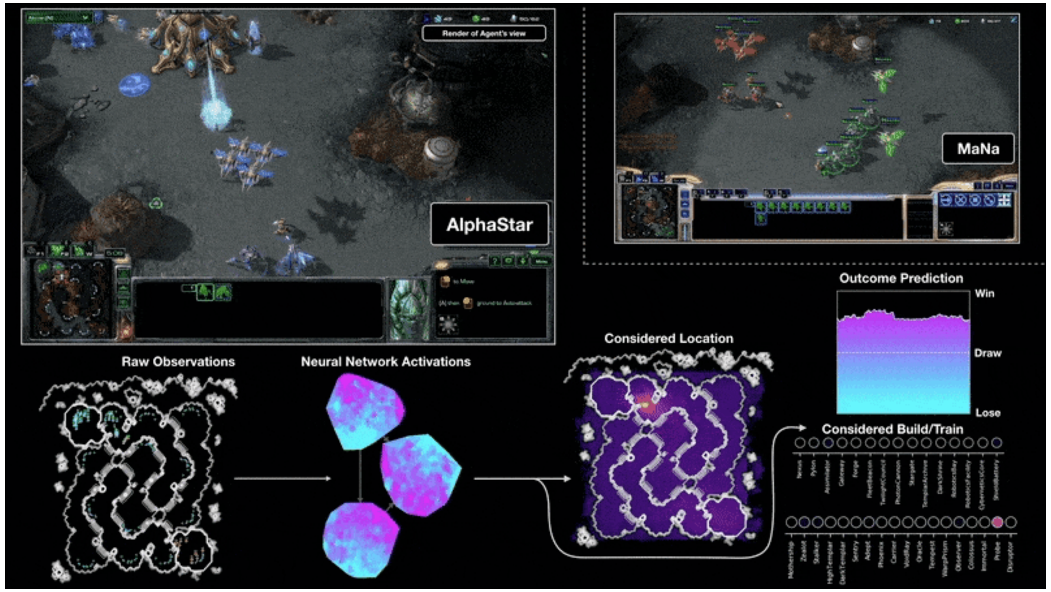
MULTI-AGENT REINFORCEMENT LEARNING

The "Sociology" of Multi-Agent Systems From Individual to Collective Intelligence Evolution

JULIA WANG

When individual intelligence meets collective behavior, magic happens – or chaos ensues."

THE ALPHA-STAR REVELATION

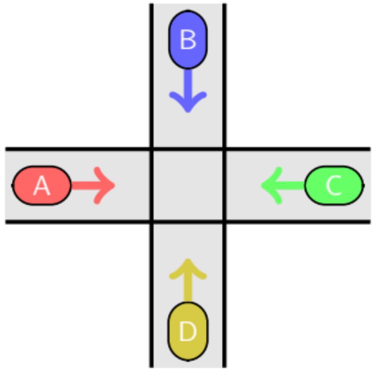


https://deepmind.google/discover/blog/alphastar-mastering-the-real-time-strategy-game-starcraft-ii/

A Moment That Changed Everything

- **2019:** AlphaStar defeats professional StarCraft II players.
- **The Twist:** Not one monolithic AI, but multiple specialized agents coordinating.
- **The Challenge:** Each unit must coordinate with others in real-time.
- **The Question:** How do you train agents that must work together?

THOUGHT EXPERIMENT: THE UNCONTROLLED INTERSECTION



The Dilemma

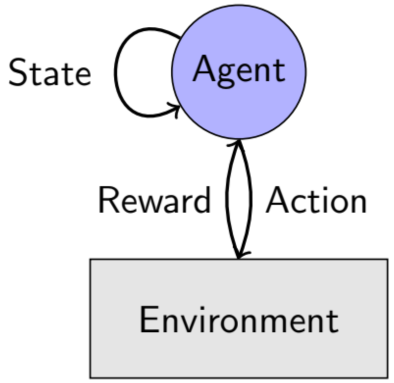
- Each autonomous car wants to cross and minimize its travel time.
- All four cars arrive at the intersection at the same time.
- There is no traffic light and no central controller.
- **Question**: What should each car do? Go? Wait?

This is the essence of Multi-Agent Reinforcement Learning!

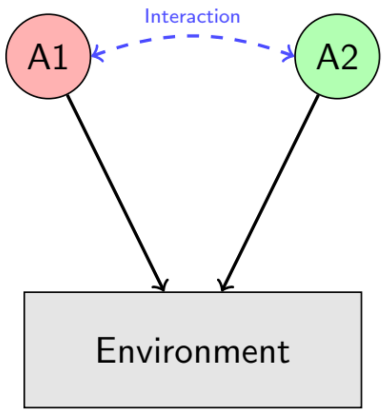
Agents must learn to either cooperate or compete without explicit instructions.

THE CORE CHALLENGE: A DYNAMIC ENVIRONMENT

Single Agent World



Multi-Agent World



The Game Changer

Your environment now includes other learning agents. Your optimal action depends on their actions, and their optimal action depends on yours.

THE EVOLUTION OF MATHEMATICAL FRAMEWORKS

A Deceptively Simple Generalization

Markov Decision Process (MDP)
(The Solitary Learner)

State: $s \in \mathcal{S}$

Action: $a \in \mathcal{A}$

Transition: $P(s'|s, a)$

Reward: $R(s, a)$

$\Rightarrow$

Markov Game (or Stochastic Game)
(The Social Network)

State: $s \in \mathcal{S}$

Actions: $\mathbf{a} = (a_1, \ldots, a_n)$

Transition: $P(s'|s, \mathbf{a})$

Rewards: $R_i(s, \mathbf{a})$ for each i

Key Insight

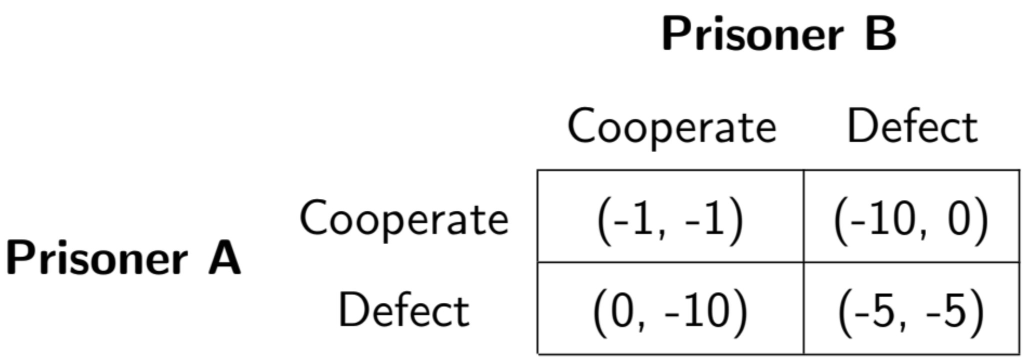
The reward and next state now depend on the **joint action** of all agents. This introduces game theory and a whole new layer of complexity.

NASH EQUILIBRIUM: WHEN NO ONE HAS AN INCENTIVE TO DEVIATE

The ”Point of No Regrets”

A set of strategies is a Nash Equilibrium if no single agent can do better by unilaterally changing their strategy, assuming everyone else’s strategy remains the same.

Classic Example: Prisoner’s Dilemma



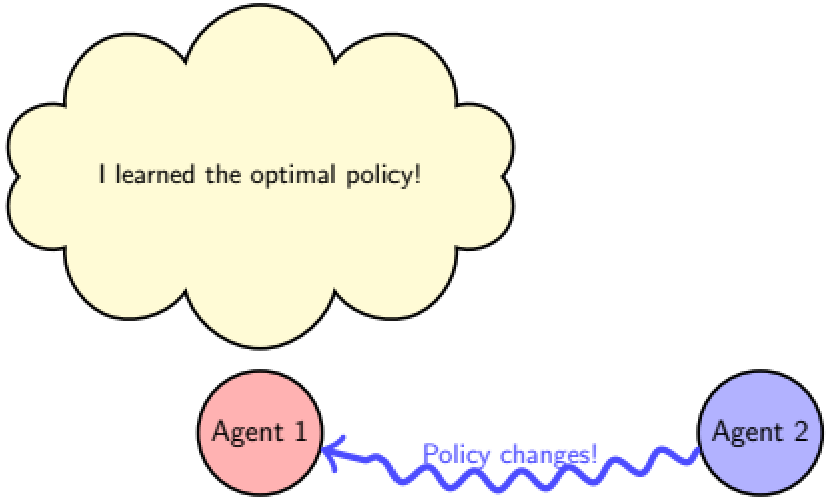
		Prisoner B	
		Cooperate	Defect
Prisoner A	Cooperate	(-1, -1)	(-10, 0)
	Defect	(0, -10)	(-5, -5)

Payoffs: (A's years in prison, B's years in prison)

The Paradox of Rationality

The only Nash Equilibrium is (Defect, Defect) , even though (Cooperate, Cooperate) is a better outcome for both. Individual rationality can lead to collective irrationality!

THE NON-STATIONARITY PROBLEM



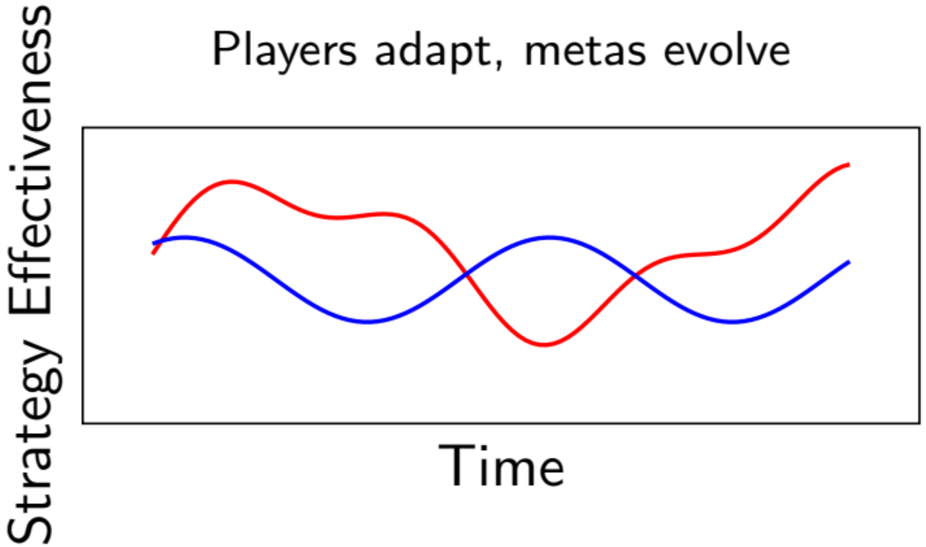
Why Standard Reinforcement Learning Fails

- In standard RL, the environment is assumed to be *stationary* (its rules don't change).
- In MARL, as other agents learn and adapt, they become part of your environment.
- From your perspective, the environment's dynamics are constantly changing!
- Your hard-earned experience can become obsolete in an instant.

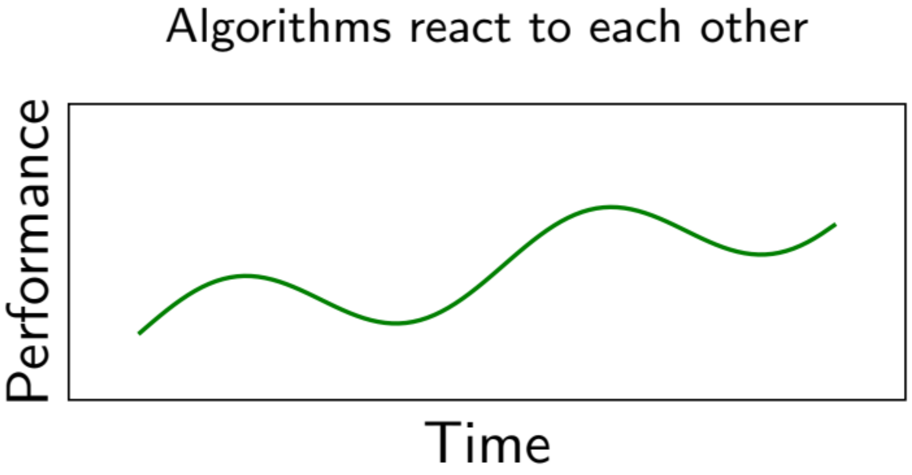
It's like learning to dance... with a partner who keeps changing the music!

VISUALIZING NON-STATIONARITY

Online Gaming ”Meta”



Algorithmic Stock Trading



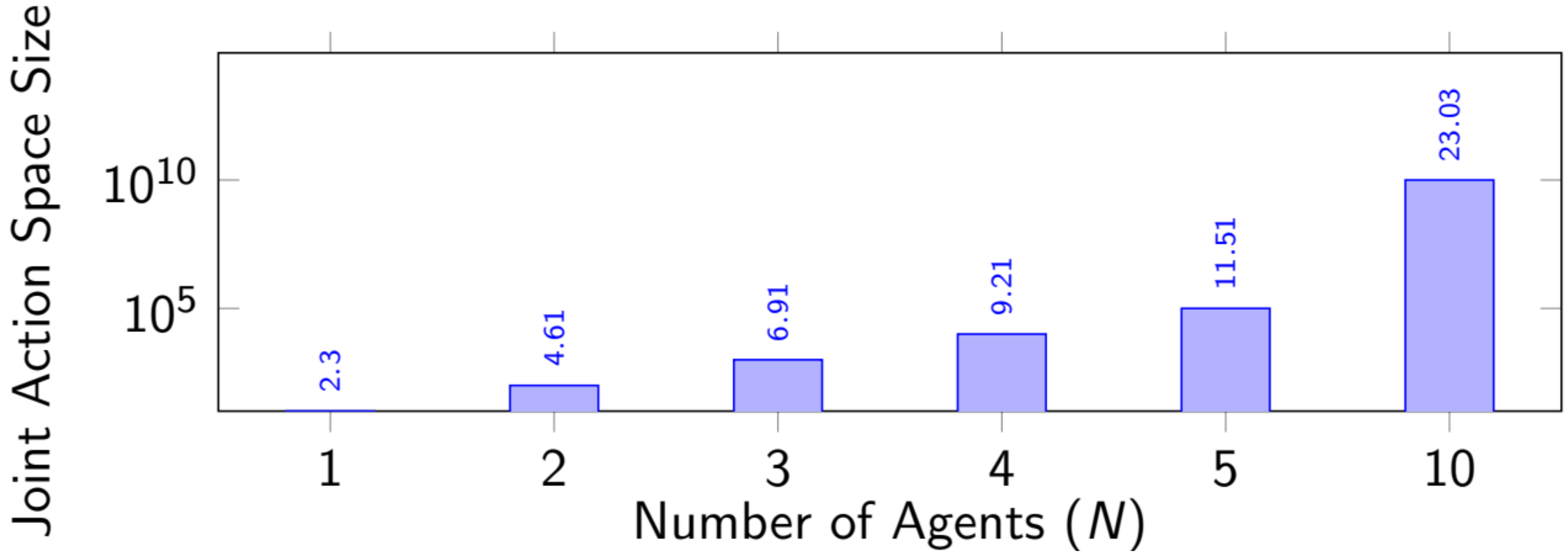
The Challenge

How can an agent learn a stable policy for a world that refuses to stand still?

THE EXPONENTIAL MONSTER

The terrifying growth of the Joint Action Space

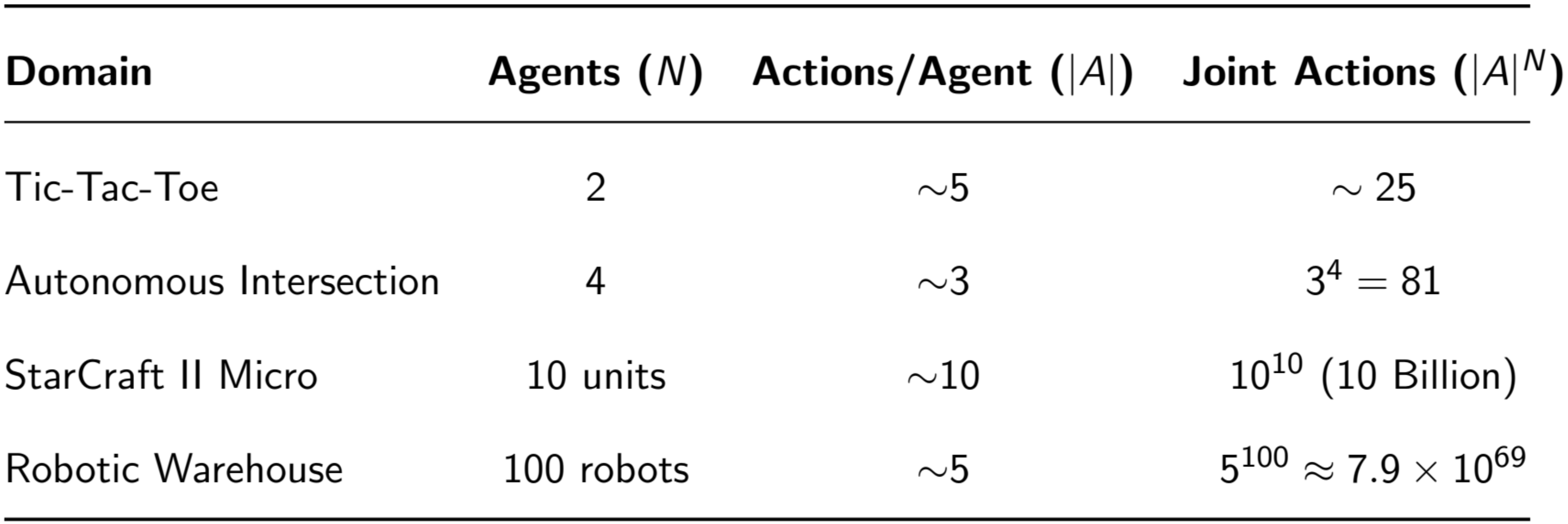
If there are N agents and each has $|A|$ actions, the total number of possible joint actions is $|A|^N$.



Assuming just 10 actions per agent.

WHEN COMBINATORICS ATTACKS IN THE REAL WORLD

Real-World Examples

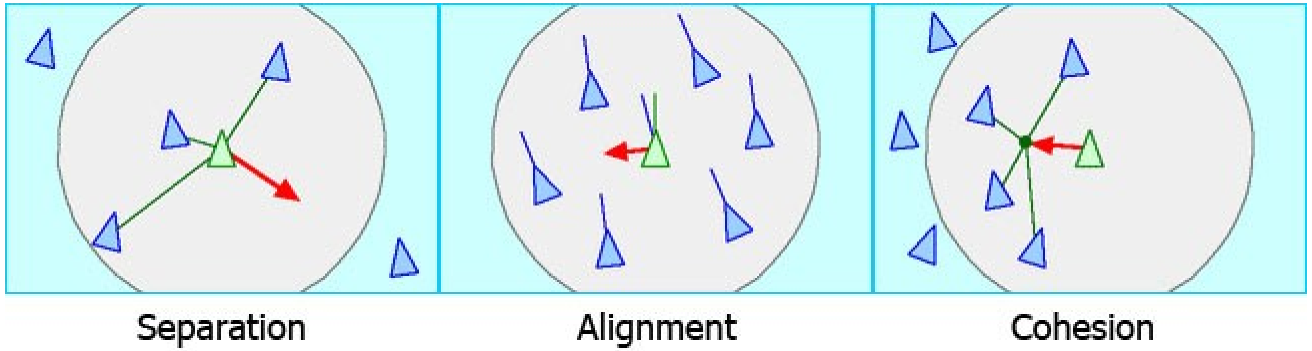


Domain	**Agents (N)**	**Actions/Agent ($\|A\|$)**	**Joint Actions ($\|A\|^N$)**
Tic-Tac-Toe	2	$\sim$5	$\sim$ 25
Autonomous Intersection	4	$\sim$3	$3^4 = 81$
StarCraft II Micro	10 units	$\sim$10	10^{10} (10 Billion)
Robotic Warehouse	100 robots	$\sim$5	$5^{100} \approx 7.9 \times 10^{69}$

The Harsh Truth

It is computationally impossible to explore the full joint action space for any non-trivial multi-agent problem. We cannot simply treat the group as one giant "meta-agent".

THE MAGIC OF EMERGENCE



https://adamprice.io/blog/boids.html

Craig Reynolds' "Boids" (1986): Simple Rules, Complex Global Behavior

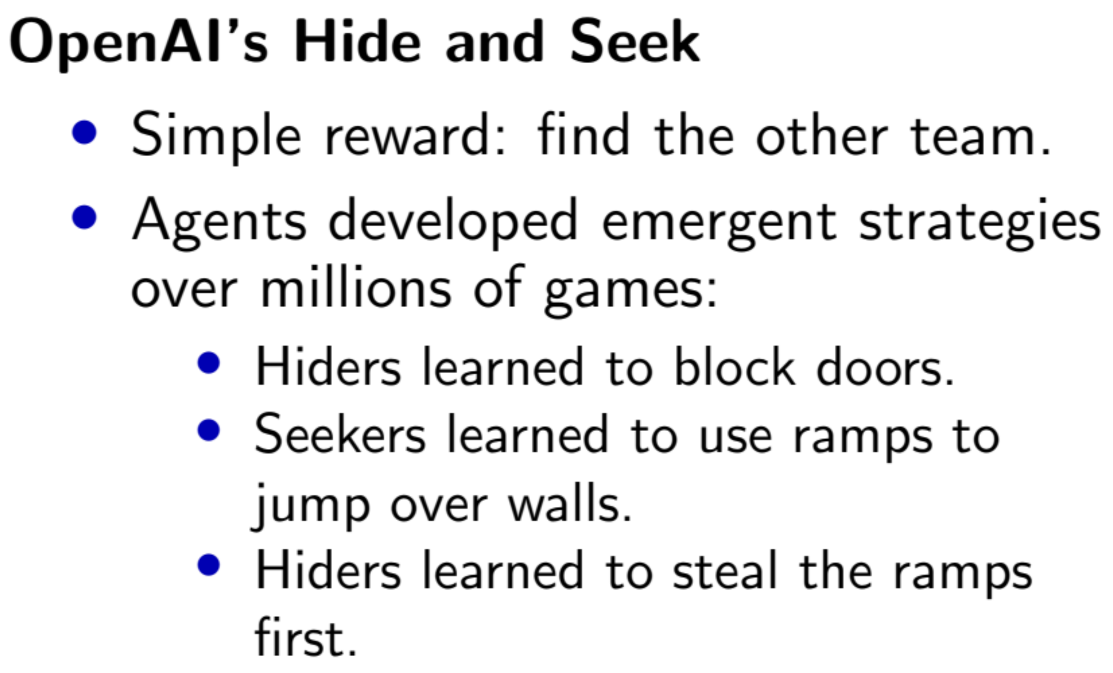
Each "boid" (bird-oid) follows only three simple rules based on its local neighbors:

1. **Separation**: Steer to avoid crowding local flockmates.
2. **Alignment**: Steer towards the average heading of local flockmates.
3. **Cohesion**: Steer to move toward the average position of local flockmates.

The Result

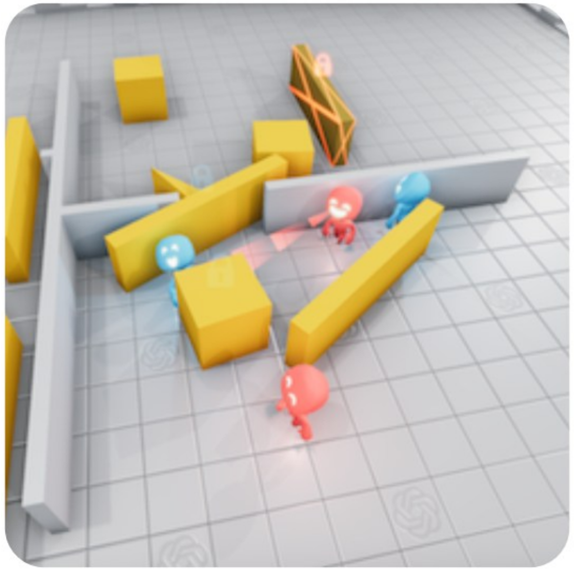
No central controller, no grand plan. Realistic, complex flocking behavior **emerges** from purely local interactions.

EMERGENCE IN MODERN MULTI-AGENT SYSTEMS



OpenAI's Hide and Seek

- Simple reward: find the other team.
- Agents developed emergent strategies over millions of games:
  - Hiders learned to block doors.
  - Seekers learned to use ramps to jump over walls.
  - Hiders learned to steal the ramps first.



https://openai.com/index/emergent-tool-use/

The agents discovered tool use on their own!

Key Insight

The goal of MARL is often not to pre-program complex behaviors, but to create a system where intelligent behaviors can **emerge** through learning.

A GLIMPSE OF ALGORITHMIC APPROACHES

How We Fight Non-Stationarity and Dimensionality

Three major paradigms have emerged, which we will explore in the next lecture.

CTCE (Centralized Training, Centralized Execution) One brain controls all agents! *Pro:* Perfect coordination with unified policy. *Con:* Single point of failure; scalability bottleneck.

CTDE (Centralized Training, Decentralized Execution) Train together, act alone. *Pro:* Stable learning with execution independence. *Con:* Training-execution mismatch; requires global information.

DTDE (Decentralized Training, Decentralized Execution) Every agent for themselves, always! *Pro:* Fully distributed; highly scalable and robust. *Con:* Non-stationary environment; potential instability.

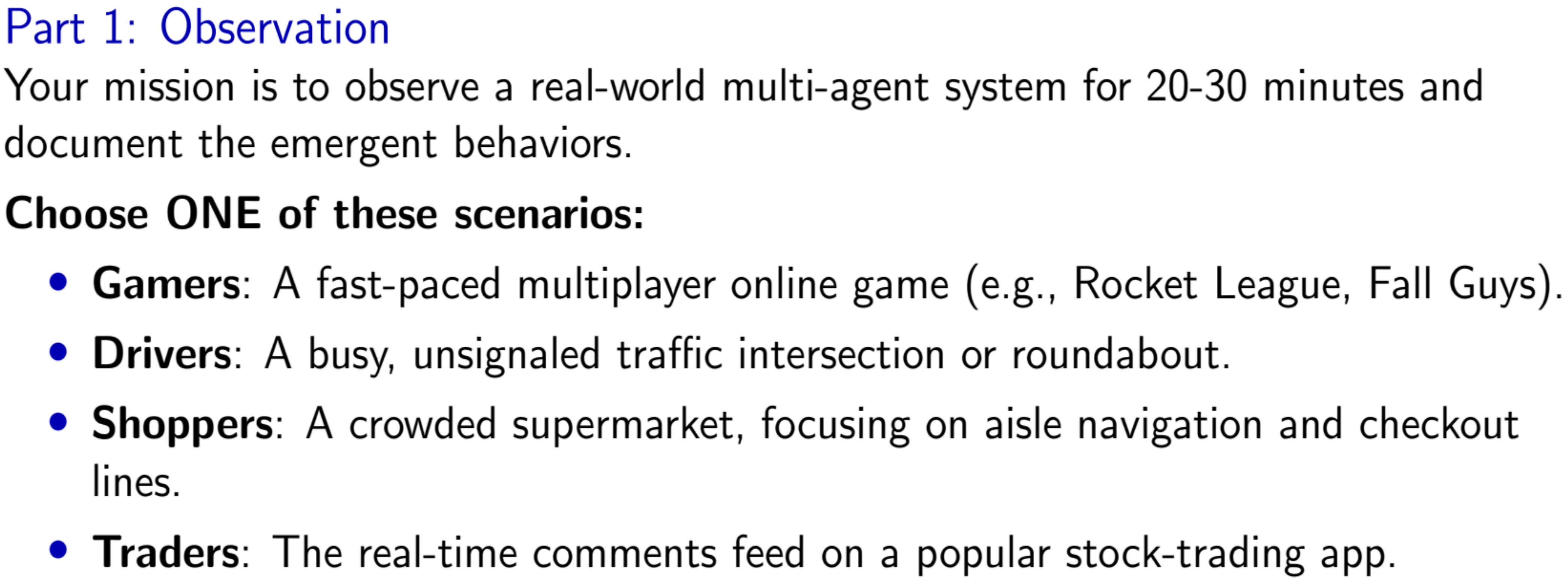
THE JOURNEY SO FAR: KEY TAKEAWAYS

1. **The Paradigm Shift**: From a static "environment" to a dynamic "social system".
2. **The Core Challenges**:
   - **Non-Stationarity**: Your world changes as others learn.
   - **Curse of Dimensionality**: The joint action space explodes.
3. **The Magic of Emergence**: Simple, local rules can create intelligent global behavior.
4. **The Goal**: Design algorithms that enable agents to coordinate towards goals.

The Big Picture

Multi-agent RL is the frontier where reinforcement learning meets game theory and complex systems science.

ASSIGNMENT: BECOME A SOCIAL INTELLIGENCE DETECTIVE



Part 1: Observation

Your mission is to observe a real-world multi-agent system for 20-30 minutes and document the emergent behaviors.

Choose ONE of these scenarios:

- **Gamers**: A fast-paced multiplayer online game (e.g., Rocket League, Fall Guys).
- **Drivers**: A busy, unsignaled traffic intersection or roundabout.
- **Shoppers**: A crowded supermarket, focusing on aisle navigation and checkout lines.
- **Traders**: The real-time comments feed on a popular stock-trading app.

ASSIGNMENT: SUBMISSION DETAILS

Part 2: Analysis and Reporting

Analyze and Report (300-400 words):

1. Describe your chosen system. Who are the ”agents”? What are their likely goals?
2. What cooperative or competitive strategies did you observe emerging naturally?
3. Where did you see evidence of agents adapting to each other’s actions?
4. Briefly describe one behavior that surprised you. Was it efficient or inefficient for the group?

Deliverable

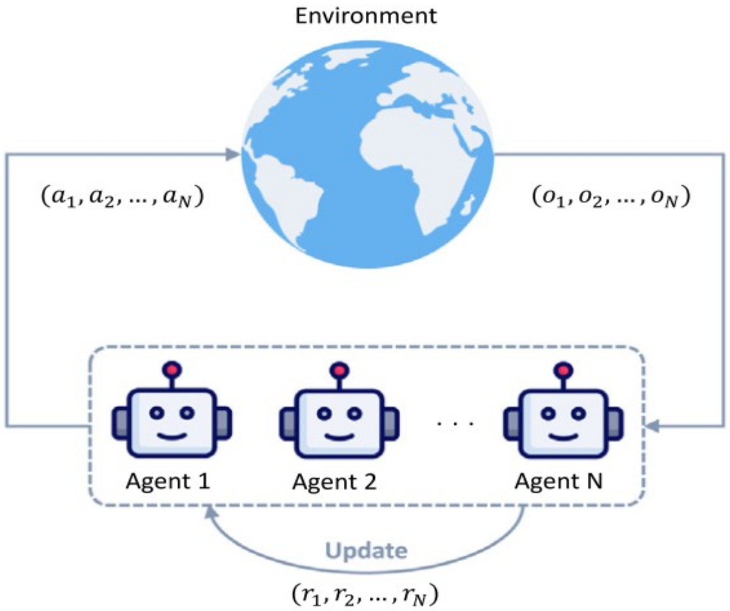
A one-page PDF report answering the four questions above. Due next week.

NEXT TIME: THE ALGORITHM WARS

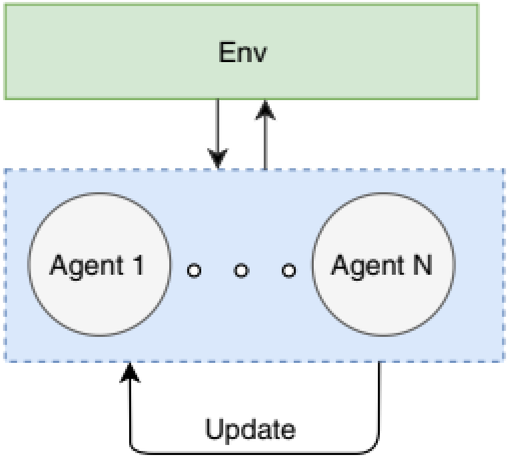
We’ll dive deep into the algorithms.

Which strategy reigns supreme?

NEXT TIME: THE ALGORITHM WARS — TEAM 1



A survey on multi-agent reinforcement learning and its application

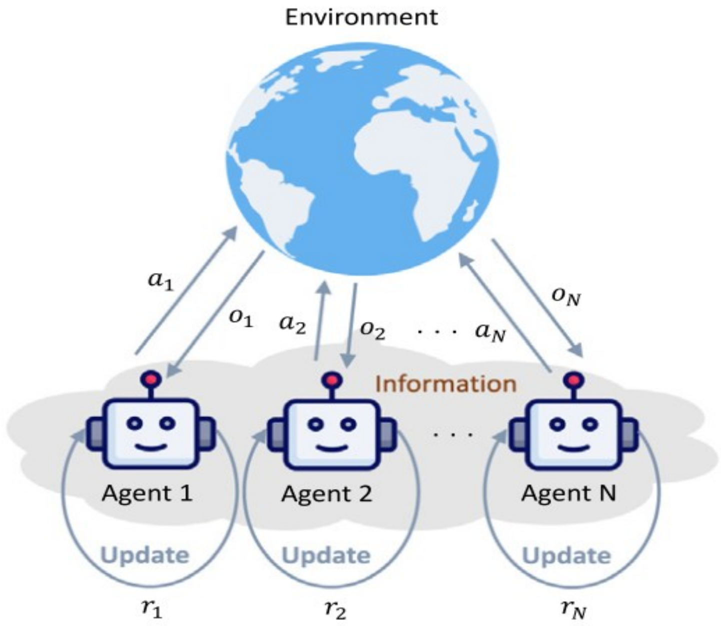


Multi-agent deep reinforcement learning: a survey

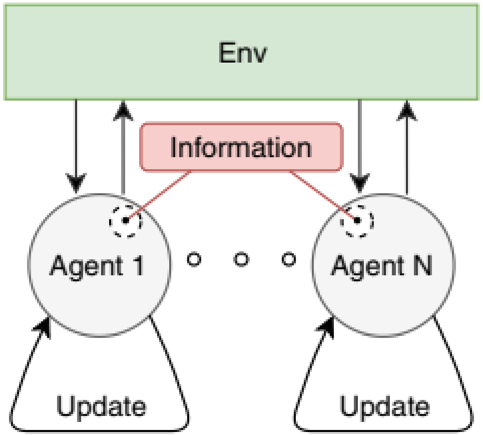
CTCE (Centralized Training, Centralized Execution)

- **Philosophy:** One brain controls all agents!
- **Strength:** Perfect coordination with unified policy.
- **Weakness:** Single point of failure; scalability bottleneck.

NEXT TIME: THE ALGORITHM WARS — TEAM 2



A survey on multi-agent reinforcement learning and its application

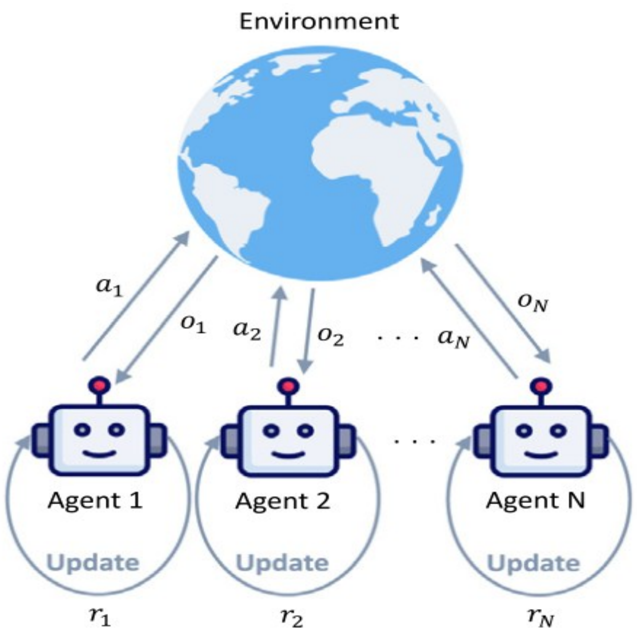


Multi-agent deep reinforcement learning: a survey

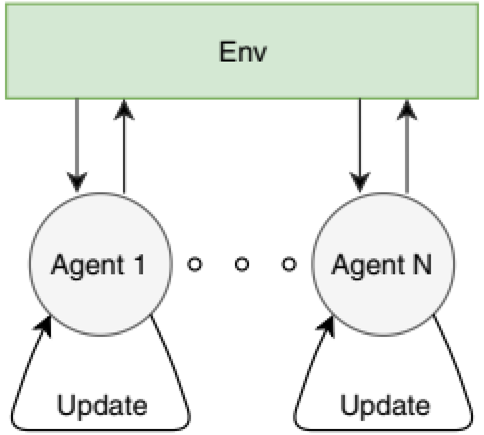
CTDE (Centralized Training, Decentralized Execution)

- **Philosophy:** Train together, act alone.
- **Strength:** Stable learning with execution independence.
- **Weakness:** Training-execution mismatch; requires global information.

NEXT TIME: THE ALGORITHM WARS — TEAM 3



A survey on multi-agent reinforcement learning and its application



Multi-agent deep reinforcement learning: a survey

DTDE (Decentralized Training, Decentralized Execution)

- **Philosophy:** Every agent for themselves, always!
- **Strength:** Fully distributed; highly scalable and robust.
- **Weakness:** Non-stationary environment; potential instability.

Questions?

The adventure into emergent intelligence has just begun!