences

Aug 16, 2021

Outline

- Motivation and Normal-form Game
- Extensive-form Game and Imperfect Information
- Bayesian Game and Incomplete Information
- Nash Equilibrium and Variants
- Repeated Game and Learning Methods
- Alternate Solution Concepts and Evolutionary Game Theory

Outline

- Motivation and Normal-form Game
- Extensive-form Game and Imperfect Information
- Bayesian Game and Incomplete Information
- Nash Equilibrium and Variants
- Repeated Game and Learning Methods
- Alternate Solution Concepts and Evolutionary Game Theory

Collective Decision Intelligence

- Progress of Artificial Intelligence

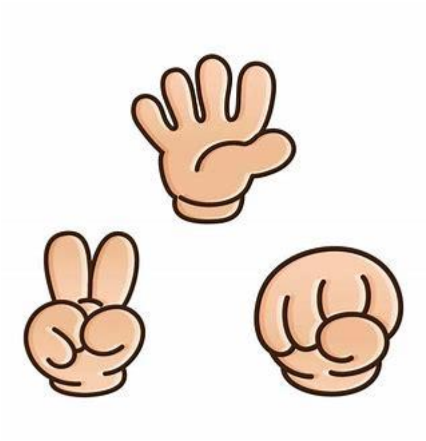
Recognition Intelligence → Decision Intelligence → Collective Decision Intelligence

Deep Learning → Deep Reinforcement Learning → Multi-agent Reinforcement Learning

Reinforcement Learning → Deep Reinforcement Learning

Game Theory → Multi-agent Reinforcement Learning

Games in Reality



Rock, Scissors, Paper

Auction

Chess

Poker

History of Game Theory

1934, Stackelberg,
Stackelberg Equilibrium[1]
1950, **Nash**,
Mixed Nash Equilibrium[2]

1951, Brown,
Fictitious Play in Repeated game[3]
1965, **Selten**,
Subgame Perfect Equilibrium in Extensive-form Game[4]

1967, **Harsanyi**,
Bayesian Nash Equilibrium in Bayesian game[5]

1973, Smith & Price,
Evolutional Game Theory[6]
1974, **Aumann**,
Correlated Equilibrium[7]

1994,
Papadimitriou,
PPAD[8]

Till now, 18 game theorists received **Nobel Prize in Economics**!

Elements of Game

- Players $N = \{1,2, \dots, n\}$
  - $N = \{1,2\}$
- Strategies (actions) $A = A_1 \times A_2 \times \cdots \times$
  - $A_1 = \{R, S, P\}$
  - $A_2 = \{R, S, P\}$

- Payoff (utility) $u = (u_1, u_2, \dots, u_n), u_i: A \to \mathbb{R}$
  - $u_1: A_1 \times A_2 \to \mathbb{R}$
  - $u_2: A_1 \times A_2 \to \mathbb{R}$

Normal-form Game

- Payoff Matrix

Column Player Actions

Row Player Actions

	R	S	P
R	0, 0	1, -1	-1, 1
S	-1, 1	0, 0	1, -1
P	1, -1	-1, 1	0, 0

- More than 2 players

Joint Actions

	p_1	p_2	p_3
R, R, R	0	-1	1
R, R, S	1	1	0
……	……	……	……

Rationality of Players

- Self-interested
  - Preference over game outcome
  - E.g. (paper, rock) is better than (rock, paper) for row player
- Utility
  - Utility of (paper, rock) is 1
  - Utility of (rock, paper) is -1
- Objective
  - Act to maximize (expected) utility

Pure Strategy and Mixed Strategy

Matching Pennies

	Heads	Tails
Heads	1, -1	-1, 1
Tails	-1, 1	1, -1

- Pure Strategy
  - $a_1 \in A_1 = \{Heads, Tails\}$
  - $a_2 \in A_2 = \{Heads, Tails\}$
- Mixed Strategy: Probability Distribution over Pure Strategy
  - $a_1 = (x_H, x_T),\ x_H \in [0,1],\ x_T \in [0,1],\ x_H + x_T = 1$
  - $a_2 = (y_H, y_T),\ y_H \in [0,1],\ y_T \in [0,1],\ y_H + y_T = 1$
- Expected Utility
  - $EU_1 = x_H y_H u_1(H,H) + x_H y_T u_1(H,T) + x_T y_H u_1(T,H) + x_T y_T u_1(T,T)$
  - $EU_2 = x_H y_H u_2(H,H) + x_H y_T u_2(H,T) + x_T y_H u_2(T,H) + x_T y_T u_2(T,T)$
- Example
  - $a_1 = (0.1,0.9), a_2 = (0.3,0.7)$
  - $EU_1 = 0.32, EU_2 = -0.32$

Classic Games

- Zero-sum Game
  - $u_1(a) + u_2(a) = 0, \forall\, a \in A$
- Cooperative Game
  - $u_i(a) = u_j(a), \forall\, a \in A, i, j \in N$
- Coordination Game
  - Multiple Nash Equilibria Exist
- Social Dilemma [9]
  - Everyone suffers in an NE

Matching Pennies

	Heads	Tails
Heads	1, -1	-1, 1
Tails	-1, 1	1, -1

Road Selection

	Left	Right
Left	1, 1	0, 0
Right	0, 0	1, 1

Battle of Sex

	Party	Home
Party	10, 5	0, 0
Home	0, 0	5, 10

Prisoner’s Dilemma

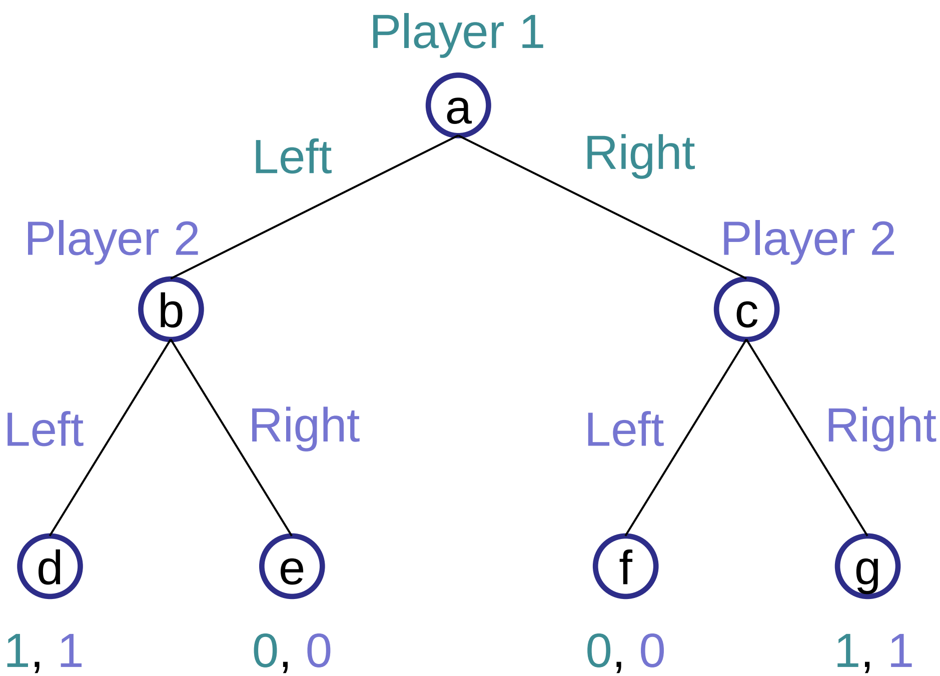
	Cooperate	Defect
Cooperate	2, 2	0, 3
Defect	3, 0	1, 1

Outline

- Motivation and Normal-form Game
- Extensive-form Game and Imperfect Information
- Bayesian Game and Incomplete Information
- Nash Equilibrium and Variants
- Repeated Game and Learning Methods
- Alternate Solution Concepts and Evolutionary Game Theory

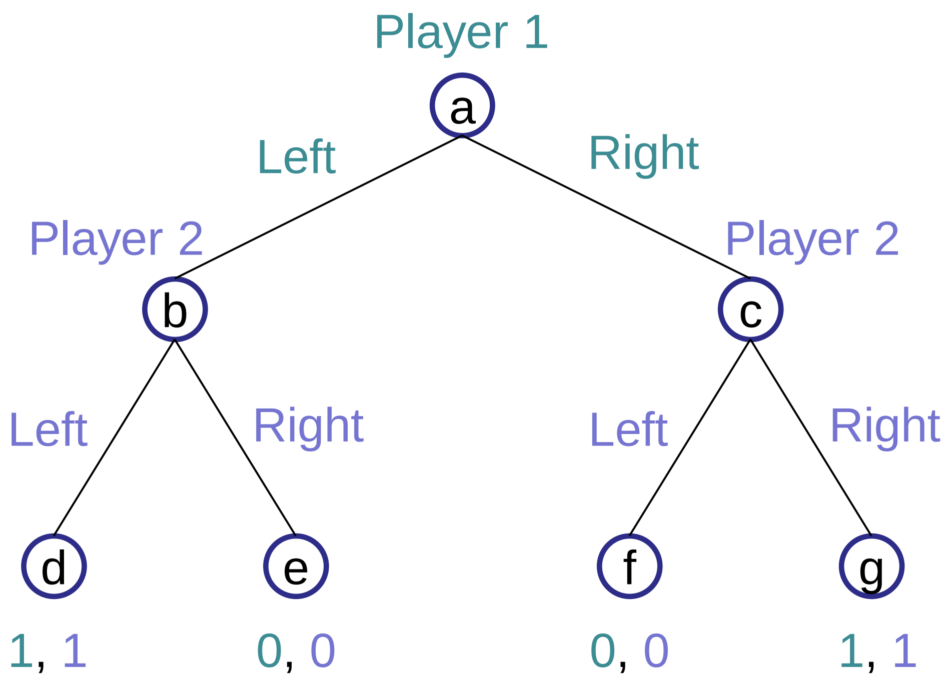
Extensive-form Game

- Game Tree
  - Node: decision point for a specified player
  - Edge: action decided by the player
  - Leaf: outcome of the game with payoff



Strategies in Extensive-form Game

- Strategy Space
  - Player 1: {Left, Right}
  - Player 2: {$\underbrace{\text{(Left, Left)}}_{\text{action in every node}}$, (Left, Right), (Right, Left), (Right, Right)}



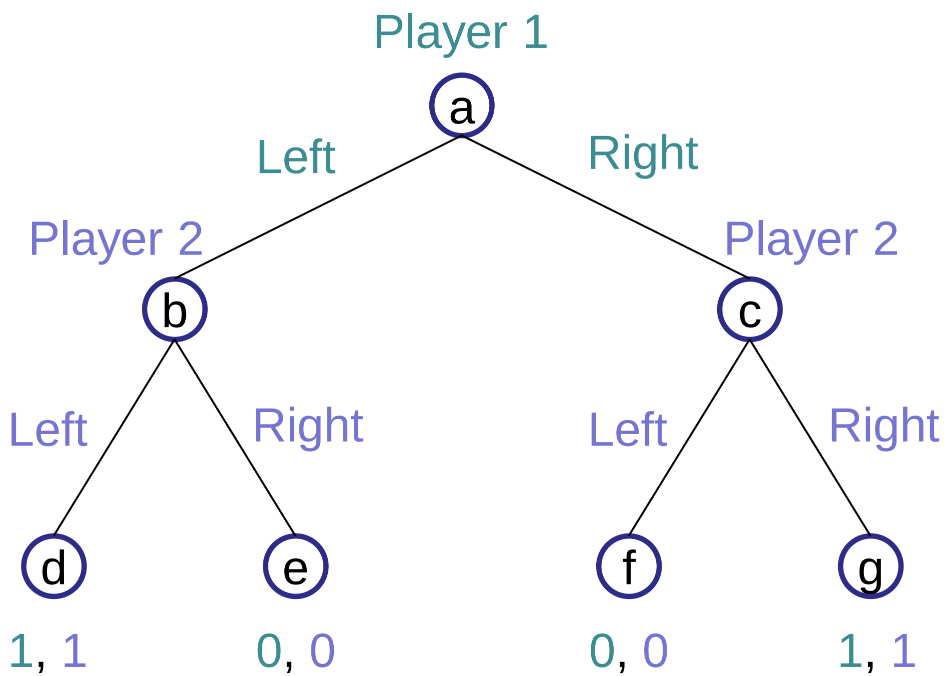
Extensive-form vs. Normal-form

- Equivalent Normal-form Game

single step/state
⇕
static

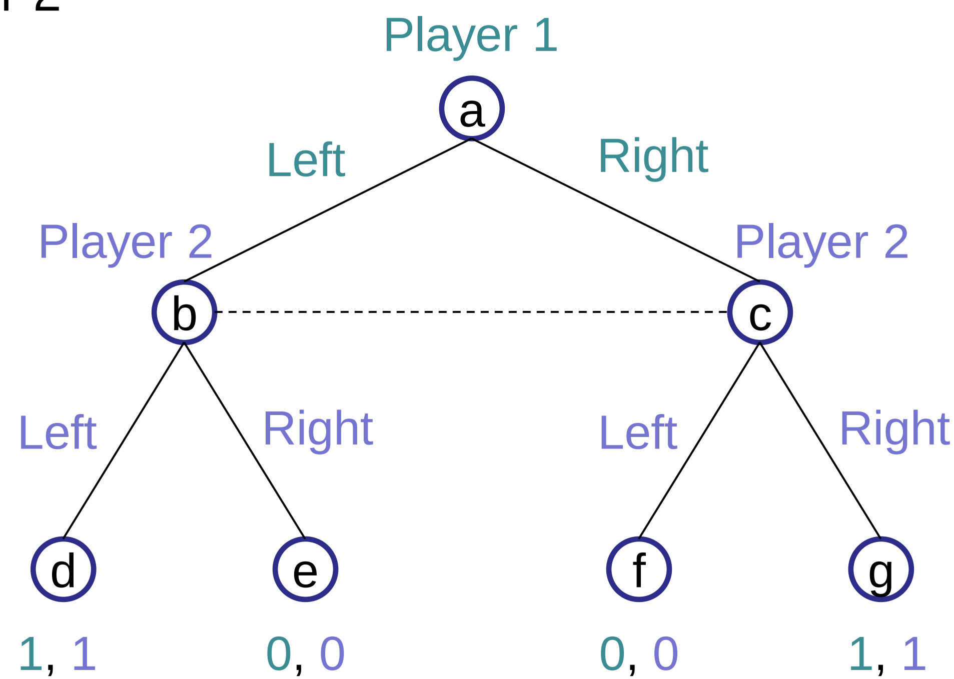
	(Left, Left)	(Left, Right)	(Right, Left)	(Right, Right)
Left	1, 1	1, 1	0, 0	0, 0
Right	0, 0	1, 1	0, 0	1, 1

multiple step/state
⇕
dynamic



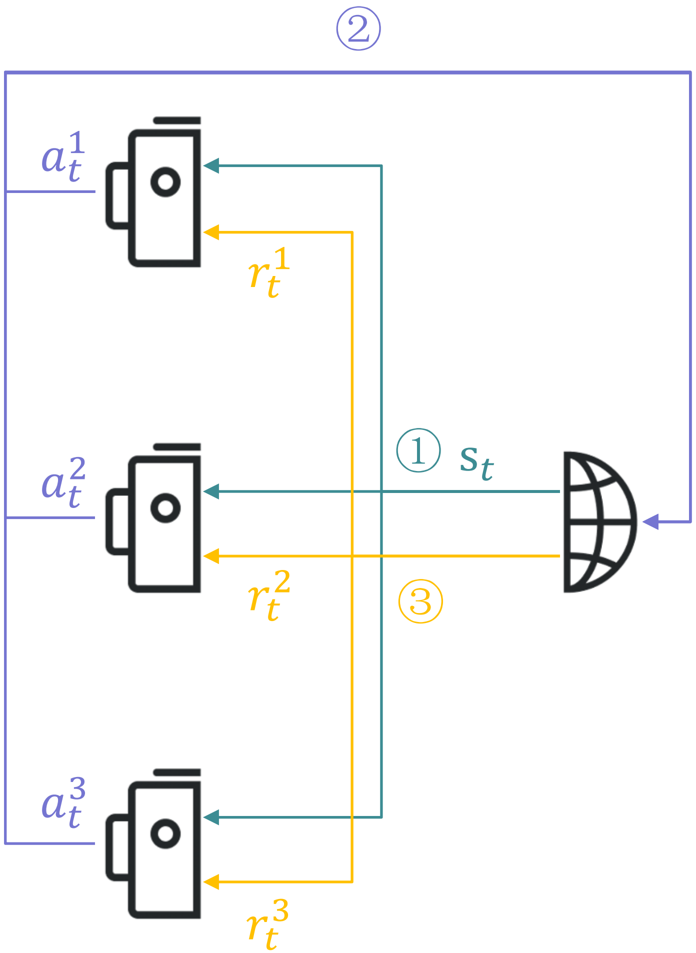
Imperfect Information

- Imperfect Information Game
  - Some historical actions are invisible by other players
- Information Set
  - A set containing undistinguishable states, e.g. {b, c} is an information set for player 2
- Strategy Space
  - Player 1: {Left, Right}
  - Player 2: $\underbrace{\text{\{Left, Right\}}}_{\text{action in every information set}}$



Markov Game (or Stochastic Game)

- Game Definition
  - State space S
  - Action space $A = A_1 \times A_2 \times \cdots \times A_n$
  - Transition function $p: S \times A \to S$
  - Reward function $r: S \times A \to \mathbb{R}^n$
- Behavioral Strategy
  - Policy $\pi_i: S \times A_i \to [0,1]$
- Properties
  - Simultaneous action (Normal-form)
  - Multiple step/state (Extensive-form)
  - Immediate reward
  - Randomness
  - Cycle



Interaction at time-step t

Summary of Strategy Representation

	Static Game (Single Step/state)	Dynamic Game (Multiple step/state)
Pure Strategy	$a_i \in A_i$	$\pi_i: S \rightarrow A_i$ or $\pi_i \in {A_i}^S$
Mixed Strategy	$a_i: A_i \rightarrow [0,1]$	$\pi_i: {A_i}^S \rightarrow [0,1]$
Behavioral Strategy	$a_i: A_i \rightarrow [0,1]$	$\pi_i: S \times A_i \rightarrow [0,1]$

Outline

- Motivation and Normal-form Game
- Extensive-form Game and Imperfect Information
- Bayesian Game and Incomplete Information
- Nash Equilibrium and Variants
- Repeated Game and Learning Methods
- Alternate Solution Concepts and Evolutionary Game Theory

Example: Auction

- Game Definition
  - Players has private value v_1, v_2
  - Players decide biddings b_1, b_2
  - Player i with higher bidding b_i has utility $v_i - b_i$
  - The other player has utility 0
- Uncertainty of Private Value
  - $v_1 = 4, v_2 = 4$
  - $b_1 \in \{1,3\}, b_2 \in \{2,4\}$

	$b_2 = 2$	$b_2 = 4$
$b_1 = 1$	0, 2	0, 0
$b_1 = 3$	1, 0	0, 0

  - $v_1 = 4, v_2 = 5$
  - $b_1 \in \{1,3\}, b_2 \in \{2,4\}$

	$b_2 = 2$	$b_2 = 4$
$b_1 = 1$	0, 3	0, 1
$b_1 = 3$	1, 0	0, 1

  - Players don't know the exact payoff matrix of the game!

Incomplete Information

- Recall the Elements of a Game
  - Players $N = \{1,2, \dots, n\}$
  - Action space $A = A_1 \times A_2 \times \cdots \times A_n$
  - Payoff functions $u = (u_1, u_2, \dots, u_n), u_i: A \to \mathbb{R}$
- Incomplete Information Game
  - Players know: N and A
  - Players don't completely know: u
  - Criteria: whether players have private information when game starts
- Example
  - Auction
  - Mahjong
  - Werewolves of Miller's Hollow

Bayesian Game

- Basic Idea
  - Payoff function p_i is unknown, but the distribution of p_i is known
- Elements of Bayesian Game
  - Players $N = \{1,2,\dots,n\}$, action space $A = A_1 \times A_2 \times \cdots \times A_n$
  - Player type space $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_n$
  - Distribution over types $d: \Theta \to [0,1]$
  - Payoff functions $u = (u_1, u_2, \dots, u_n), u_i: \Theta \times A \to \mathbb{R}$
- Strategy
  - Pure strategy $\pi_i: \Theta_i \to A_i$
  - Mixed strategy $\pi_i: \Theta_i \times A_i \to [0,1]$
- Example
  - $\Theta_1 = \{4\}, \Theta_2 = \{4,5\}$
  - $d(4,4) = 0.3, d(4,5) = 0.7$

$v_2 = 4$
0.3

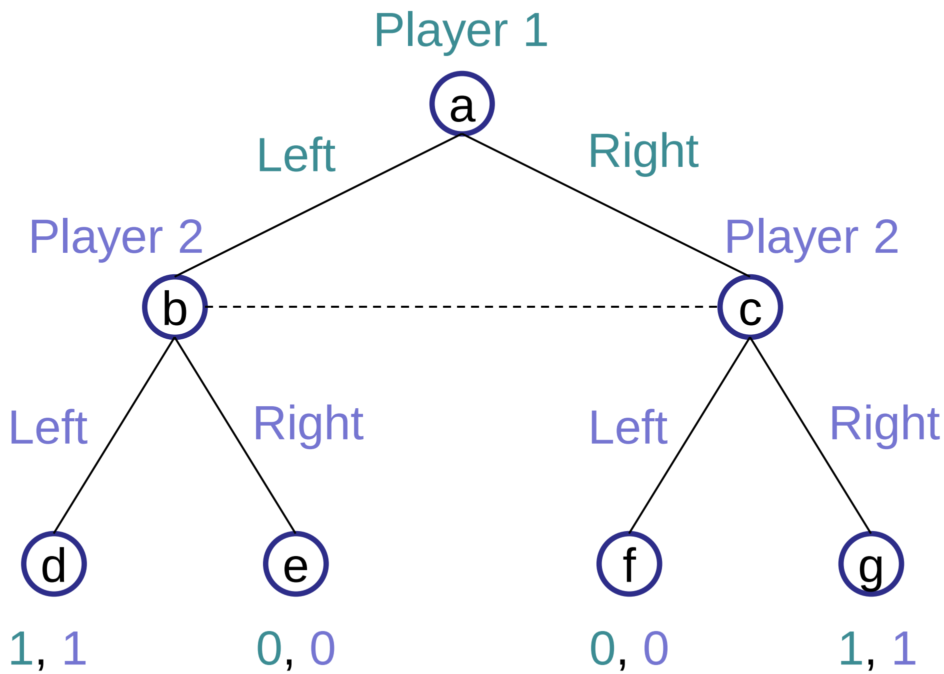
	$b_2 = 2$	$b_2 = 4$
$b_1 = 1$	0, 3	0, 1
$b_1 = 3$	1, 0	0, 1

$v_2 = 5$
0.7

	$b_2 = 2$	$b_2 = 4$
$b_1 = 1$	0, 2	0, 0
$b_1 = 3$	1, 0	0, 0

Dynamic Bayesian Game

- Belief System in Imperfect Information Extensive-form Game
  - Distribution over the states in an information set $b_i: S \to [0,1]$
- Strategy
  - Pure strategy $\pi_i: S \to A_i$
  - Behavioral strategy $\pi_i: S \times A_i \to [0,1]$



Summary of Game Representation

		Complete	**Incomplete**
Static		Normal-form Game, e.g. Prisoner’s Dilemma	Bayesian Game, e.g. Auction
Dynamic	**Perfect**	Extensive-form Game, e.g. Chess	Texas Hold’em Poker
	Imperfect	StarCraft	Mahjong

Dynamic Bayesian game

- Harsanyi Transformation
  - Incomplete Information → Imperfect Information
  - Introduce a nature player who decides the type of each player

Outline

- Motivation and Normal-form Game
- Extensive-form Game and Imperfect Information
- Bayesian Game and Incomplete Information
- Nash Equilibrium and Variants
- Repeated Game and Learning Methods
- Alternate Solution Concepts and Evolutionary Game Theory

Game Solution Reasoning

- Best Response (BR)
  - Given $a_{-i} \in A_1 \times \cdots \times A_{i-1} \times A_{i+1} \times \cdots \times A_n$
  - a_i is best response to a_{-i} ⇔ $u_i(a_i, a_{-i}) \geq u_i(a_i', a_{-i}), \forall a_i{}' \in A_i$
- Dominant Strategy (DS)
  - a_i is dominant strategy ⇔ Given any a_{-i}, a_i is best response
- Example

Prisoner's Dilemma

	Cooperate (C)	Defect (D)
Cooperate (C)	2, 2	0, 3
Defect (D)	3, 0	1, 1

$\underbrace{\text{3>2, D is BR to C} \quad \text{1>0, D is BR to D}}_{\text{D is DS}}$

Game Solution Concept: Nash Equilibrium

- Definition
  - A joint strategy (or strategy profile) $a \in A$ is a Nash Equilibrium $\Leftrightarrow$ a_i is best response to a_{-i} holds for every player i
- Example

Matching Pennies

	Heads	Tails
Heads	1, -1	-1, 1
Tails	-1, 1	1, -1

Battle of Sex

	Party	Home
Party	10, 5	0, 0
Home	0, 0	5, 10

Road Selection

	Left	Right
Left	1, 1	0, 0
Right	0, 0	1, 1

Prisoner's Dilemma

	Cooperate	Defect
Cooperate	2, 2	0, 3
Defect	3, 0	1, 1

Pareto Optimality vs. Nash Equilibrium

- Pareto Optimality (PO)
  - A joint strategy (or strategy profile) $a \in A$ achieves Pareto optimality $\Leftrightarrow \nexists a' \in A\ s.t.$ ① $\forall i, u_i(a') \geq u_i(a)$, ② $\exists i, u_i(a') > u_i(a)$
  - A Pareto optimality is not necessarily a Nash equilibrium
  - A Nash equilibrium is not necessarily a Pareto optimality

Chicken

	C	D
C	3, 3	1, 4
D	4, 1	0, 0

C-D is PO and NE

Stag Hunt

	C	D
C	3, 3	0, 2
D	2, 0	1, 1

D-D is NE but not PO

Prisoner's Dilemma

	C	D
C	2, 2	0, 3
D	3, 0	1, 1

C-C is PO but not NE

Mixed-Strategy Nash Equilibrium

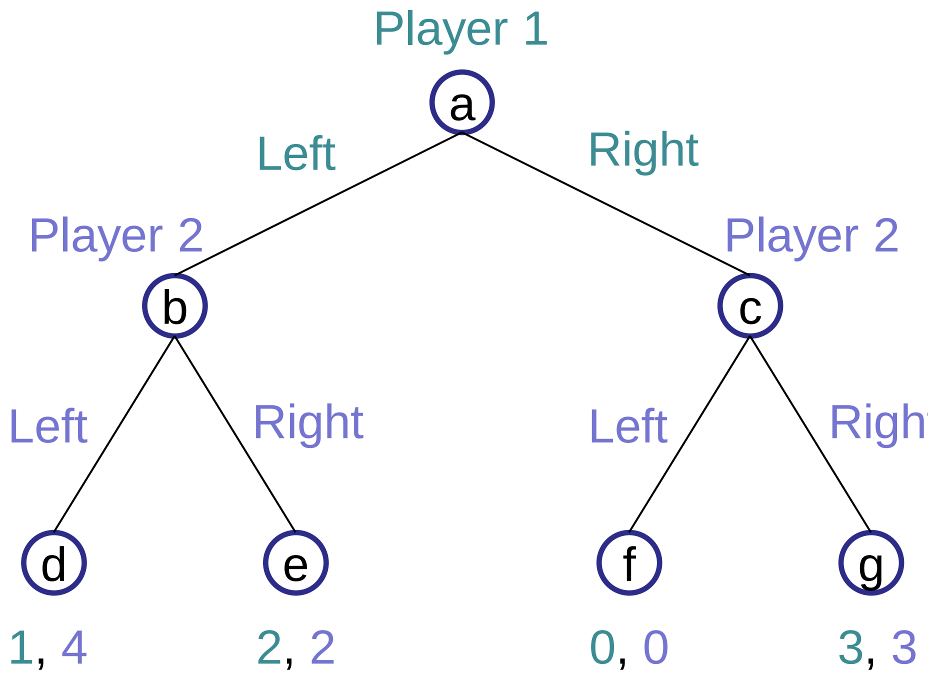
- Definition
  - A mixed-strategy profile $(a_1, a_2, \ldots, a_n), a_i \in \mathrm{PD}(A_i)$ is a Nash Equilibrium ⇔ a_i is best response to a_{-i} holds for every player i
- Example (Rock-Scissors-Paper)
  - $a_1 = ({}^1/_3, {}^1/_3, {}^1/_3), a_2 = ({}^1/_3, {}^1/_3, {}^1/_3)$
  - $EU_1(a_1, a_2) = 0 \geq EU_1(a_1', a_2) = 0, \forall a_1' \in A_1$
  - $EU_2(a_1, a_2) = 0 \geq EU_2(a_1, a_2') = 0, \forall a_2' \in A_2$

		${}^1/_3$	${}^1/_3$	${}^1/_3$
		R	S	P
${}^1/_3$	R	0, 0	1, -1	-1, 1
${}^1/_3$	S	-1, 1	0, 0	1, -1
${}^1/_3$	P	1, -1	-1, 1	0, 0

Nash Equilibrium in Extensive-form Game

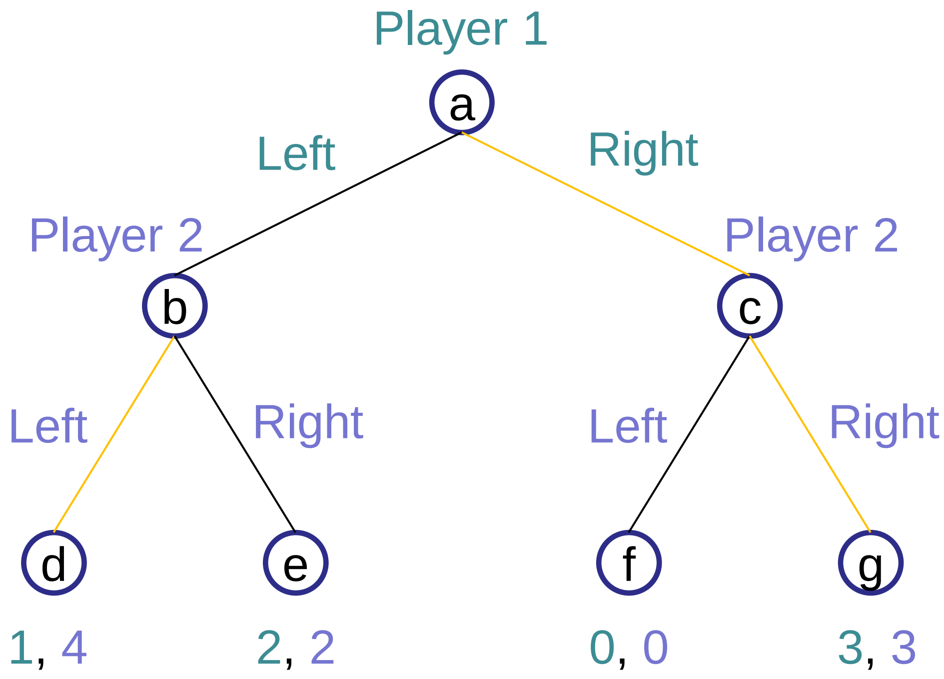
- Incredible Threat

	(Left, Left)	(Left, Right)	(Right, Left)	(Right, Right)
Left	1, 4 ?	1, 4	2, 2	2, 2
Right	0, 0	3, 3	0, 0	3, 3 ?



Subgame Perfect Nash Equilibrium (SPNE)

- Definition
  - An NE is SPNE ⇔ the NE holds in every subgame
- Solution
  - Backward induction: Right - (Left, Right)



Bayesian Nash Equilibrium

- Recall Bayesian Game
  - Player type space $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_n$
  - Distribution over types $d: \Theta \rightarrow [0,1]$
  - Payoff functions $u = (u_1, u_2, \dots, u_n), u_i: \Theta \times A \rightarrow \mathbb{R}$
- Strategy in Bayesian Game
  - Pure strategy $\pi_i: \Theta_i \rightarrow A_i$
  - Mixed strategy $\pi_i: \Theta_i \times A_i \rightarrow [0,1]$
- Bayesian Nash Equilibrium (BNE)
  - Assume each player i knows her own type $\theta_i \in \Theta_i$
  - Set expected utility $\mathbb{E}[u_i|\pi, \theta] = \sum_{a \in A}(\prod_{j \in N} \pi_j\big(\theta_j, a_j\big)\, u_i(a, \theta))$
  - π is BNE $\Leftrightarrow$ $\pi_i \in \operatorname{argmax}_{\pi_i{}'} \sum_{\theta_{-i} \in \Theta_{-i}} d(\theta_i, \theta_{-i}) \mathbb{E}[u_i \big| \pi_i', \pi_{-i}, \theta_i, \theta_{-i}]$ holds for each player i with her own type θ_i

Bayesian Nash Equilibrium: Example

- Auction
  - $A_1 = \{1,3\}, A_2 = \{2,4\}, \Theta_1 = \{4\}, \Theta_2 = \{4,5\}, d(4,4) = 0.3, d(4,5) = 0.7$
- Strategy
  - $\pi_1(4,1) = \mathrm{x}, \pi_1(4,3) = 1 - \mathrm{x}$
  - $\pi_2(4,2) = y_1, \pi_2(4,4) = 1 - y_1$
  - $\pi_2(5.2) = y_2, \pi_2(5,4) = 1 - y_2$

$v_2 = 4$
0.3

	$b_2 = 2$	$b_2 = 4$
$b_1 = 1$	0, 3	0, 1
$b_1 = 3$	1, 0	0, 1

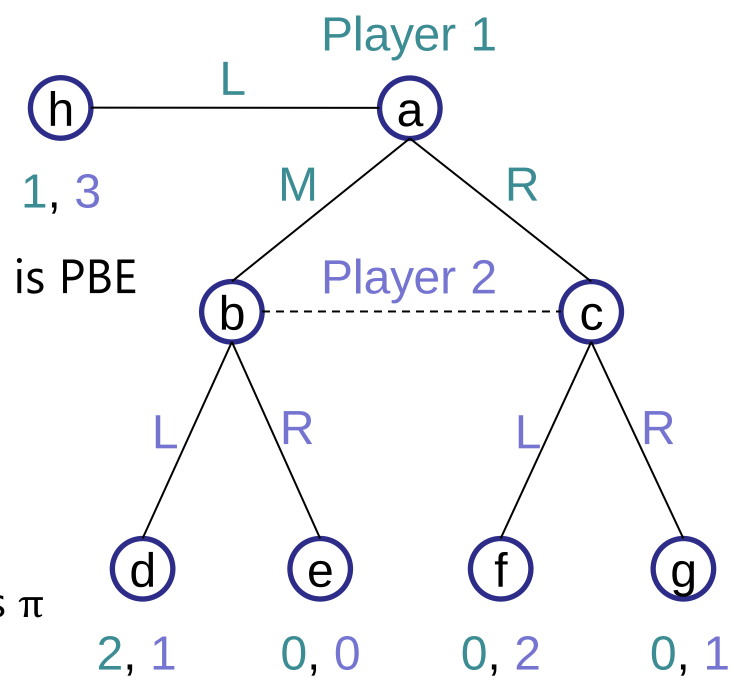
$v_2 = 5$
0.7

	$b_2 = 2$	$b_2 = 4$
$b_1 = 1$	0, 2	0, 0
$b_1 = 3$	1, 0	0, 0

- Equilibrium
  - $\mathbb{E}[u_1|\pi_1, \pi_2, 4,4] = (1 - x)y_1$
  - $\mathbb{E}[u_1|\pi_1, \pi_2, 4,5] = (1 - x)y_2$
  - $\mathbb{E}[u_2|\pi_1, \pi_2, 4,4] = 3xy_1 + x(1 - y_1) + (1 - \mathrm{x})(1 - y_1)$
  - $\mathbb{E}[u_2|\pi_1, \pi_2, 4,5] = 2xy_2$
  - (x, y_1, y_2) satisfies $x = \mathrm{argmax}_x 0.3(1 - x)y_1 + 0.7(1 - x)y_2$ and $y_1 = \mathrm{argmax}_{y_1} 3xy_1 + x(1 - y_1) + (1 - \mathrm{x})(1 - y_1)$ and $y_2 = \mathrm{argmax}_{y_2} 2xy_2$

Perfect Bayesian (Nash) Equilibrium

- Motivation
  - SPNE is not enough for some imperfect information game
  - Example: (L,R) is an SPNE but is incredible
- Recall Dynamic Bayesian Game
  - Belief function $b_i: S \to [0,1]$
  - Behavioral strategy $\pi_i: S \times A_i \to [0,1]$
- Perfect Bayesian Equilibrium (PBE) [10]
  - A strategy profile π with a belief system b is PBE
  - **Sequential rationality**
    - Each player has best expected utility in each information set following b and π
  - **Consistency of beliefs with Strategies**
    - Beliefs b are correct according to strategies π



Summary of Nash Equilibrium

<table>
<tr><th colspan="2"></th><th>Complete</th><th>Incomplete</th></tr>
<tr><td colspan="2">Static</td><td>Nash Equilibrium</td><td>Bayesian Nash Equilibrium</td></tr>
<tr><td rowspan="2">Dynamic</td><td>Perfect</td><td>Subgame Perfect Nash Equilibrium</td><td rowspan="2">Perfect Bayesian Nash Equilibrium</td></tr>
<tr><td>Imperfect</td><td></td></tr>
</table>

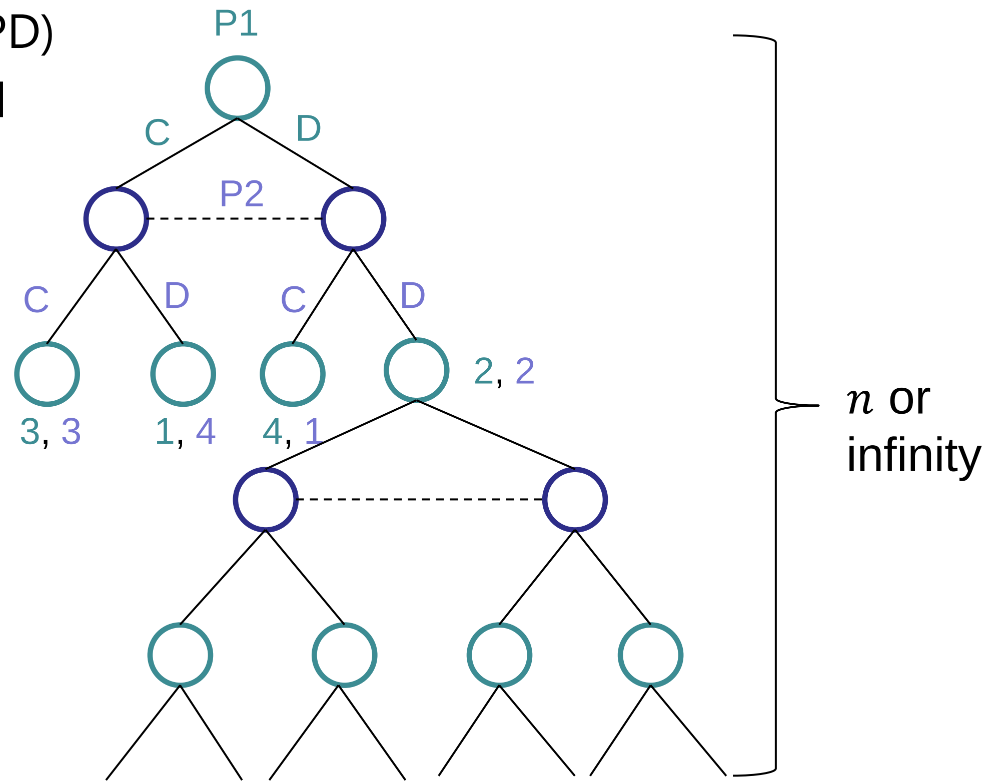
- Harsanyi Transformation
  - Incomplete Information → Imperfect Information
  - Introduce a nature player who decides the type of each player

Outline

- Motivation and Normal-form Game
- Extensive-form Game and Imperfect Information
- Bayesian Game and Incomplete Information
- Nash Equilibrium and Variants
- Repeated Game and Learning Methods
- Alternate Solution Concepts and Evolutionary Game Theory

Repeated Game

- Definition
  - A normal-form game is played over and again by the same players
  - The game repeated in each period is referred to as the **stage game**
- Example
  - Iterated Prisoners' Dilemma (IPD)
  - Reward: average or discounted
  - Memory: perfect recall



Memory

- Historical Behavior
  - At stage t, the action profile is a_t
  - Each player remembers the action profiles at last k stages
  - We say the players have k-memory
- Relation to Markov Game
  - Memory is regarded as state

1-memory

	1	2	3	…	m
P1	C	C	D	…	*D*
P2	D	D	C	…	C
…					
Pn	D	C	D	…	C

as state

k-memory

	1	2	3	…	m
P1	C	C	D	…	*D*
P2	D	D	C	…	C
…					
Pn	D	C	D	…	C

as state

Tit-for-tat

- Idea [11]
  - The Tit-for-tat strategy copies what the other player previously choose.
  - Nice: start by cooperating.
  - Clear: be easy to understand and adapt to.
  - Provocable: retaliate against anti-social behavior.
  - Forgiving: cooperate when faced with pro-social play.

	C	D
C	2, 2	0, 3
D	3, 0	1, 1

Prisoner's Dilemma

	1	2	3	4	…
P1	C	C	C	C	…
P2	C	C	C	C	…

cooperate by playing strategy (C,C)

$$\text{payoff} = 2 + 2\gamma + 2\gamma^2 + 2\gamma^3 + \cdots = 2\frac{\gamma^n - 1}{\gamma - 1} = \frac{2}{1-\gamma}$$

	1	2	3	4	…
P1	C	D	C	D	…
P2	D	C	D	C	…

a player deviates to defecting (D)

$$\text{payoff} = 3 + 0\gamma + 3\gamma^2 + 0\gamma^3 + \cdots = \frac{3}{1-\gamma^2}$$

Win-stay, lose-shift

	C	D
C	2, 2	0, 3
D	3, 0	1, 1

Prisoner's Dilemma

- Idea [12]
  - **Repeat** if it was rewarded by 2 or 3
  - **Shift** if it was punished by 0 or 1
  - Advantage: tolerant, one round of mutual defection followed by a return to cooperation
  - Disadvantage: fares poorly against inveterate defectors

	1	2	3	4	…
P1	C	D	C	C	…
P2	D	D	C	C	…

P2 deviates to defecting(D) initially

$$\text{Payoff} > \frac{3}{1-\gamma^2}$$

	1	2	3	4	…
P1	C	C	D	C	…
P2	C	D	D	D	…

P2 is an inveterate defectors

$$\text{payoff}_1 = 2 + 0\gamma + 1\gamma^2 + 0\gamma^3 + \cdots = \frac{1\gamma}{1-\gamma^2} + 1$$

$$\wedge$$

$$\text{payoff}_2 = 2 + 3\gamma + 1\gamma^2 + 3\gamma^3 + \cdots = \frac{1+3\gamma}{1-\gamma^2} + 1$$

Strategies in Iterated Prisoner's Dilemma

	C	D
C	2, 2	0, 3
D	3, 0	1, 1

Prisoner's Dilemma

Action profile at time t	Player 1 strategies at time $t+1$ with 1-memory															
(a_1, a_2)=(C, C)	C	C	C	C	C	C	C	C	D	D	D	D	D	D	D	D
(a_1, a_2)=(C, D)	C	C	C	C	D	D	D	D	C	C	C	C	D	D	D	D
(a_1, a_2)=(D, C)	C	C	D	D	C	C	D	D	C	C	D	D	C	C	D	D
(a_1, a_2)=(D, D)	C	D	C	D	C	D	C	D	C	D	C	D	C	D	C	D

Tit-for-tat strategy

Win-stay, lose-shift strategy

Folk Theorem

- Game Setting
  - n-player infinitely-repeated game $G = (N, A, u)$ with average reward
- Enforceable
  - A payoff profile r is **enforceable** if $r_i \geq v_i, v_i = \min_{s_{-i} \in S_{-i}} \max_{s_i \in S_i} u_i(s_{-i}, s_i)$
- Feasible
  - A payoff profile r is **feasible** if there exist rational, non-negative values α_a such that for all i, we can express r_i as $\sum_{a \in A} \alpha_a u_i(a)$, with $\sum_{a \in A} \alpha_a = 1$
- Folk Theorem
  - r is **feasible** and **enforceable** $\Rightarrow$ r is the payoff in some Nash equilibrium

	C	D
C	2, 2	0, 3
D	3, 0	1, 1

Prisoner's Dilemma

(v_1, v_2) =(1, 1)
(-1, -1) is not enforceable, not feasible
(0.5, 2) is not enforceable, **feasible**
(5, 5) is **enforceable**, not feasible
(2, 2) is **enforceable**, **feasible**

Fictitious Play

- Definition
  - Each player plays a best response to **assessed** strategy of the opponent and observe the opponent's actual play and update **beliefs**.

Matching Pennies

	Heads	Tails
Heads	1, -1	-1, 1
Tails	-1, 1	1, -1

Round	1's action	2's action	1's beliefs	2's beliefs
0			(1.5, 2)	(2, 1.5)
1	**T**	T	(1.5, 3)	(2, **2.5**)
2	T	H	(2.5, 3)	(2, 3.5)
3	T	**H**	(**3.5**, 3)	(2, 4.5)
4	H	H	(4.5, 3)	(3, 4.5)
…	…	…	…	…

Convergence of Fictitious Play

- Fictitious Play → Convergence
  - Each of the following are a sufficient conditions for the empirical frequencies of play to converge in fictitious play:
    - The game is zero sum;
    - The game is solvable by iterated elimination of strictly dominated strategies;
    - The game is a potential game;
    - The game is 2 n and has generic payoffs.
- Convergence → Nash Equilibrium
  - If the empirical distribution of each player’s strategies converges in fictitious play, then it converges to a Nash equilibrium.
- Results in Extensive-form Game with Imperfect Information
  - Fictitious self-play converges to approximate Nash equilibrium [13]
  - AlphaStar for StarCraft [14]

No-regret Learning

- Regret
  - Let a^t be the action profile played at time t
  - Regret of player i for not playing action a_i' at time t is $R^t(a_i') = u_i(a_i', a_{-i}^t) - u_i(a^t)$
  - Regret cumulated from time 1 to T is $CR^T(a_i') = \sum_{t=1}^{T} R^t(a_i')$
- Regret Matching
  - At each time step, each action is chosen with probability proportional to its cumulated regret: $\sigma_i^{t+1}(a_i) = \frac{CR^t(a_i)}{\sum_{a_i' \in A_i} CR^t(a_i')}$
  - Converge to correlated equilibrium
- No-regret learning in Extensive-form Game
  - Counterfactual Regret Minimization (CFR)
  - DeepStack for Texas Hold'em poker [15]

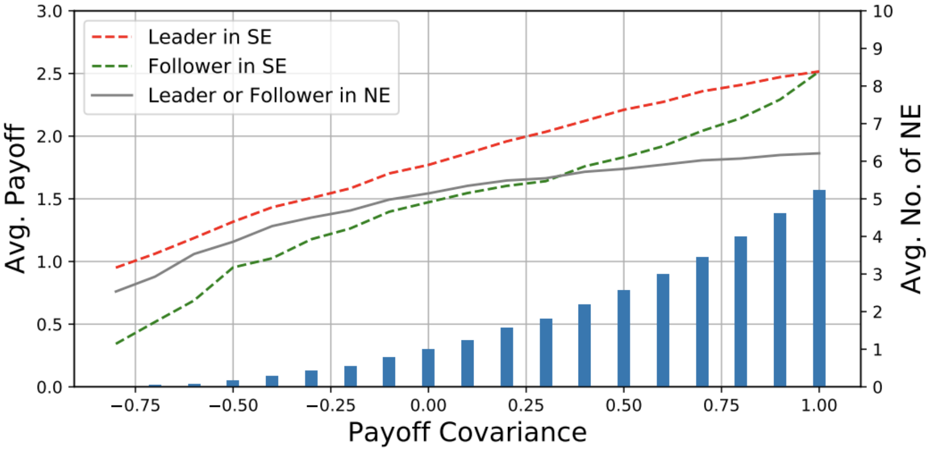
Outline

- Motivation and Normal-form Game
- Extensive-form Game and Imperfect Information
- Bayesian Game and Incomplete Information
- Nash Equilibrium and Variants
- Repeated Game and Learning Methods
- Alternate Solution Concepts and Evolutionary Game Theory

Stackelberg Equilibrium

- Stackelberg Game
  - A **leader** moves first
  - The **follower(s)** move after the leader
- Equilibrium
  - Subgame perfect Nash equilibrium
- Compared with Nash Equilibrium
  - Order is good in highly cooperative games
  - Bi-level Actor-critic RL [16]

	X	Y	Z
A	20, 15	0, 0	0, 0
B	30, 0	10, 5	0, 0
C	0, 0	0, 0	5, 10



Correlated Equilibrium

- Motivation
  - Equilibrium selection
- Basic Idea
  - Introduce a **public signal**
  - Sample from a probability distribution over **action profiles**
  - Each player is informed with her own action
  - No player has incentive to deviate
- Example
  - Pr[(Party, Party)]=0.5
  - Pr[(Home, Home)]=0.5
  - Pr[(Home, Party)]=0
  - Pr[(Party, Home)]=0

Battle of Sex

	Party	Home
Party	10, 5	0, 0
Home	0, 0	5, 10

Evolutional Game Theory

- Motivation
  - Nash equilibrium is static, the dynamic of strategy is not described
  - Players are not fully rational
- Basic Idea
  - Strategy is inherent and player can not select strategy by herself
  - Player with high payoff is has more chance to be reproduced
- Evolutionary Stable Strategy (ESS)
  - If almost every member of the population follows a strategy, no mutant (that is, an individual who adopts a novel strategy) can successfully invade.

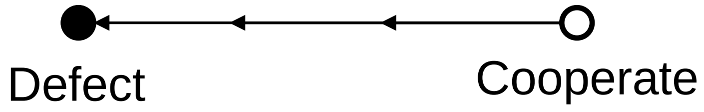
Ref: https://plato.stanford.edu/entries/game-evolutionary/

Replicator Dynamics

- Definition
  - $\dot{x}_i = x_i[f_i(x) - \varphi(x)], \varphi(x) = \sum_{j=1}^{n} x_j f_j(x)$
  - x is distribution of types(strategies) over the population
  - $f_i(x)$ is the fitness for type i in population x
  - $\varphi(x)$ is the average fitness of the population

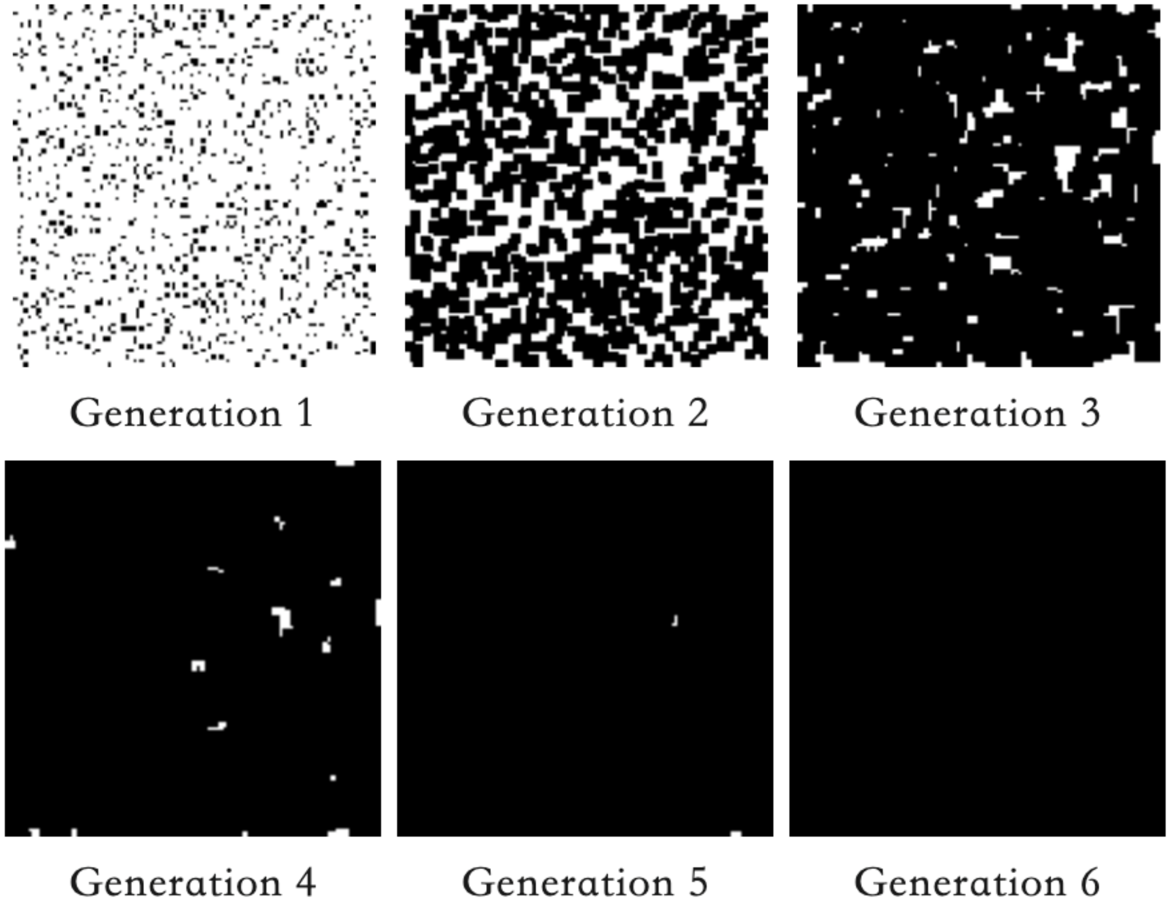
	C	D
C	2, 2	0, 3
D	3, 0	1, 1

Prisoner's Dilemma



Replicator Dynamics: Experiment

- On a local interaction model [17]
  - T = 2.8, R = 1.1, P = 0.1, and S = 0 **All Defect**



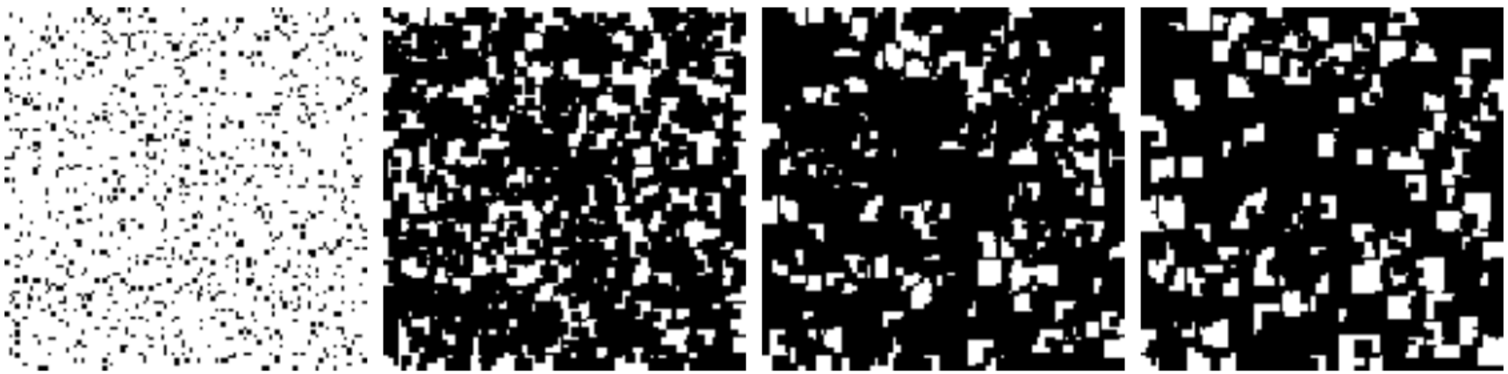
Replicator Dynamics: Experiment

- On a local interaction model
  - T = 1.2, R = 1.1, P = 0.1, and S = 0 **Cooperate**

Generation 1 Generation 2 Generation 19 Generation 20

Replicator Dynamics: Experiment

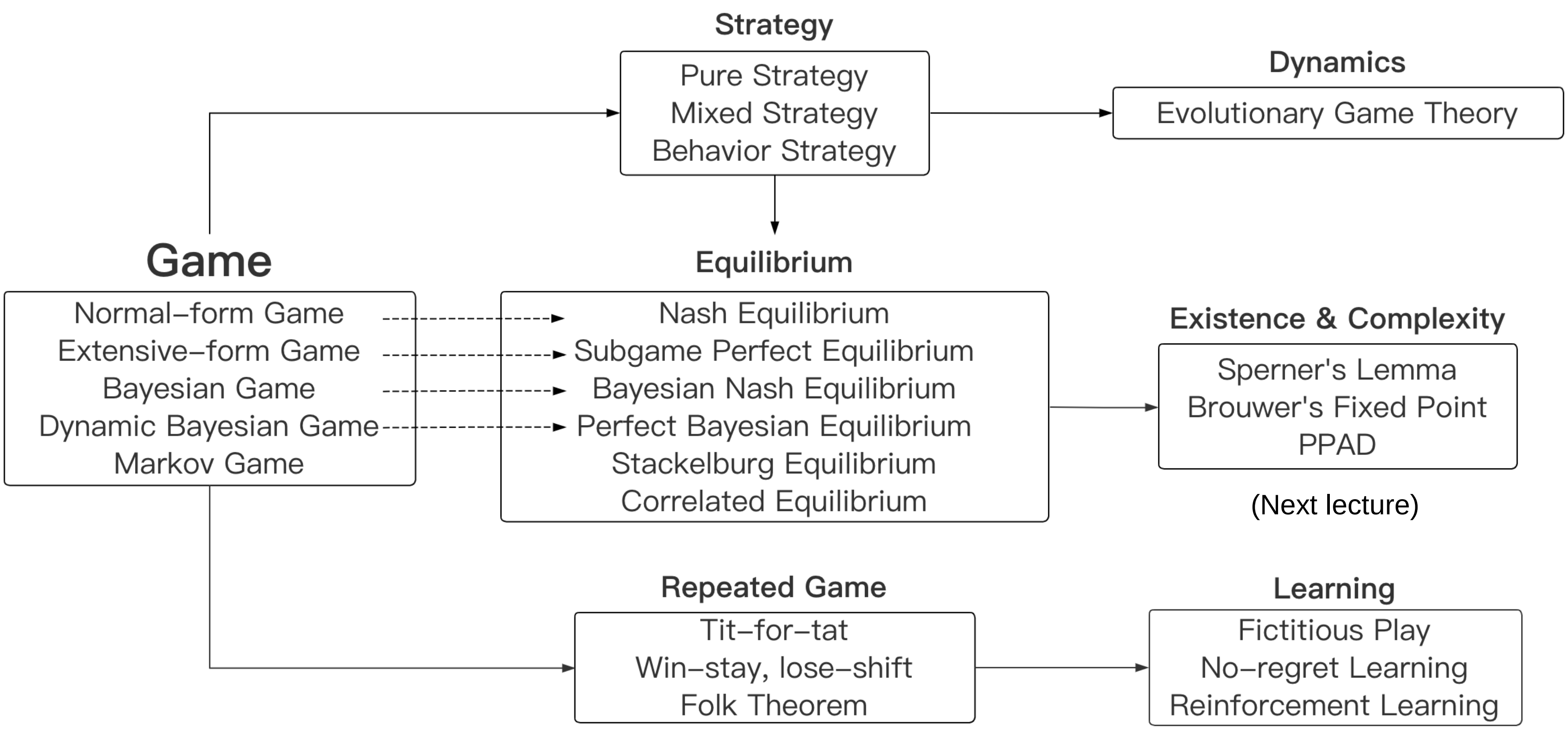
- On a local interaction model
  - T = 1.61, R = 1.01, P = 0.01, and S = 0 **Chaotic**



Generation 1 Generation 3 Generation 5 Generation 7

Generation 9 Generation 11 Generation 13 Generation 15

Summary


Strategy
Pure Strategy
Mixed Strategy
Behavior Strategy
Dynamics
Evolutionary Game Theory
Game
Normal–form Game
Extensive–form Game
Bayesian Game
Dynamic Bayesian Game
Markov Game
Equilibrium
Nash Equilibrium
Subgame Perfect Equilibrium
Bayesian Nash Equilibrium
Perfect Bayesian Equilibrium
Stackelburg Equilibrium
Correlated Equilibrium
Existence & Complexity
Sperner's Lemma
Brouwer's Fixed Point
PPAD
(Next lecture)
Repeated Game
Tit–for–tat
Win–stay, lose–shift
Folk Theorem
Learning
Fictitious Play
No–regret Learning
Reinforcement Learning

Reference

[1] Stackelberg, H. "Market Structure and Equilibrium. Translation into English, Bazin, Urch & Hill, Springer, 2011." (1934).

[2] Nash, John. "Non-cooperative games." Annals of mathematics (1951): 286-295.

[3] Harsanyi, John C. "Games with incomplete information played by “Bayesian” players, I–III Part I. The basic model." Management science 14.3 (1967): 159-182.

[4] Selten. R. (1965), “Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit – Teil I Bestimmung des dynamischen Preisgleichgewichts”, Zeitschrift für die gesamte Staatswissenschaft 121: 301–24.

[5] Brown, G.W. "Iterative Solutions of Games by Fictitious Play" In Activity Analysis of Production and Allocation, T. C. Koopmans (Ed.), New York (1951) : Wiley.

[6] Maynard-Smith, J.; Price, G. R. "The Logic of Animal Conflict". Nature (1973). 246 (5427): 15–18.

[7] Aumann, Robert "Subjectivity and correlation in randomized strategies". Journal of Mathematical Economics (1974). 1 (1): 67–96.

[8] Christos Papadimitriou "On the complexity of the parity argument and other inefficient proofs of existence". Journal of Computer and System Sciences (1994). 48 (3): 498–532.

[9] Macy, Michael W., and Andreas Flache. "Learning dynamics in social dilemmas." Proceedings of the National Academy of Sciences 99.suppl 3 (2002): 7229-7236.

[10] Fudenberg, Drew, and Jean Tirole. "Perfect Bayesian equilibrium and sequential equilibrium." journal of Economic Theory 53.2 (1991): 236-260.

Reference

[11] Nowak, Martin A. and Sigmund, K. (1992). “Tit For Tat in Heterogenous Populations,” Nature, 359: 250–253.

[12] Nowak, Martin, and Karl Sigmund. "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game." Nature 364.6432 (1993): 56-58.

[13] Heinrich, Johannes, Marc Lanctot, and David Silver. "Fictitious self-play in extensive-form games." International Conference on Machine Learning. 2015.

[14] Vinyals, Oriol, et al. "Grandmaster level in StarCraft II using multi-agent reinforcement learning." Nature 575.7782 (2019): 350-354.

[15] Moravčík, Matej, et al. "Deepstack: Expert-level artificial intelligence in heads-up no-limit poker." Science 356.6337 (2017): 508-513.

[16] Zhang, Haifeng, et al. "Bi-level Actor-Critic for Multi-agent Coordination." AAAI 2020.

[17] Nowak, Martin A. and May, Robert M. “The Spatial Dilemmas of Evolution,” International Journal of Bifurcation and Chaos (1993), 3: 35–78.