Ideas sparked by
Game Theory* & *Deep Learning

Expert Student Talk on CS228

Game Theoretical Methodology
and Technique for Internet Protocols

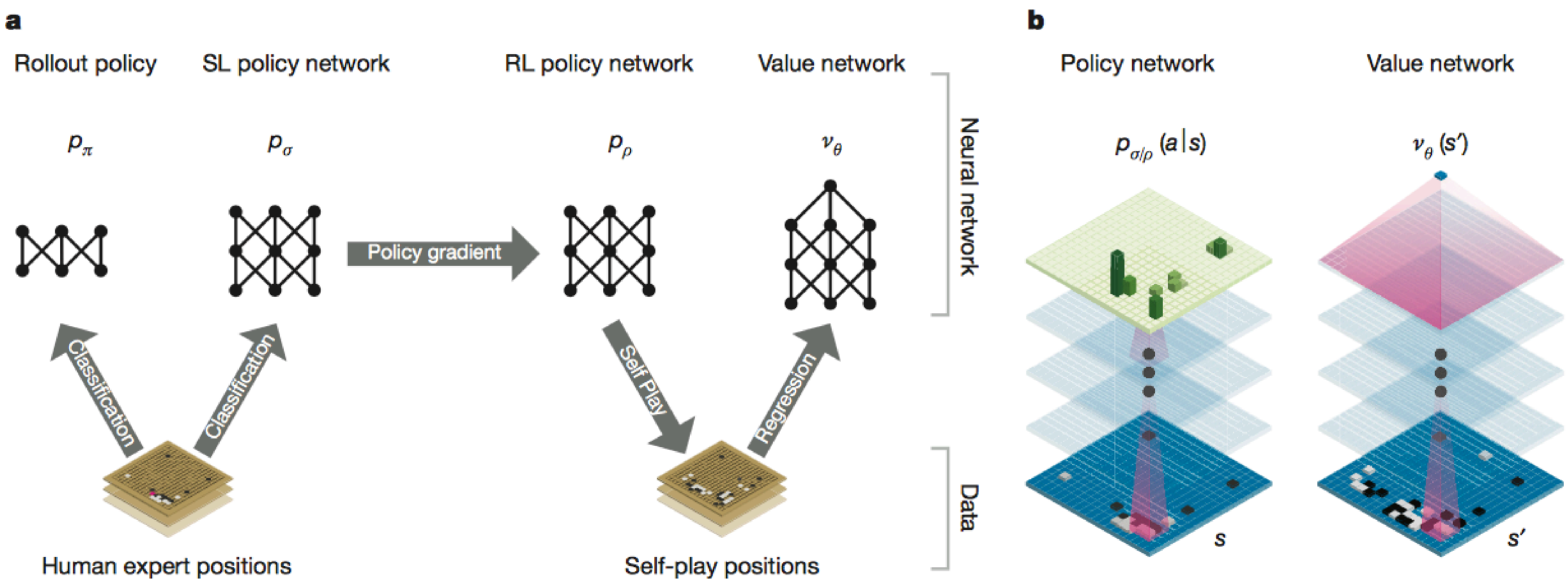
Runzhe Yang @ SJTU ACM CLASS

What’s happening in AI community?

Intro: Deep learning in Game

Alpha Go v.s. Lee Sedol, from youtube.com

Intro: Deep learning in Game

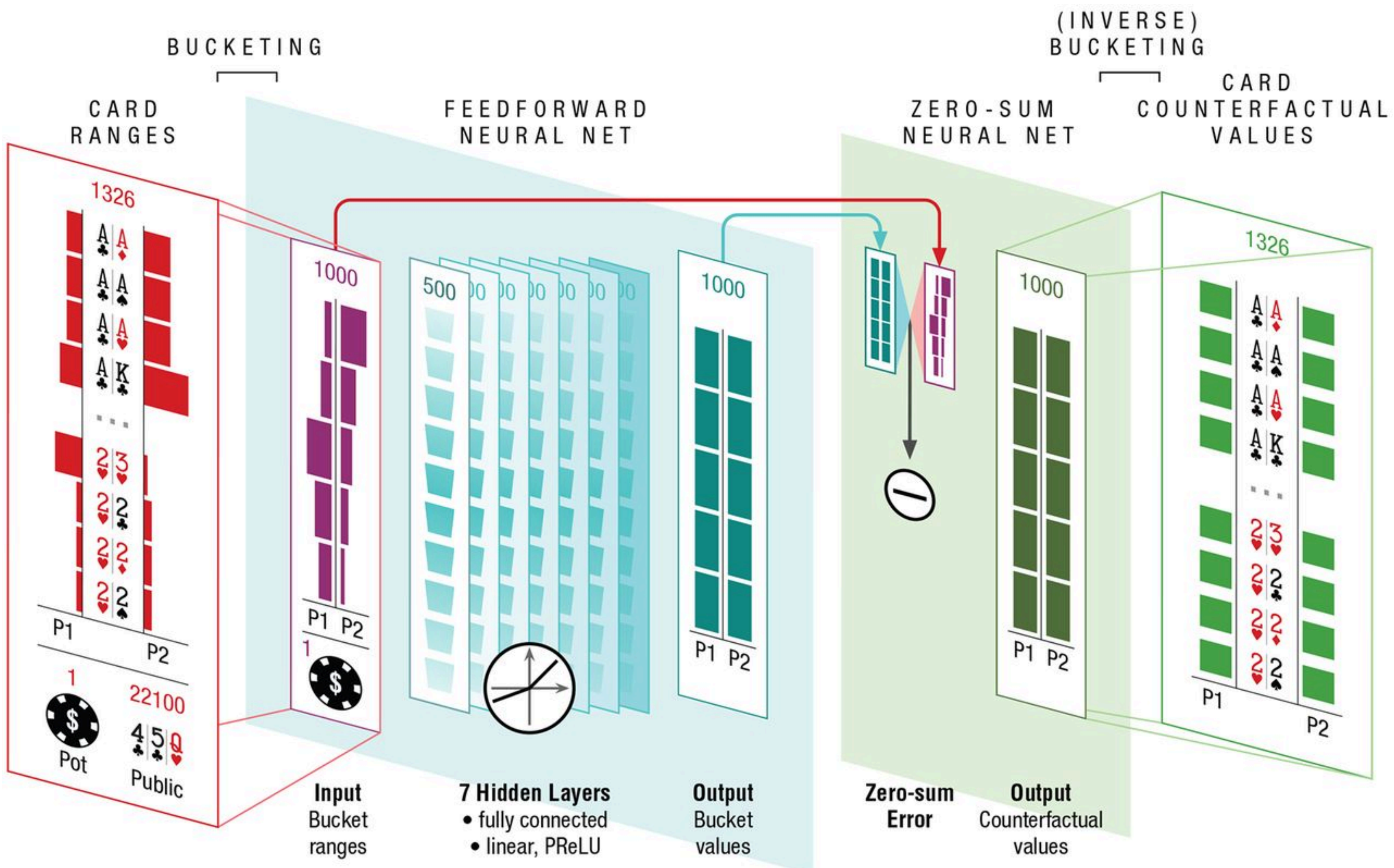


Mastering the game of Go with deep neural networks and tree search, from *Nature*

Intro: Deep learning in Game

The artificial intelligence **Libratus** always knows when to hold 'em and when to fold 'em, from slate.com

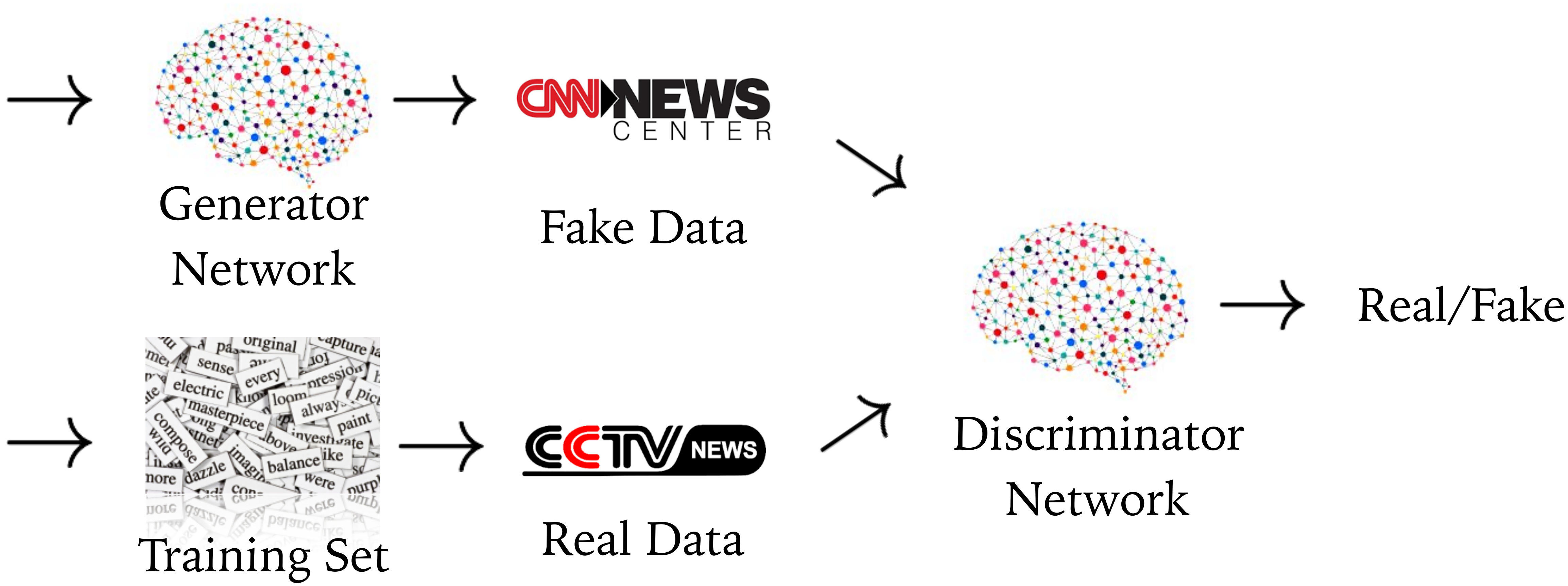
Intro: Deep learning in Game



DeepStack: Expert-level artificial intelligence in heads-up no-limit poker, from *Science*

Intro: Game theory in Learning

Generative Adversarial Nets

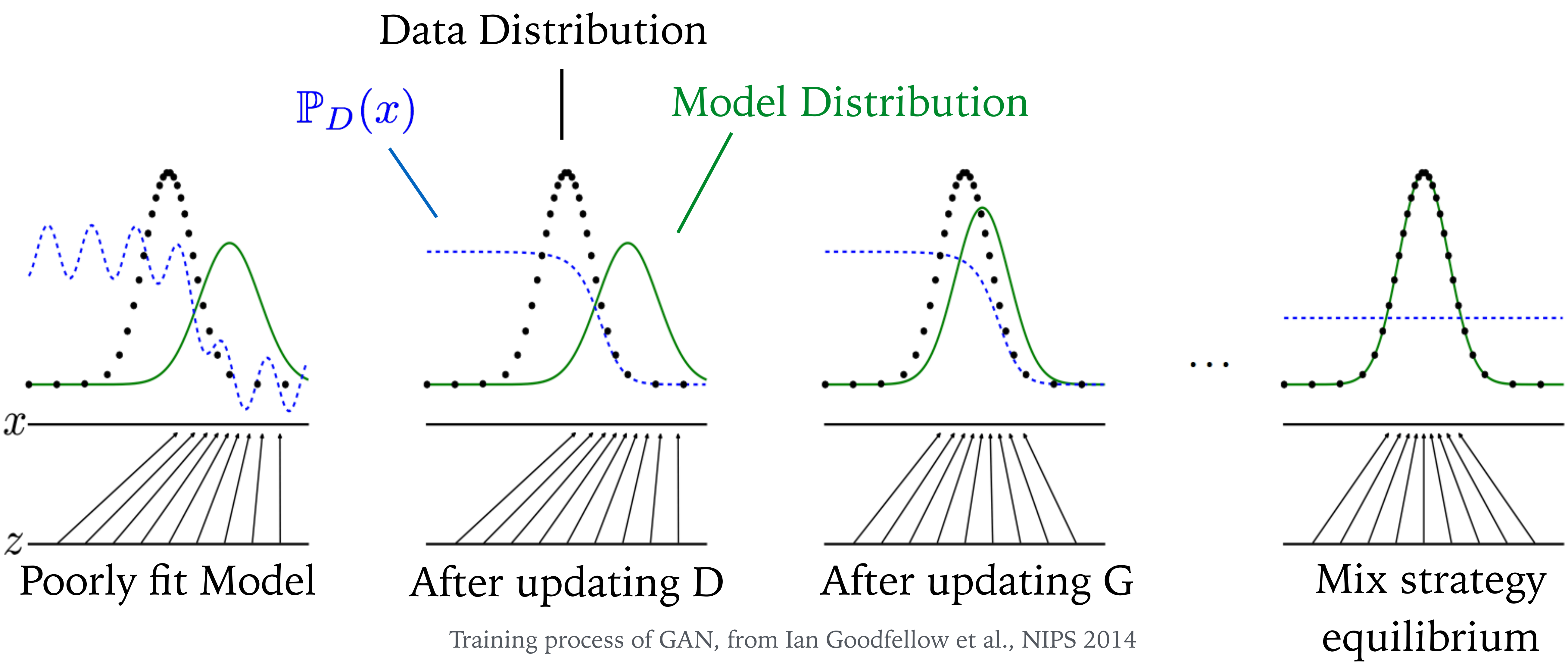


$$\mathcal{L}_{\text{discriminator}}(D) = \mathbb{E}_{\boldsymbol{x}\sim p}[-\log D(x)] + \mathbb{E}_{\boldsymbol{x}\sim G}[-\log(1 - D(x))]$$

$$\mathcal{L}_{\text{generator}}(G) = \mathbb{E}_{\boldsymbol{x}\sim G}[-logD(x)] + \mathbb{E}_{\boldsymbol{x}\sim G}[log(1 - D(x))]$$

Intro: Game theory in Learning

Generative Adversarial Nets

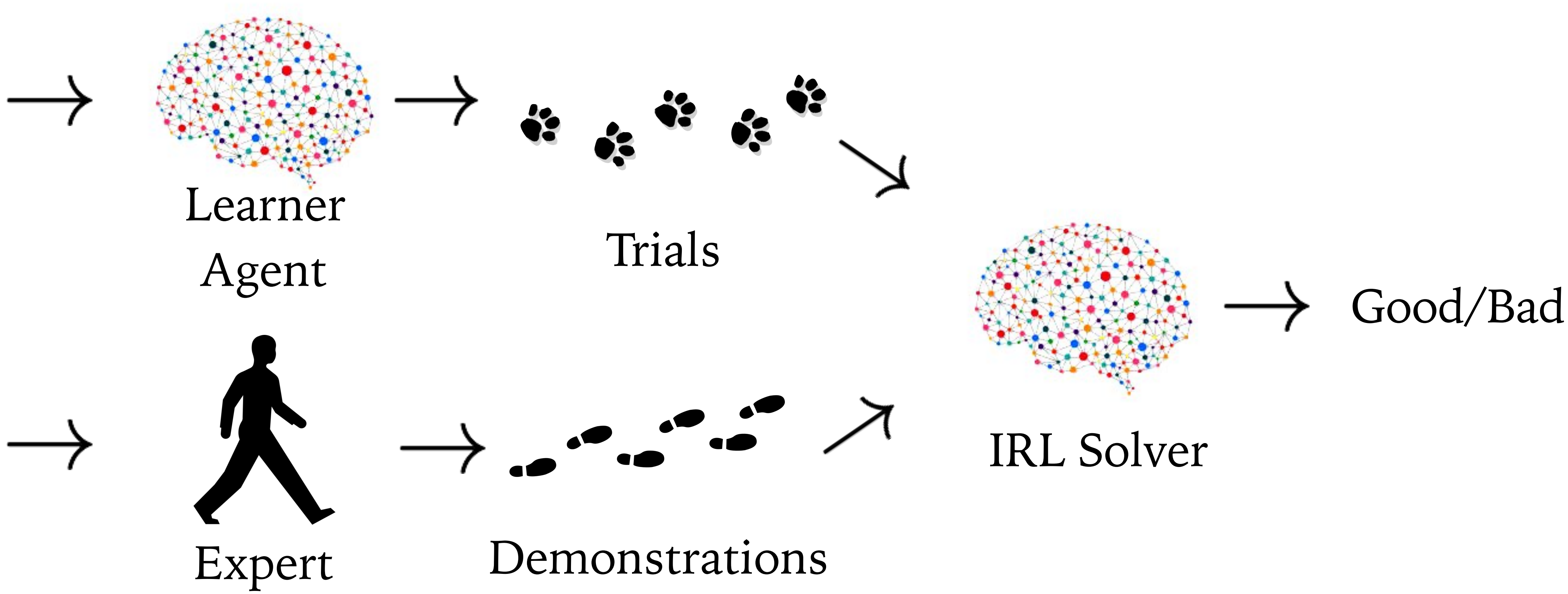


Training process of GAN, from Ian Goodfellow et al., NIPS 2014

$$\min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} \mathop{\mathbb{E}}_{x \sim \mathcal{D}_{real}} [\log D_v(x)] + \mathop{\mathbb{E}}_{h \sim \mathcal{D}_h} [\log(1 - D_v(G_u(h)))]$$

Intro: Game theory in Learning

Max Entropy Inverse Reinforcement Learning

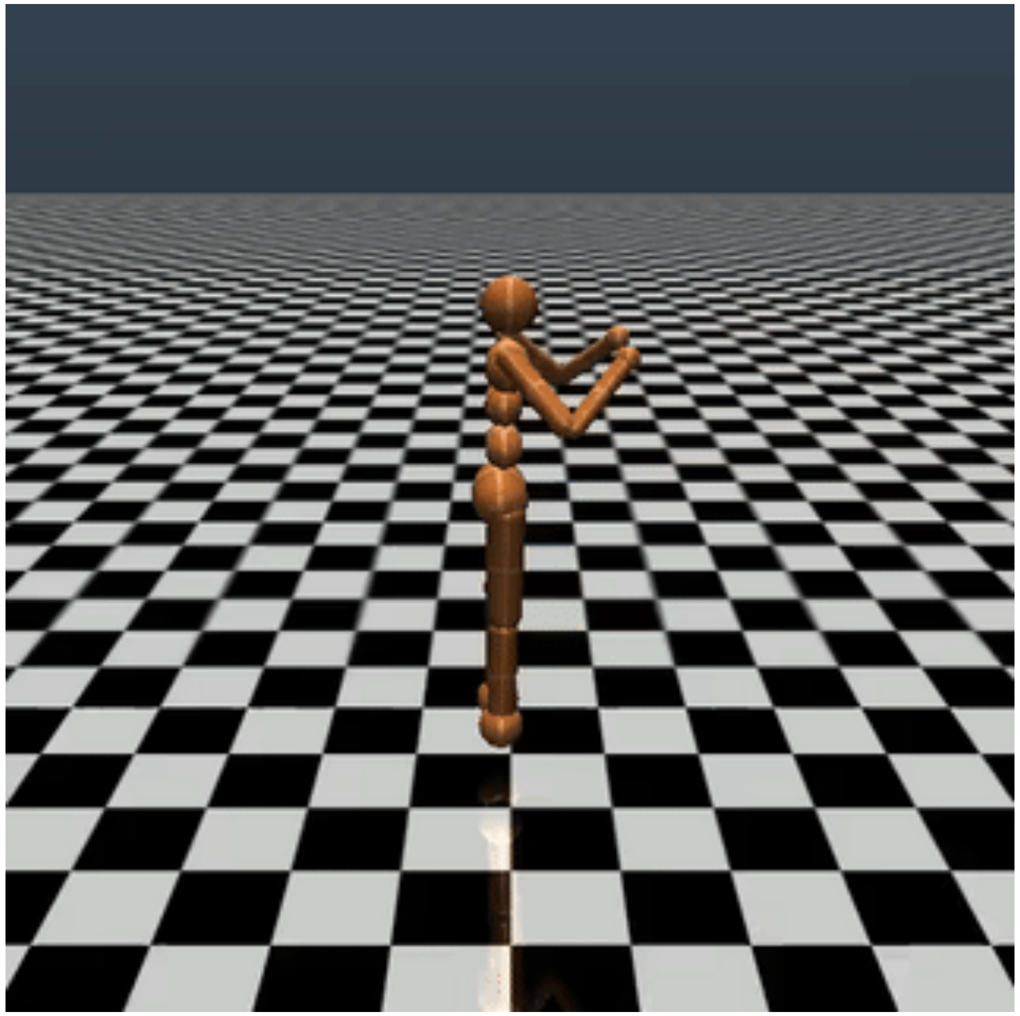


$$\mathcal{L}_{\text{cost}}(\theta) = \mathbb{E}_{\tau\sim\pi^e}[c_\theta(\tau)] + \mathbb{E}_{\tau\sim\pi^\ell}\left[\log\left(\frac{\exp(-c_\theta(\tau))}{\pi^\ell(\tau)}\right)\right]$$

$$\mathcal{L}_{\text{learner}}(\pi^\ell) = \mathbb{E}_{\tau\sim\pi^\ell}[c_\theta(\tau)] + \mathbb{E}_{\tau\sim\pi^\ell}[\log \pi^\ell(\tau)]$$

Intro: Game theory in Learning

Generative Adversarial Imitation Learning




Generative Adversarial Imitation Learning, from Ermon Group, NIPS 2016

Game Theory is elegant but hard to solve.

Game Theory is elegant but hard to solve.

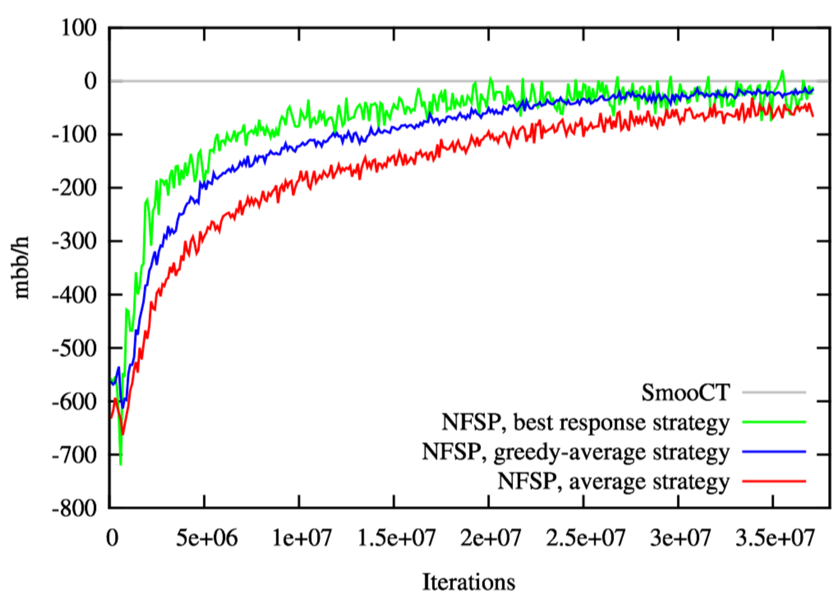
- Plan in Markov Decision Process or POMDP

Game Theory is elegant but hard to solve.

- Plan in Markov Decision Process or POMDP
- Solve Nash Equilibrium with Imperfect Information
  - Counterfactual regret minimization (CFR)
  - Neural Fictitious Self-Play (NFSP)

Game Theory is elegant but hard to solve.

- Plan in Markov Decision Process or POMDP
- Solve Nash Equilibrium with Imperfect Information
  - Counterfactual regret minimization (CFR)
  - Neural Fictitious Self-Play (NFSP)



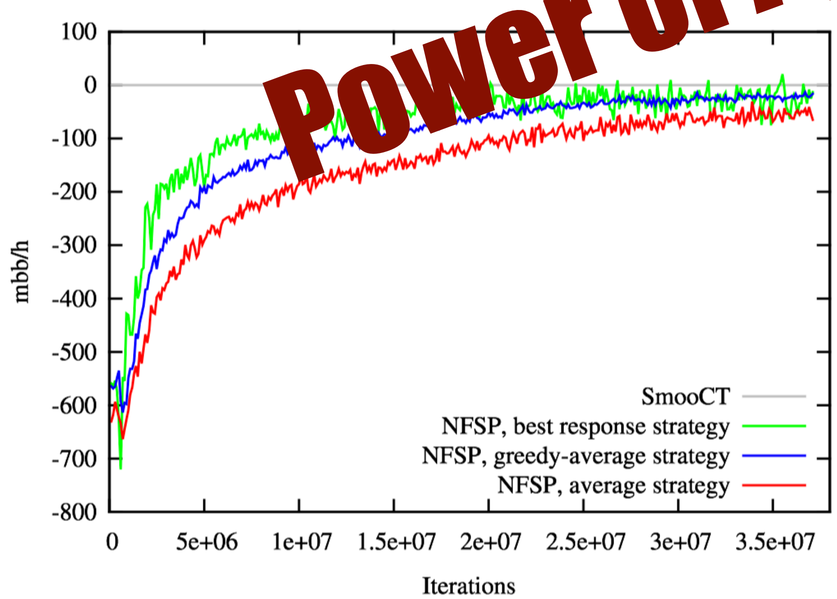
Match-up	Win rate (mbb/h)
escabeche	-52.1 ± 8.5
SmooCT	-17.4 ± 9.0
Hyperborean	-13.6 ± 9.2

Performance of NFSP in Limit Texas Hold'em. David Silver et al.

Game Theory is elegant but hard to solve.

- Plan in Markov Decision Process or POMDP
- Solve Nash Equilibrium with Imperfect Information
  - Counterfactual regret minimization (CFR)
  - Neural Fictitious Self-Play (NFSP)

Power of Approximation



Match-up	Win rate (mbb/h)
escabeche	-52.1 ± 8.5
SmooCT	-17.4 ± 9.0
Hyperborean	-13.6 ± 9.2

Performance of NFSP in Limit Texas Hold'em. David Silver et al.

Deep Learning is pragmatic but lacks theoretical guarantee.

Deep Learning is pragmatic but lacks theoretical guarantee.

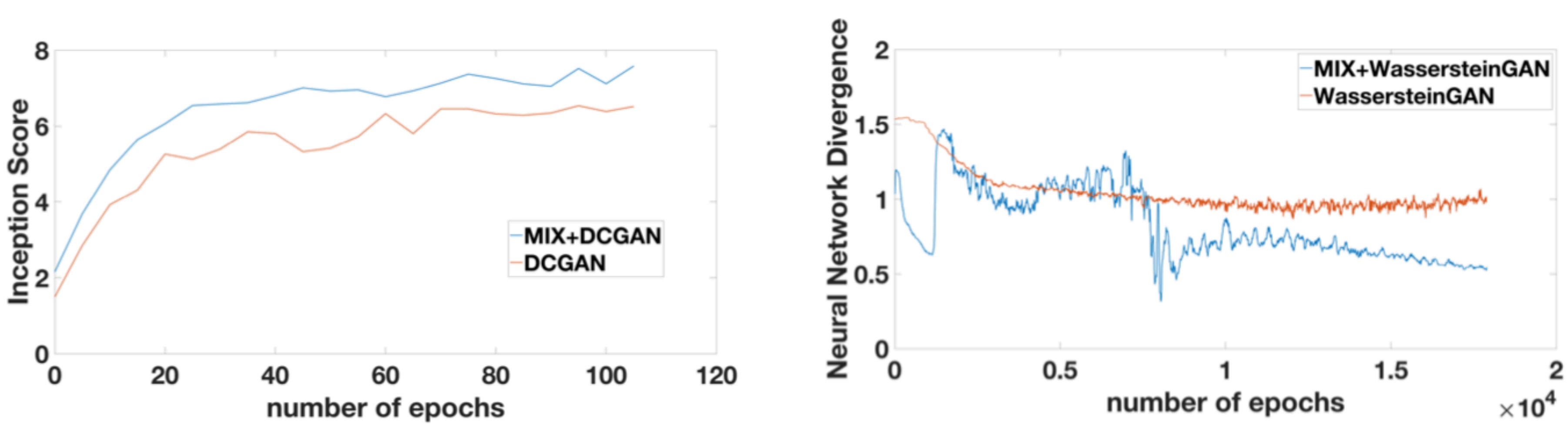
- Use game theoretical methods to explain and design DL model

Deep Learning is pragmatic but lacks theoretical guarantee.

- Use game theoretical methods to explain and design DL model
- GAN & Imitation Learning

Deep Learning is pragmatic but lacks theoretical guarantee.

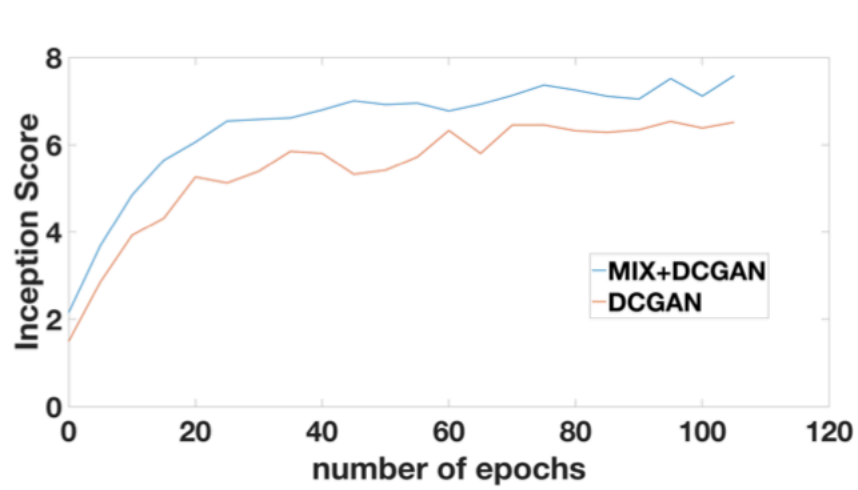
- Use game theoretical methods to explain and design DL model
- GAN & Imitation Learning
  - Sanjeev Arora et la. **Generalization and Equilibrium in Generative Adversarial Nets**. arXiv.org. (2017, March 2)

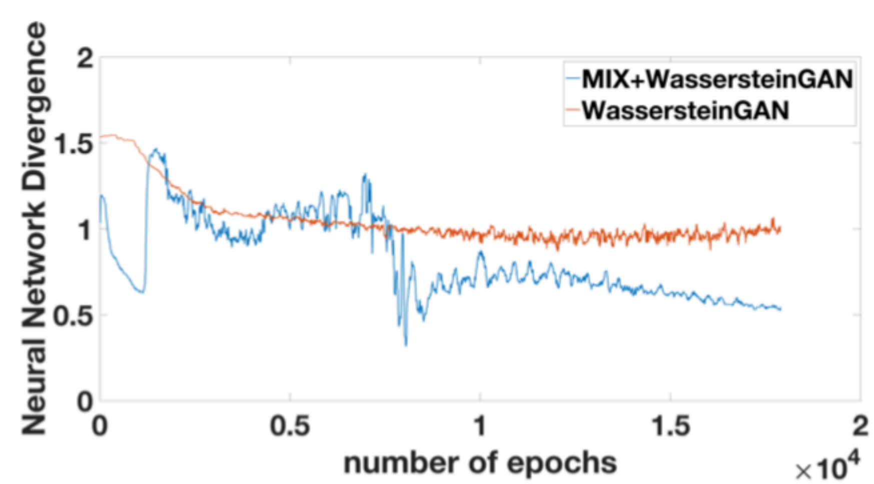


Deep Learning is pragmatic but lacks theoretical guarantee.

- Use game theoretical methods to explain and design DL model
- GAN & Imitation Learning
  - Sanjeev Arora et la. **Generalization and Equilibrium in Generative Adversarial Nets**. arXiv.org. (2017, March 2)

Power of Analysis





Artificial Intelligence

?

Deep Learning $\longleftrightarrow$ Game Theory

“Humans nowadays completely dominate the planet not because the individual human is **far smarter** and **more nimble-fingered** than the individual chimp or wolf, but because Homo sapiens is the only species on earth **capable of co-operating flexibly in large numbers**.”

Excerpt From: Yuval Noah Harari.
Homo Deus: A Brief History of Tomorrow

Understanding Agent Cooperation

DeepMind

Runzhe Yang @ SJTU ACM CLASS

Matrix Game Social Dilemmas (MGSD)

Matrix Game Social Dilemmas (MGSD)

	C	D
C	R, R	S, T
D	T, S	P, P

R **reward** of mutual cooperation

P **punishment** arising from mutual defection

S **sucker outcome** obtained by the player who cooperates with a defecting partner

T **temptation outcome** achieved by defecting against a cooperator

Matrix Game Social Dilemmas (MGSD)

	C	D
C	R, R	S, T
D	T, S	P, P

social dilemma inequalities

(1) R > P Mutual cooperation is preferred to mutual defection.

(2) R > S Mutual cooperation is preferred to being exploited by a defector.

(3) 2R > T + S This ensures that mutual cooperation is preferred to an equal probability of unilateral cooperation and defection.

- either ***greed***: T > R Exploiting a cooperator is preferred over mutual cooperation
- or ***fear***: P > S Mutual defection is preferred over being exploited.

Matrix Game Social Dilemmas (MGSD)

three canonical examples:

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

Stag Hunt	C	D
C	4, 4	0, 3
D	3, 0	1, 1

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

Matrix Game Social Dilemmas (MGSD)

Temporal Extension: **Sequential Social Dilemmas**

$$r_i : \mathcal{S} \times \mathcal{A}_1 \times \mathcal{A}_2 \to \mathbb{R} \qquad \mathcal{T} : \mathcal{S} \times \mathcal{A}_1 \times \mathcal{A}_2 \to \Delta(\mathcal{S})$$

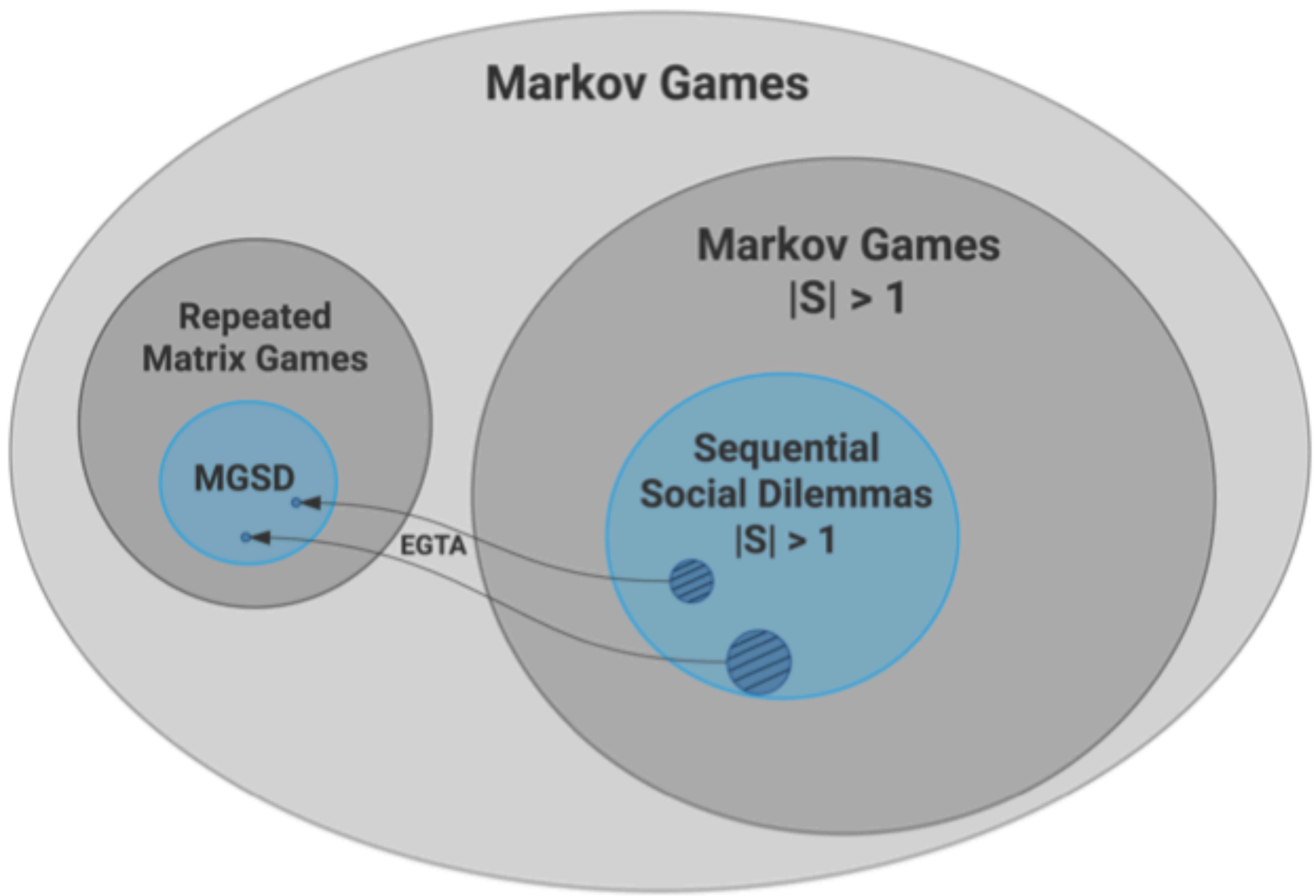
$$(\pi^C \in \Pi^C, \pi^D \in \Pi^D) \Rightarrow (R(s), P(s), S(s), T(s))$$

long-term pay-off:

$$V_i^{\vec{\pi}}(s_0) = \mathbb{E}_{\vec{a}_t \sim \vec{\pi}(O(s_t)), s_{t+1} \sim \mathcal{T}(s_t, \vec{a}_t)} \left[\sum_{t=0}^{\infty} \gamma^t r_i(s_t, \vec{a}_t) \right]$$

Matrix Game Social Dilemmas (MGSD)

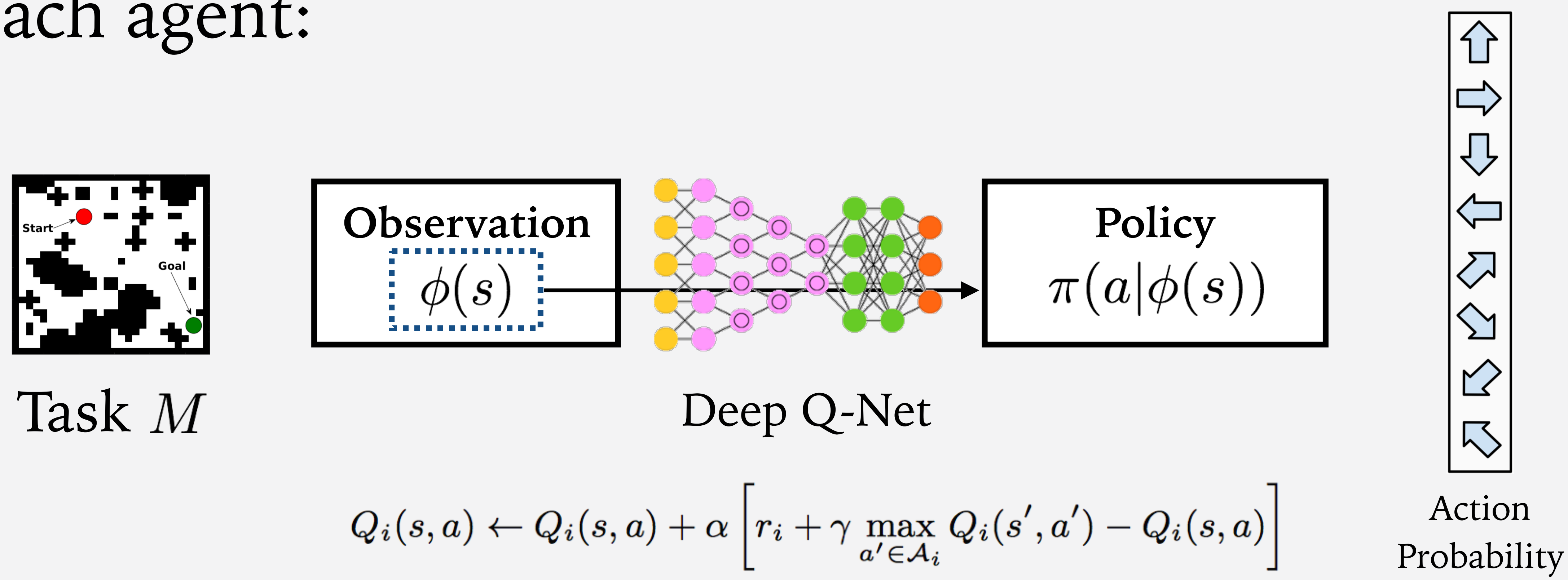
Temporal Extension: **Sequential Social Dilemmas**



Sequential Social Dilemmas (SSD)

Deep Multi-agent Reinforcement Learning

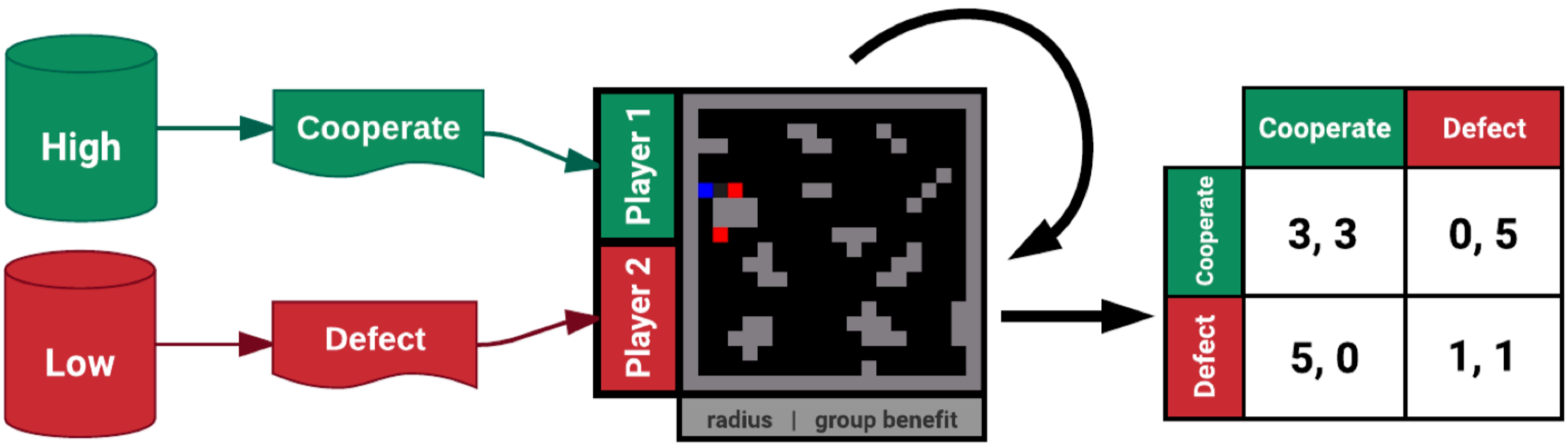
Each agent:



$$Q_i(s,a) \leftarrow Q_i(s,a) + \alpha \left[r_i + \gamma \max_{a' \in \mathcal{A}_i} Q_i(s',a') - Q_i(s,a) \right]$$

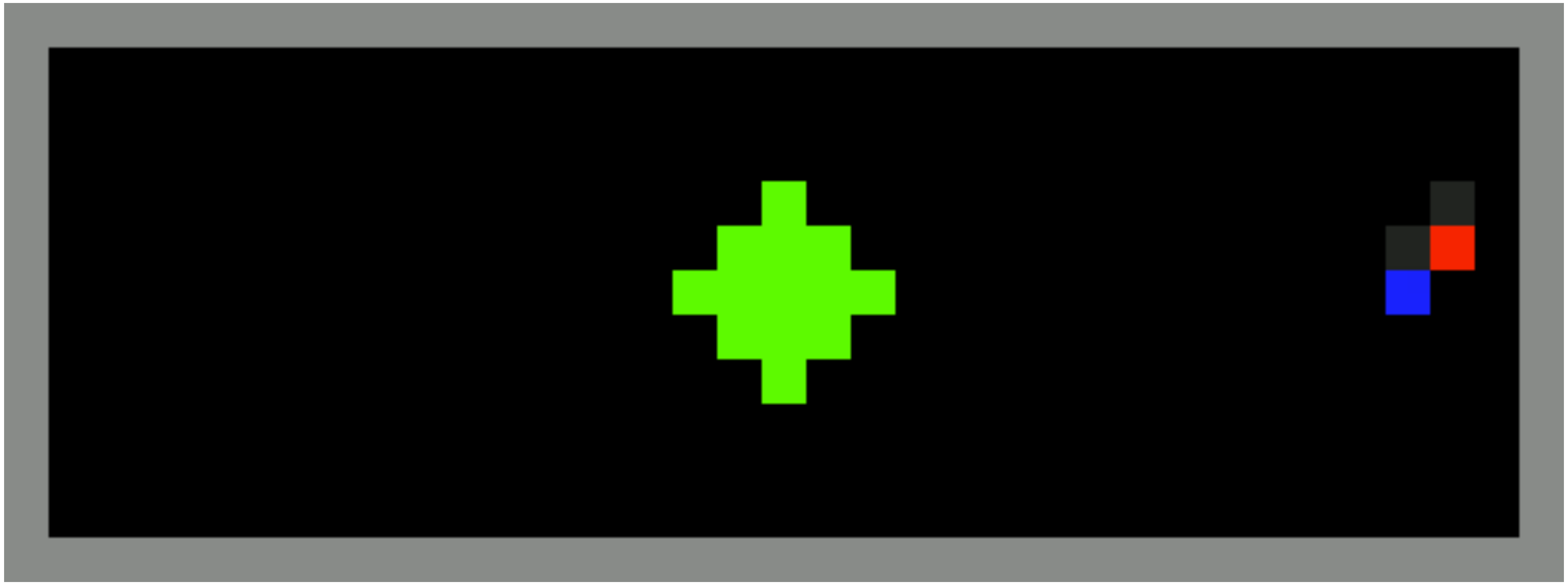
Sequential Social Dilemmas (SSD)

Empirical payoff matrices



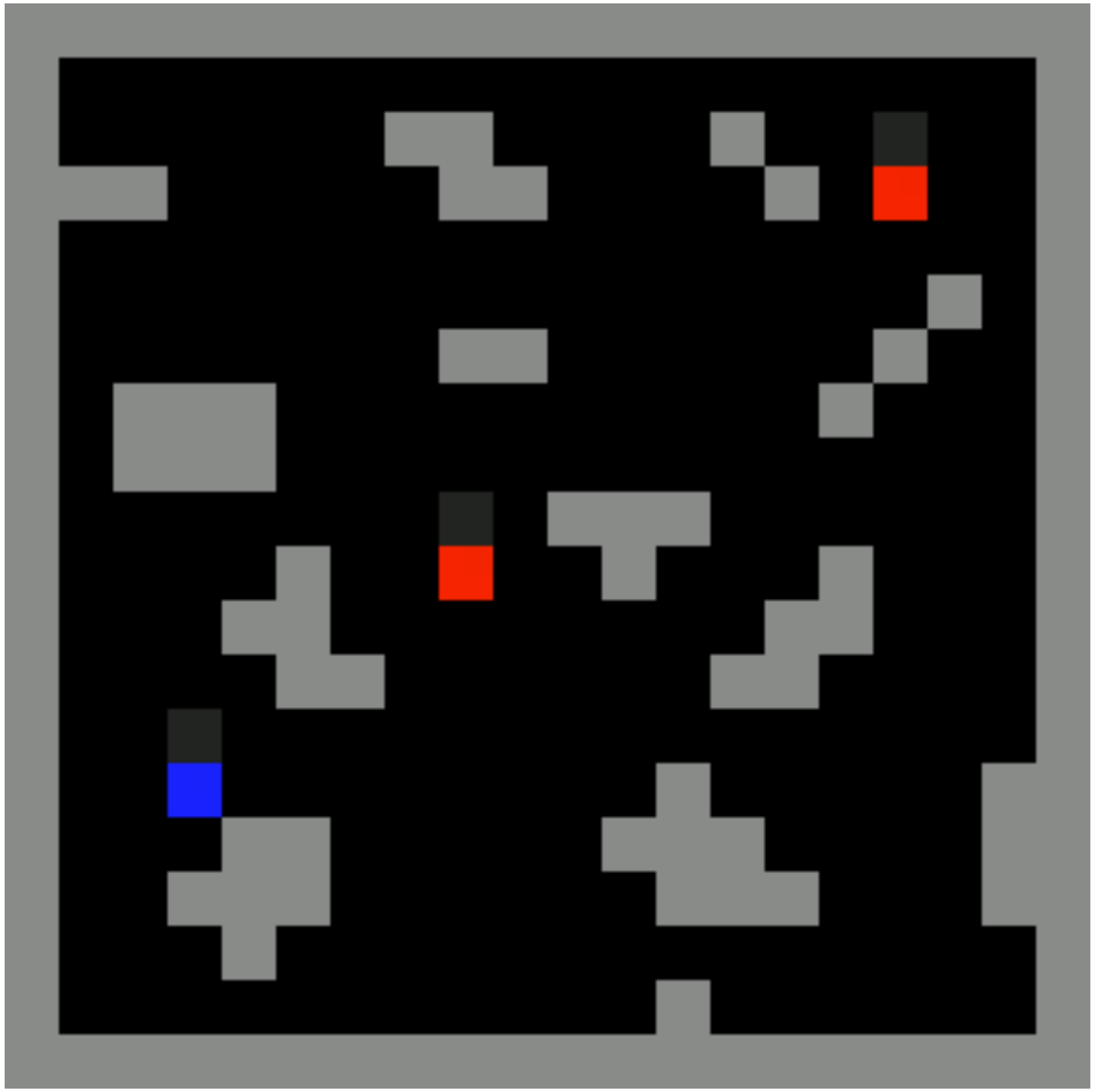
Sequential Social Dilemmas (SSD)

Gathering

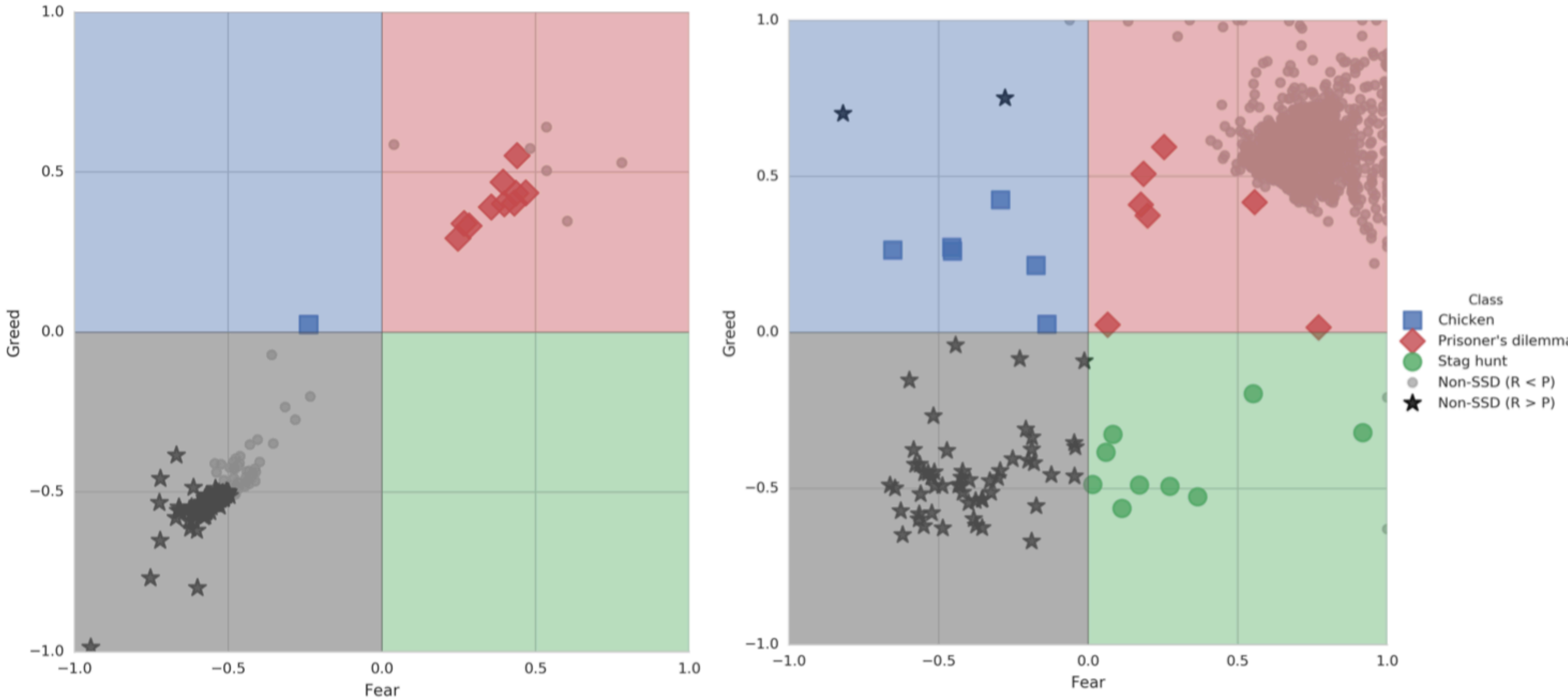


Sequential Social Dilemmas (SSD)

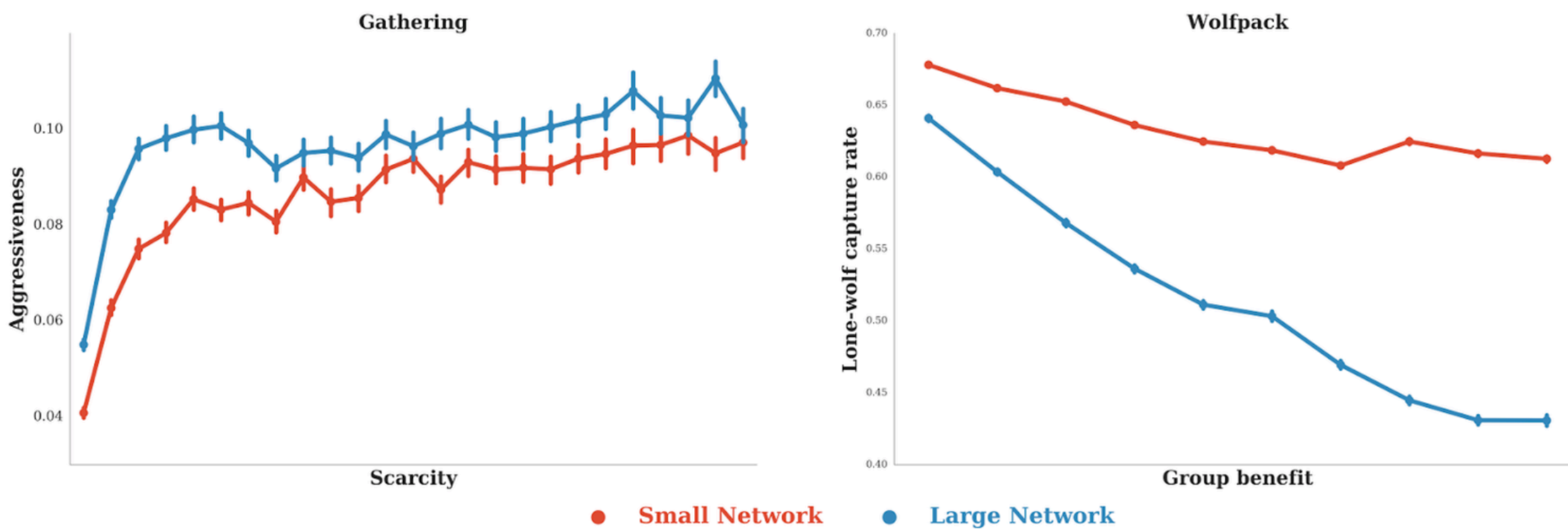
Wolfpack



Sequential Social Dilemmas (SSD)



Sequential Social Dilemmas (SSD)



Sequential Social Dilemmas (SSD)

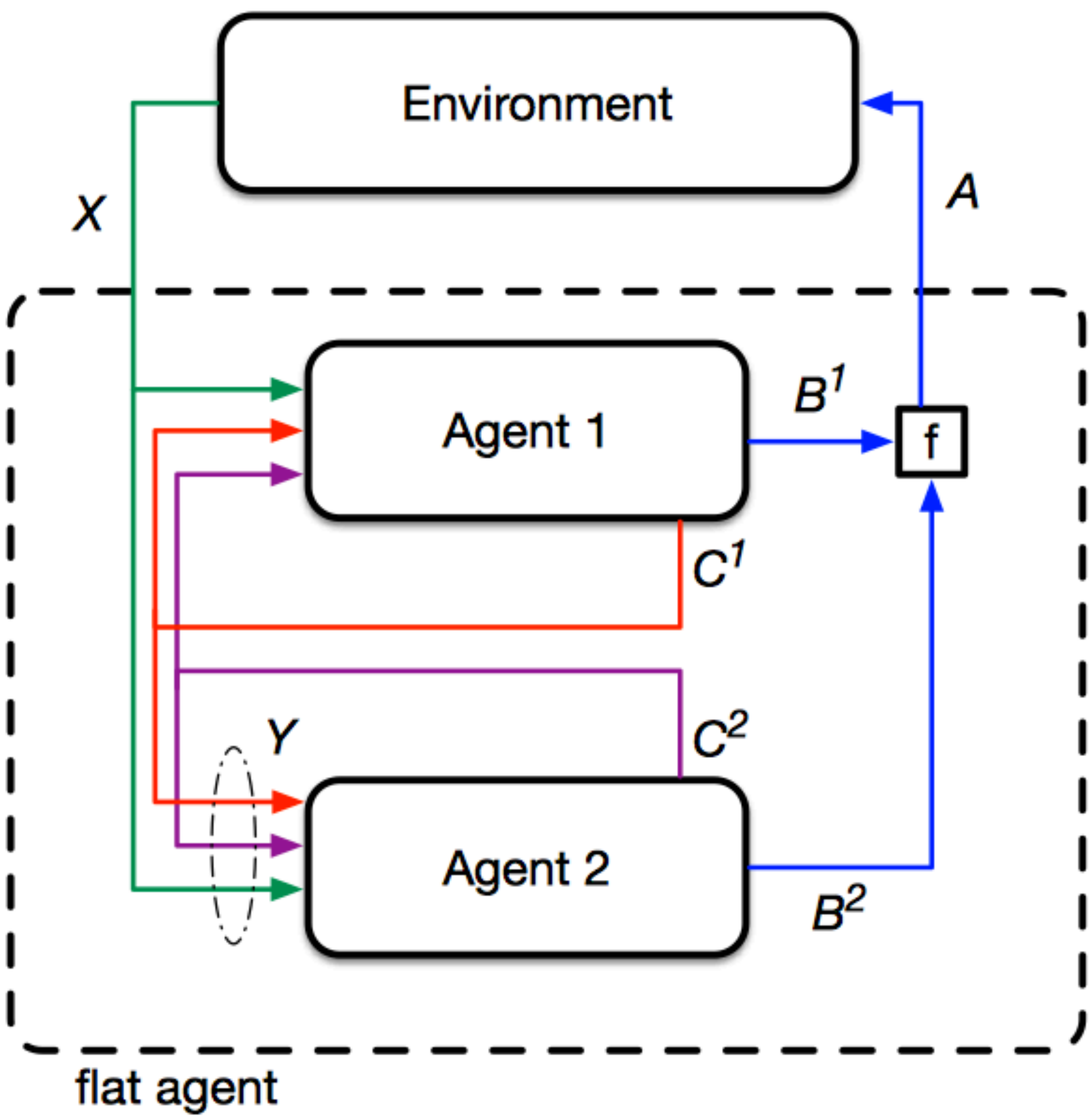
“Homo Economicus”

Specialization:
Improve Scalability of RL

Maluba

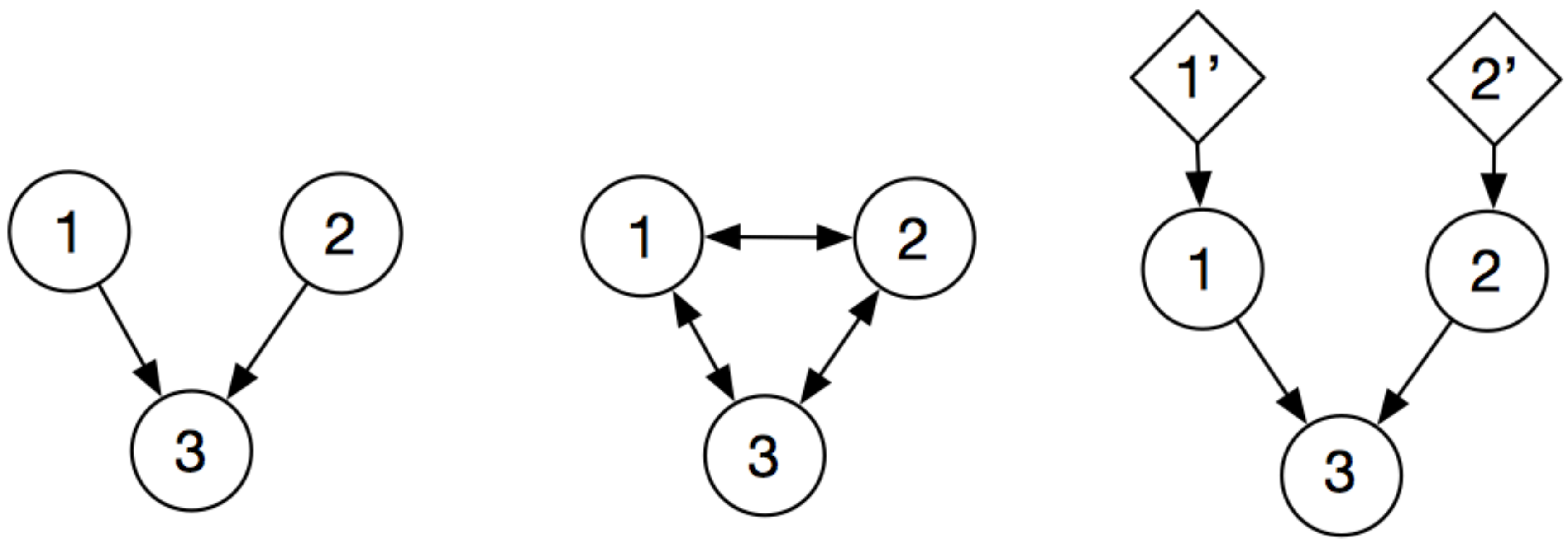
Scalable Reinforcement Learning

Separation of Concerns Model



Scalable Reinforcement Learning

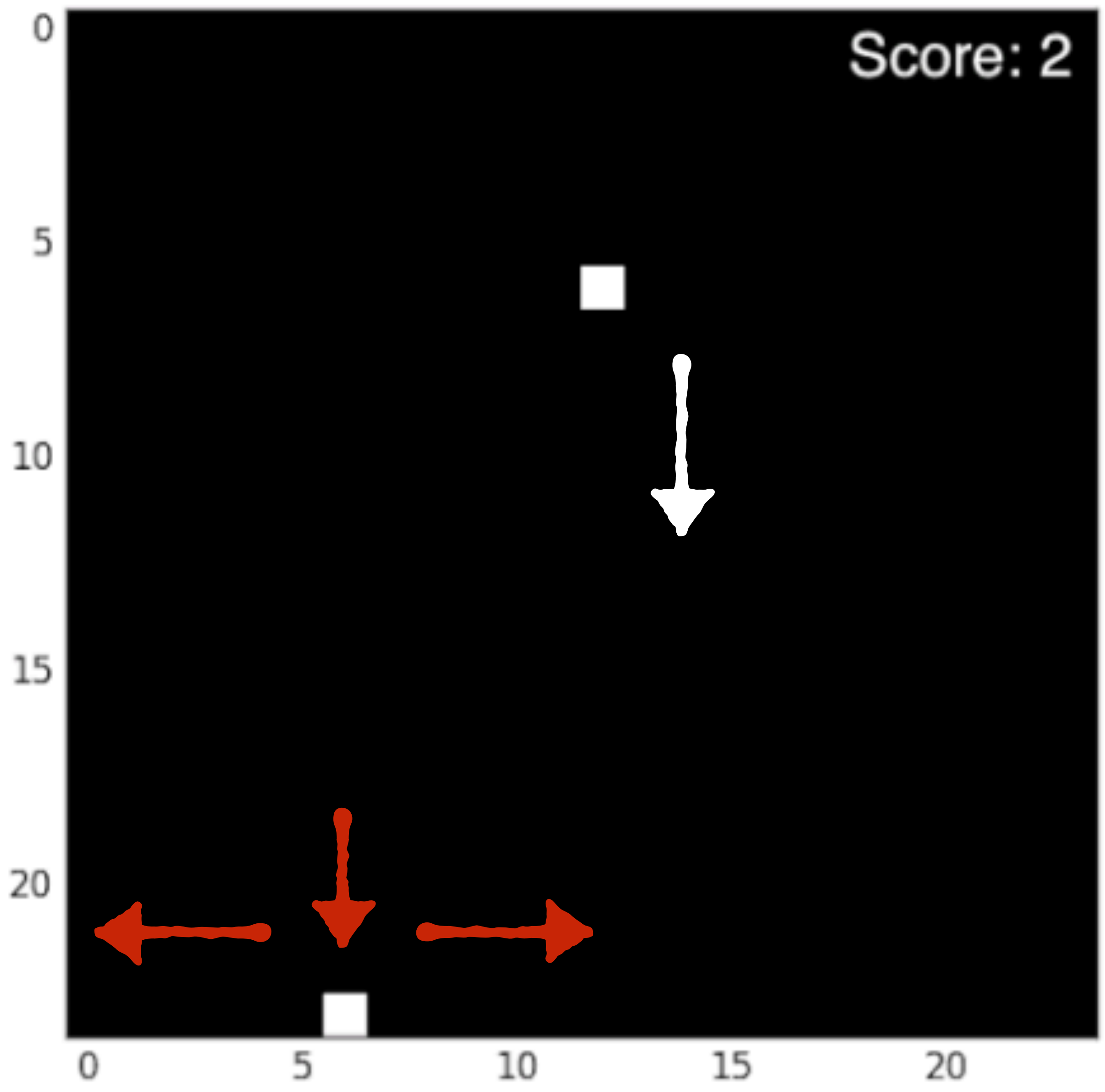
Convergence



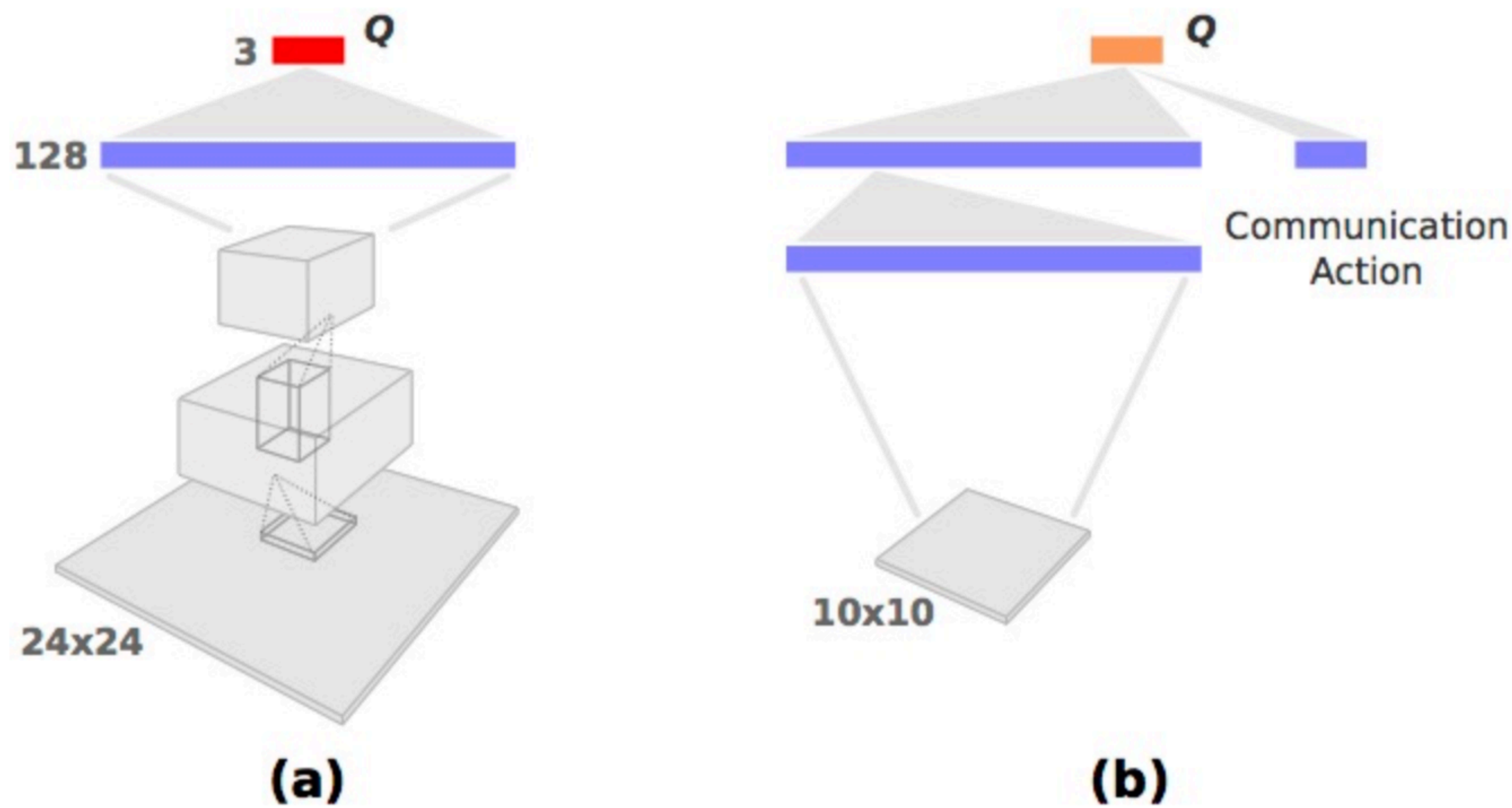
(a) (b) (c)

Scalable Reinforcement Learning

Catch



Scalable Reinforcement Learning

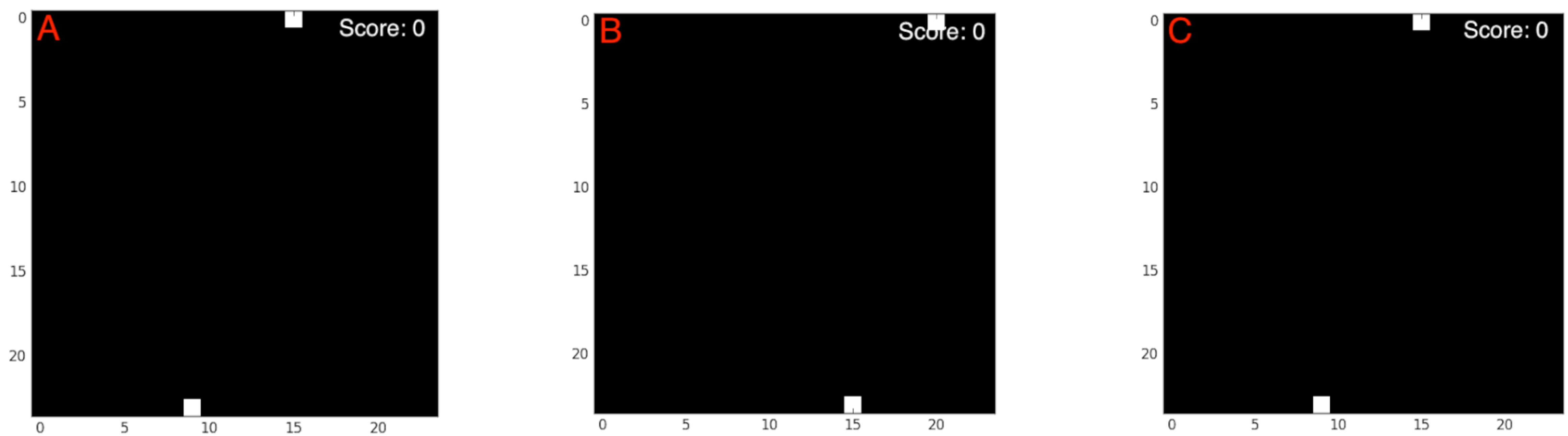


high-level agent:
high discount factor (adapts slowly)
accesses to the full screen

low-level agent:
low discount factor (adapts fast)
only sees part of the screen

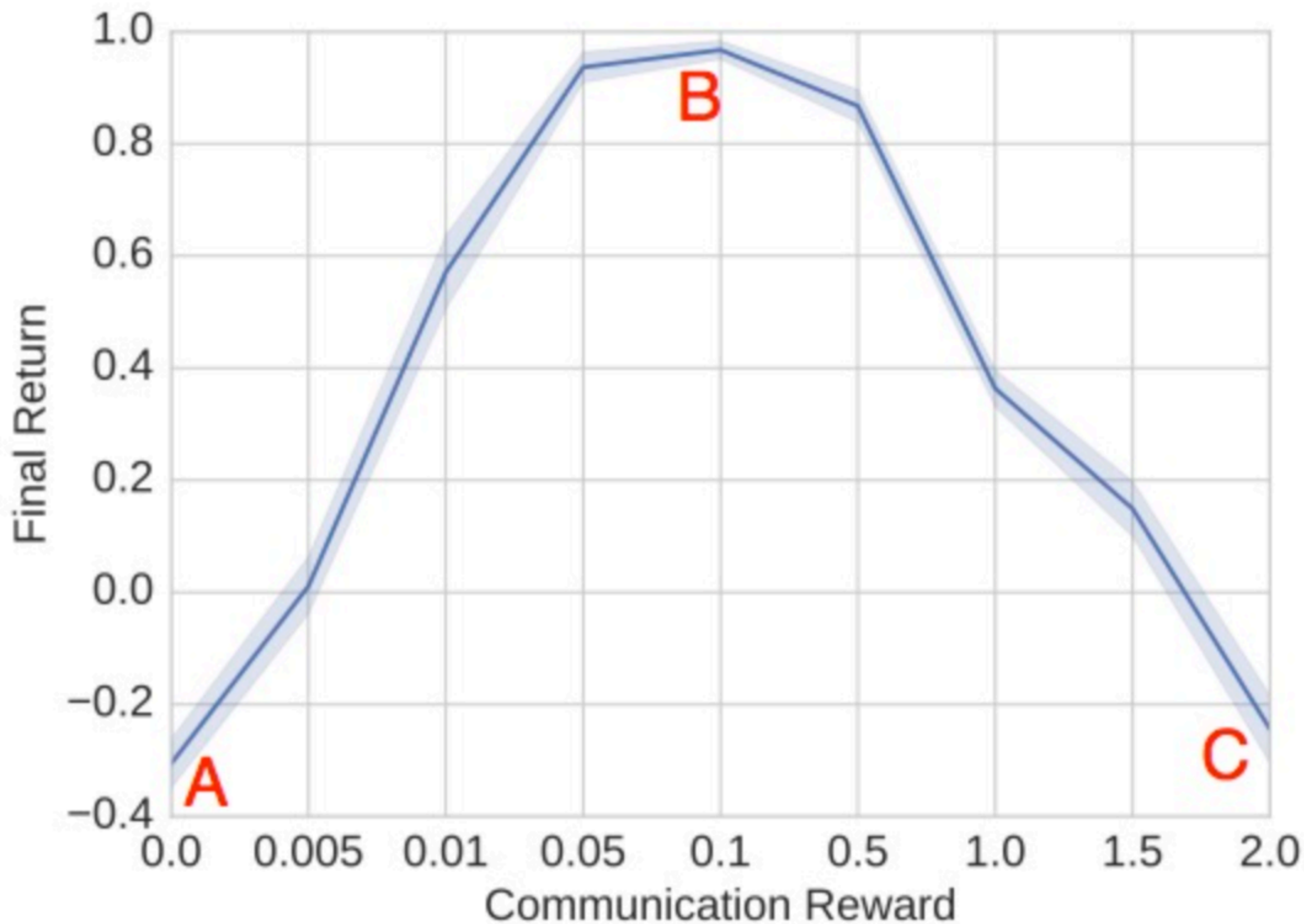
Scalable Reinforcement Learning

Catch



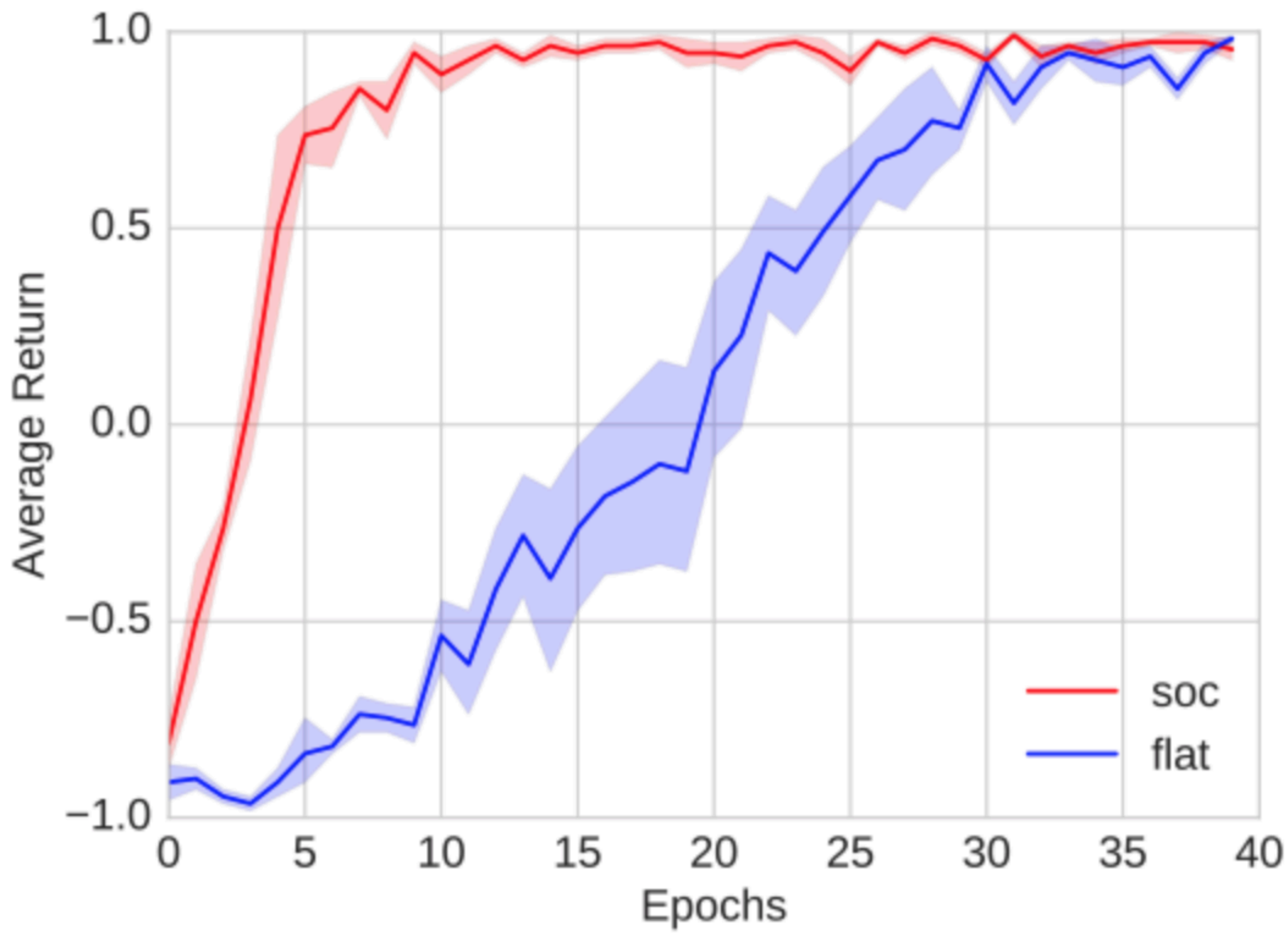
Scalable Reinforcement Learning

Catch



Scalable Reinforcement Learning

Catch



My Vision

Artificial Intelligence

?

Deep Learning

Game Theory

My Vision

AI Cooperation, a Cool Future!

↑ ↑

Deep Learning + Game Theory

References:

- Arora, S., Ge, R., Liang, Y., Ma, T., & Zhang, Y. (2017, March 2). Generalization and Equilibrium in Generative Adversarial Nets (GANs). arXiv.org.
- Ho, J., & Ermon, S. (2016). Generative Adversarial Imitation Learning. *Nips*.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative Adversarial Nets. *Nips*.
- Leibo, J. Z., Zambaldi, V., Lanctot, M., Marecki, J., & Graepel, T. (2017, February 10). Multi-agent Reinforcement Learning in Sequential Social Dilemmas. arXiv.org.
- Heinrich, J., & Silver, D. (2016, March 3). Deep Reinforcement Learning from Self-Play in Imperfect-Information Games. arXiv.org.
- Seijen, H., Fatemi M. & Romoff, J. (2016, Dec 15) . Improving Scalability of Reinforcement Learning by Separation of Concerns. arXiv.org.
- Moravčík, M., Schmid, M., Burch, N., Lisý, V., Morrill, D., & Bard, N., et al. (2017). Deepstack: expert-level artificial intelligence in no-limit poker. Science
- Gibney, E. (2016). Google AI algorithm masters ancient game of Go. *Nature*, 529(7587), 445-446.
- Finn, C., Christiano, P., Abbeel, P., & Levine, S. (2016, November 12). A Connection between Generative Adversarial Networks, Inverse Reinforcement Learning, and Energy-Based Models. *arXiv.org*.