Multiagent Systems

James R. Wright
University of Alberta

Deep Learning & Reinforcement Learning Summer School 2019

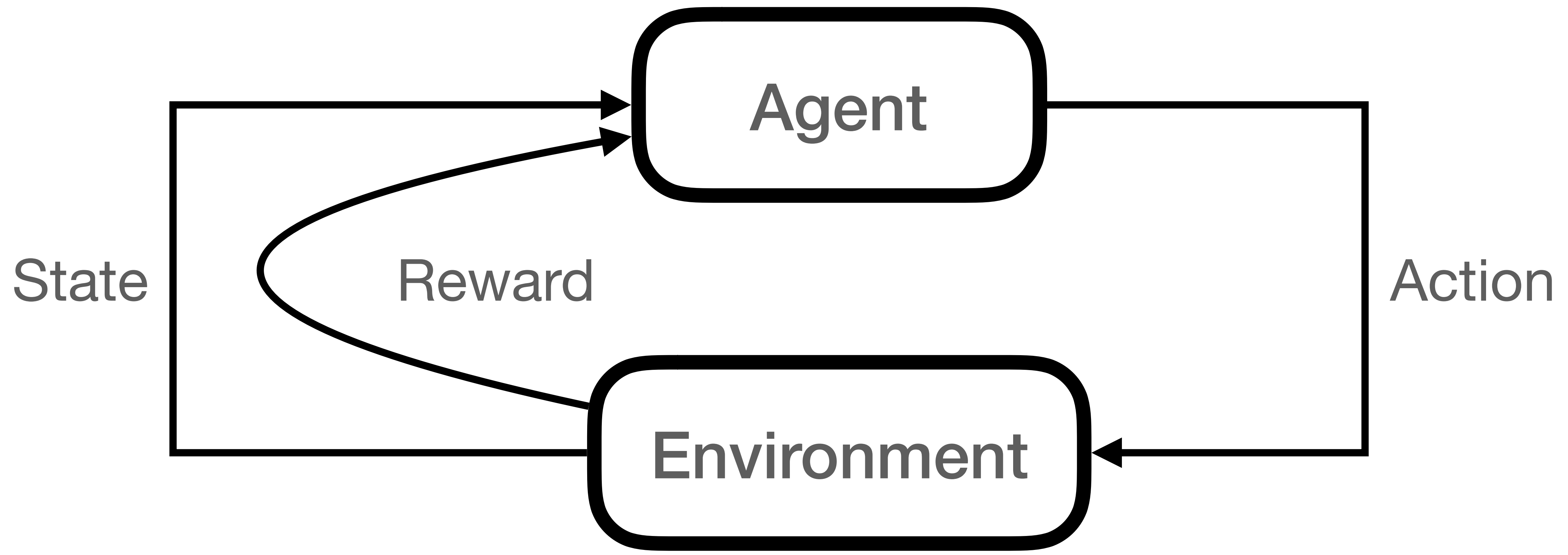
Mo' Agents,
Mo' Problems

Nobody ever says "maybe I should solve my problem with multiagent systems!"

Multiagent systems aren't solutions. They are **problems**.

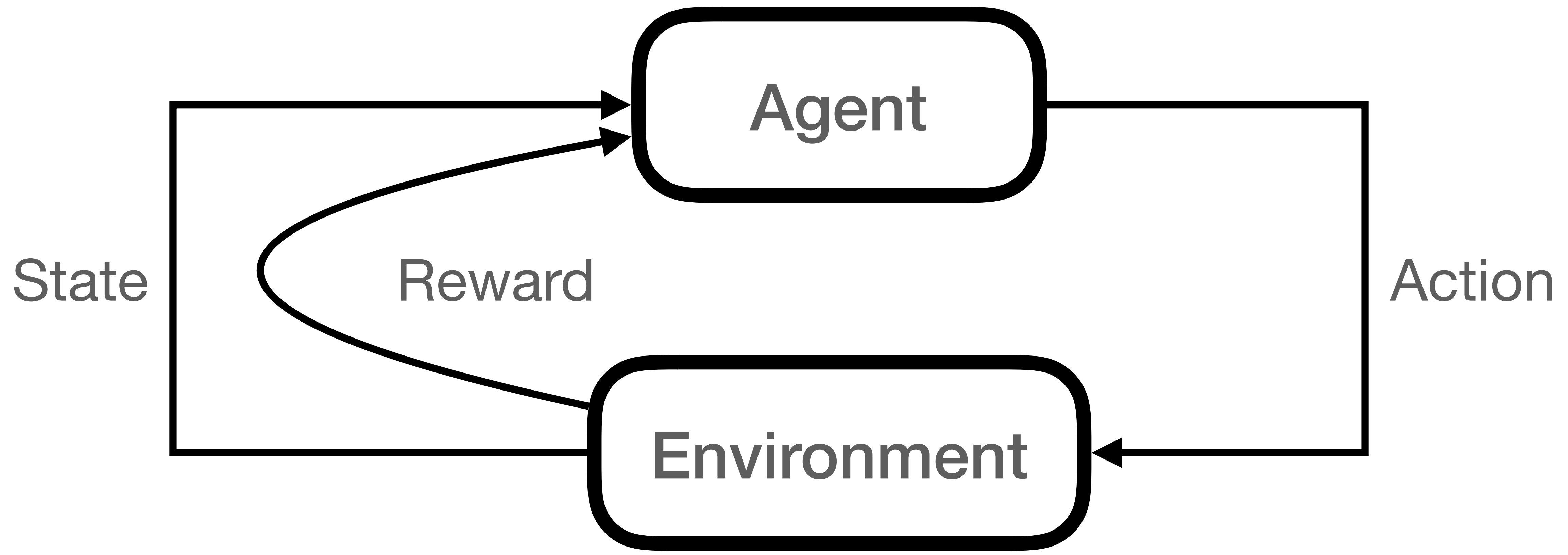
But they are problems that are important to solve.

Single Agent Systems

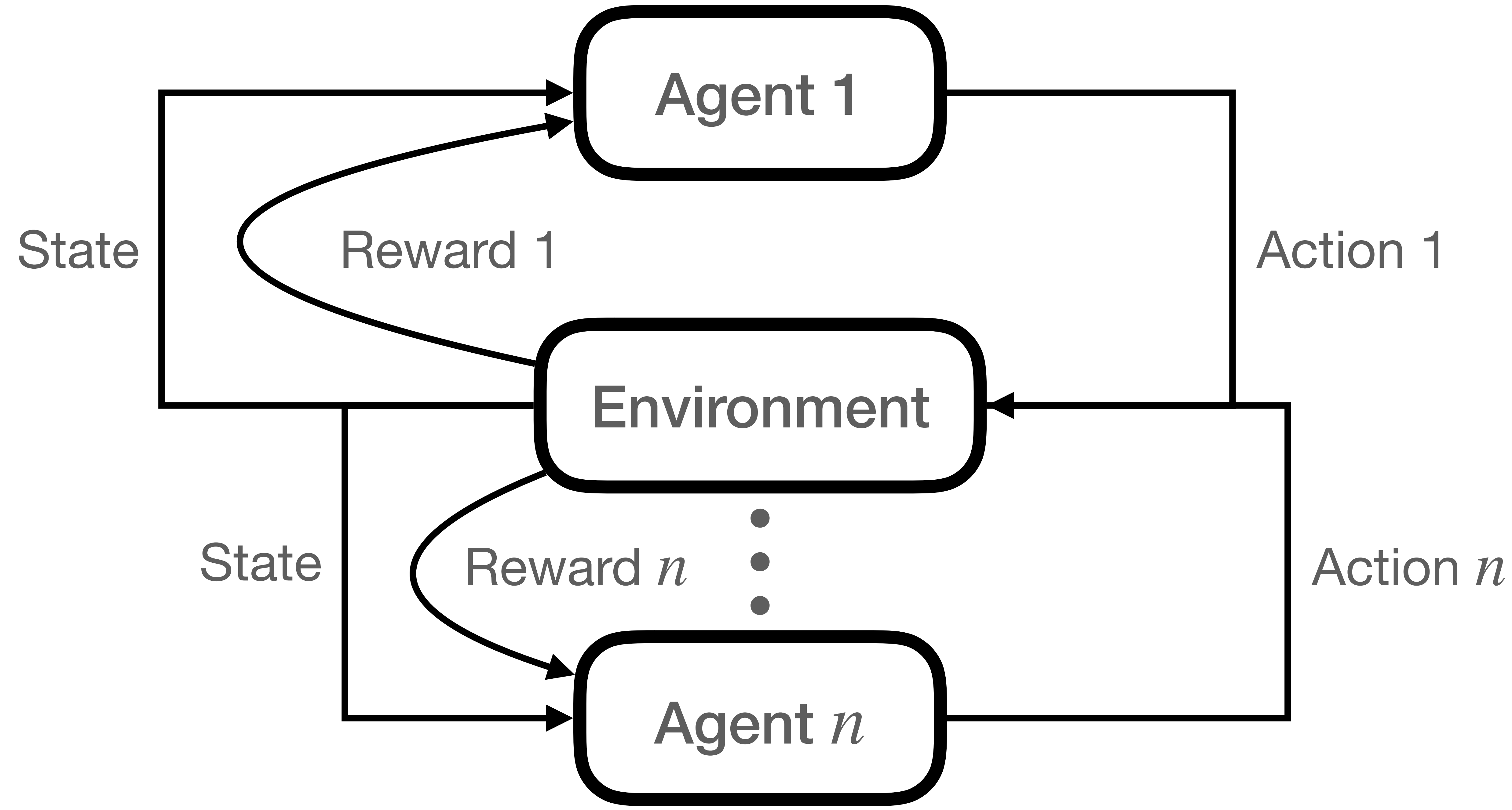


1. An agent interacts with an environment by taking an **action**
2. Agent's action and environment's previous state determine:
    - Agent's reward
    - Environment's next state

Multiagent Systems



Multiagent Systems

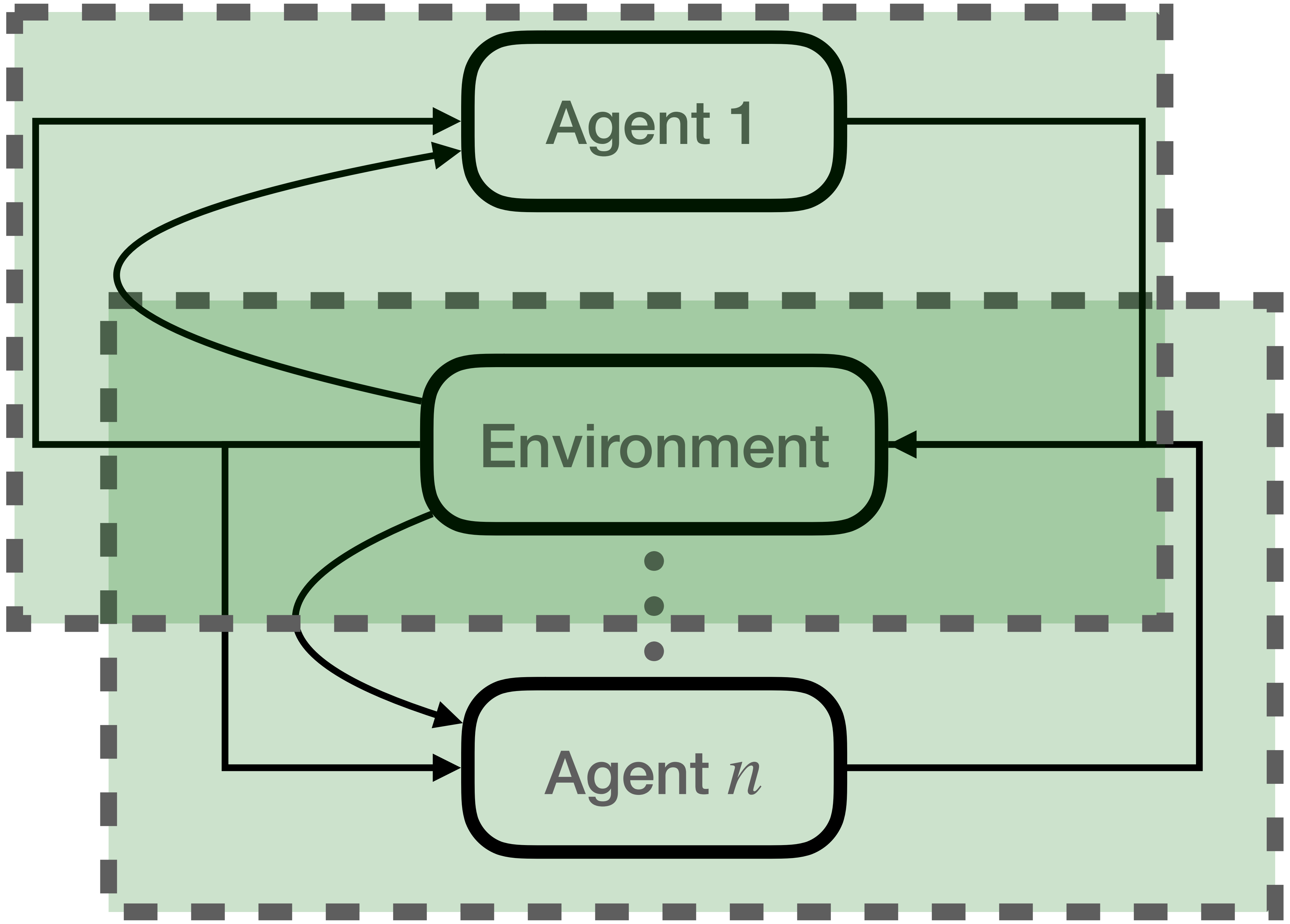


1. Multiple **autonomous** agents
2. With **distinct** goals / priorities / rewards

Multiagent Systems are Widespread

- Games (Poker, Go, Shogi, StarCraft)
- Auctions (ad auctions, spectrum auctions)
- Online platforms (sharing economy, crowdsourcing)
- Cooperative interaction (language, advice, decision support)
- Resource allocation (packet routing, server allocation)
- Kidney exchange
- Urban planning

Multiagent Systems are Nonstationary



- Every agent is part of every other agent's environment
- Each agent may try to **game** the others (special kind of nonstationarity)
  - Environment changes in **anticipation** of agent's actions

Multiagent Objectives are Often Unclear

In a single-agent system:

$$\max_{\pi} \mathbb{E}\left[R^{(1)} + \gamma R^{(2)} + \ldots + \gamma^{T-1} R^{(T)} \mid p\right]$$

Maximize total reward over all the agents:

$$\max_{\boxed{\pi_1,\ldots,\pi_n}} \mathbb{E}\left[\sum_{1 \le i \le n} R_i^{(1)} + \gamma R_i^{(2)} + \ldots + \gamma^{T-1} R_i^{(T)} \middle| p\right]$$

Multiagent Objectives

Agent Design: Maximize individual agent's reward:

$$\max_{\pi_i} \mathbb{E}\left[R_i^{(1)} + \gamma R_i^{(2)} + \ldots + \gamma^{T-1} R_i^{(T)} \mid p, \boxed{\pi_1, \ldots, \pi_{i-1}, \pi_{i+1}, \ldots, \pi_n}\right]$$

Mechanism Design: Maximize total reward *by optimizing the environment*

$$\max_{p} \mathbb{E}\left[\sum_{1 \le i \le n} R_i^{(1)} + \gamma R_i^{(2)} + \ldots + \gamma^{T-1} R_i^{(T)} \,\middle|\, \boxed{\pi_1, \ldots, \pi_n}\right]$$

Both of these depend on the (other) agents' policies!

In This Talk

1. Multiagent systems are **important** and **hard**
2. **Game theory**: Multiagent mathematics
3. Thinking realistically about multiagent systems
4. Predicting human behaviour in normal-form games

(Noncooperative) Game Theory

Game theory is the mathematical study of interaction between multiple rational, self-interested agents.

1. **Rational:** Each agent maximizes their expected utility

2. **Self-interested:** Agents pursue only their **own preferences**

    - *Not* the same as "agents are psychopaths"! Their preferences may include the well-being of other agents.

    - Rather, the agents are **autonomous**: they decide on their own priorities independently.

Fun Game: Prisoner's Dilemma

	Cooperate	Defect
Cooperate	-1,-1	-5,0
Defect	0,-5	-3,-3

Two suspects are being questioned separately by the police.

- If they both remain silent (**cooperate** -- i.e., with each other), then they will both be sentenced to **1 year** on a lesser charge
- If they both implicate each other (defect), then they will both receive a reduced sentence of **3 years**
- If one defects and the other cooperates, the defector is given immunity (0 years) and the cooperator serves a full sentence of **5 years**.

Play the game with someone near you. Then find a new partner and play again. Play 3 times in total, against someone new each time.

Normal-Form Games

The Prisoner's Dilemma is an example of a **normal-form game**.

Agents make a single decision **simultaneously**, and then receive a payoff depending on the **profile** of actions.

There are n players: $N = \{1,2,\ldots,n\}$, indexed by i.

Each player i has a set A_i of available actions.

A tuple $a \in A = A_1 \times \ldots \times A_n$ giving an action a_i for each player i is an an action profile.

Each player i has a **utility function** $u_i : A \to \mathbb{R}$ (their preferences)

Maximize Individual Agent's Rewards: Best Response

In game theory, a rational agent is one that maximizes their **expected utility** given their **beliefs** π_{-i} about the other agents' policies:

$$a_i^* \in \arg\max_{a_i} \mathbb{E}[u_i(a_i, a_{-i}) \mid \pi_{-i}]$$

$\pi_j \in \Delta(A_j)$ is a distribution over j's actions
π_{-i} is a tuple of beliefs about every agent **other than** i

A shorter way to say this: a_i^* is a **best response**:

$$a_i^* \in BR_i(\pi_{-i})$$

Nash Equilibrium

Where do the beliefs π_{-i} come from?

Standard game theoretic **solution concept**:

1. Every agent is perfectly rational
2. Therefore every agent is best responding to the actual policies of the other agents
3. Therefore $\pi_i(a_i) > 0 \implies a_i \in BR_i(\pi_{-i})$ for all i

Any **profile** π^* of policies that satisfies (3) is a **Nash equilibrium**.

Nash Equilibrium Predictions

Every agent is **perfectly rational**, therefore
Every agent **best responds** to every other agent simultaneously
(Otherwise they're behaving irrationally!)

	Ballet	Soccer
Ballet	2, 1	0, 0
Soccer	0, 0	1, 2

Nash equilibria:

1. **(Ballet, Ballet): Neither agent gains by *deviating* (choosing Soccer)**
2. (Soccer, Soccer): Neither agent gains by *deviating* (choosing Ballet)
3. $\left(\left[\frac{2}{3}: Ballet, \frac{1}{3}: Soccer\right], \left[\frac{1}{3}: Ballet, \frac{2}{3}: Soccer\right]\right)$

Maximize Total Reward?

Mechanism Design: Maximize total reward *by optimizing the environment*

$$\max_p \mathbb{E}\left[\sum_{1\le i\le n} R_i^{(1)} + \gamma R_i^{(2)} + \ldots + \gamma^{T-1} R_i^{(T)} \,\middle|\, \pi_1, \ldots, \pi_n\right]$$

- Utility values between agents cannot necessarily be added!
- Agent i with utility u_i and agent j with utility $u_j(a) = 10u_i(a)$ will behave in exactly the same way:

$$\arg\max_a \mathbb{E}[10u_i(a)] = \arg\max_a \mathbb{E}[u_i(a)]$$

- So how should we reason about mechanism design?

Pareto Optimality

1. A profile of utilities $(x_1, \ldots, x_n)$ **Pareto dominates** another profile $(y_1, \ldots, y_n)$ if:
    - $x_i \geq y_i$ for every $i \in N$, and
    - There exists an $i \in N$ such that $x_i > y_i$.
    - Every agent is at least as happy with x as with y, and at least one agent strictly prefers x to y.
2. An outcome $(x_1, \ldots, x_n)$ is **Pareto optimal** if it is not Pareto dominated by any other outcome.
    - Any outcome that is better for one agent must be worse for another

Individually Rational Behaviour is not Always Globally Optimal

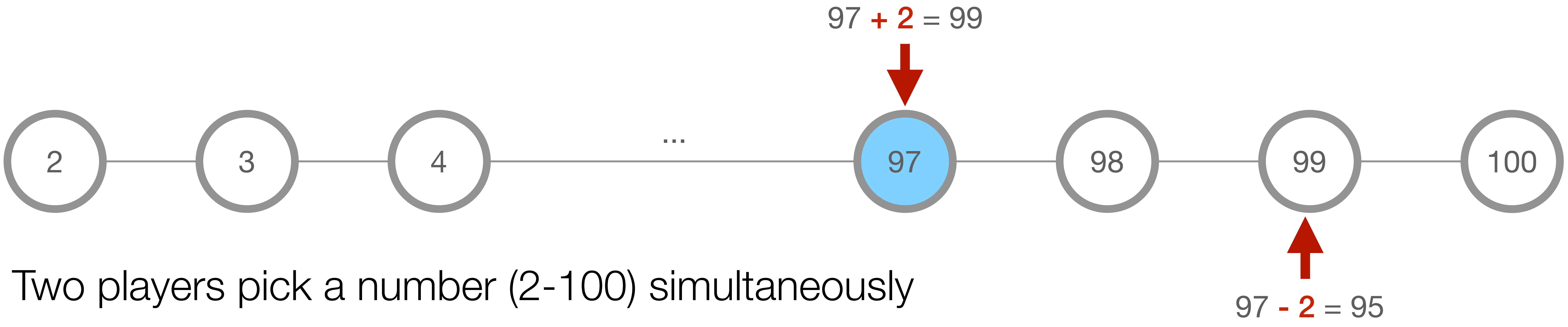
The only equilibrium of Prisoner's Dilemma is also the *only* outcome that is Pareto-dominated!

	Cooperate	Defect
Cooperate	-1,-1	-5,0
Defect	0,-5	-3,-3

1. Which outcome is Pareto optimal?
2. What is the Nash equilibrium?

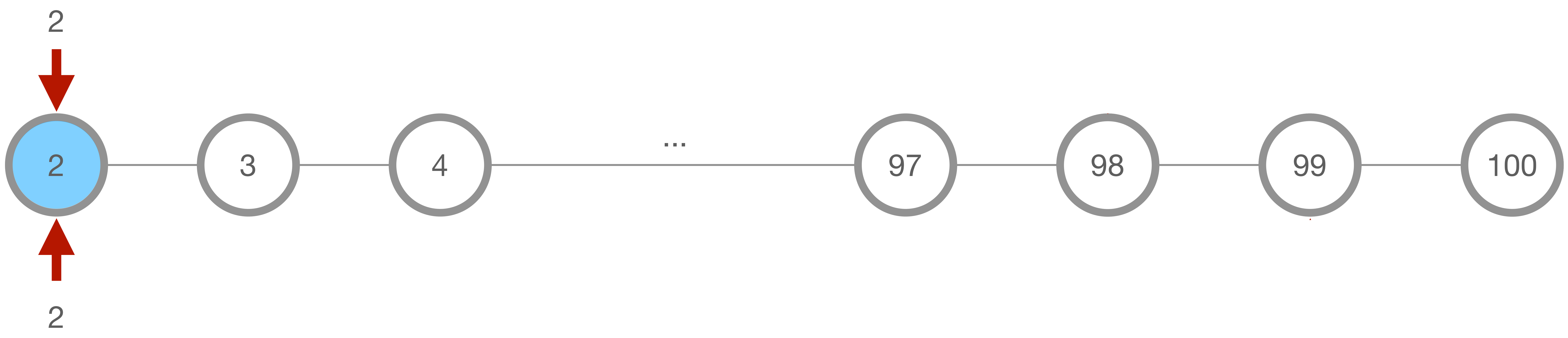
Individual optimization can lead to the **worst-possible** global outcome.

Fun Game: Traveller's Dilemma



- Two players pick a number (2-100) simultaneously
- If they pick the same number *x*, then they both get $*x* payoff
- If they pick different numbers:
    - Player who picked **lower** number gets **lower** number, plus **bonus** of $2
    - Player who picked higher number gets **lower** number, minus **penalty** of $2
- Play against the person next to you! Each write down a number, reveal simultaneously.

Traveller's Dilemma



Traveller's Dilemma has a unique Nash equilibrium

Individual optimization can lead to the nearly worst-possible **individual** outcome.

Nash Equilibrium vs. Humans

- Nash equilibrium often makes **counterintuitive predictions.**
  - In Traveller's Dilemma: The vast majority of human players choose **97–100**. The Nash equilibrium is **2**.
- Modifications to a game that don't change Nash equilibrium predictions at all **can cause large changes** in how human subjects play the game [Goeree & Holt 2001].
  - In Traveller's Dilemma: When the penalty is large, people play much closer to Nash equilibrium.
  - But the size of the penalty does not change the equilibrium!
- Clearly Nash equilibrium **does not capture the whole story** (for humans)

Multiple Nash Equilibria

Multiple equilibria are a problem even in **perfectly cooperative** games

	Left	Right
Left	1, 1	-100, -100
Right	-100, -100	1, 1

Nash equilibria:

1. (Left, Left): Neither agent gains by *deviating* (choosing Right)
2. (Right, Right): Neither agent gains by *deviating* (choosing Left)
3. $\left(\left[\frac{1}{2} : Left, \frac{1}{2} : Right\right], \left[\frac{1}{3} : Left, \frac{2}{3} : Right\right]\right)$

Best Response is Expensive

$$a_i^* \in BR_i(\pi_{-i})$$

- All of the previous problems arise in very simple games
- But in many settings you can't even reliably compute BR_i
    - Poker, Go, Chess

What's Going Wrong Here?

1. We can't do what game theory prescribes, and neither can the other players

    - Finding a Nash equilibrium is computationally hard [Chen et al., 2009]

2. Nash equilibrium is about what satisfies certain **conditions**, not how you get there

    - Hence multiple, unordered equilibria

3. Unrestricted individual optimization can cause even small misalignments in interests to snowball

Thinking About the Problem

Three ways of viewing a multiagent problem:

1. Normative
2. Descriptive
3. Positive

Normative Modelling

- Start from axioms (assumptions) and derive consequences:
  - Every agent **maximizes expected utility**
  - Every agent has **accurate beliefs**
- Game theory problems come from:
  - Violation of these assumptions
  - Focus on consequences rather than process

Descriptive Modelling

Construct a model of how agents **actually behave**

- Need a descriptive model of other players to evaluate algorithm performance:
    - $\max_{\pi_i} \mathbb{E}\left[R_i^{(1)} + \gamma R_i^{(2)} + \ldots + \gamma^{T-1} R_i^{(T)} \mid p, \boldsymbol{\pi_1, \ldots, \pi_{i-1}, \pi_{i+1}, \ldots, \pi_n}\right]$
    - $\max_{p} \mathbb{E}\left[\sum_{1 \le i \le n} R_i^{(1)} + \gamma R_i^{(2)} + \ldots + \gamma^{T-1} R_i^{(T)} \,\middle|\, \boldsymbol{\pi_1, \ldots, \pi_n}\right]$
- Understanding human behavior can give insight into **bounded rationality**

Positive Modelling

Given a descriptive model of behaviour and environment, figure out how to achieve a goal

1. **Agent Design:** Design an agent that will achieve its **individual goals**:
    - Maximum-value bundle of goods, winning a game
2. **Mechanism Design:** Design an environment for agents such that the agents' **collective actions** will achieve a goal:
    - Markets, online platforms, auctions, Internet protocols

Thinking About the Problem

Three ways of viewing a multiagent problem:

1. Normative
2. **Descriptive**
3. Positive

Predicting Human Behaviour In Normal-Form Games

- Predicting human behaviour is a **prediction problem**

  Features: Description of the setting (i.e., the game)
  Labels: Actions taken by agents

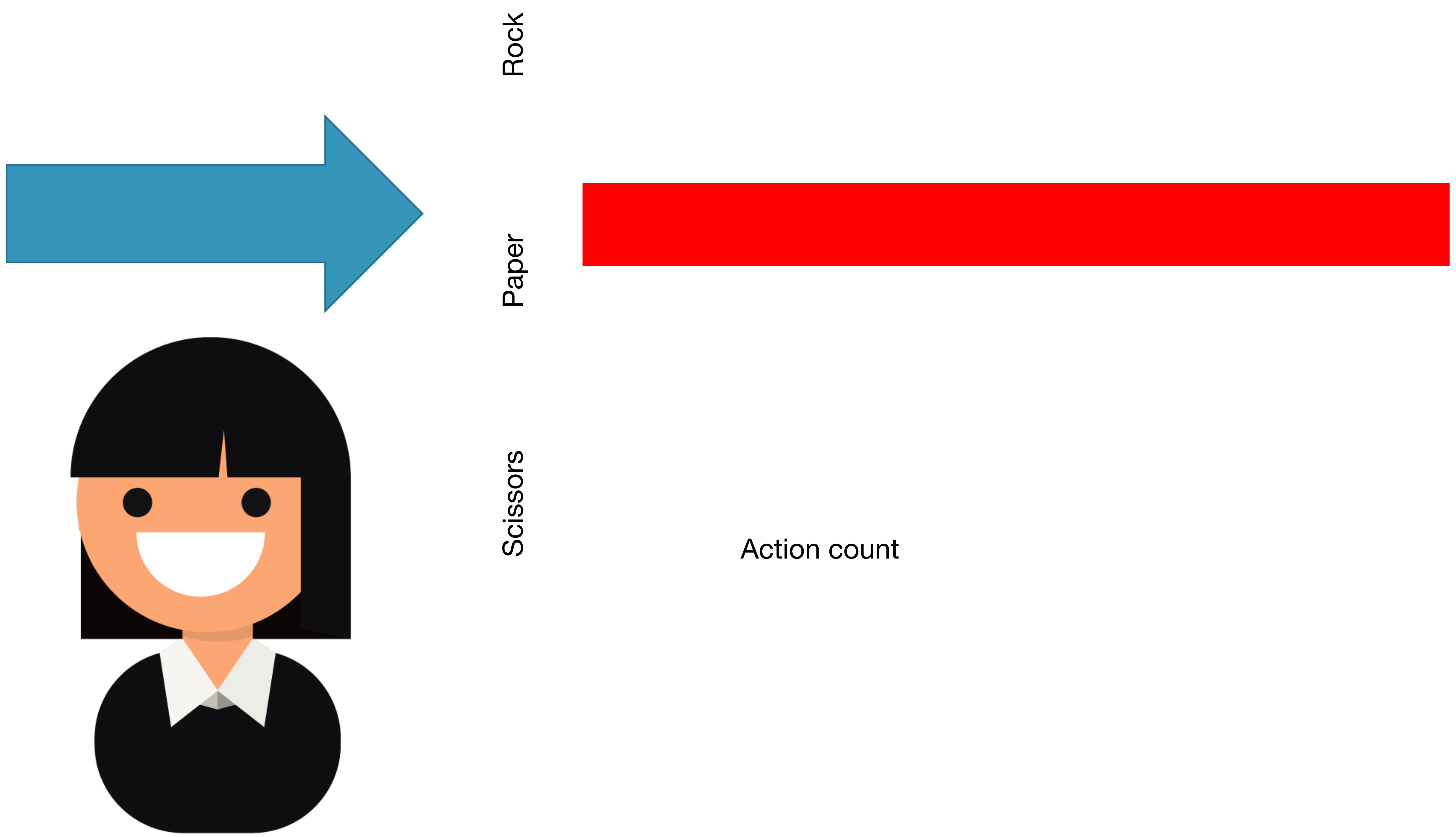
- Machine learning is **really good** at prediction problems

- Why not take a direct machine learning approach?

Normal-Form Games

Column player's actions

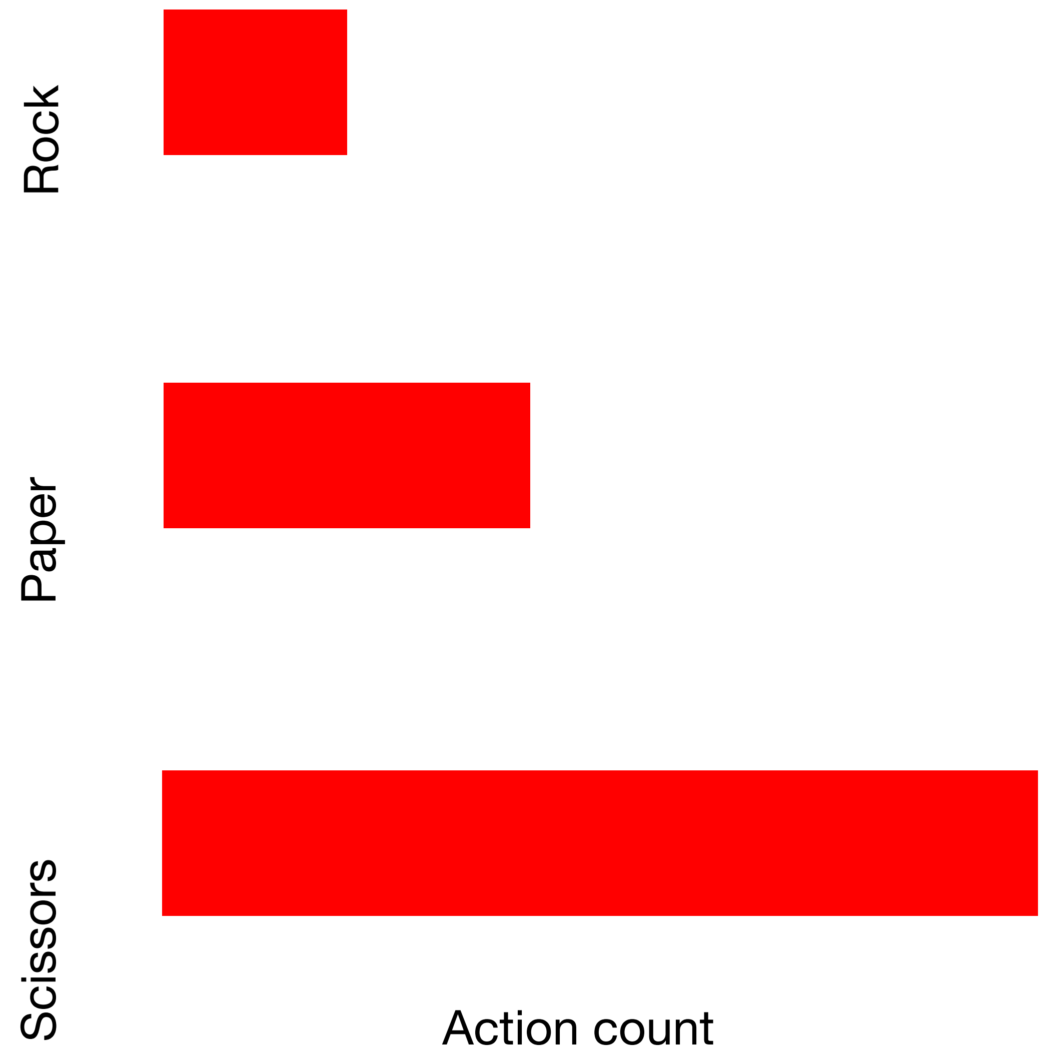
Row player's actions	Rock	Paper	Scissors
Rock	0, 0	-1, 1	2, -2
Paper	1, -1	0, 0	-1, 1
Scissors	-2, 2	1, -1	0, 0



Normal-Form Games

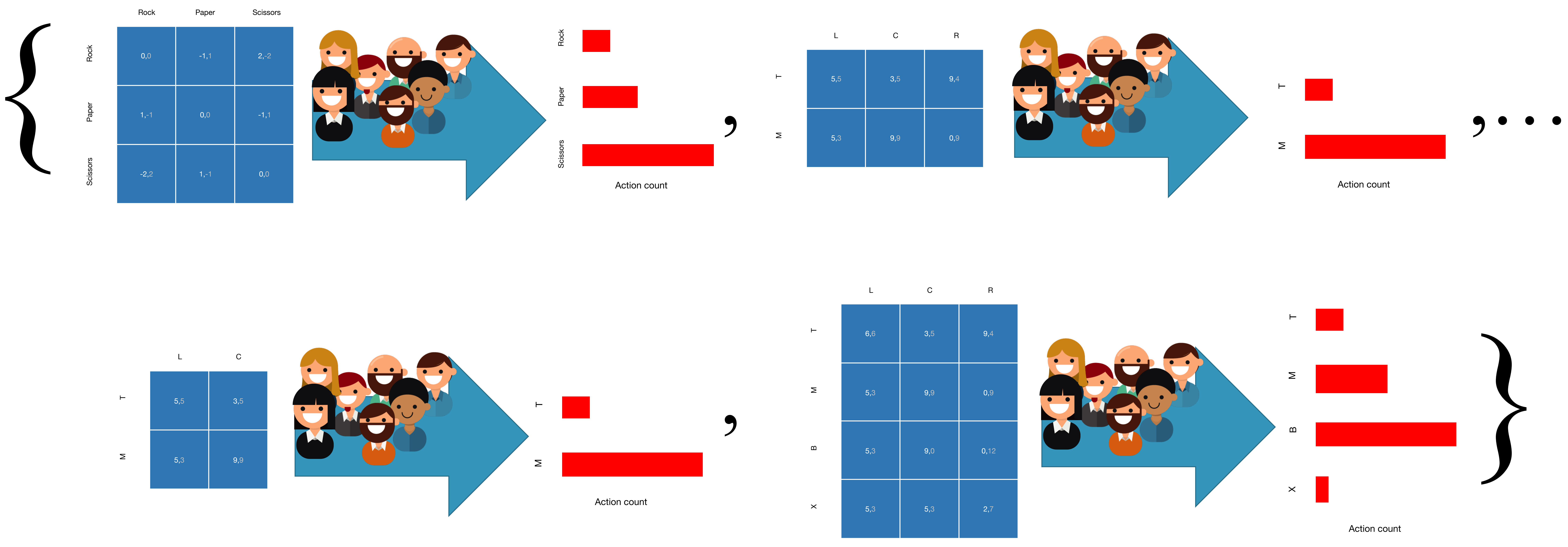
Column player's actions

Row player's actions	Rock	Paper	Scissors
Rock	0, 0	-1, 1	2, -2
Paper	1, -1	0, 0	-1, 1
Scissors	-2, 2	1, -1	0, 0



Learning Problem

Given a dataset of games, each with observed action counts:

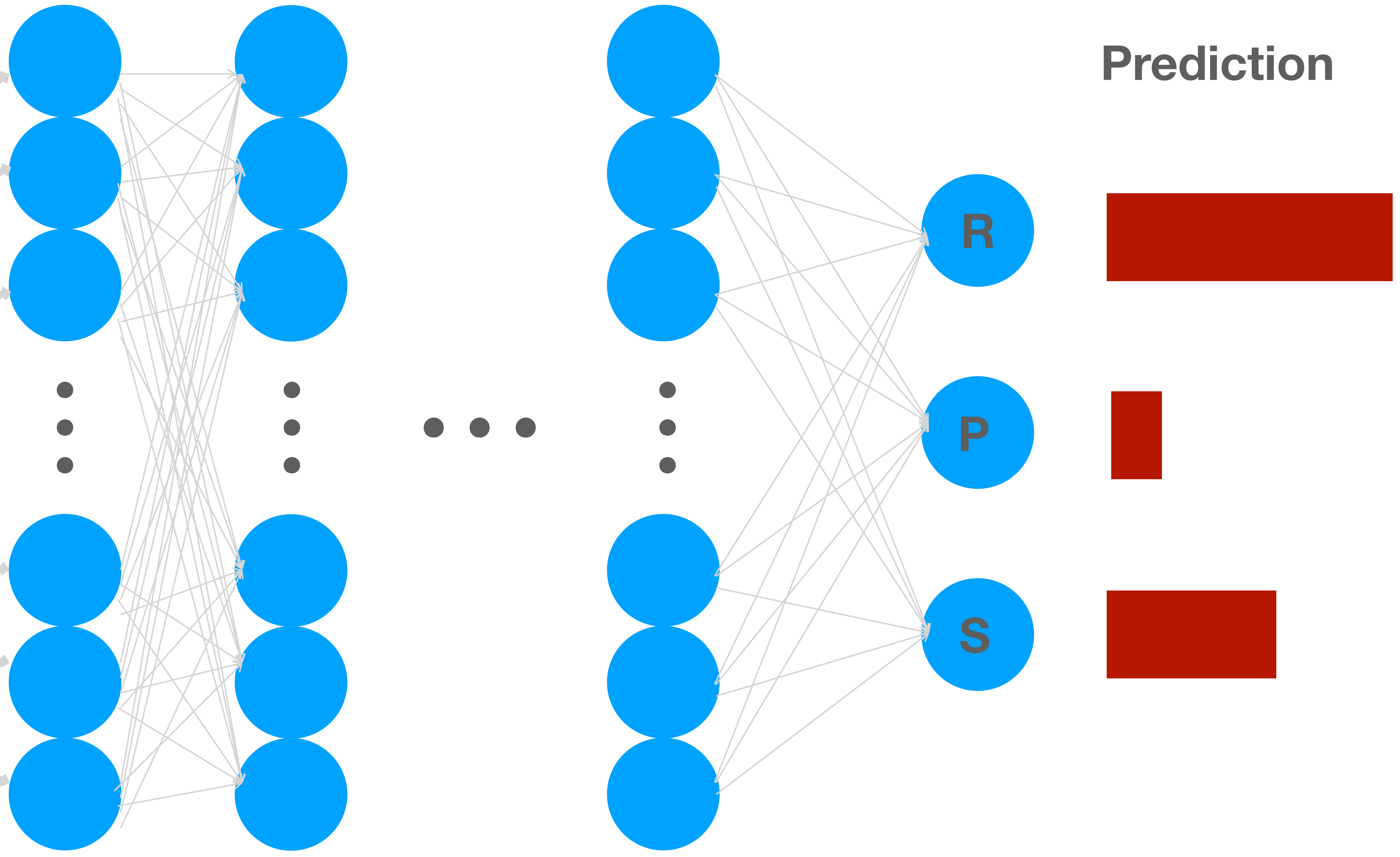


... learn a model that predicts players' distribution over actions

Deep Learning for Human Strategic Behavior

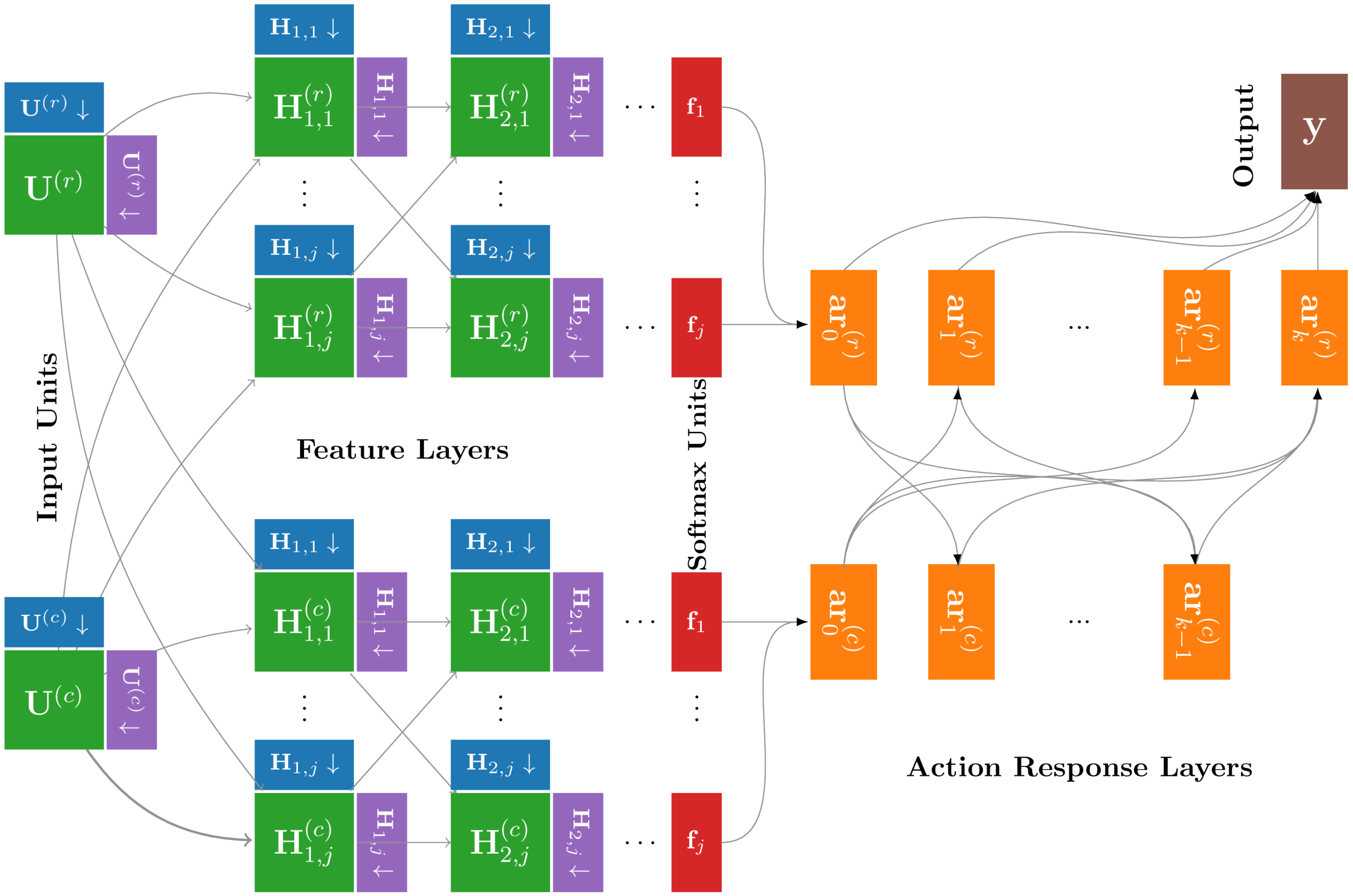
[Hartford, Wright, Leyton-Brown: NeurIPS 2016]

	Rock	Paper	Scissors
Rock	0, 0	-1, 1	1, -1
Paper	1, -1	0, 0	-1, 1
Scissors	-1, 1	1, -1	0, 0
Nuclear	-9, -9	-9, -9	-9, -9



Deep Learning for Human Strategic Behaviour

[Hartford, Wright, Leyton-Brown: NeurIPS 2016]



Interpreting the Deep Model

[Smith, Hartford, Leyton-Brown, Wright: *in progress*]

Activation maximization: Synthesize games (via optimization) that maximally activate a **particular unit**

- These games will exemplify the structure that unit detects
- Can add the discovered features to our cognitive models

Relevance propagation: [Bach et al, 2015]

- *Images:* Which pixels were most relevant?
- *Games:* Which **outcomes** were most influential?

Final Thoughts

- Our lives are increasingly part of **novel** multiagent systems
  - It is imperative that we understand their properties so we can design them properly
- This is a **wide-open research area**
  - We don't even have a coherent notion of optimal behaviour yet
- Game theory helps us reason precisely about multiagent interactions,
  but it isn't enough
- Need real-world descriptive models and a theory of how to use them

Thanks!