Fourteenth Session

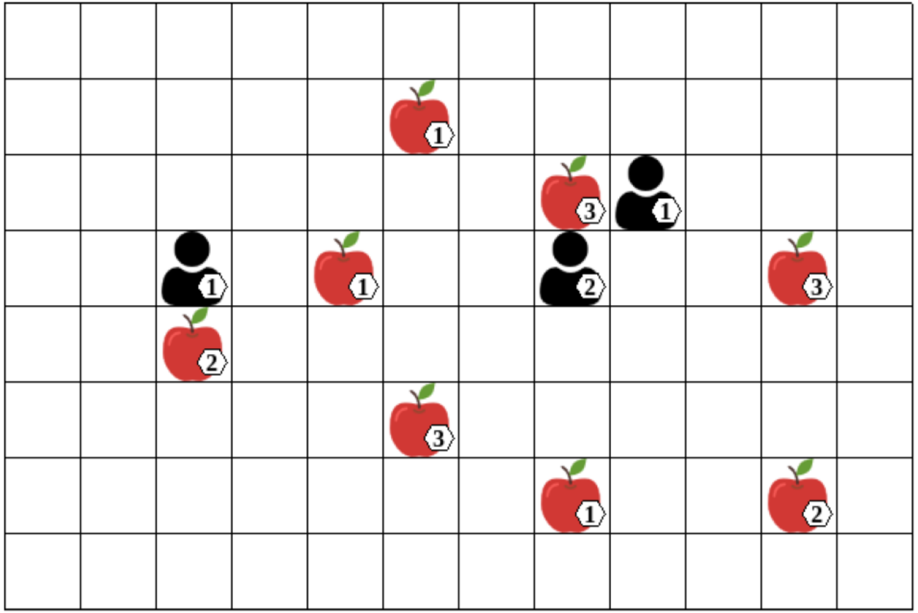
Alireza Moradi

Multi-agent systems (MAS)

- A **multi-agent system** consists of multiple decision-making agents which interact in a shared environment to achieve common or conflicting goals.

- Additional complexity arises in such systems.

- In this example we have two kinds of agents with different skills:



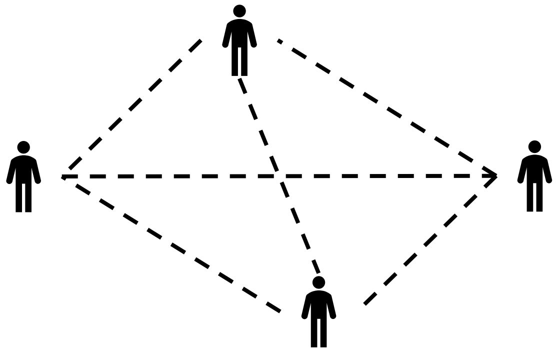
Multi-Agent Reinforcement Learning (MARL)

- **Multi-agent reinforcement learning** algorithms learn optimal policies for a set of agents in a multi-agent system.

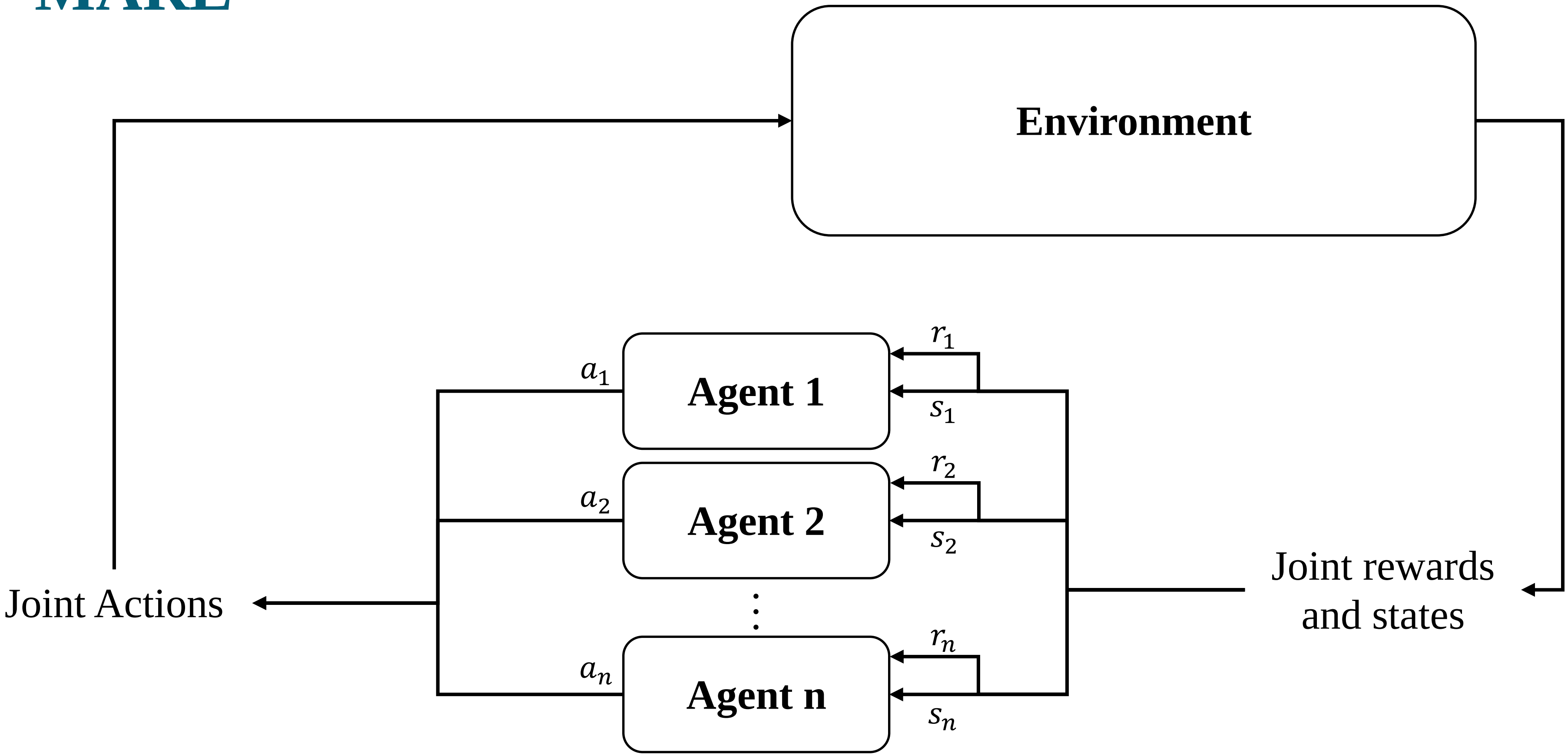
MARL = Reinforcement Learning(**RL**) in Multi-Agent Systems(**MAS**)

- In last session, we saw how RL allows a single agent to learn a policy that maximizes a possibly delayed reward signal in its environment.

- However, when multiple agents apply RL in a shared environment, the optimal policy of an agent depends not only on the environment, but on the policies of the other agents as well.



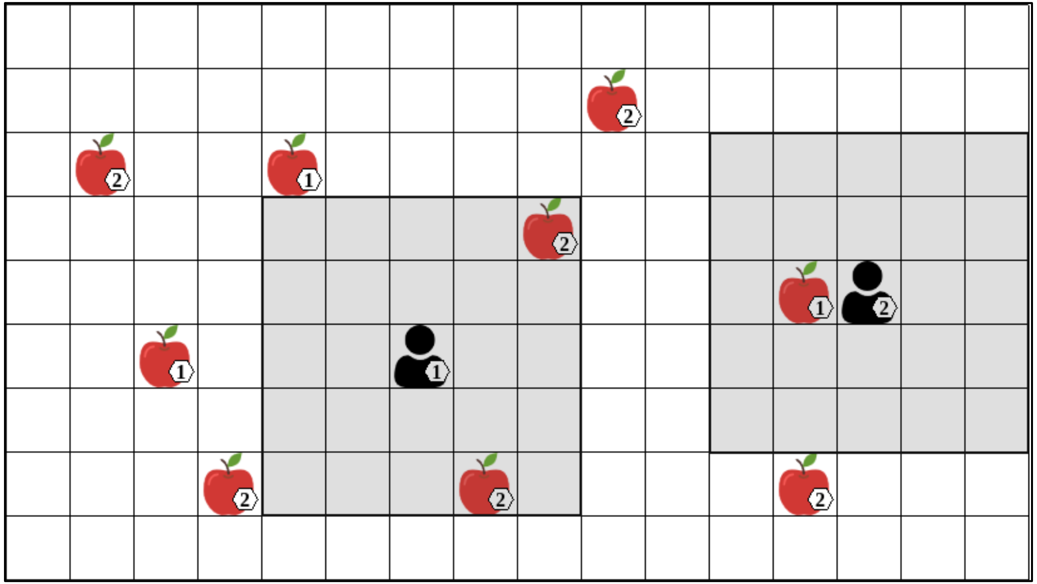
MARL


Environment
r_1
a_1
Agent 1
s_1
r_2
a_2
Agent 2
s_2
Joint Actions
Joint rewards
and states
r_n
a_n
Agent n
s_n

Partial observability

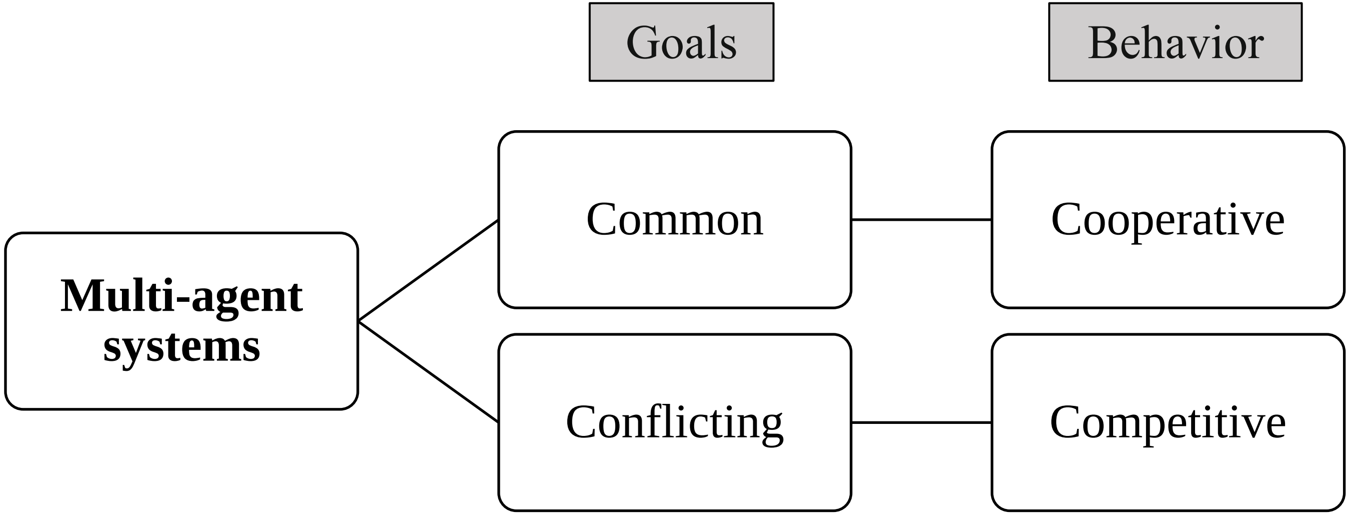
- Agents observe the world through their local vision fields (shown as gray rectangles around the agents).

State ≥ Observation



Cooperative & Competitive behavior

- In MAS, multiple agents are acting in the same environment to achieve **common** or **conflicting** goals.



- **Reward Structure** heavily influences agents' behavior. A reward system that emphasizes collective success promotes cooperation.

Cooperative Behavior

- With cooperative behavior, agents work together to achieve a common goal that benefits everyone.

- By collaborating, agents can achieve more complex tasks, solve problems more efficiently, and potentially gain access to resources they couldn't obtain individually.

- Examples:

    - Swarm robots working together to build a structure.
    - Ants cooperating to find and transport food back to their colony.
    - Multi-Robot Warehouse Management

Competitive behavior

- Agents compete for limited resources or rewards, aiming to outperform each other.

- Examples:
  - Auction houses where bidders compete for items.
  - Predators competing for prey in an ecosystem.
  - Trading bots in a financial market trying to buy low and sell high.

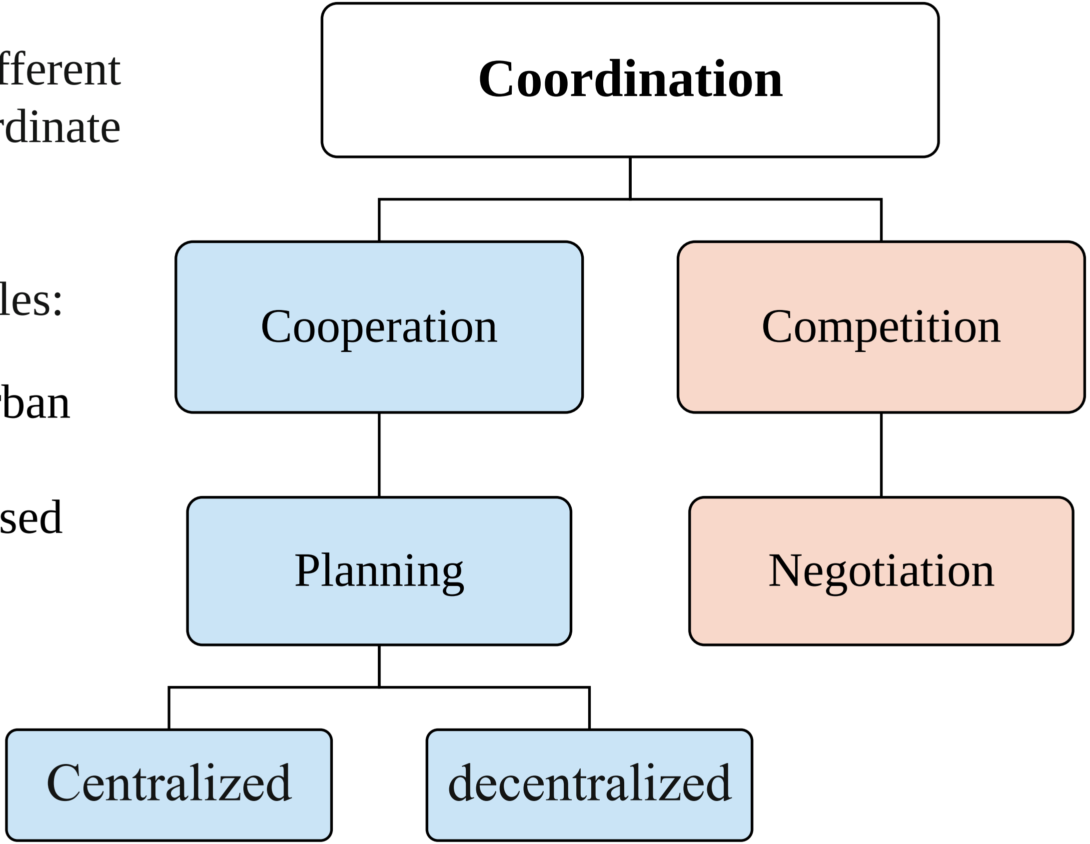
✓ What other examples?

Communications

- A taxonomy of some of the different ways in which agents can coordinate their behavior and activities:

- Decentralized planning examples:

    - autonomous driving in urban environments
    - a team of mobile robots used in search-and-rescue



Game theory

Game theory

- **Game theory** provides a theoretical basis to the field of multi-agent systems.
- Games = Models of Multi-Agent Interaction
- Some examples:

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

(a) Rock-Paper-Scissors

	A	B
A	10	0
B	0	10

(b) Coordination Game

	C	D
C	-1,-1	-5,0
D	0,-5	-3,-3

(c) Prisoner's Dilemma

Game theory

Game theory	RL	Description
game	environment	Model specifying the possible actions, observations, and rewards of agents, and the dynamics of how the state evolves over time and in response to actions.
player	agent	An entity which makes decisions.
payoff, utility	reward	Scalar value received by an agent/player after taking an action.
strategy	policy	Function used by an agent/player to assign probabilities to actions.

Game theory

- **Game theory** is the study of mathematical models of **strategic** interactions among **rational** agents.

- **Strategic behavior** is basically decisions that consider the possible reactions of others. What the opponent does also depends upon what he thinks the first player will do.

- Recall that **rationality** is doing the right thing.

- We assume "common knowledge" of rationality.

✓ What implications does this common knowledge have for players?

Keynes Beauty Contest Game

- Each player names an integer between 1 and 100

- The player who names the integer closest to two thirds of the average integer wins a prize, the other players get nothing.

✓ What is your respond? (Google forum link)

Keynes Beauty Contest Game

- Suppose a player believes the average play will be X (including his or her own integer)

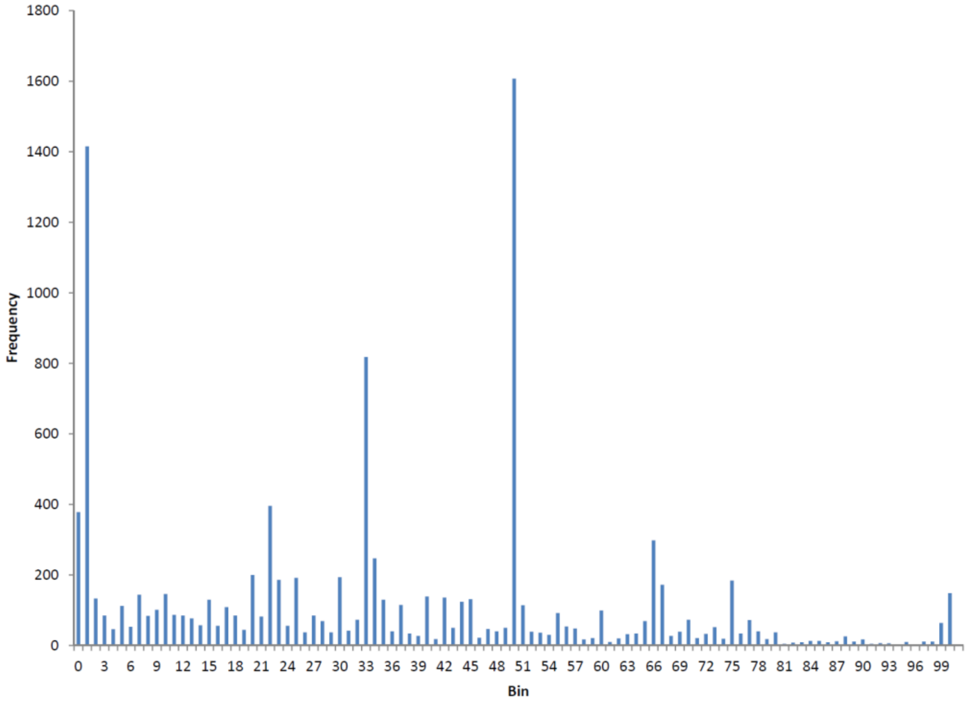
- That player's optimal strategy is to say the closest integer to $\frac{2}{3}X$.

  X must be less than 100, so the optimal strategy of any player must be no more than 67.
  If X is no more than 67, then the optimal strategy of any player must be no more than $\frac{2}{3} \times 67$.
  …

✓ Optimal answer?

Keynes Beauty Contest Game

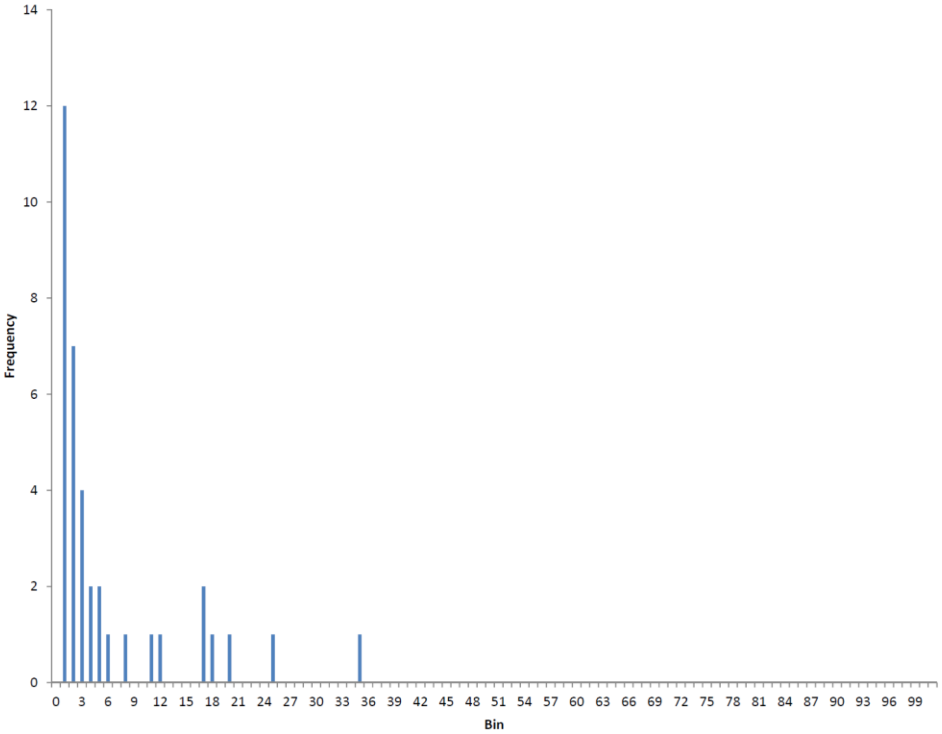
- Experiment results:



Mean	Mode	Median	Winner
34	50	33	23

Keynes Beauty Contest Game

- Experimenting again:



Mean	Mode	Median	Winner
6	1	2	4

Game

- A game has 3 components:
  1. **Players**
  2. **Actions**
  3. **Utilities** => The preferences of each player over outcomes

- In **normal-form games** (also called **matrix games**), we represent the payoffs each player receives as a function of the actions they simultaneously choose.

- **Extensive Form** includes timing of moves. Players move sequentially, represented as a tree

✓ Example?

Matrix representation form

- A 2-player game as a matrix form(normal form):

		Player B	
		Action 1	Action 2
Player A	Action 1	u_{11}^{A}, u_{11}^{B}	u_{12}^{A}, u_{12}^{B}
	Action 2	u_{21}^{A}, u_{21}^{B}	u_{22}^{A}, u_{22}^{B}

- Player A is row player and player B is column player.

Game theory

- Game theory is now a standard tool in **economics.**

- For example, Game theory is a way of modeling the economic activity of competitive firms as a simple game.

		Firm B	
		Stable price	Price war
Firm A	Stable prices	40,40	0,60
	Price war	60,0	3,3

Zero-Sum vs. Non-Zero-Sum Games

- **Zero-sum** is a situation, in which one player's gain is equivalent to another's loss, so the net change in utilities is zero.

- **Non-Zero-Sum** means there's at least one outcome in which $u_{ij}^{A} + u_{ij}^{B} \neq 0$.

- A **zero-sum game** is also called a strictly competitive game, while **non-zero-sum games** can be either competitive or non-competitive.

✓ Examples:
  - Chess
  - Trade agreement
  - Financial derivatives

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

Rock-Paper-Scissors

Prisoner’s Dilemma

- The **prisoner's dilemma** presents a situation where two parties, separated and unable to communicate, must each choose between cooperating with the other or not.

	B Cooperates	B Defects
A Cooperates	**-1 , -1** A’s payoff B’s payoff	**-5 , 1** A’s payoff B’s payoff
A Defects	**1 , -5** A’s payoff B’s payoff	**-2 , -2** A’s payoff B’s payoff

Prisoner’s Dilemma

- Every game in this form is a **Prisoner’s Dilemma**:

		Player B	
		I	II
Player A	I	a, a	b, c
	II	c, b	d, d

$$if\ c > a > d > b$$

- **Prisoner’s Dilemma** is a non-zero sum game.

Prisoner’s Dilemma

- A **solution** for a game is a joint strategy that dictates the optimal actions for all players.

✓ What action do you think each player should choose in Prisoner’s Dilemma?

		Player B	
		Cooperates	Defects
Player A	Cooperates	-1,-1	-5,1
	Defects	1,-5	-2,-2

Infinitely repeated game

- An **infinitely repeated game** consists of an infinite sequence of repetitions of a game.

- Players observe each other's action choices in each period and have perfect recall.

Iteration	1	2	3	4	…
A payoffs	p_1^A	p_2^A	p_3^A	p_4^A	…
B payoffs	p_1^B	p_2^B	p_3^B	p_4^B	…

- The players discount payoffs with discount factor $\gamma \in (0,1)$.

$$R = \sum_t \gamma^t \times p(t)$$

Iterated prisoner's dilemma

- If two players play the prisoner's dilemma more than once in succession, remember their opponent's previous actions, and are allowed to change their strategy accordingly, the game is called the **iterated prisoner's dilemma**.

		Player B	
		Cooperates	Defects
Player A	Cooperates	-1,-1	-5,1
	Defects	-5,1	-2,-2

Iteration	1	2	3	…	M
A payoff					
B payoff					

✓ What strategy is most effective here?

Iterated prisoner's dilemma

- Some possible strategies:
  - Defection
  - Cooperation
  - Tit for Tat
  - Tit for Two Tats
  - Trigger
  - Random
  - Tit for Tat with Occasional Surprise D
  - …

		Player B	
		Cooperates	Defects
Player A	Cooperates	0,0	-1,0
	Defects	0,-1	-3,-3

✓ Think about what happens to defectors in a society of cooperators or cooperators in a society of defectors. Who will be able to succeed?

Iterated prisoner's dilemma

- You can play this game here.

- You can also see here for battle of residents and invaders

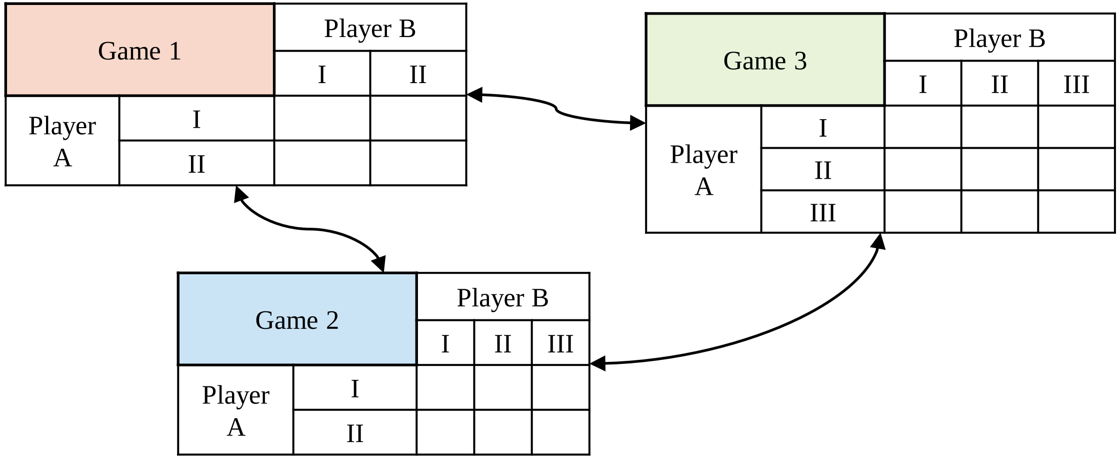
Resident		
Last Move	**Last Opponent**	**Strategy**
Cooperate	Cooperate	Cooperate
Cooperate	Defect	Defect
Defect	Cooperate	Cooperate
Defect	Defect	Cooperate
	Initial	Cooperate

Invader		
Last Move	**Last Opponent**	**Strategy**
Cooperate	Cooperate	Defect
Cooperate	Defect	Defect
Defect	Cooperate	Cooperate
Defect	Defect	Defect
	Initial	Defect

Stochastic Games

✓ What if we didn't always repeat back to the same game?

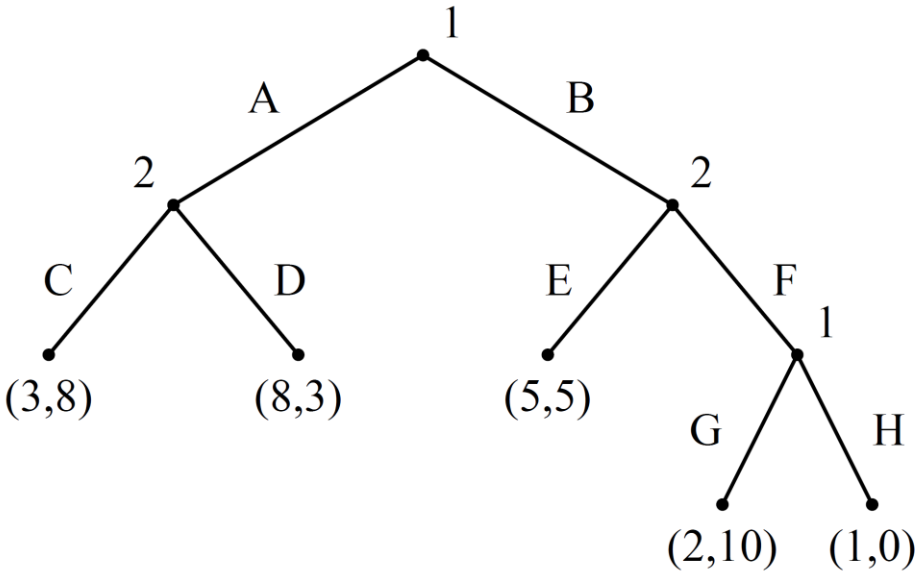
- A **stochastic game** is a generalization of repeated games. In a stochastic game, agents repeatedly play games from a set of normal-form games. The game played at any iteration depends on the previous game played and the actions taken by all agents in that game.



Extensive form Games

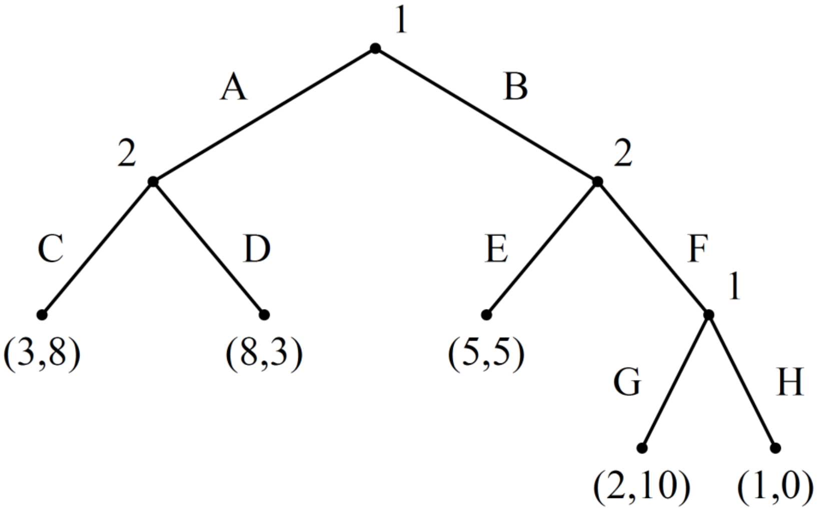
- The normal form game representation does not incorporate any notion of sequence, or time, of the actions of the players.

- The **extensive form** is an alternative representation that makes the temporal structure explicit.



Extensive form Games

- we can convert an extensive-form game into normal form:



	CE	CF	DE	DF
AG	$3,8$	$3,8$	$8,3$	$8,3$
AH	$3,8$	$3,8$	$8,3$	$8,3$
BG	$5,5$	$2,10$	$5,5$	$2,10$
BH	$5,5$	$1,0$	$5,5$	$1,0$

Solution to a game

- A **solution** for a game is a joint strategy.

- There exists a series of **equilibrium solution** concepts:

  - Nash equilibrium
  - Minimax
  - Correlated equilibrium
  - …

- These equilibrium solutions are all based on the idea that every agent is best-responding to all other agents under the equilibrium, and hence no agent can unilaterally deviate from the equilibrium to increase its returns.

Dominant and dominated strategy

- A **dominant strategy** is when one choice gives better result than other, no matter what the opponents do.

- If one of a player's strategies is never the right thing to do, no matter what the opponents do, then it is **strictly dominated**.

- Example:

Game		Player B		
		I	II	III
Player A	I	1,2	1,5	2,4
	II	2,3	3,7	1,1
	III	2,4	4,9	1,3

Nash equilibrium

- **Nash equilibrium** is a pattern of behavior, whereby everybody is doing as well as they can, given what everybody else is doing.

- The Nash equilibrium solution concept applies the idea of a mutual best response to general-sum games with two or more agents. That such a solution exists in any general-sum non-repeated normal-form game was first proven in the celebrated work of Nash (1950).

- **Nash equilibrium** => where each player has nothing to gain by changing

- Nash’s theorem only applies to finite games.

Game		Player B		
		I	II	III
Player A	I			
	II			
	III			

Nash equilibrium

- We can sometimes find Nash equilibrium with dominant strategies.

- What is the solution here?

Game		Player B		
		I	II	III
Player A	I	1,2	1,5	2,7
	II	2,3	3,7	1,1
	III	2,4	4,9	1,3

E-Nash equilibrium

- in many applications, reaching a strict equilibrium may be **too computationally costly**.

- Instead, it may be good enough to compute a solution that is sufficiently close to a strict equilibrium, meaning that agents could technically deviate to improve their returns, but any such gains are sufficiently small.

- The e-Nash equilibrium relaxes the strict Nash equilibrium by requiring that no agent can improve its expected returns by more than epsilon > 0 when deviating from its policy in the equilibrium.

- This is basically a **trade off** between computation and optimality.

Minimax

- **Minimax** is a solution concept defined for two-agent zero-sum games, in which one agent's reward is the negative of the other agent's reward.

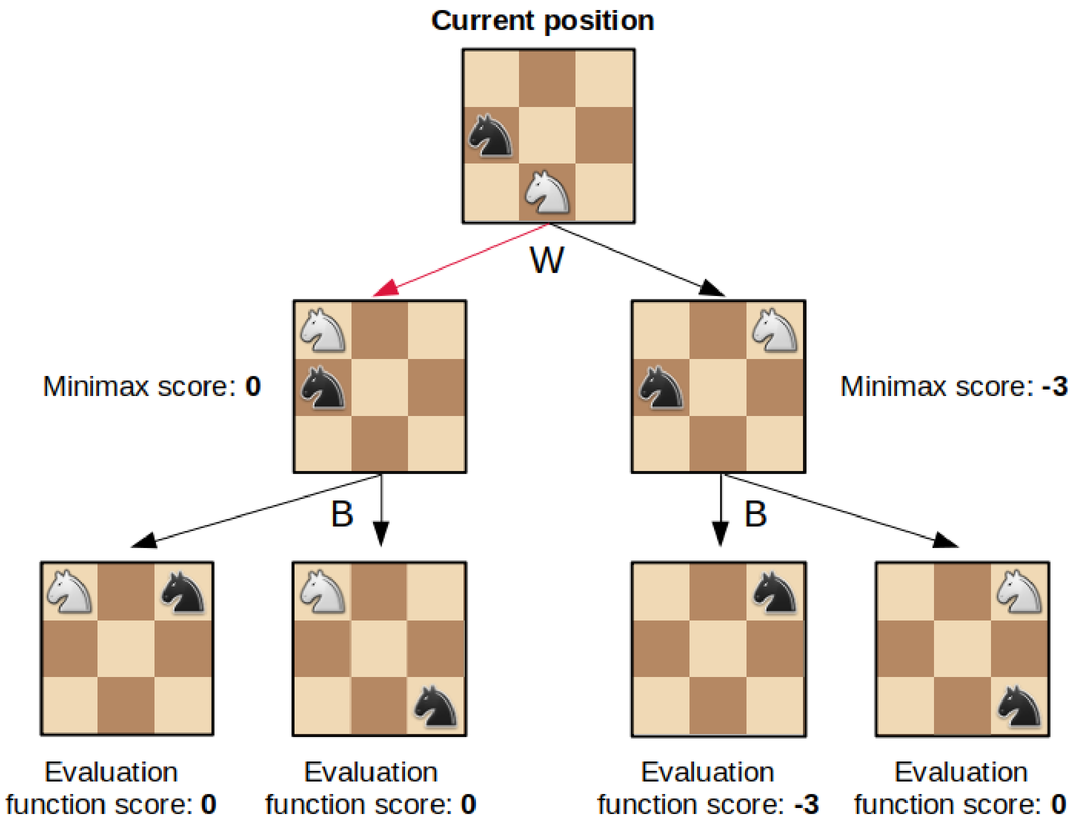
- **Minimax** is a decision rule used to minimize the worst-case potential loss.

- In other words, a player considers all the best opponent responses to his strategies and selects the strategy such that the opponent's best strategy gives a payoff as large as possible.

- Stockfish uses min-max as its core.

Minimax

- Example:



Minimax

- Another example: