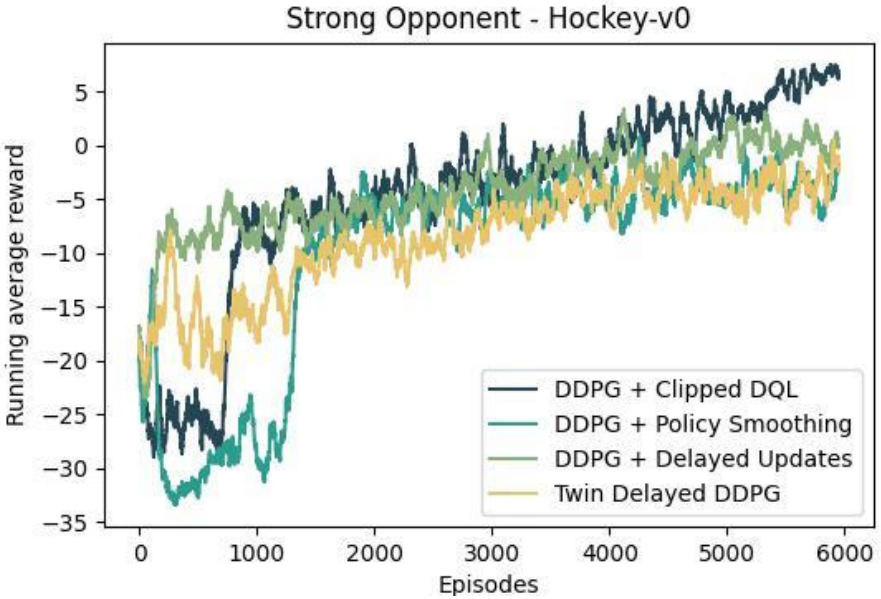
Reinforcement Learning Project Evaluation

Electric Dream Machine

Implemented Algorithms

- D4PG
- TD3
- DDQN

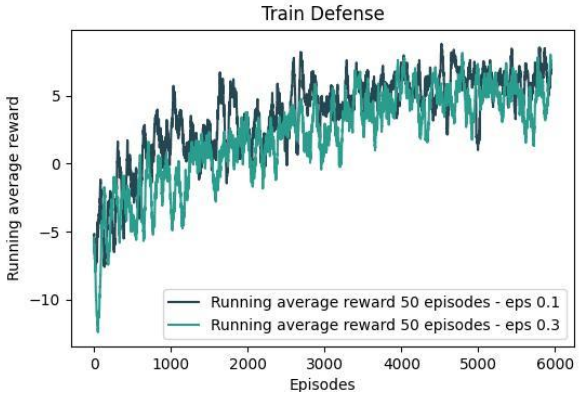
Twin Delayed DDPG (TD3)

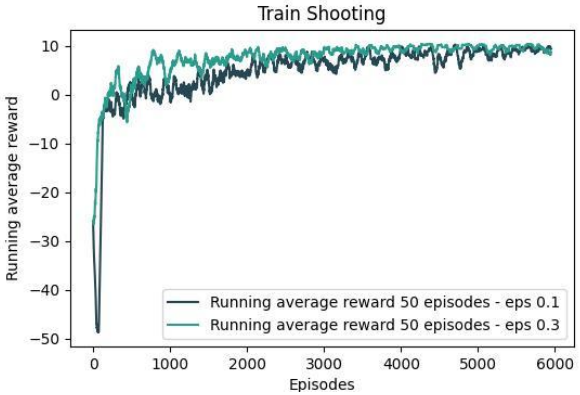


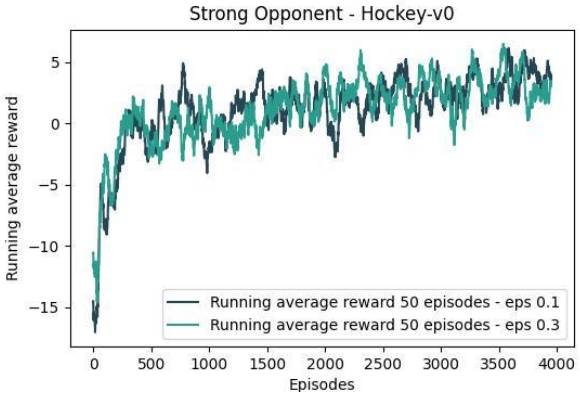
Adjustments based on DDPG implementation

- Clipped Double Q-Learning
- Policy Smoothing
- Delayed Target and Policy Updates

Training on Hockey environments (TD3)



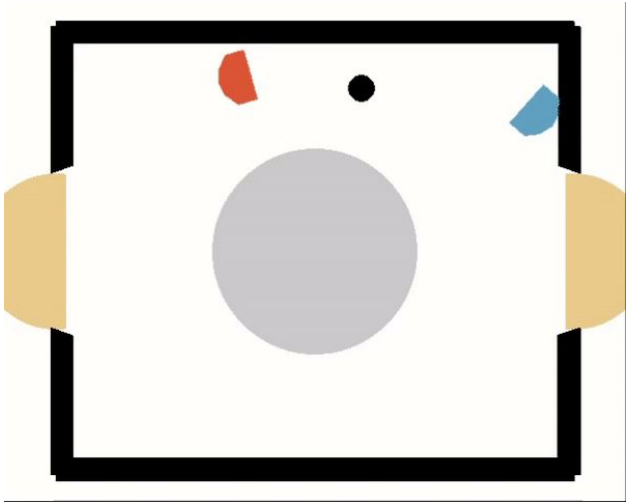




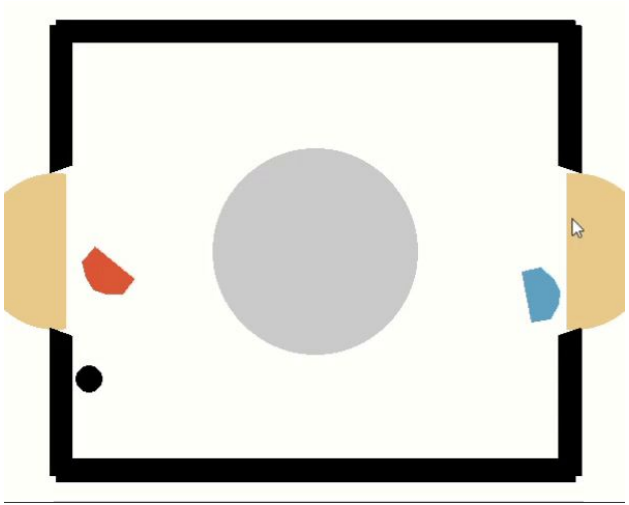
Models were trained respectively for their environment

Trained Model against Strong Opponent (TD3)

Gameplay against Strong Opponent



Gameplay against Strong Opponent



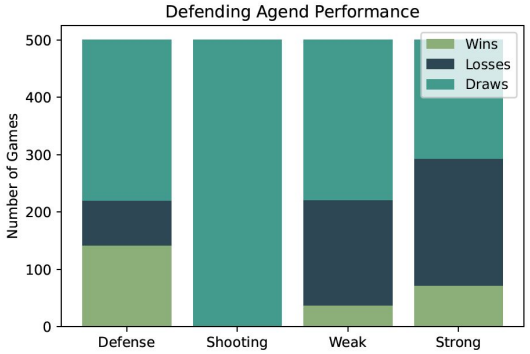
Dueling Double DQN (D3QN) with PER

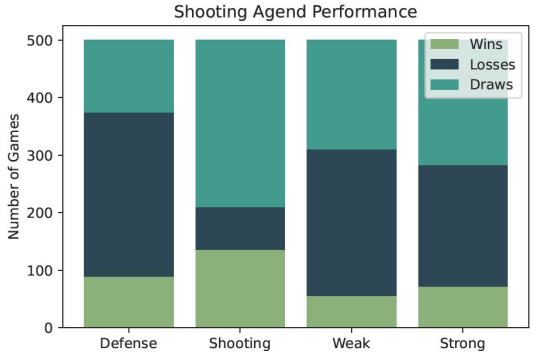
Rewards achieved during Training in all 4 Modes of the Hockey-v0 environment

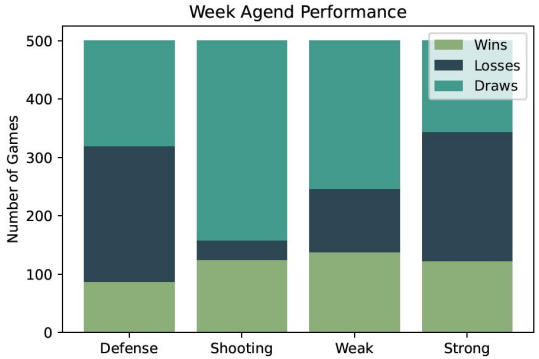
Adjustments based on DQN implementation

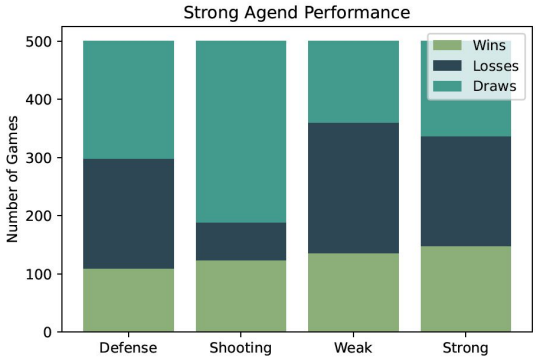
- Dueling Network Architecture
- Double Q Learning
- Prioritized Experience Replay

Dueling Double DQN (D3QN) with PER







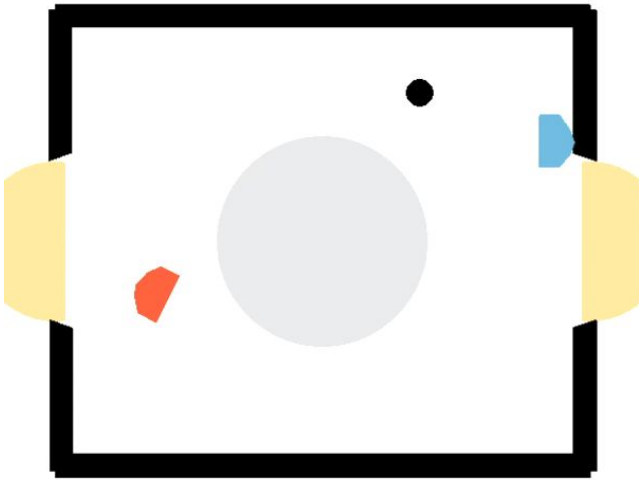


Training:

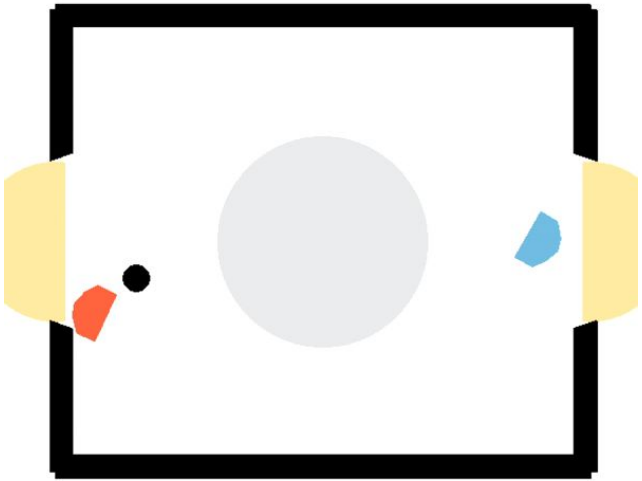
- Mixed Training lead to buggy behavior and poor performance
- Analysis: Agents trained only on one specific Mode but evaluated in each
- Winner: **Strong Agent**

Dueling Double DQN (D3QN) with PER

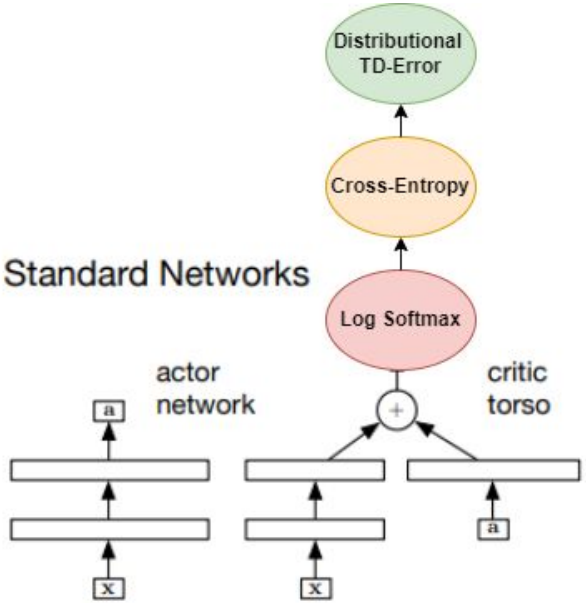
Mixed training on all 4 modes



Only trained on strong opponent



Distributed Distributional DDPG (D4PG)



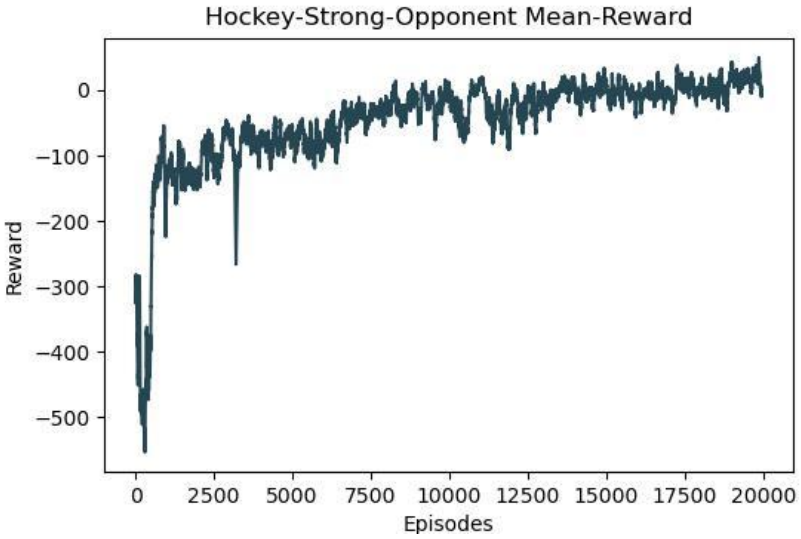
Adjustments based on DDPG implementation

- Critic Output & Learning
- Distributional Updates
- Hard Target Updates

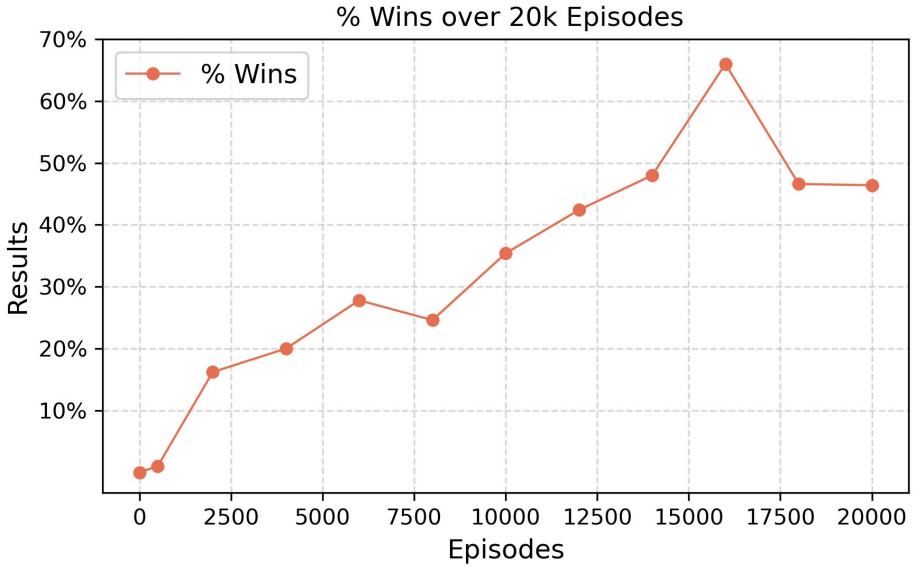
Barth-Maron, G., Hoffman, M.W., Budden, D., Dabney, W., Horgan, D., Tb, D., Muldal, A., Heess, N. and Lillicrap, T., 2018. Distributed distributional deterministic policy gradients. *arXiv preprint arXiv:1804.08617*.

Distributional DDPG (D4PG)

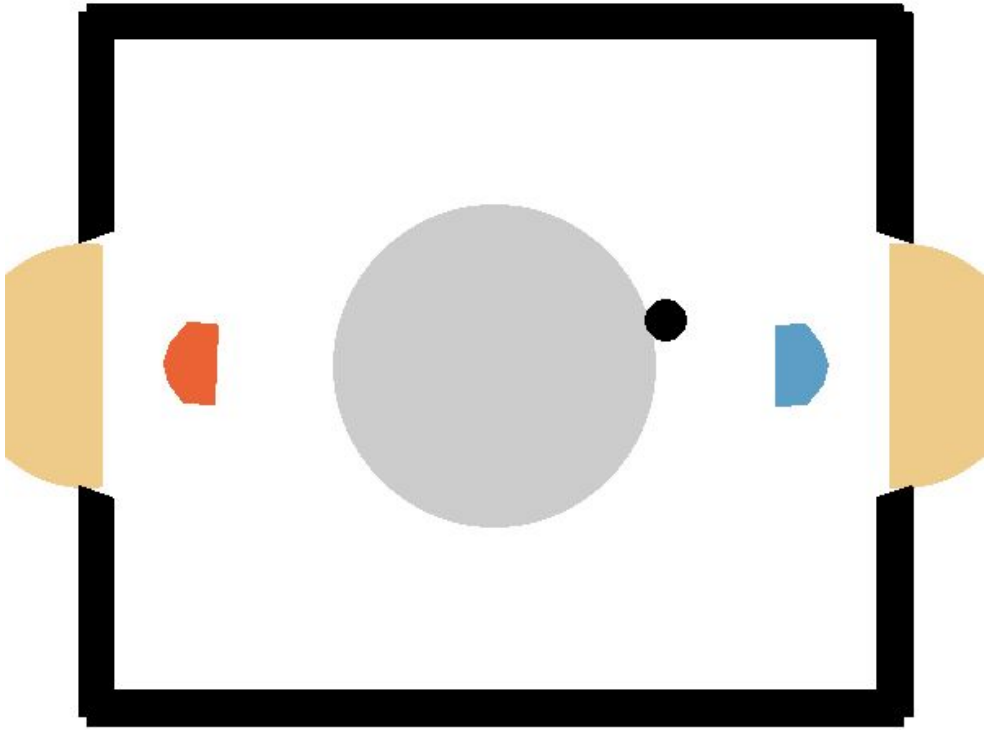
Reward obtained over time:



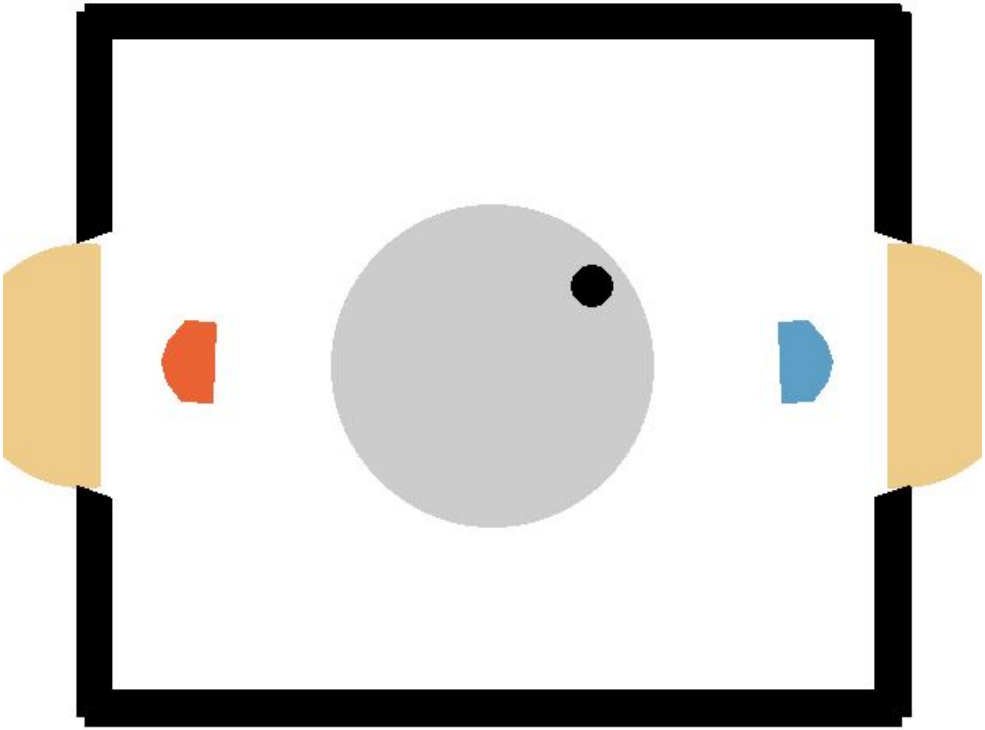
Win Percentage at different points in time or 500 eval episodes:



Example Attack Agent



Example Defense Agent



Lessons Learned

- Training against one Opponent not sufficient
- For complex environments many episodes needed to yield good results