



Imperfect Information Games & Multi-Agent Reinforcement Learning

Presenter: Jérémy Scheurer

Going Beyond Alpha Zero

1

Imperfect Information



Going Beyond Alpha Zero

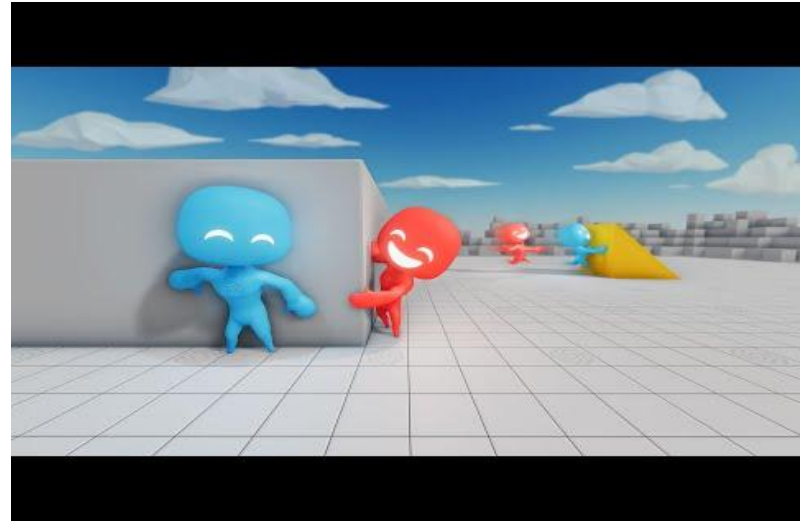
1

Imperfect Information



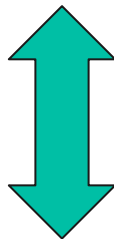
2

Multiple Agents & Rewards



Fundamental Questions

How can we solve imperfect information games?



How does Multi-Agent Reinforcement Learning relate to imperfect information games?

Perfect vs Imperfect environments

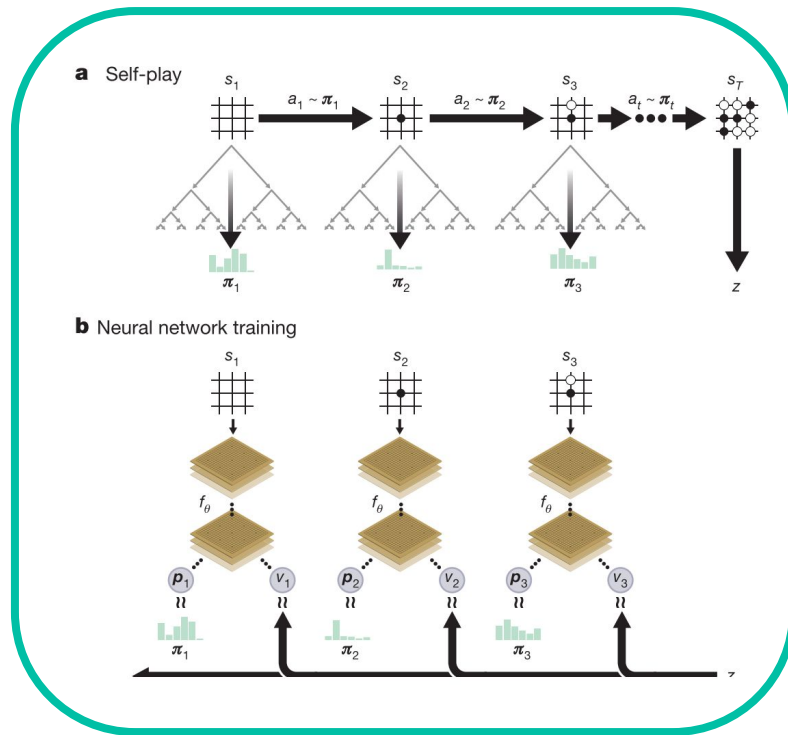
Perfect Information Game



Imperfect Information Game

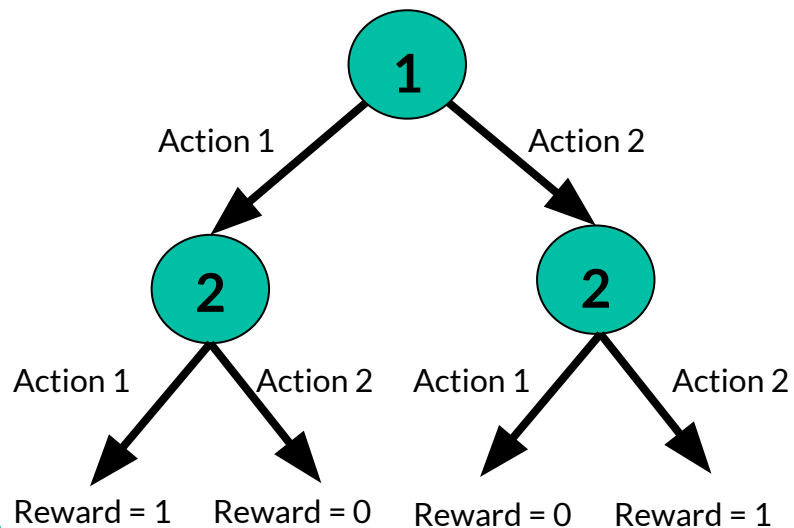


Can we use Alpha Zero for Imperfect Information Games?

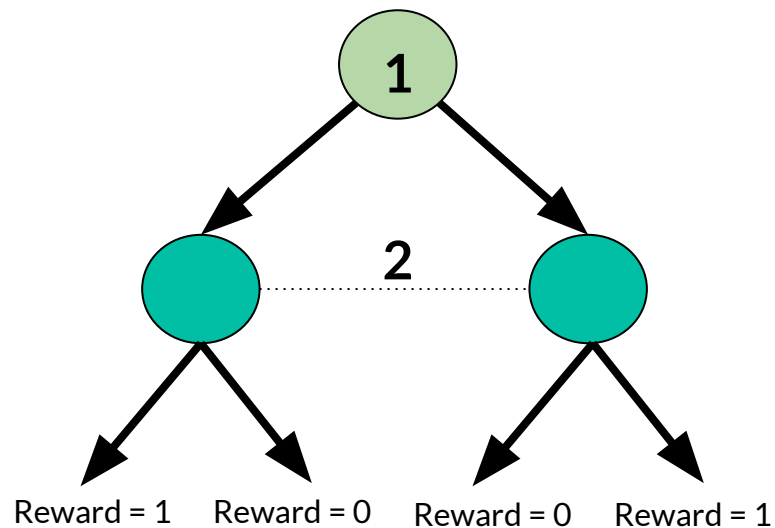


MCTS Assumes Perfect Information

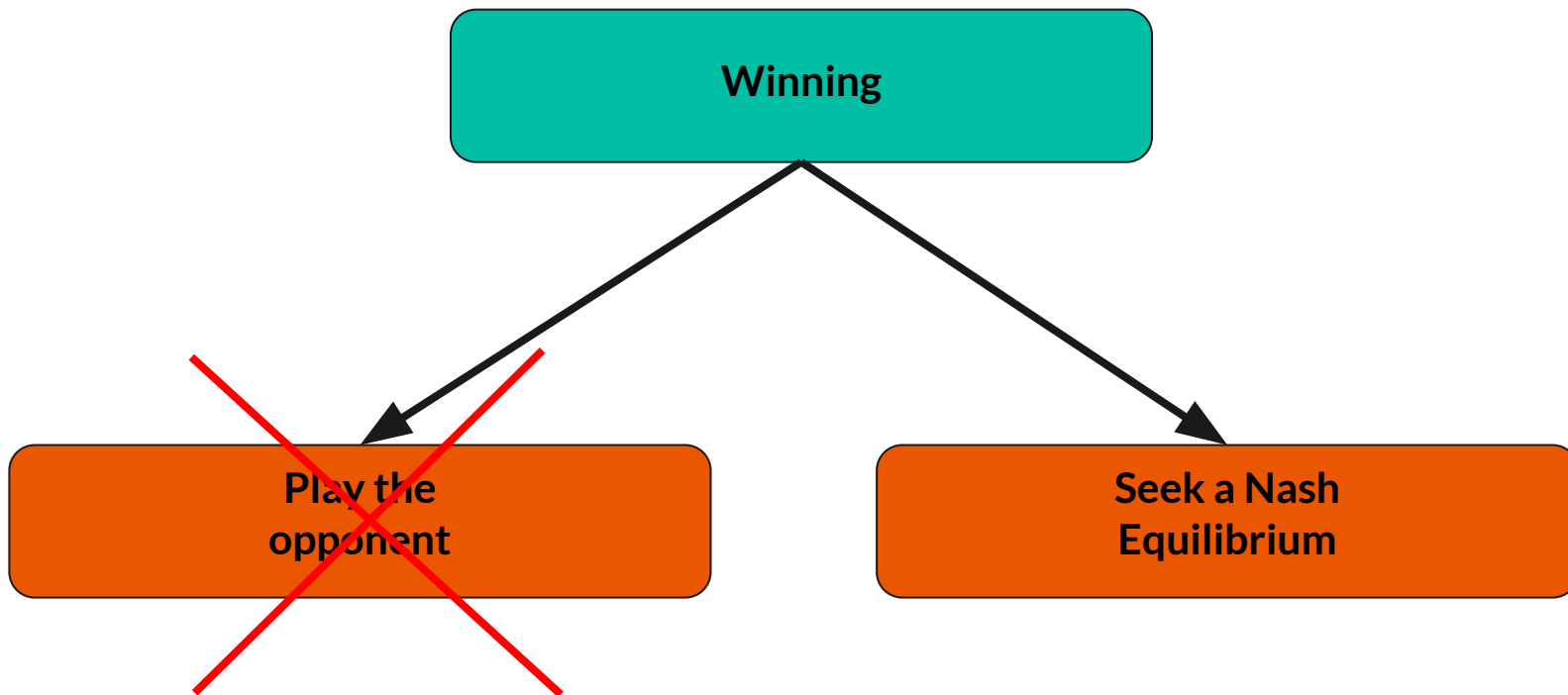
Perfect Information Game



Imperfect Information Game



What is the goal in any game?

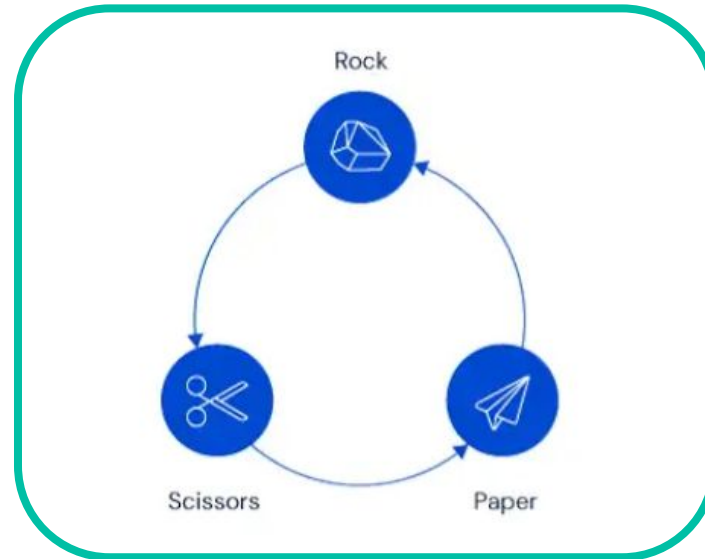


What is a Nash Equilibrium

If each player has chosen a strategy, and no player can benefit by changing strategies while the other players keep theirs unchanged.

	Betray	Remain Silent
Betray	2/2	0/3
Remain Silent	3/0	1/1

Best Response



Fictitious Play [1]

Round	Player 1's action	Player 2's action	Player 1's beliefs	Player 2's beliefs
0			(R=3, P=1, S=1)	(1, 1, 1)
1				
2				
...				
n				

Fictitious Play [1]

Round	Player 1's action	Player 2's action	Player 1's beliefs	Player 2's beliefs
0			(R=3, P=1, S=1)	(1, 1, 1)
1	Paper	Paper	(3, 2, 1)	(1, 2, 1)
2				
...				
n				

Fictitious Play [1]

Round	Player 1's action	Player 2's action	Player 1's beliefs	Player 2's beliefs
0			(R=3, P=1, S=1)	(1, 1, 1)
1	Paper	Paper	(3, 2, 1)	(1, 2, 1)
2	Paper	Scissor	(3, 2, 2)	(1, 3, 1)
...				
n				

Fictitious Play [1]

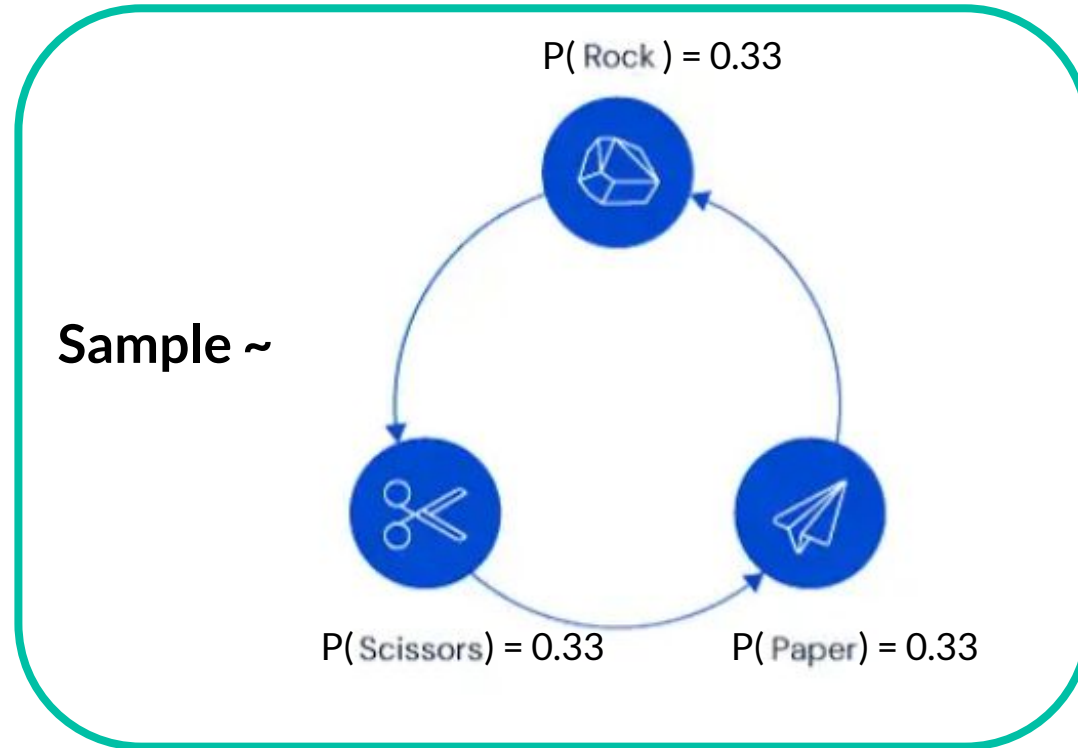
Round	Player 1's action	Player 2's action	Player 1's beliefs	Player 2's beliefs
0			(R=3, P=1, S=1)	(1, 1, 1)
1	Paper	Paper	(3, 2, 1)	(1, 2, 1)
2	Paper	Scissor	(3, 2, 2)	(1, 3, 1)
...
n			(0.33, 0.33, 0.33)	(0.33, 0.33, 0.33)

Fictitious Play Converges to Nash Equilibrium

Theorem 1: Fictitious play converges to a Nash Equilibrium in two-player zero sum game.

...
n			(0.33, 0.33, 0.33)	(0.33, 0.33, 0.33)

Play Mixed Strategy

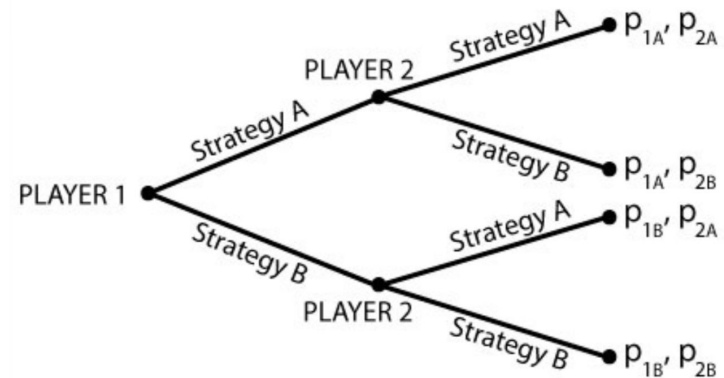


Normal Form vs Extensive Form Games

Normal Form Games

	Betray	Remain Silent
Betray	2/2	0/3
Remain Silent	3/0	1/1

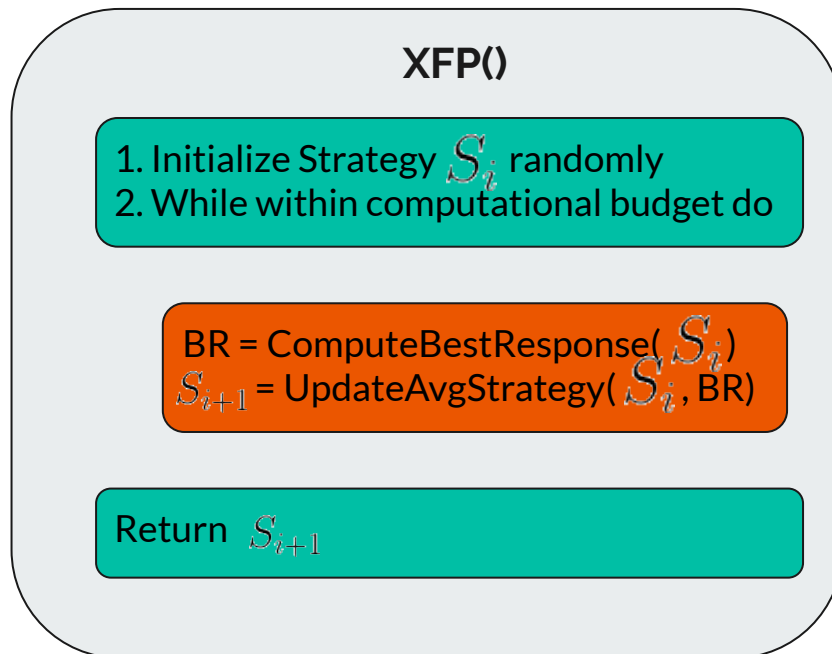
Extensive Form Games (Behavioural Strategies)



Extensive Form Fictitious Play Convergences

Extensive Form fictitious play inherits properties of fictitious play i.e. Average strategy converges to Nash-Equilibrium (in certain games).

Extensive Form Fictitious Play (XFP) [2]



2 Policies to Play

Best Response Policy

vs.

**Average Policy
(converges to Nash
Equilibrium)**



Problem of XFP

Computationally Intractable

Approximate XFP

function **ComputeBestResponse**(S):



MDP



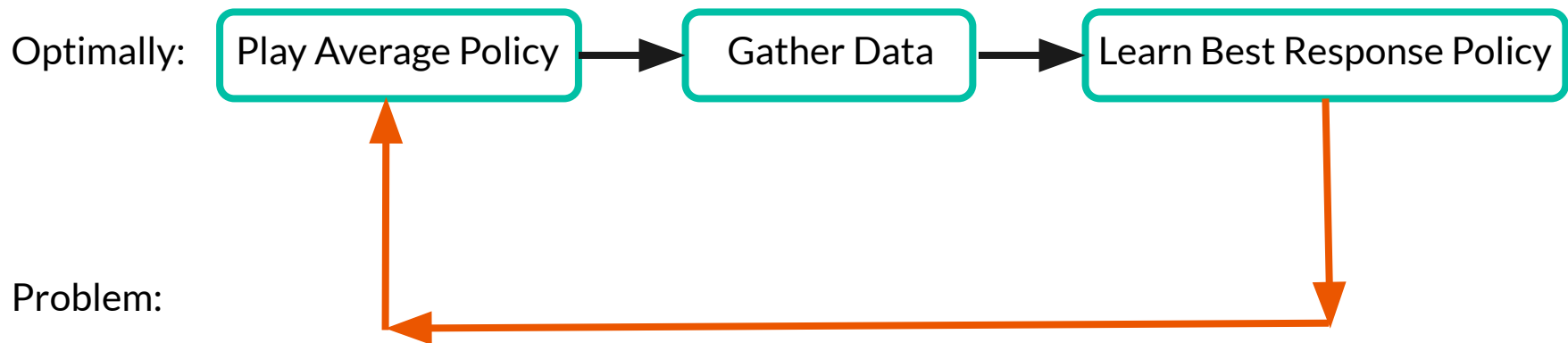
Reinforcement Learning

Function **UpdateAvgStrategy**(S, BR):

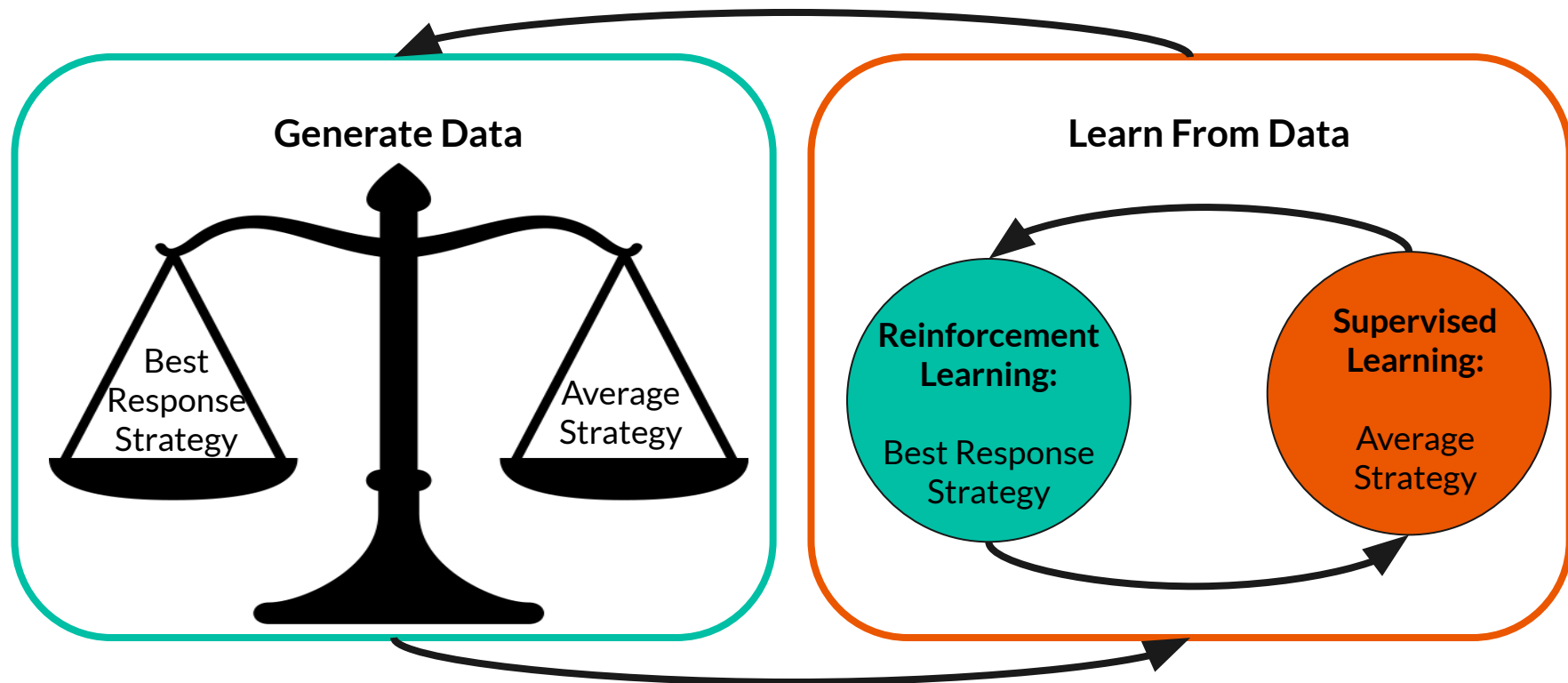


Supervised Learning

Dilemma: What policy to play?

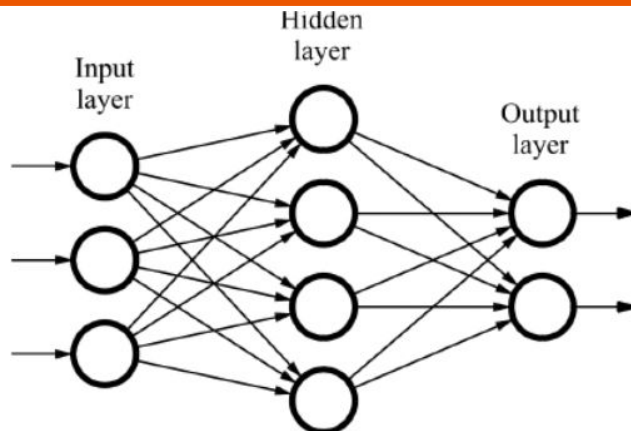


Fictitious Self Play [2]



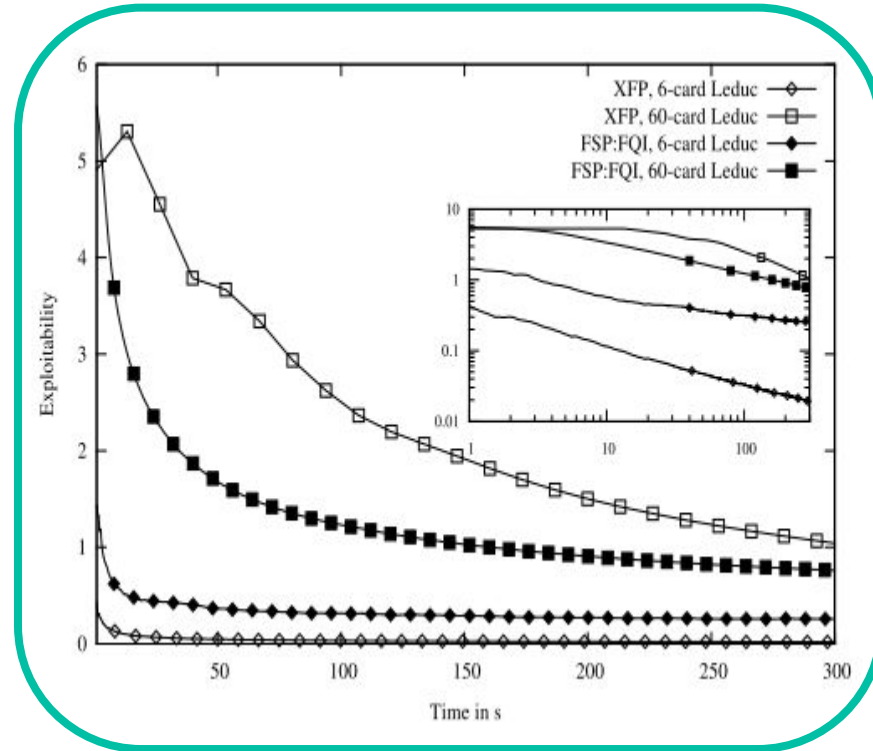
Neural Fictitious Self Play [3]

Fictitious Self Play

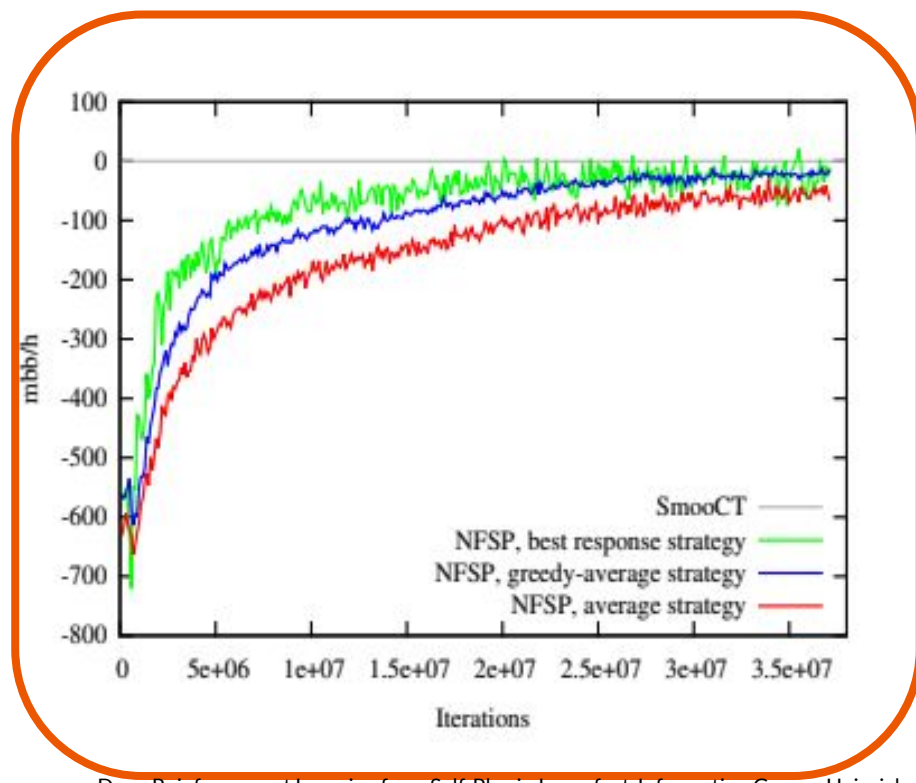
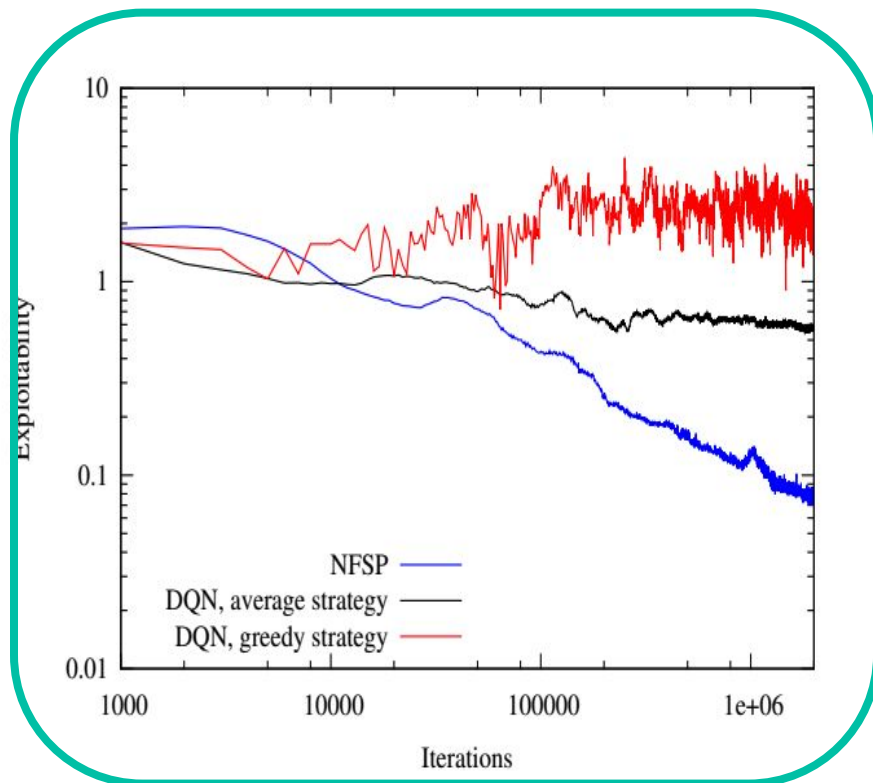


Neural Fictitious Self Play

Results: XFP vs FSP



Results: NFSP



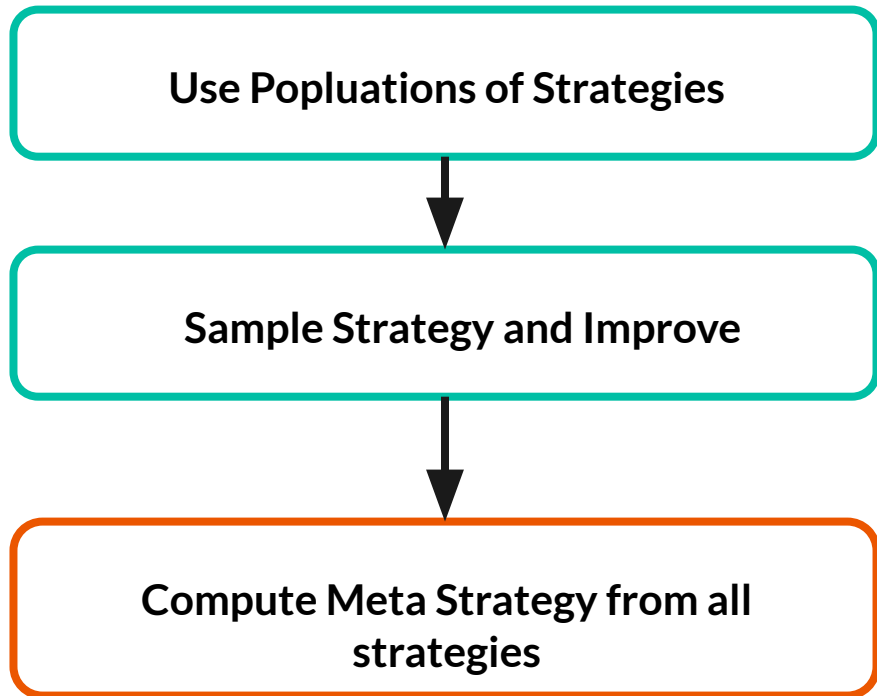


Drawbacks of NFSP

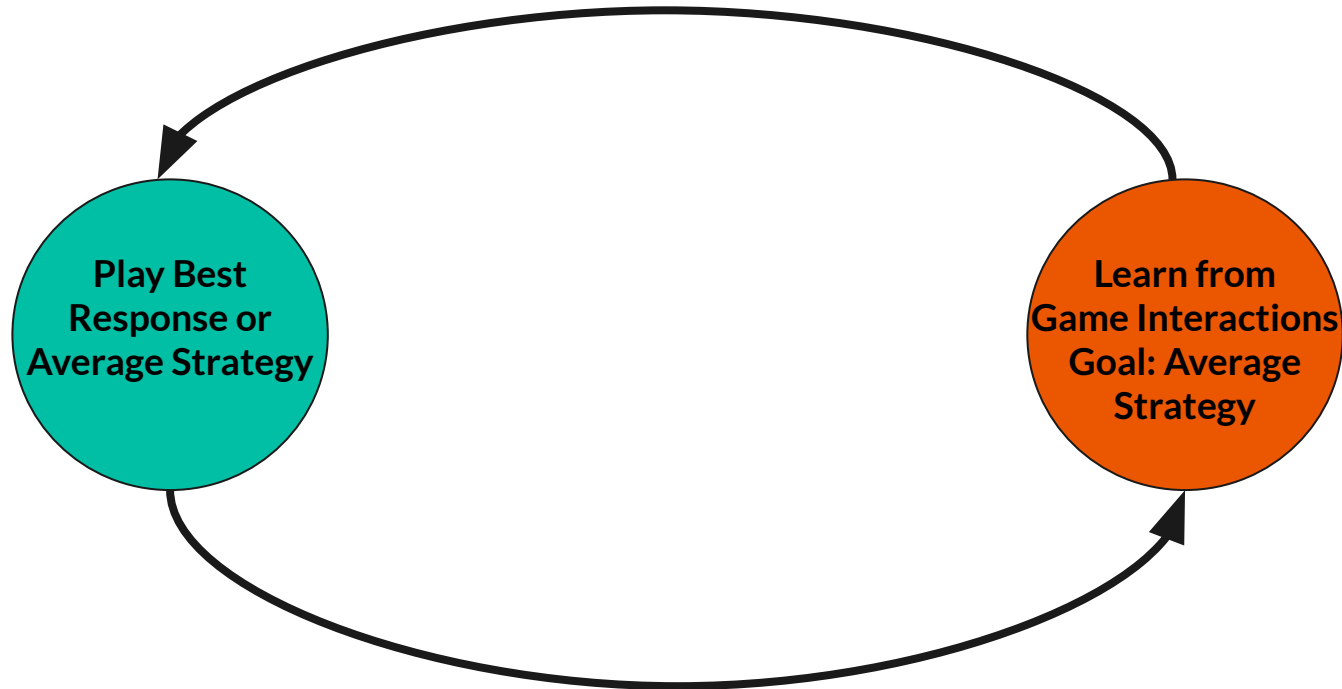
- It requires off-policy Reinforcement Learning Algorithms (DQN)

Still has trouble to learn in environments with very large strategic spaces

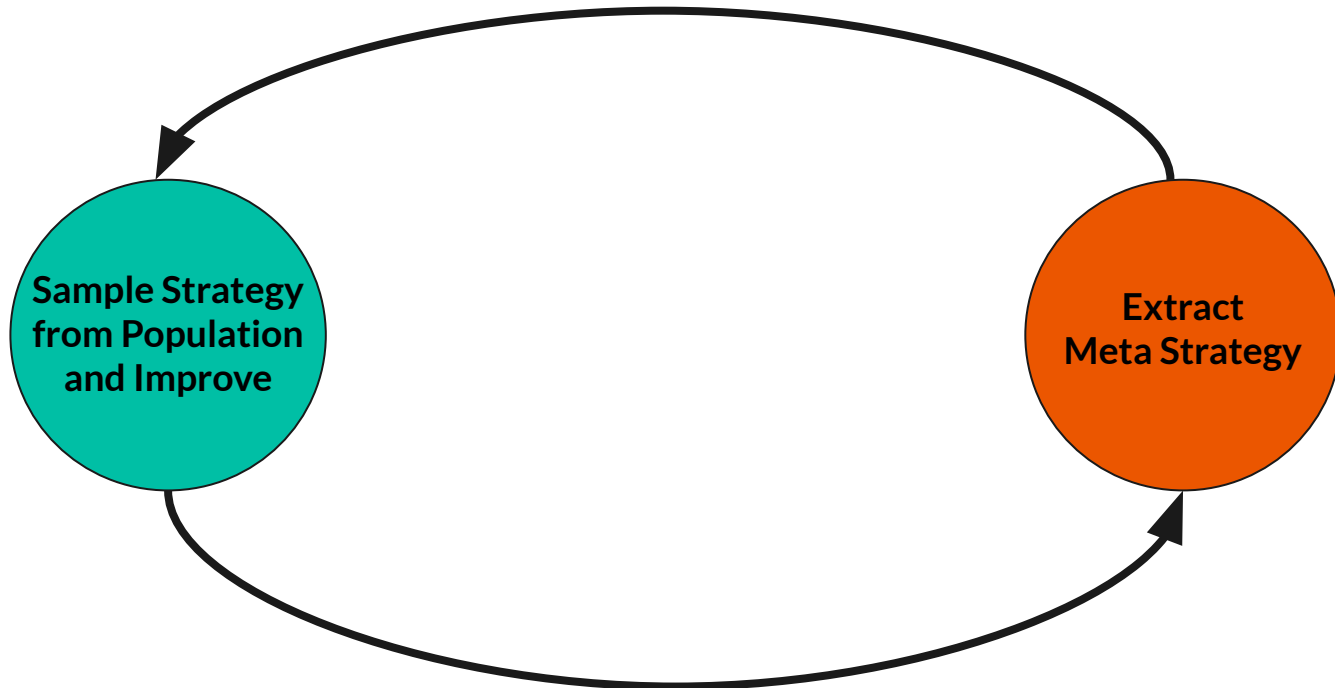
A unified Game Theoretic Approach to Multi-Agent RL - PSRO Algorithm [4]



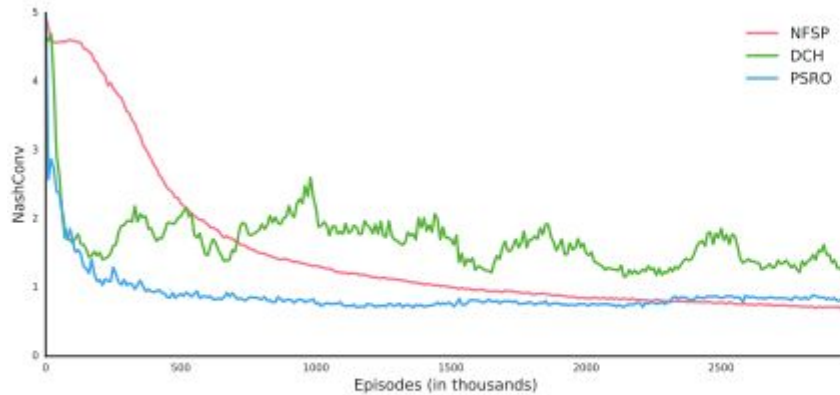
NFSP



Policy Space Response Oracles (PSRO)



Results: PSRO



(a) 2 players

Alphastar for Starcraft [5]



[5] Grandmaster level in StarCraft II using multi-agent reinforcement learning, Vinyals et. al.



Starcraft

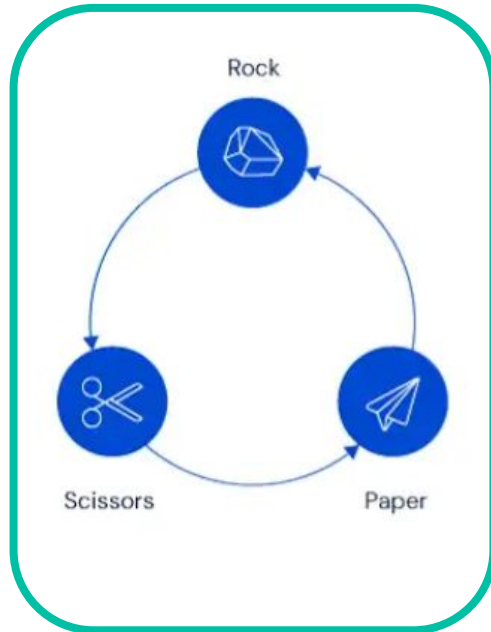
- What is Starcraft and how does it differ from Go?
 - Imperfect Information
 - Many more actions
 - Longer episodes (long-term planning)
 - Real-time

Alpha Star [6]

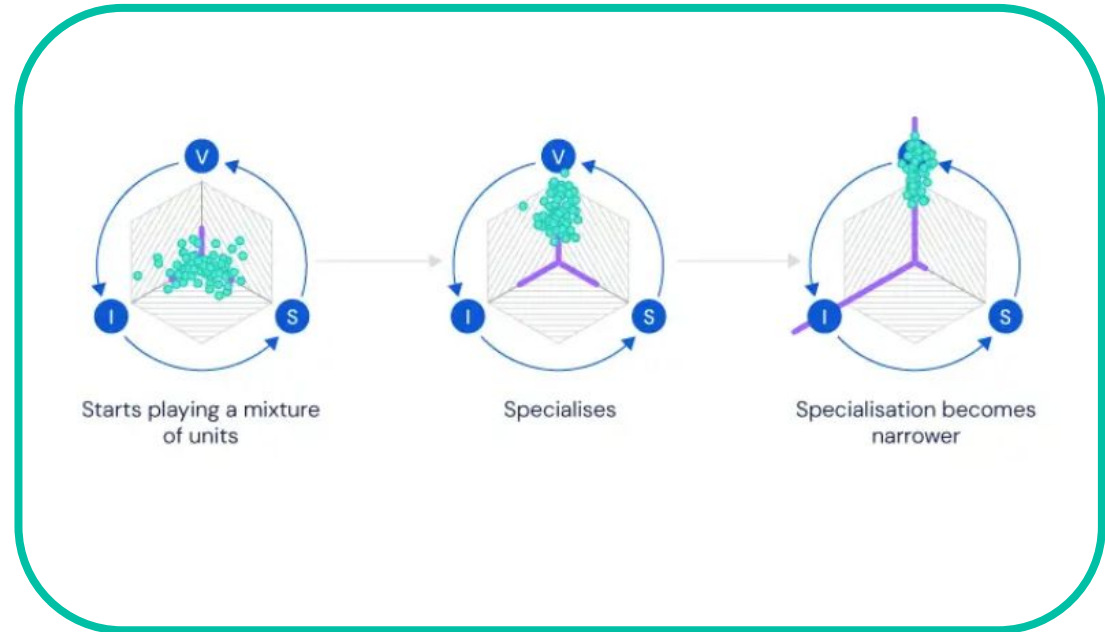
- Complicated Model
- Imitation Learning
- RL-Algorithm is similar to advantage Actor Critic

Does this work?

Forgetting



Focus on easy strategies



When do cycles arise?

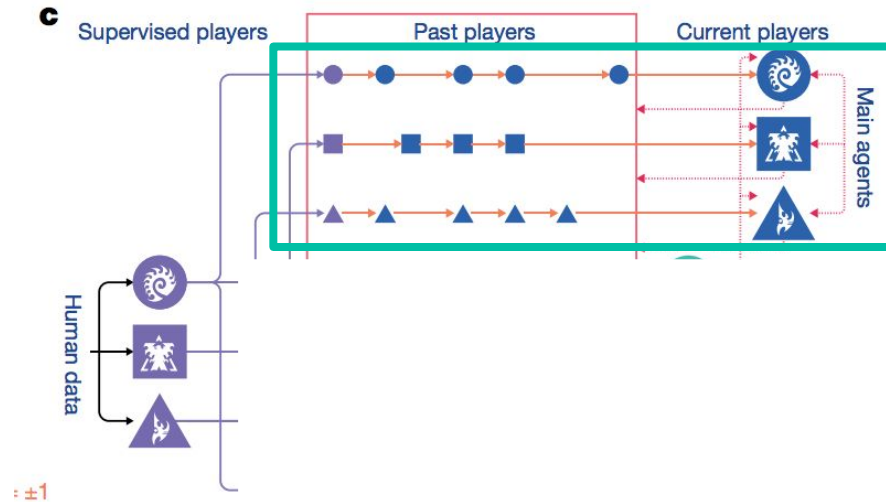
Strategic cycles often arise when agents play simultaneous move or imperfect information games such as rock-paper-scissors, poker, or StarCraft.[6]

Potentially also:

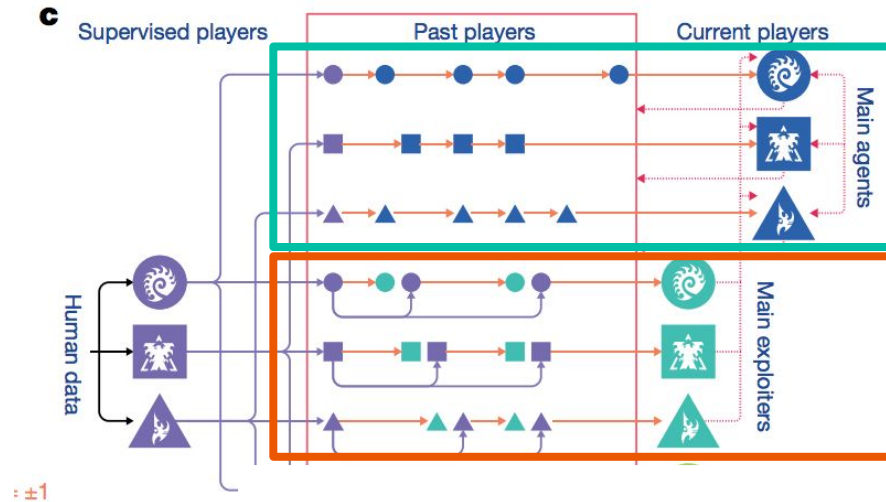
Large episodes

Large Action Spaces

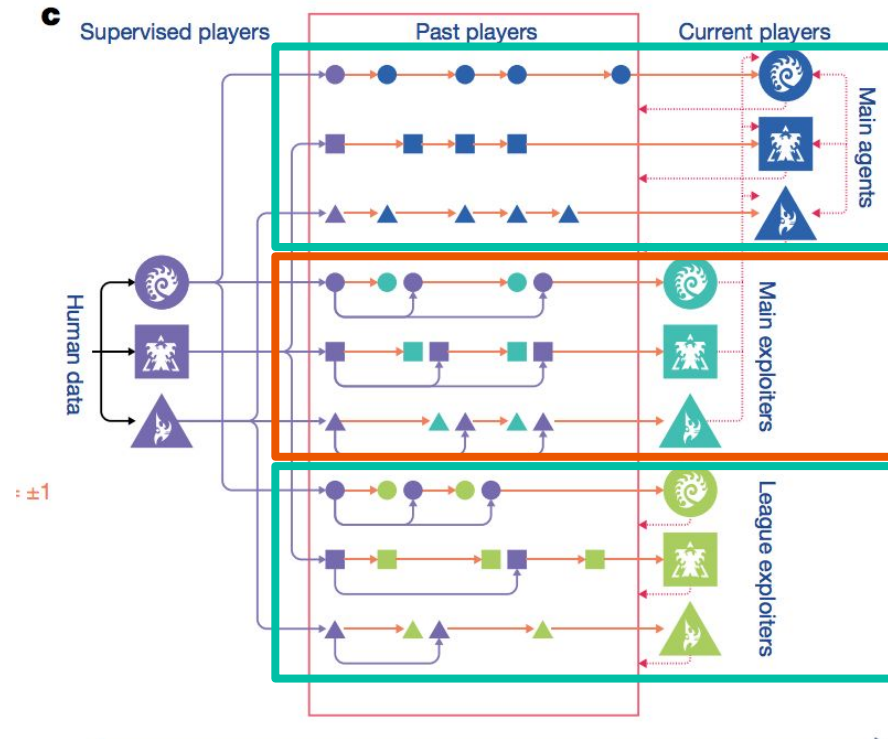
AlphaStar League



AlphaStar League

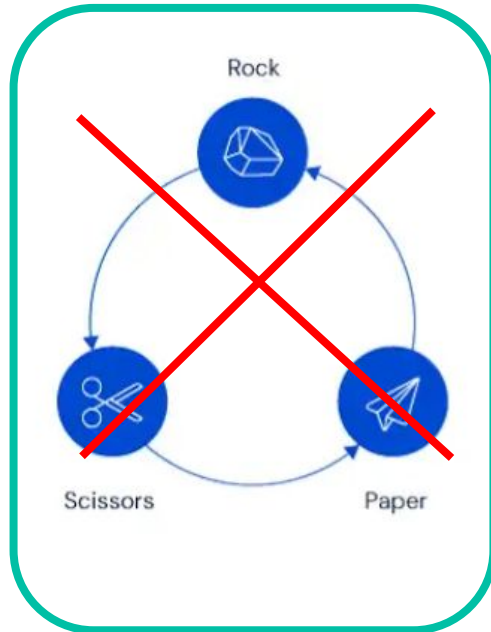


AlphaStar League

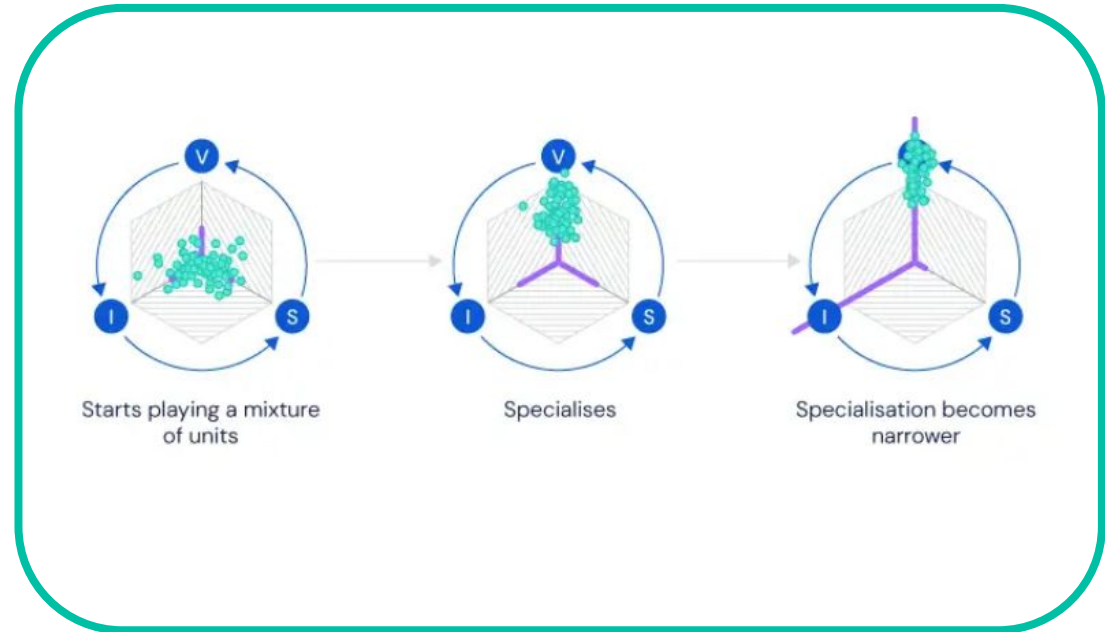


Does this work?

Avoid Forgetting

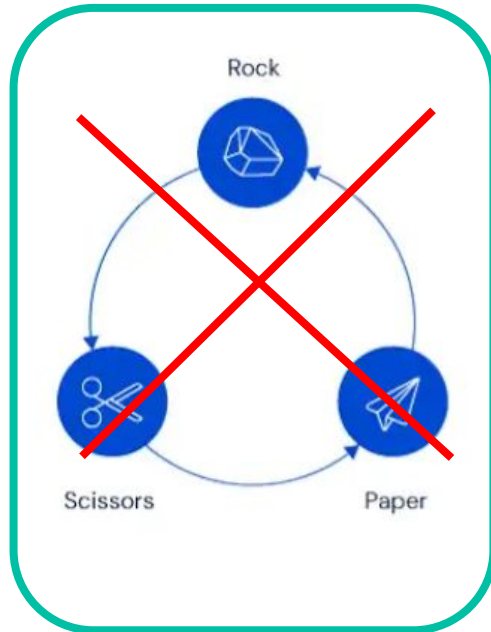


Focus on easy strategies

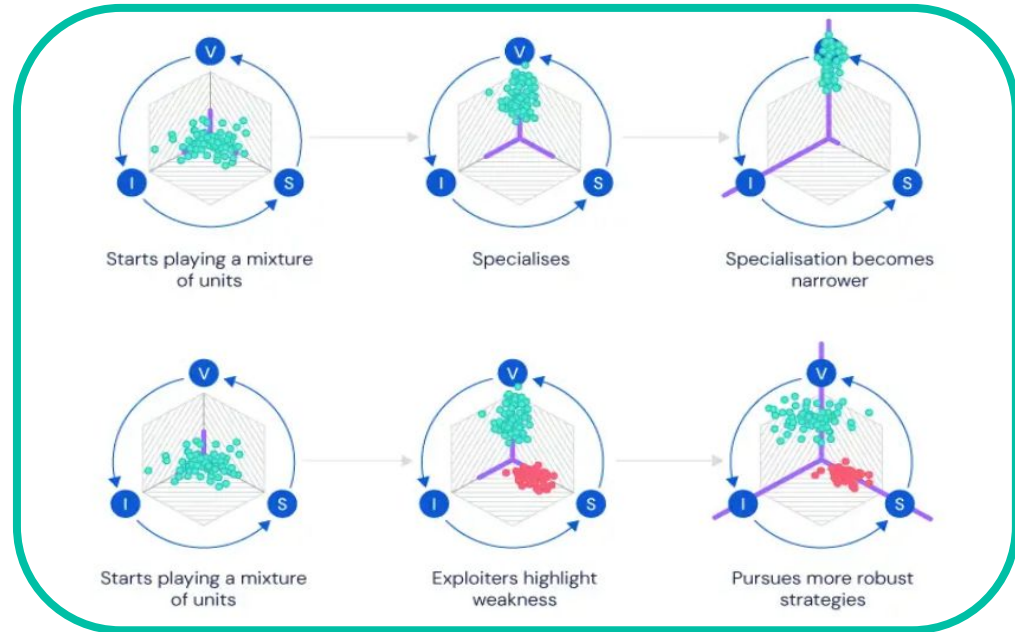


Does this work?

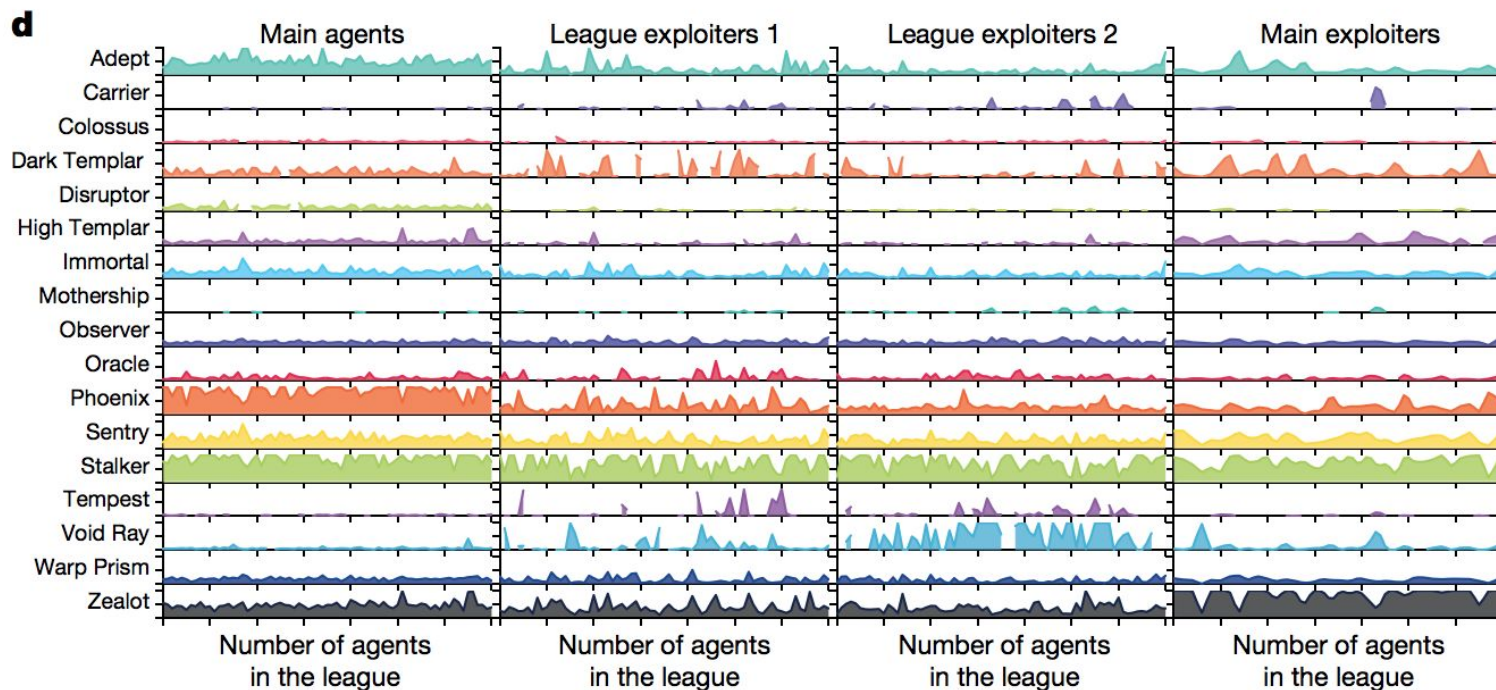
Avoid Forgetting



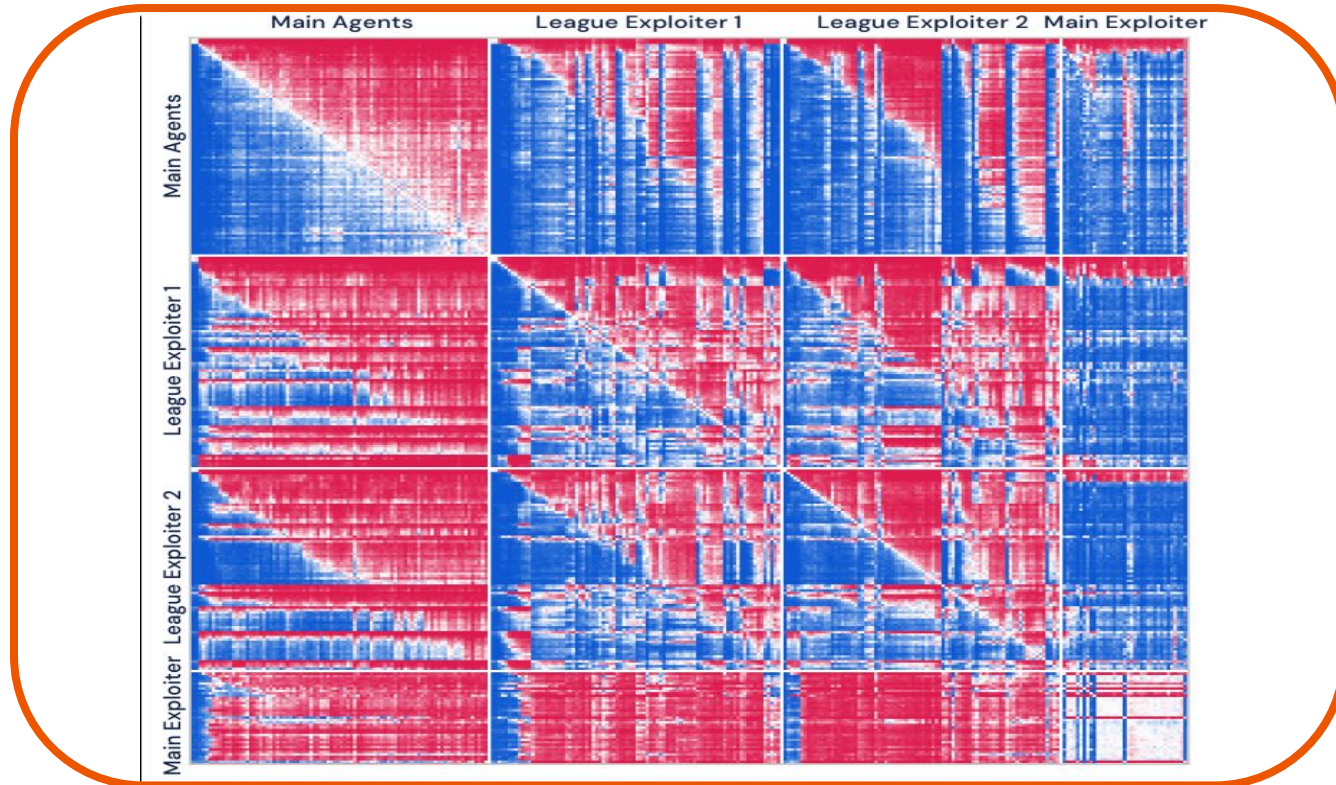
Explore Difficult Strategies



Unit distribution for types of agents



Payoff Matrix



Why no Multi-Agent RL for AlphaZero?

Neural Nets and Self Play are very prone to cycles, how can we avoid them?

We do keep older versions to train against

MCTS to the rescue (probably)

Finding an algorithm for stable learning in GO is a HUGE achievement!



Lessons Learned

- Limitations of Alpha Zero
- How to solve Imperfect Information Games
- How to leverage Multiple Agents/Leagues



Thank You



References - Images/Videos

- Planet Image: Deepmind, Alphastar
<https://deepmind.com/blog/article/AlphaStar-Grandmaster-level-in-StarCraft-II-using-multi-agent-reinforcement-learning>, accessed 22.3.2020
- Poker image: <https://www.scinexx.de/news/technik/ki-gewinnt-gegen-pokerprofis/>, accessed 12.3.2020
- Multi Agent Seek and Hide, Openai, <https://www.youtube.com/watch?v=kopoLzvH5jY&t=69s>, accessed 21.3.2020
- Chess Image: <https://www.chess.com/article/view/game-of-the-century-chess>, accessed 12.3.2020
- Alpha Zero MCTS: Mastering the Game of Go without human knowledge, Silver et. al.
- Rock Paper Scissors: Deepmind, Alphastar,
<https://deepmind.com/blog/article/AlphaStar-Grandmaster-level-in-StarCraft-II-using-multi-agent-reinforcement-learning>, accessed 22.3.2020
- Extensive Form Games: <https://towardsdatascience.com/fictitious-self-play-30b76e30ec6a>, accessed 13.3.2020
- Neural Network: <https://databricks.com/glossary/neural-network>, accessed 22.3.2020
- Starcraft Video: <https://www.youtube.com/watch?v=6EQAsrfUlyo&feature=youtu.be>, accessed 22.3.2020
- Self-Play & Exploiters Development:
<https://deepmind.com/blog/article/AlphaStar-Grandmaster-level-in-StarCraft-II-using-multi-agent-reinforcement-learning>, accessed 22.3.2020