Aishik Pyne
Harshavardhan Abhichandani
Niharika Shrivastava

Deceptive Poker

Agents learns to cheat via collaboration in an competitive multi-agent RL environment

Introduction

Problem Statement

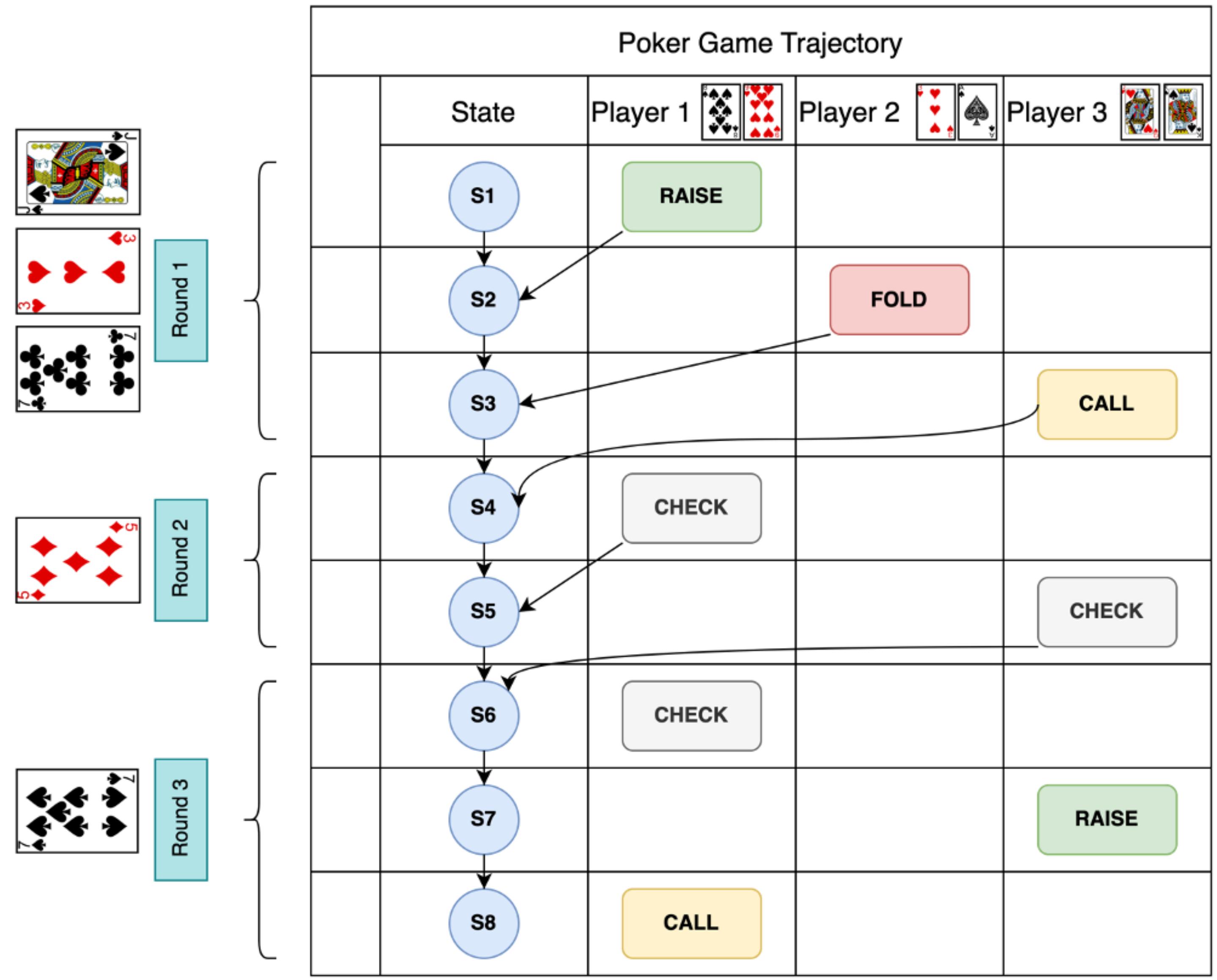
Texas Hold’em Poker

A popular skill and luck based gambling game

Modelling Poker in RL Setting

- Partially Observable
- Multi-Agent Zero-Sum
- Extensive Form game
- Intractable State-Space: 10^{14}

One Round of Poker



Winner get’s all chips

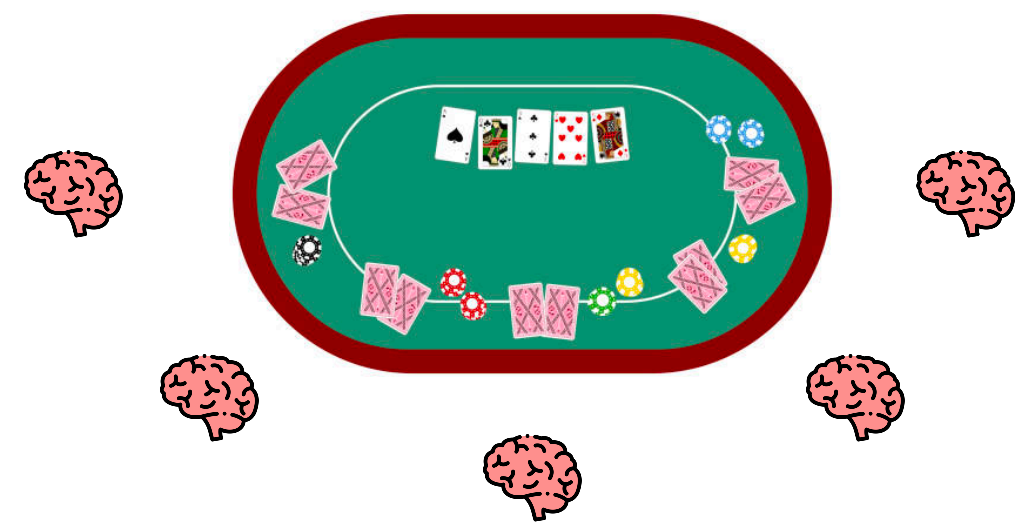
Introduction

Objectives

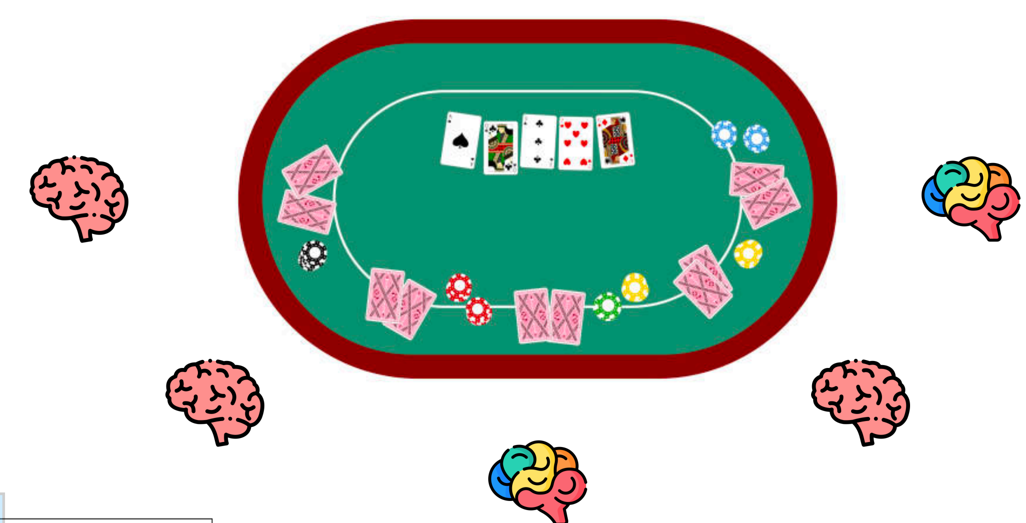
- Train fully competitive agent to play poker.
- Train agents which learn to cheat via collaborating.
- Create a classifier which can detect agents that are cheating

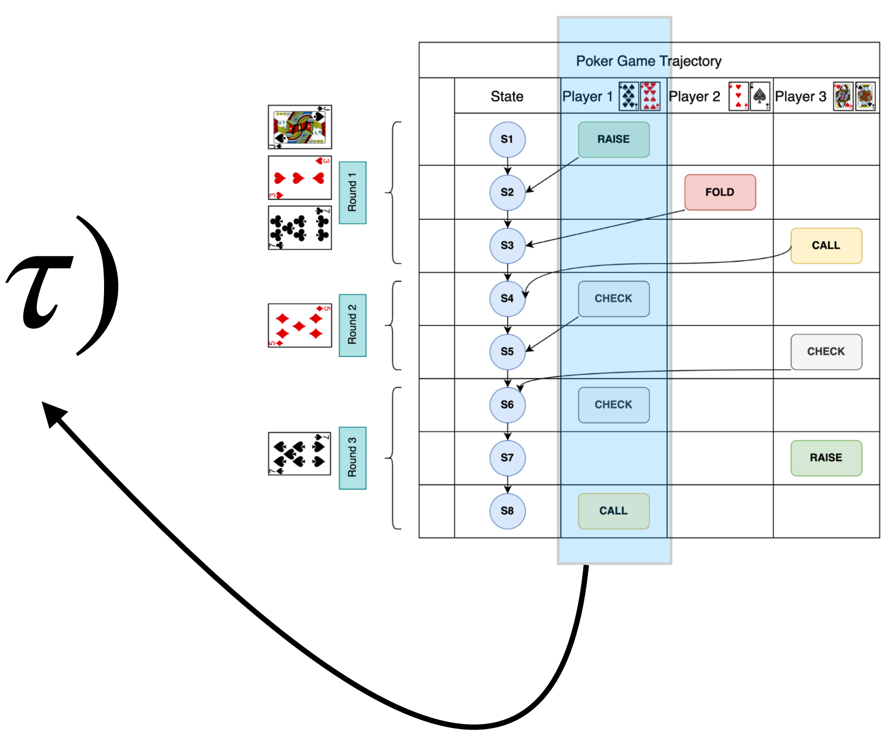
$$p(\{B_1, B_2\} \mid \tau)$$

Training Brain 1



Training Brain 2



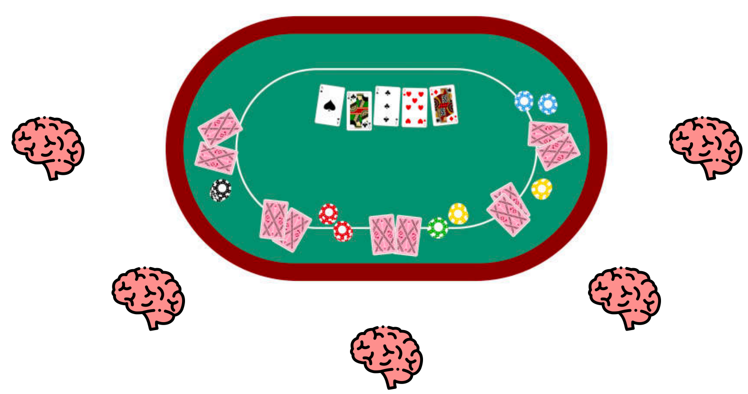



Brain 1:
Training Fully
Competitive Agents

Brain 1: Competitive Agents

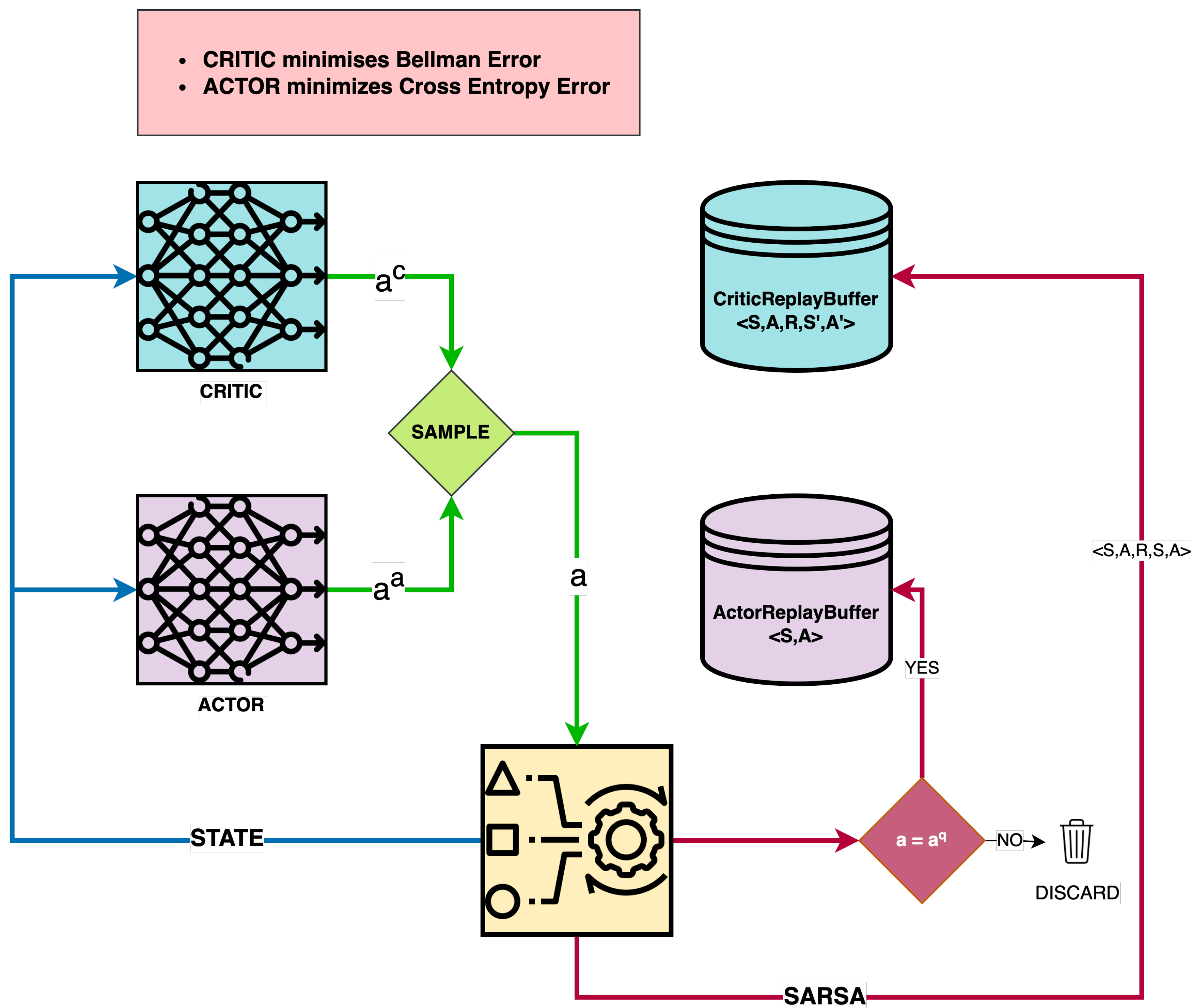
Algorithm

Training Brain 1



Note: We use a 3 player setting despite the image.

Neural Fictitious Self Play



Evaluation

- NSFP consistently beats a Rule Based agent with win-ratio 3.65


Brain 2:
Training
Collaborative Agents

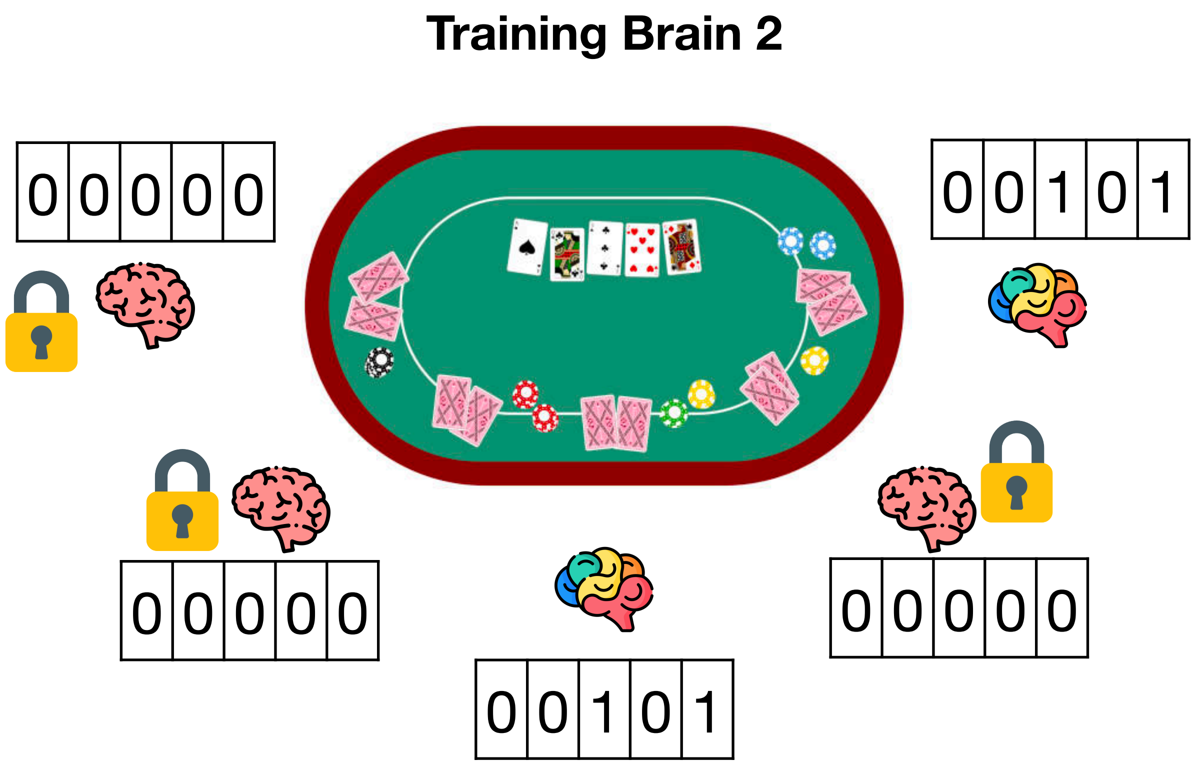
Brain 2: Collaborative Agents

Setup

- Add context to the agents as to who their partner is by indicating which historical raises were made by the agent's partner

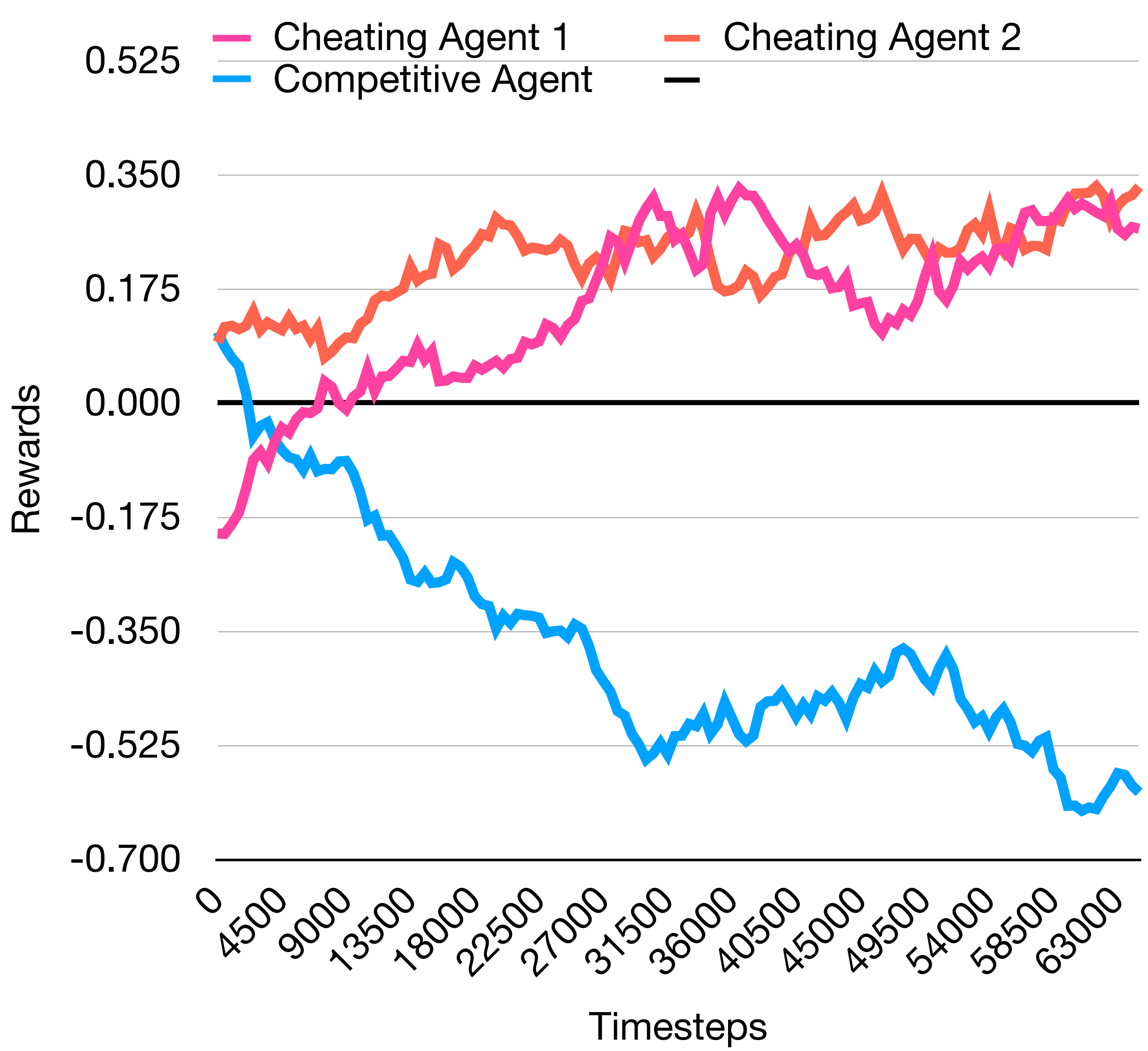
- Update the reward function:
  - Either agent wins: Max Winning
  - Both agents lose: Avg Loss

- Freeze the competitive agents during training



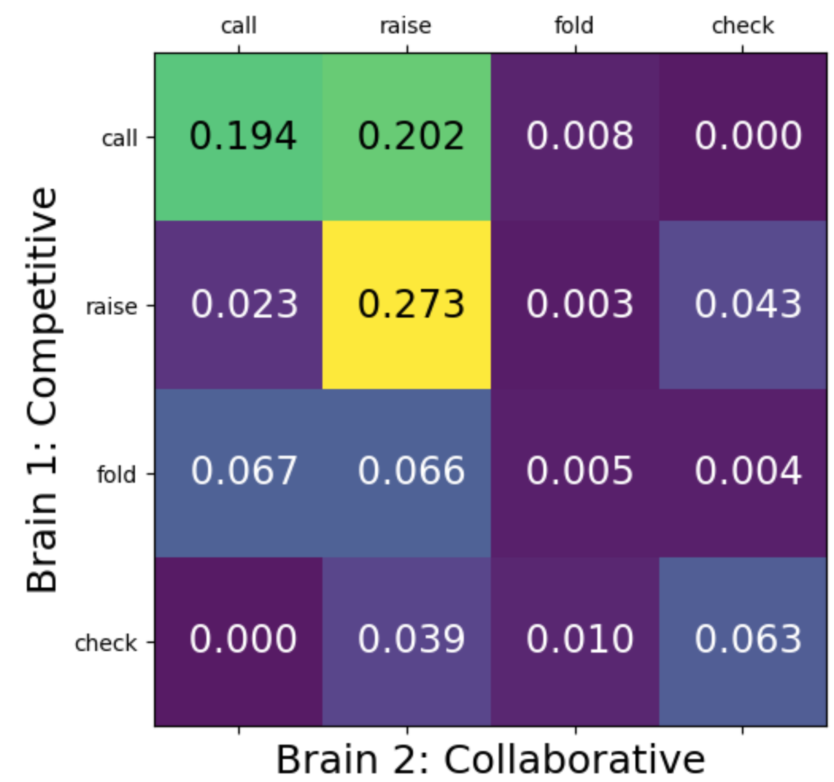
Brain 2: Collaborative Agents

Training and Results



Training Brain 2

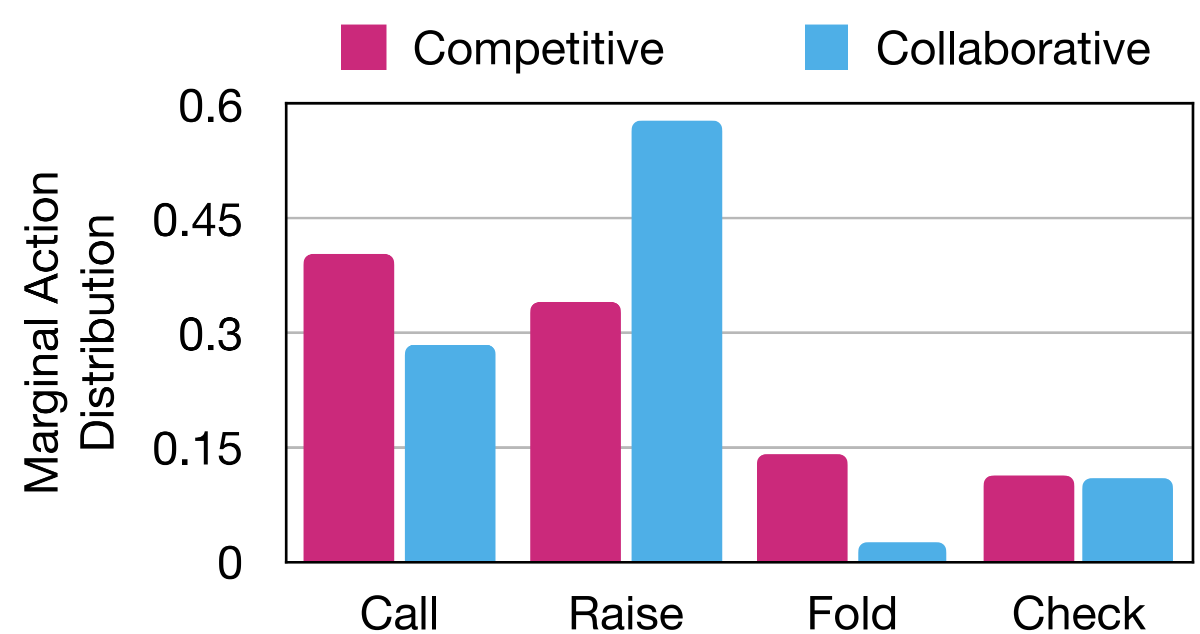
Action Distribution: $p(B_1, B_2)$



Probability of Action Similarity among the Brains

$$p(B_1 == B_2) = \frac{\sum_{i=j} A_{i,j}}{\sum A_{i,j}}$$

= 0.535



Discriminator: Detecting who is cheating

Discriminator

Setup and Challenges

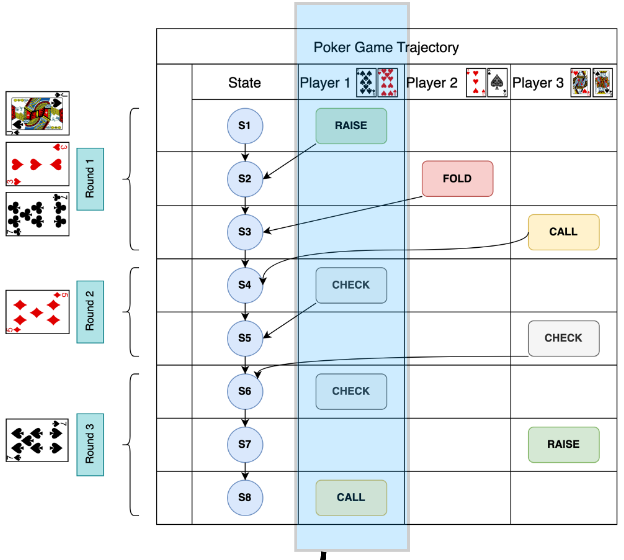
- $f_{B_1}(s) \rightarrow a$

- $f_{B_2}(s) \rightarrow a$

- Action Space is small

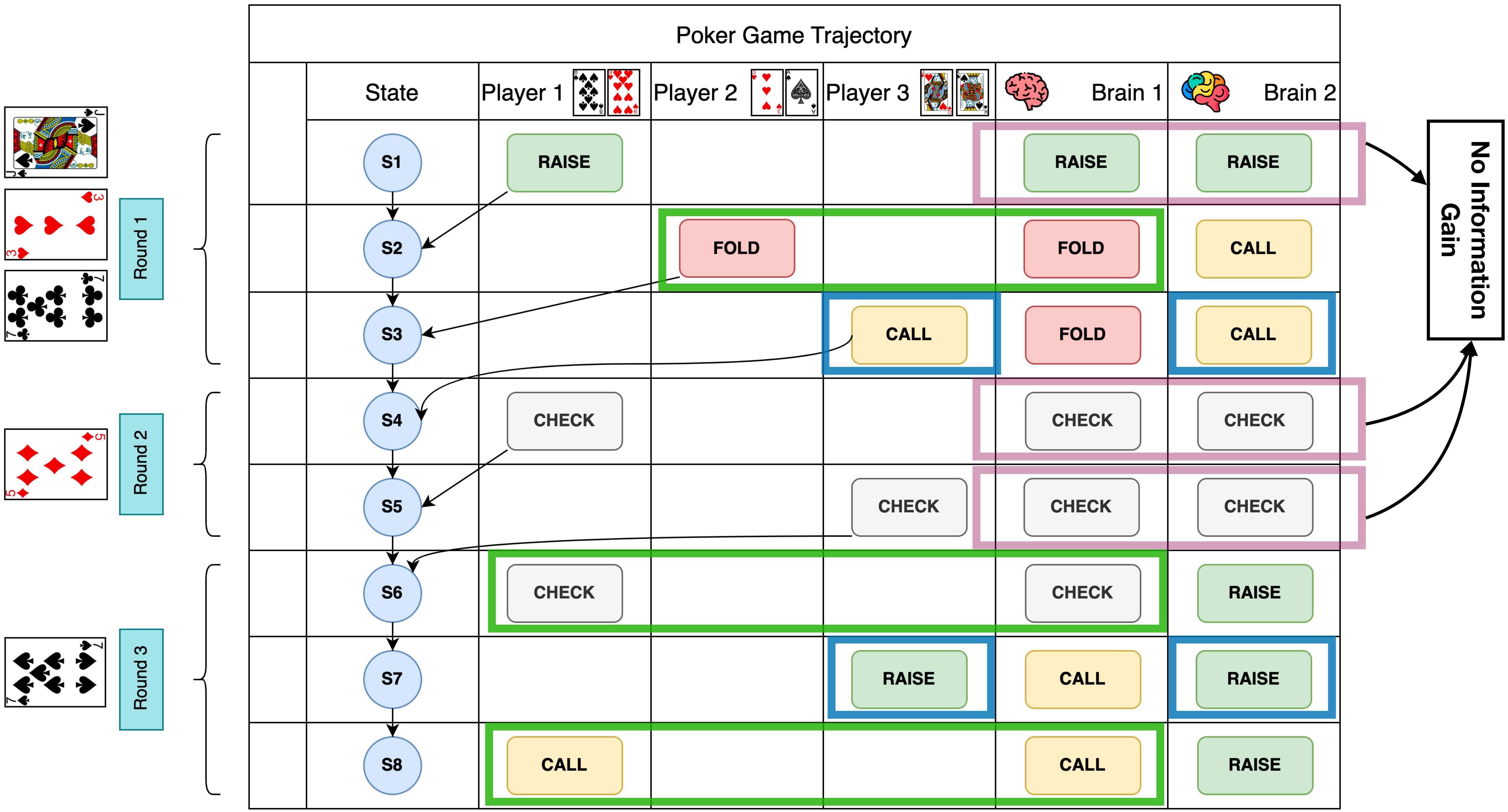
- State Space is intractable

$$p(\{B_1, B_2\} \mid \tau)$$



Can we take advantage of small action space to create a discriminator which doesn't enumerate the state space?

Discriminator Logic



Match B1	2	1	0
Match B2	0	0	2

Takeaways

- Pro:
- Con: Our discriminator assumes that in real life scenarios, agents who are cheating deploys strategies similar to strategies learned by our policy. However, this assumption can be broken at times.

References

1. Bowling, M., Burch, N., Johanson, M., & Tammelin, O. (2015). Heads-up limit hold'em poker is solved. *Science*, *347*(6218), 145–149. https://doi.org/10.1126/science.1259433
2. Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., & Whiteson, S. (2017). *Counterfactual Multi-Agent Policy Gradients* (arXiv:1705.08926). arXiv. https://doi.org/10.48550/arXiv.1705.08926
3. Heinrich, J., Lanctot, M., & Silver, D. (2015). Fictitious self-play in extensive-form games. Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, 805–813. http://proceedings.mlr.press/v37/heinrich15.pdf
4. Zhang, K., Yang, Z., & Başar, T. (2021). Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms (arXiv:1911.10635). arXiv. http://arxiv.org/abs/1911.10635
5. Icons images from FlatIcon.