Chase Tag Game(MARL)

CSE 510 Reinforcement Learning
Final Project

Gaurav Vijay Vergiya

UB ID gvijayve UBNO:50344611

Chase Tag Game

https://gfycat.com/reflectinguntidyatlanticblackgoby-professional-tag-world-chase-tag

Introduction

In this project we aim at exploring MARL in an competitive environment like Chase Tag and the challenges faced in Training a competitive MARL setup where the agents have different learning potential. We use A2C algorithm to train our agents

The project explores various issues while training a competitive MARL setup where one agent has an inherent advantage over the other, like cold start problem, batch training to avoid overpowering the agent with advantage etc.

Introduction

In the game teams score points for every successful evasion and the rate of evasion is generally low as the chaser has to just keep running towards the runner but the runner who tries to run away from the chaser can’t continue to do that as he would hit the boundary and at my instances in the game the runner has to take a step towards the chaser or be stationary (which is against it objective) the while the chaser is already moving towards the runner hence by the concept of relative velocity two point moving towards each other reduces distances between them with a factor of 2 causing runner to have inherent disadvantage and the challenges of training such a environment with MARL are explored in this project

Chase Tag Game

Just like the rules of the real chase tag game In chase tag game simulation we have a fixed space with 2 agents and time limit of 20 steps

1. Runner Agent: The goal of the agent is to avoid the or escape the chaser for 20 time steps
2. Chaser Agent: The goal of the agent is to catch the Runner within 20 steps

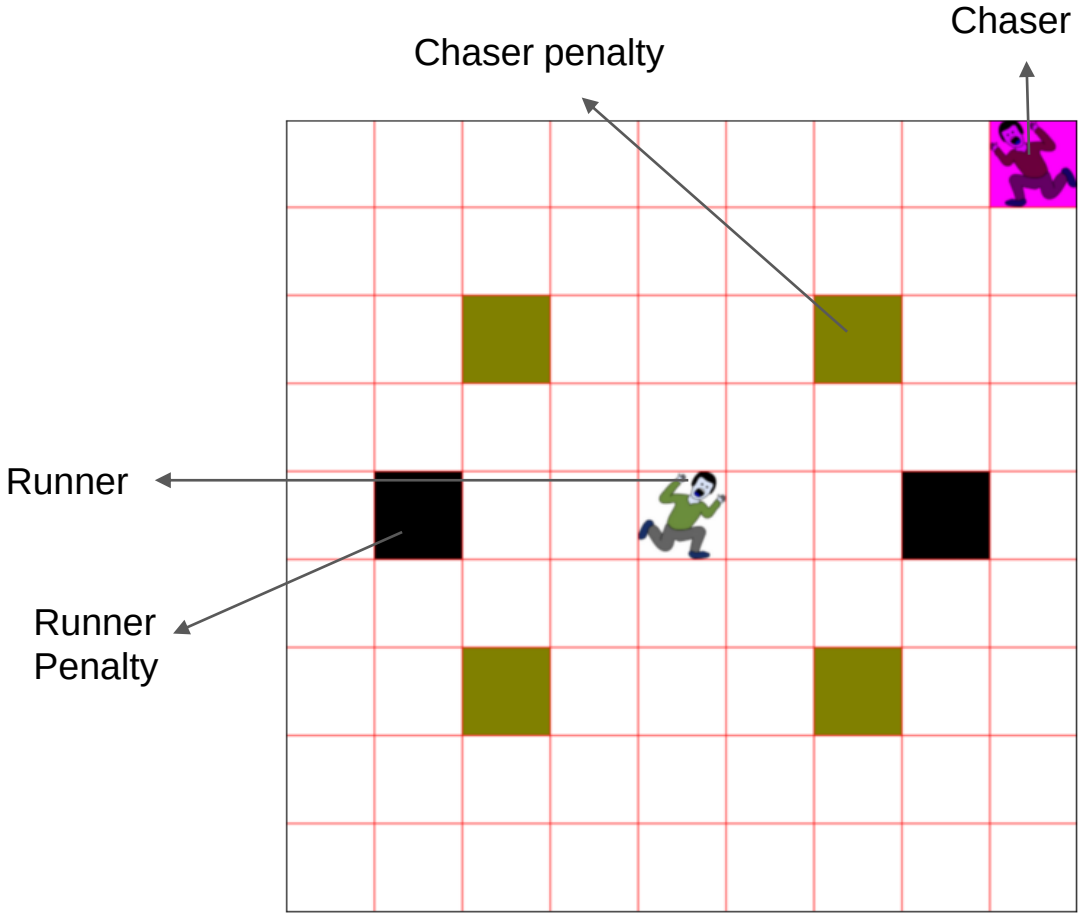
Environment

The arena is a 9x9 size square

Chaser starts at any of the four corners of the Square

Runner always starts from the center

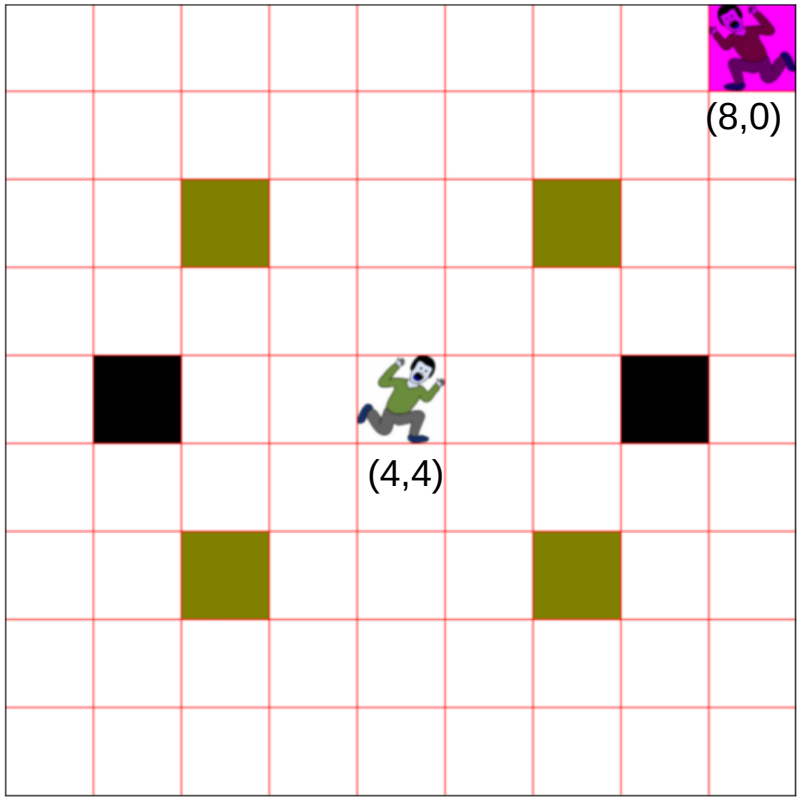
Game ends when the chaser catches the Runner(distance is one unit) or 20 timesteps are done



State/Observation

State is defined as :

(X_runner,Y_runner,X_chaser,Y_chaser)

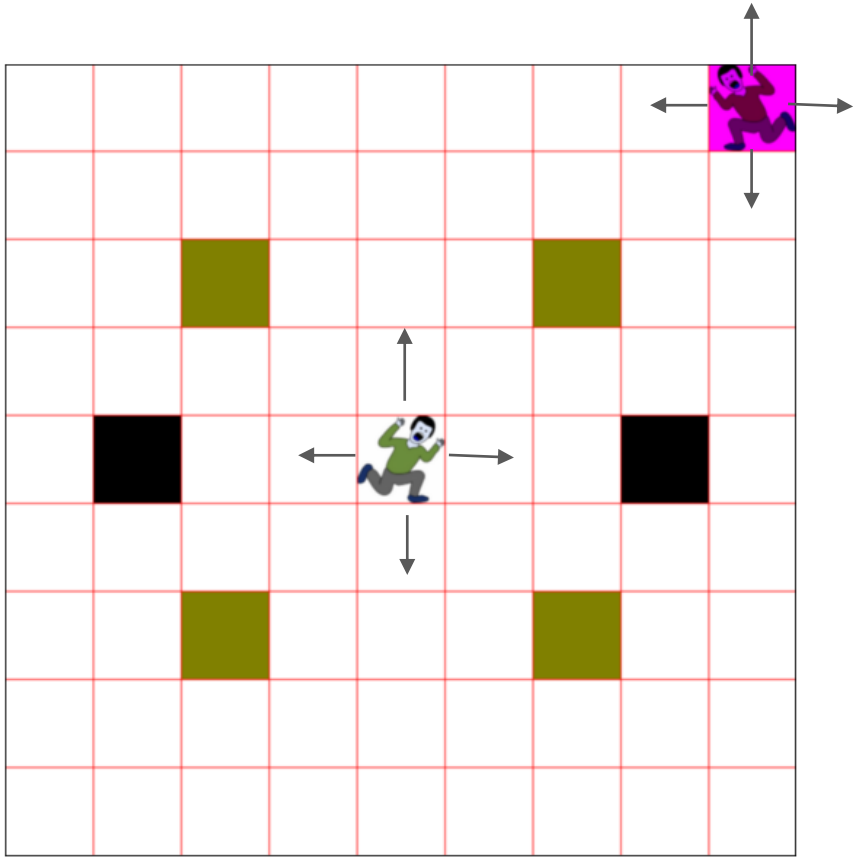


State = (4,4,8,0)

Action Space

Both the Runner and Chaser can move in 4 directions up, down, left, right

Agent can’t move outside the arena on hitting the boundary agent doesn’t change its position



Reward Dynamics: Runner

The runner gets a reward of 0.2 for moving away from the chaser and reward of 0 for moving towards the chaser

The runner gets a penalty of -2 for running into the boundary

The runner gets a penalty of -1 for moving into a penalty spot

The runner gets a penalty of -5 on losing (chaser catches the runner)

The Runner gets 50 point for winning (escaping the chaser for 20 timesteps)

Reward Dynamics: Chaser

The chaser gets a reward of -1 for moving away from the runner and reward of 0 for moving towards the runner

The chaser gets a penalty of -2 for running into the boundary

The chaser gets a penalty of -1 for moving into a penalty spot

The chaser gets 5 point for winning (catching the runner within 20 timesteps)

Algorithm: Advantage Actor Critic (A2C)

Actor: Dense neural network with 3 hidden layers of size 128, 256 and 128 respectively with output layer with 4 nodes with softmax to predict the probabilities of actions

Critic: Dense neural network with two hidden layers of size 256,256 respectively with single node in output layer to predict the value of the state

Both the runner and chaser agents use the same architecture

learning_rate = 0.0001

gamma = 0.9

max_episodes = 300000

Training Process

As we have two agents to train and the behaviour of each agents depends on the previous actions of the other agents specially for the runner because unlike the chaser which has to always move towards the runner, the runner just can’t keep running away from the chaser as it would hit the boundary and hence it has to at few moments decide move closer to chaser or wait at the corner.

On starting to train the Runner against a random chaser it essentially learns nothing as the random chaser doesn’t challenge the runner and hence it also does random movements and wins a lot of the games as its difficult to catch the runner within 20 timesteps with random movements and thus the runner agent learns to not move towards any objective as it wins anyways

Training Process

Plus we know it’s easy to train a chaser as it has an easy objective to learn when compared to a runner

So we know we have to challenge the runner for it to learn and also its easier to train the chaser we go for a strategy where we train the chaser for 200 steps and fix the chaser and train the runner for next 1800 steps and we only train the agents if they fall below their reward thresholds which are not set to very ambitious values for any agent but to intendent to reach an equilibrium.

Once we train the agents in this fashion we avoid the cold start problem and we now have very basic agents which still follow their respective objective

Training Process

Now the we have two basic agents which we call runner_easy and chaser_easy and we train these agents against each other, i.e fixing one agent and improving the other agent and this time we set our objectives to be slightly ambitious as we intent to train the agents to perform well at their objectives and not towards an equilibrium. Hence we now have two agents which are better than the easy version of them and have started to learn and improve towards their objectives.Let’s call them runner_medium and chaser_medium.

We repeat the same process with slightly more ambitious reward thresholds and generate runner_hard,chaser_hard and runner_veryhard,chaser_veryhard.

This style of training was important because for runner/chaser to improve we need to give it new challenges and make them learn to avoid this challenges with well calibrated thresholds as both the agents have different learning capabilities and chaser can easily learn to over power the runner and potentially hamper the the possible learning opportunities for the runner.

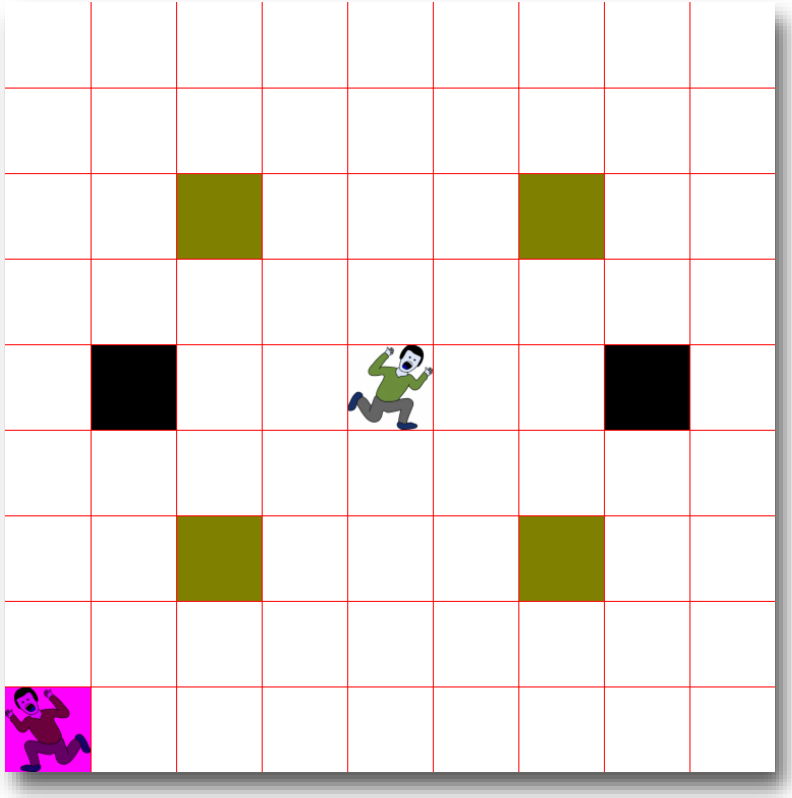
Results

After 28045 episodes of training both the runner and chaser in the fashion as mentioned before it reached the equilibrium. The threshold for runner was 35(decent reward for which ensure agent learns to win) and the threshold for chaser was -1 (not overpowering the chaser)

Runner Easy vs Chaser Easy

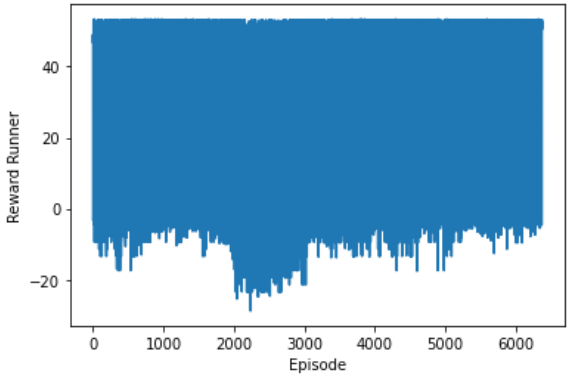
Round-0 ##chaser won
Round-1 ##chaser won
Round-2 ##Runner won
Round-3 ##chaser won
Round-4 ##chaser won
Round-5 ##chaser won
Round-6 ##Runner won
Round-7 ##chaser won
Round-8 ##Runner won
Round-9 ##Runner won
Round-10 ##Runner won

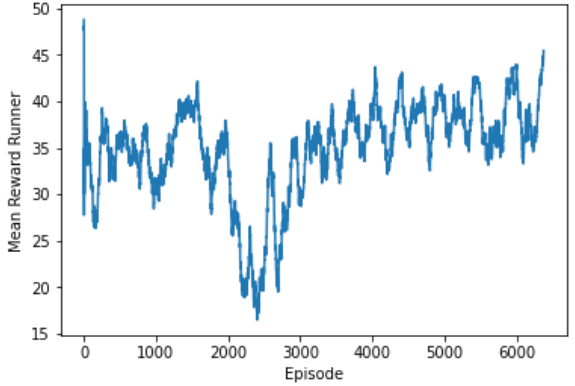
final scores runner: 5, chaser: 6

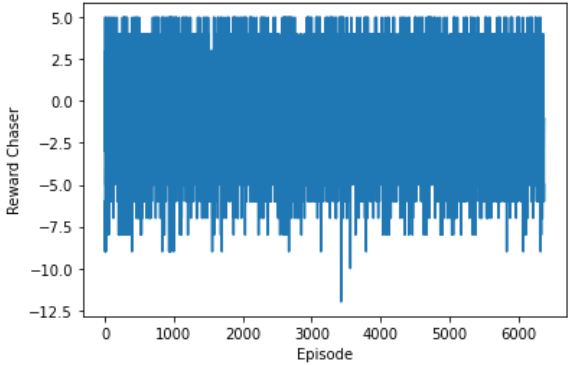


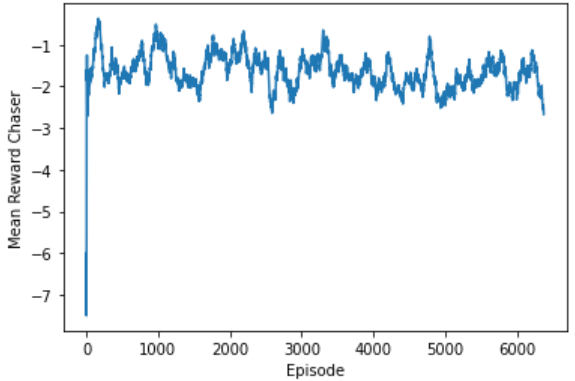
Runner trained with Easy Chaser---> Medium Runner

We train the Runner Easy agent against the chaser Easy agent with a reward threshold for run set to 45
Which is met after 6374 episodes



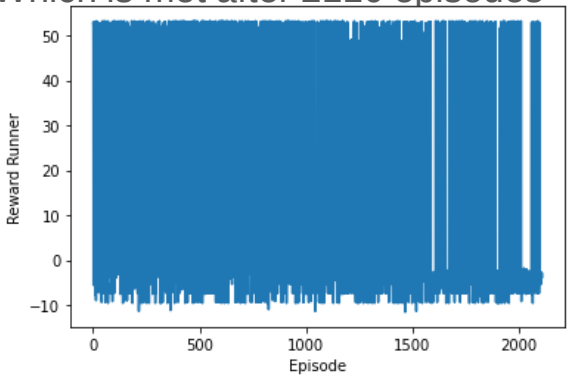


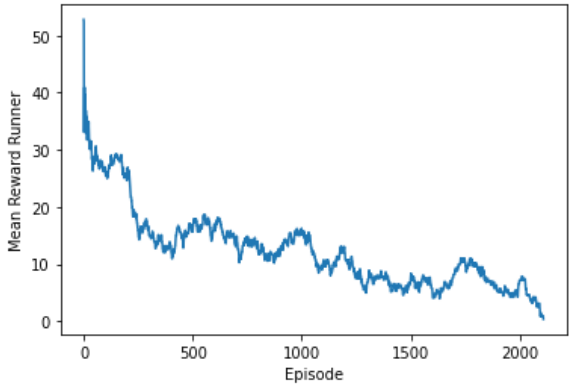


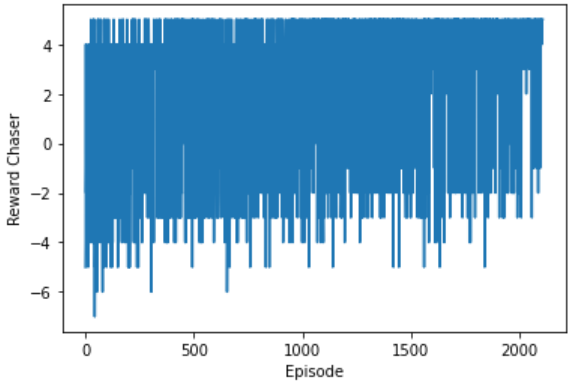


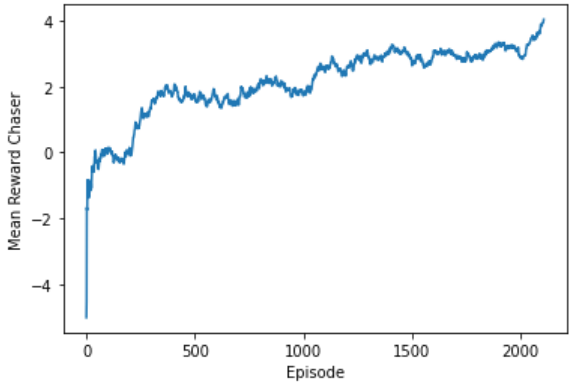
Chaser trained with Easy Runner ---> Medium chaser

We train the Chaser Easy agent against the Runner Easy agent with a reward threshold for run set to 4
Which is met after 2110 episodes





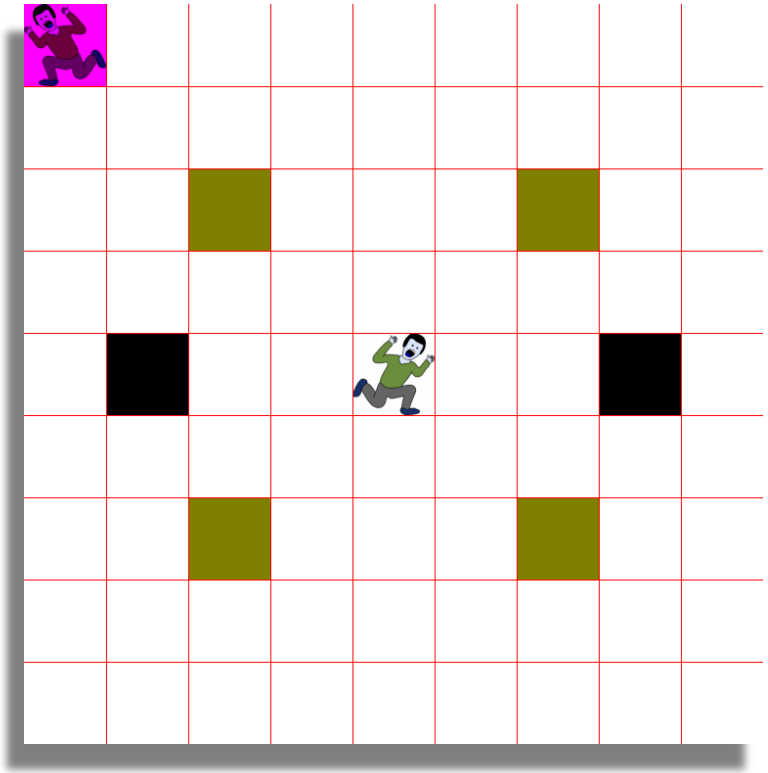




Runner Medium vs Chaser Easy

Round-0 ##Runner won
Round-1 ##chaser won
Round-2 ##Runner won
Round-3 ##Runner won
Round-4 ##Runner won
Round-5 ##Runner won
Round-6 ##chaser won
Round-7 ##Runner won
Round-8 ##Runner won
Round-9 ##Runner won
Round-10 ##Runner won

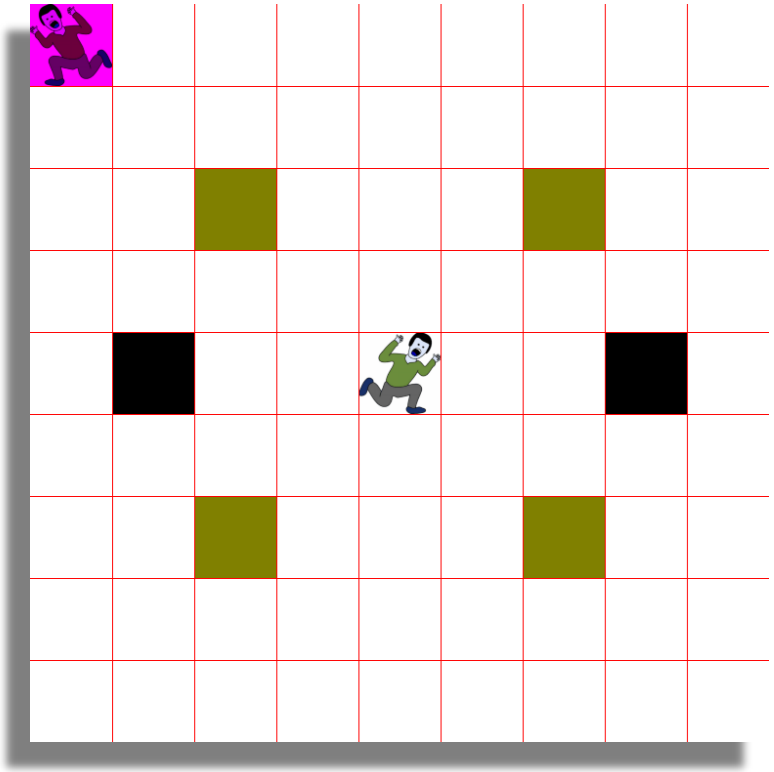
final scores runner: 9, chaser: 2



Runner Easy vs Chaser Medium

Round-0 ##chaser won
Round-1 ##chaser won
Round-2 ##chaser won
Round-3 ##chaser won
Round-4 ##chaser won
Round-5 ##chaser won
Round-6 ##Runner won
Round-7 ##chaser won
Round-8 ##chaser won
Round-9 ##chaser won
Round-10 ##chaser won

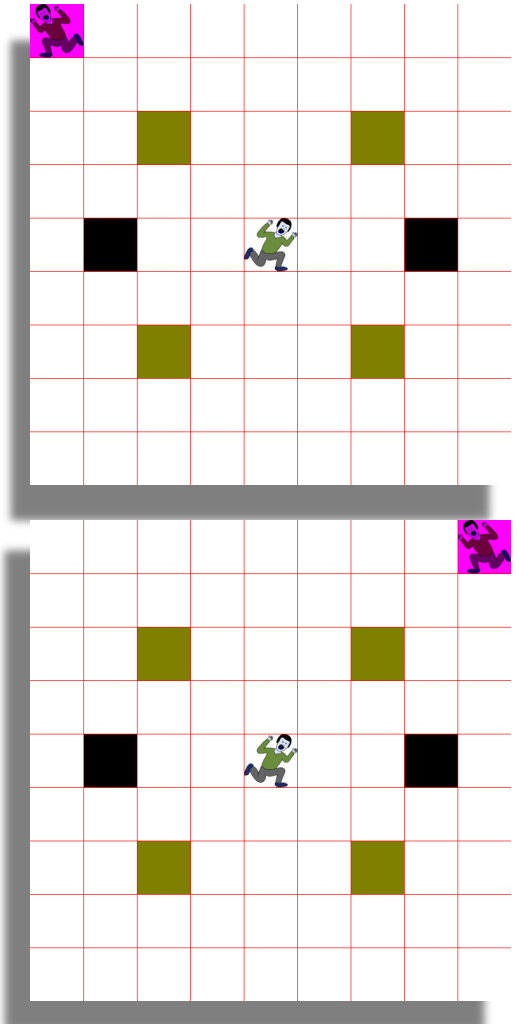
final scores runner: 1, chaser: 10



Runner Medium vs Chaser Medium

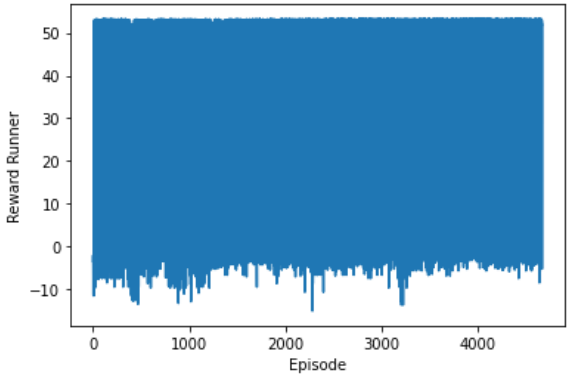
Round-0 ##chaser won
Round-1 ##chaser won
Round-2 ##chaser won
Round-3 ##chaser won
Round-4 ##Runner won
Round-5 ##chaser won
Round-6 ##Runner won
Round-7 ##chaser won
Round-8 ##chaser won
Round-9 ##chaser won
Round-10 ##chaser won

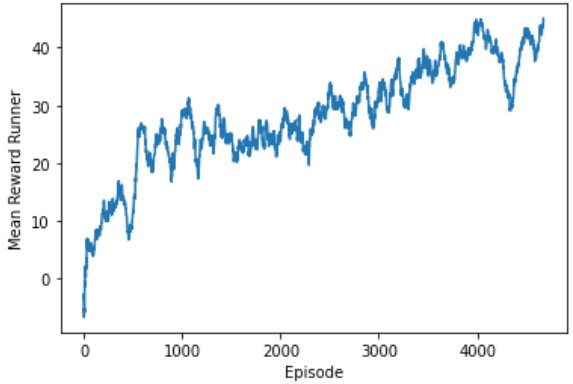
final scores runner: 2, chaser: 9

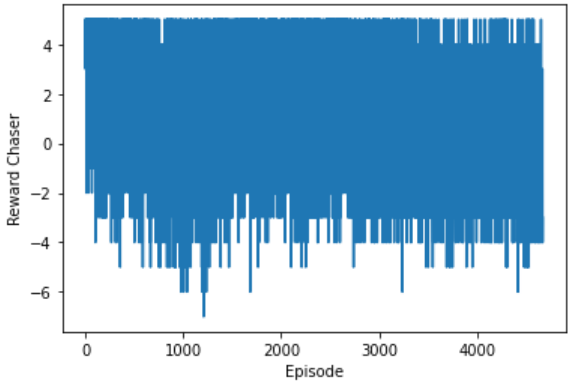


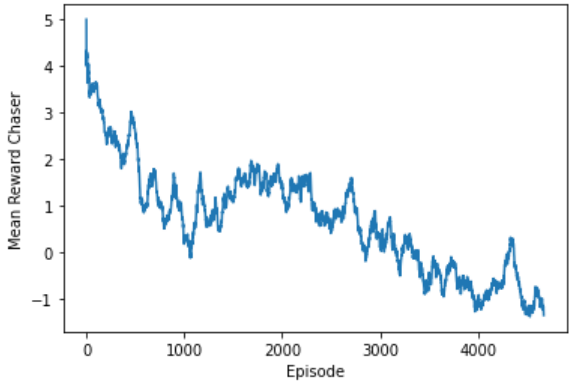
Runner trained with Medium Chaser ---> Hard Runner

We train the Runner Medium agent against the chaser Medium agent with a reward threshold for run set to 45 which is met after 4674 episodes





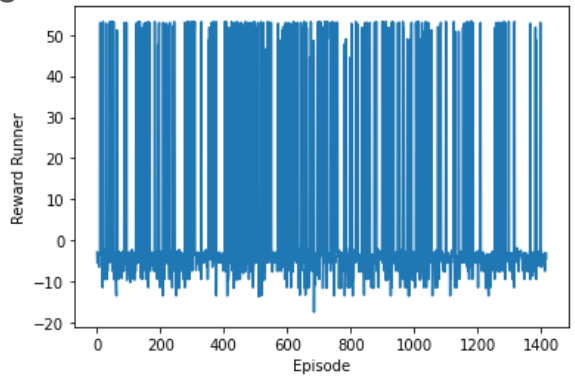


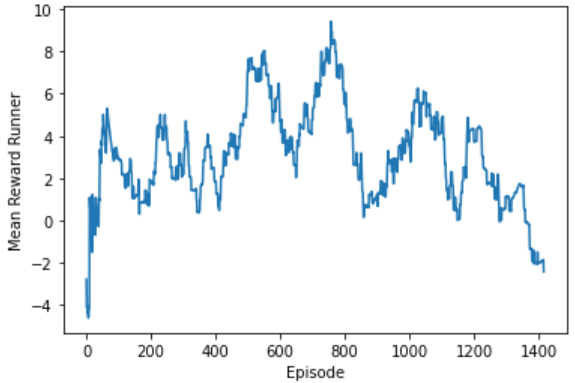


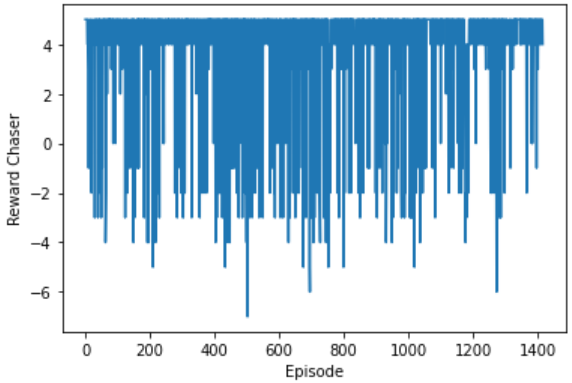
Chaser trained with Medium Runner ---> Hard Chaser

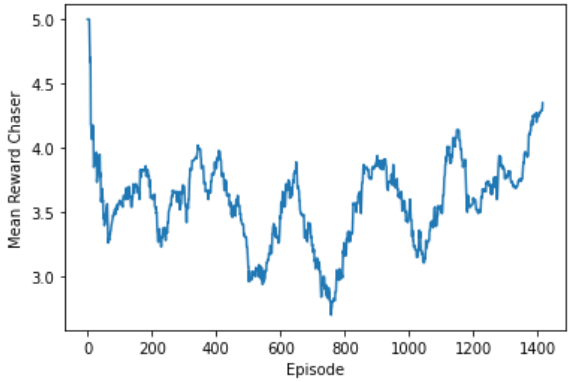
We train the Chaser Medium agent against the Runner Medium agent with a reward threshold for run set to 4.3 Which is met after 1418 episodes

3





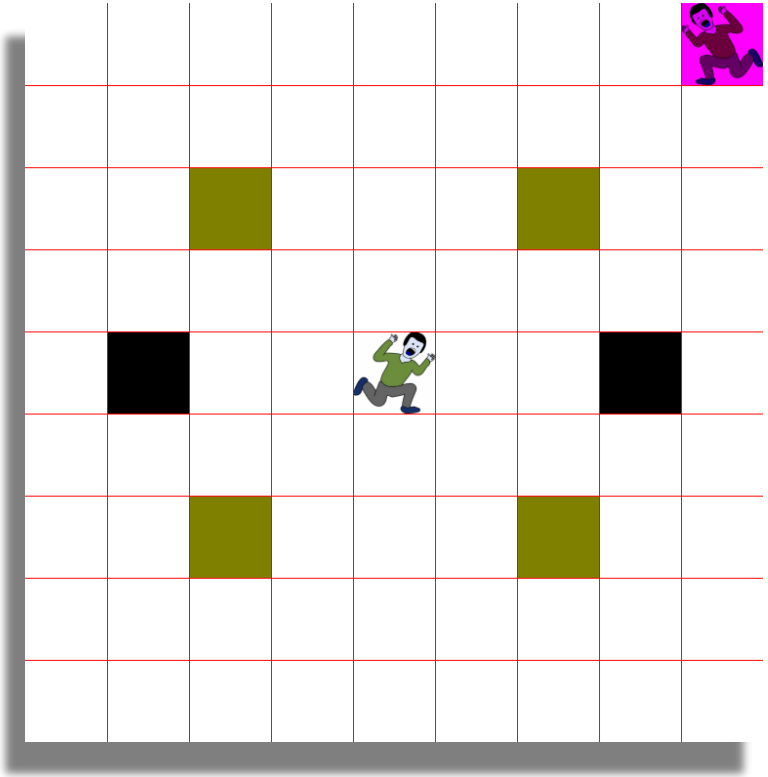




Runner Hard vs Chaser Medium

Round-0 ##Runner won
Round-1 ##Runner won
Round-2 ##Runner won
Round-3 ##Runner won
Round-4 ##Runner won
Round-5 ##Runner won
Round-6 ##Runner won
Round-7 ##chaser won
Round-8 ##Runner won
Round-9 ##Runner won
Round-10 ##chaser won

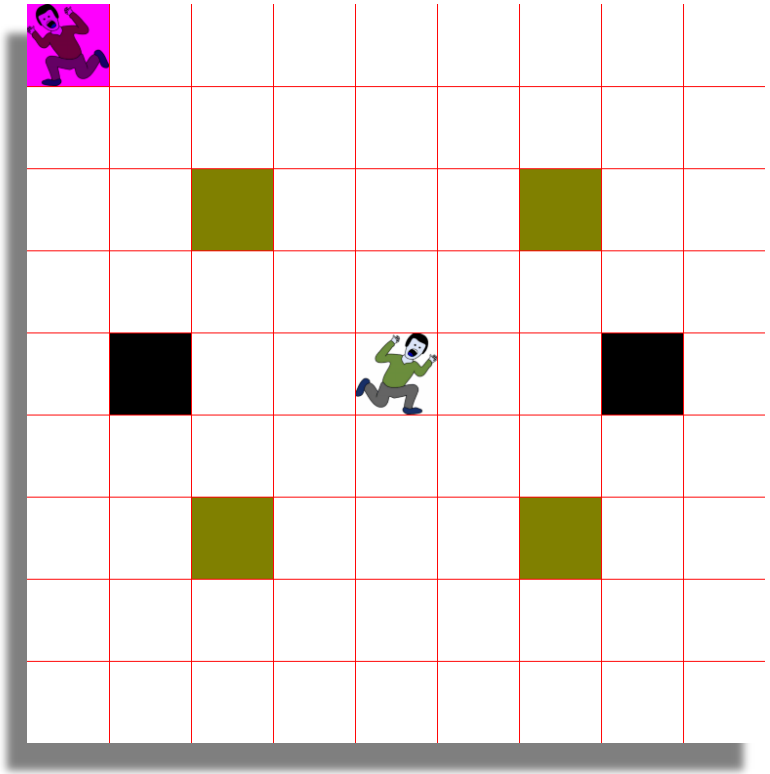
final scores runner: 9, chaser: 2



Runner Medium vs Chaser Hard

Round-0 ##chaser won
Round-1 ##chaser won
Round-2 ##chaser won
Round-3 ##chaser won
Round-4 ##chaser won
Round-5 ##chaser won
Round-6 ##chaser won
Round-7 ##chaser won
Round-8 ##chaser won
Round-9 ##Runner won
Round-10 ##chaser won

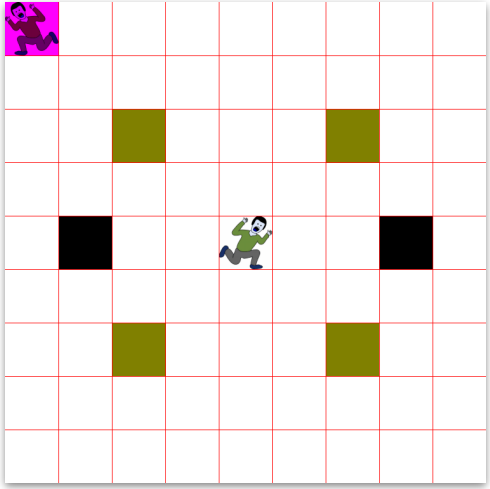
final scores runner: 1, chaser: 10

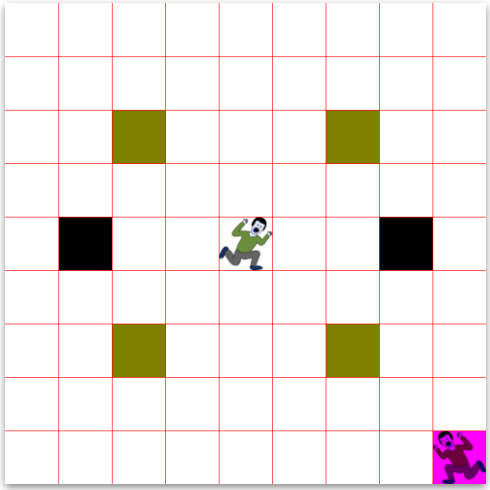


Runner Hard vs Chaser Hard

Round-0 ##Runner won
Round-1 ##Runner won
Round-2 ##Runner won
Round-3 ##Runner won
Round-4 ##Runner won
Round-5 ##chaser won
Round-6 ##Runner won
Round-7 ##Runner won
Round-8 ##chaser won
Round-9 ##chaser won
Round-10 ##Runner won

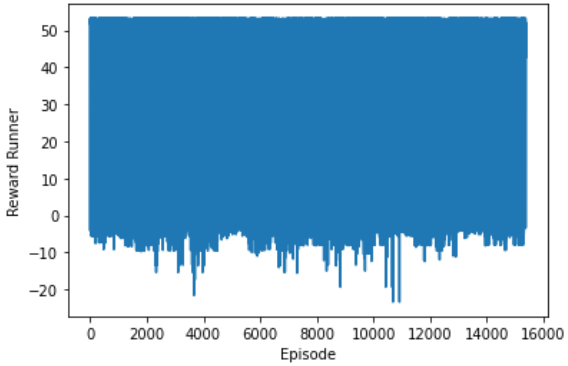
Final scores runner: 8, chaser: 3

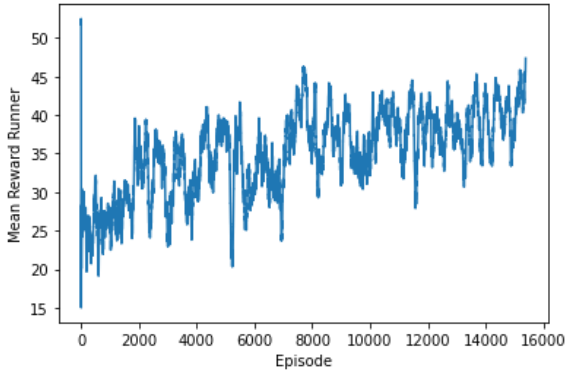


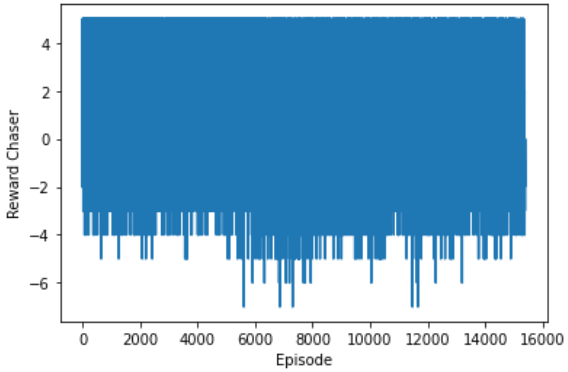


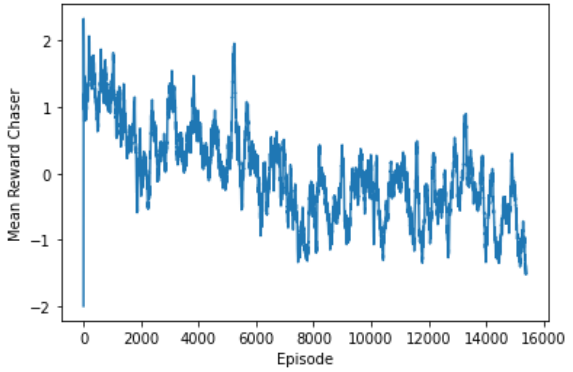
Runner trained with Hard Chaser ---> Very Hard Runner

We train the Runner Hard agent against the chaser Hard agent with a reward threshold for run set to 47 which is met after 15387 episodes



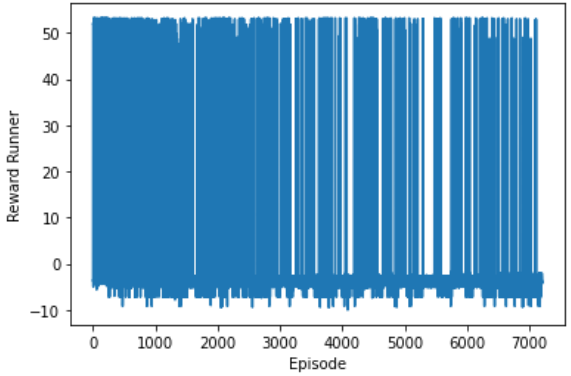


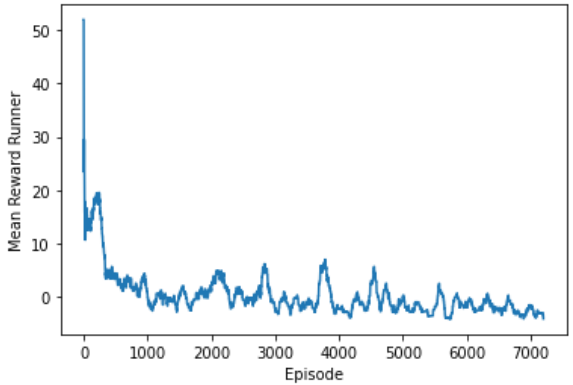


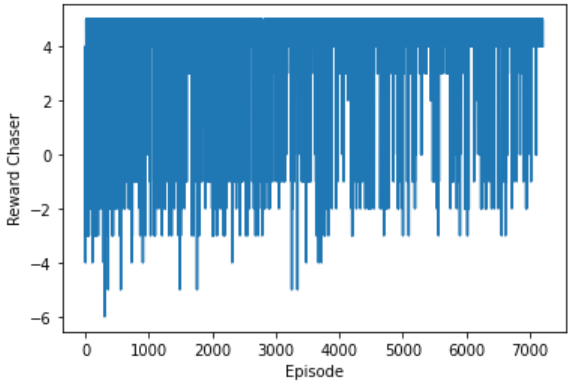


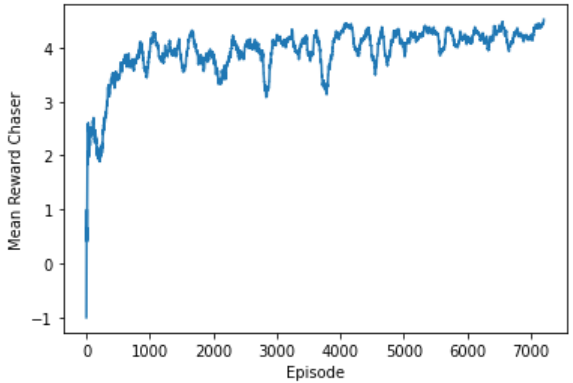
Chaser trained with Hard Runner ---> Very Hard Chaser

We train the Chaser Hard agent against the Runner Hard agent with a reward threshold for run set to 4.5
Which is met after 7207 episodes





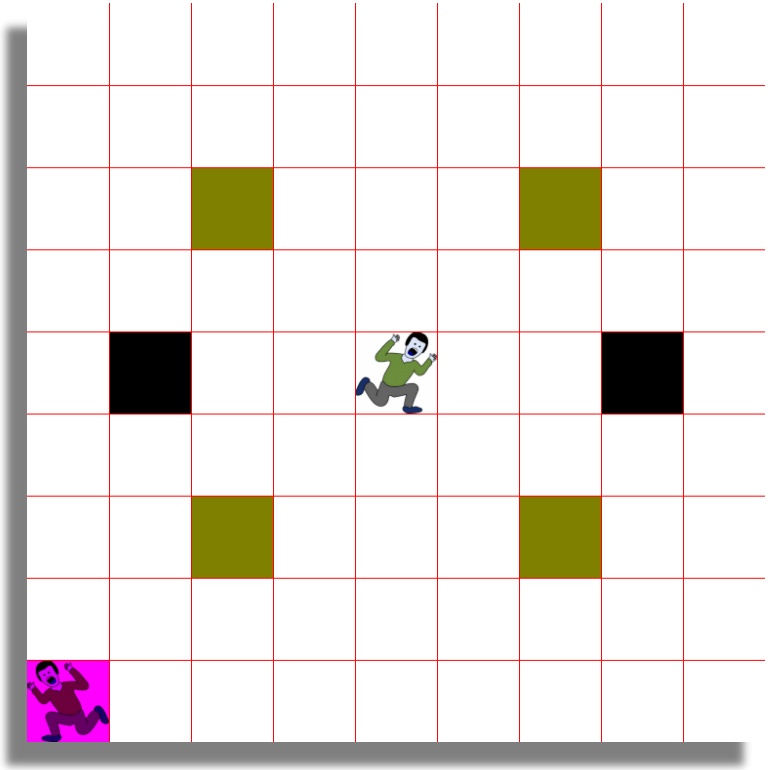




Runner VeryHard vs Chaser Hard

Round-0 ##Runner won
Round-1 ##Runner won
Round-2 ##Runner won
Round-3 ##Runner won
Round-4 ##Runner won
Round-5 ##Runner won
Round-6 ##Runner won
Round-7 ##Runner won
Round-8 ##Runner won
Round-9 ##Runner won
Round-10 ##Runner won

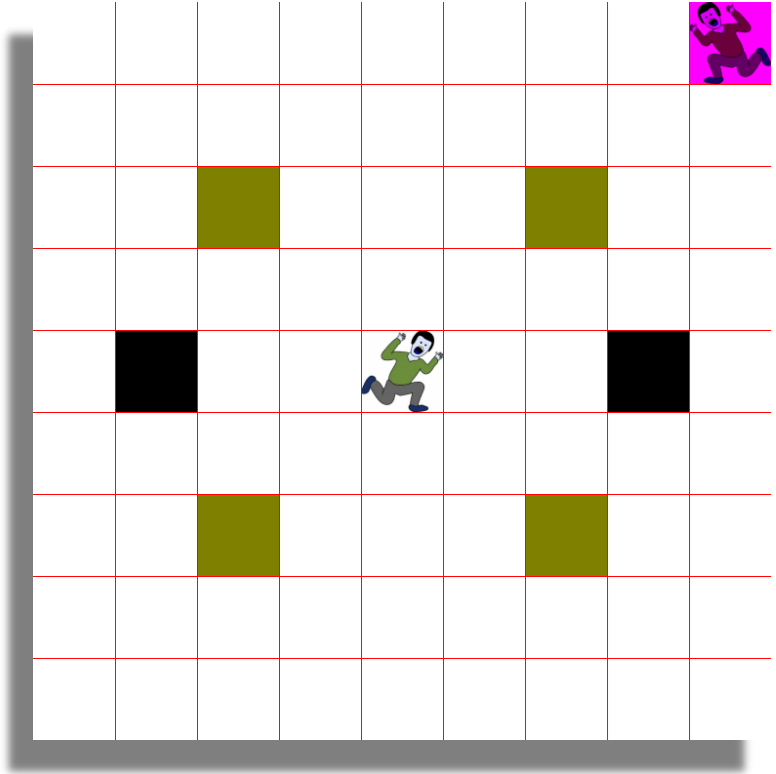
final scores runner: 11, chaser: 0



Runner Hard vs Chaser VeryHard

Round-0 ##chaser won
Round-1 ##chaser won
Round-2 ##chaser won
Round-3 ##chaser won
Round-4 ##chaser won
Round-5 ##chaser won
Round-6 ##chaser won
Round-7 ##chaser won
Round-8 ##chaser won
Round-9 ##chaser won
Round-10 ##chaser won

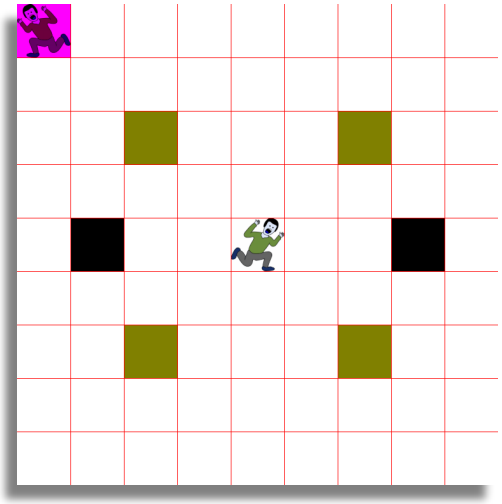
final scores runner: 0, chaser: 11

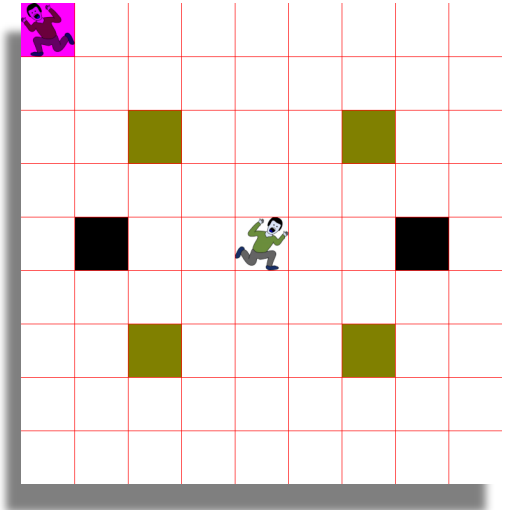


Runner VeryHard vs Chaser VeryHard

Round-0 ##chaser won
Round-1 ##chaser won
Round-2 ##chaser won
Round-3 ##chaser won
Round-4 ##chaser won
Round-5 ##chaser won
Round-6 ##chaser won
Round-7 ##chaser won
Round-8 ##chaser won
Round-9 ##chaser won
Round-10 ##chaser won

final scores runner: 0, chaser: 11





Conclusion

It is challenging to train a Competitive MARl model but its get even more difficult when we have agent with differences like inherent advantages and disadvantages we have to be extra careful while training the MARL setup as empowering one agent can leave the other agent with almost no opportunity to learn and hence a balance on improving both at different rates is important as both learn there objectives and also learn to avoid mistakes.

Thank You