PROYECTO JAVIVI

Grupo 6

QUIÉNES SOMOS

JAVIER CORREA MARICHAL

VIREN SAJJU DHANWANI DHANWANI

GABRIEL GARCÍA JAUBERT

TABLA DE CONTENIDOS

01 Propuesta

02 Escenario sencillo

03 Escenario con obstáculos

04 Escenario colaborativo

01

PROPUESTA

ESCENARIOS

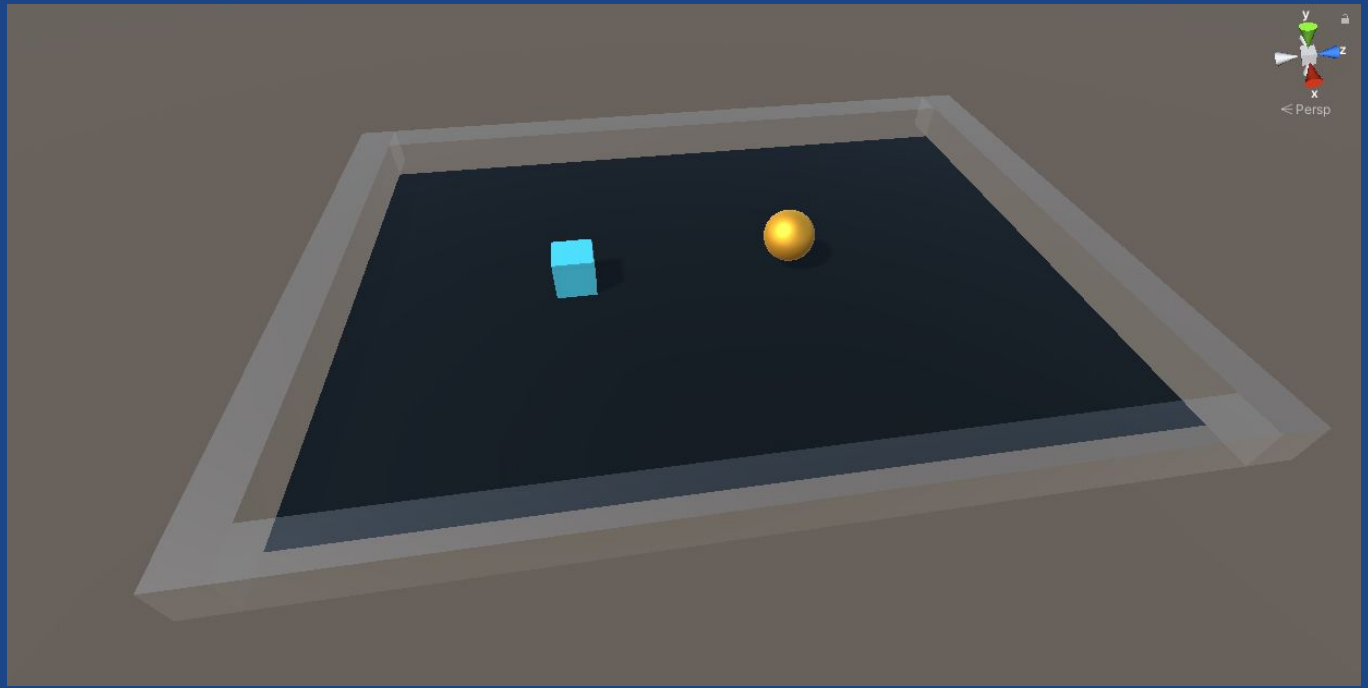
3 escenarios

- Llegar a un objetivo simple
- Esquivar obstáculos
- Colaborar

02

ESCENARIO SENCILLO

OBJETIVO



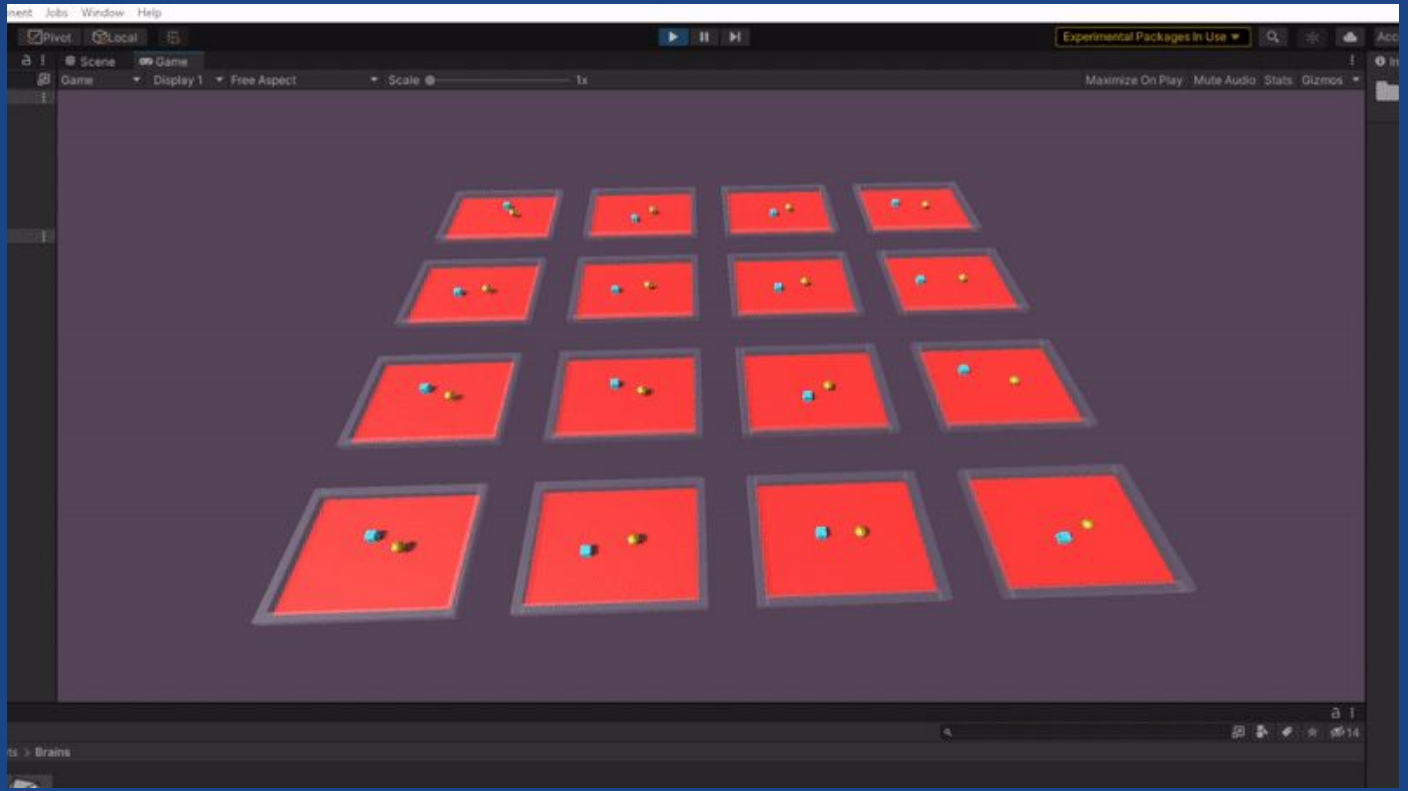
DESARROLLO

Posición del agente

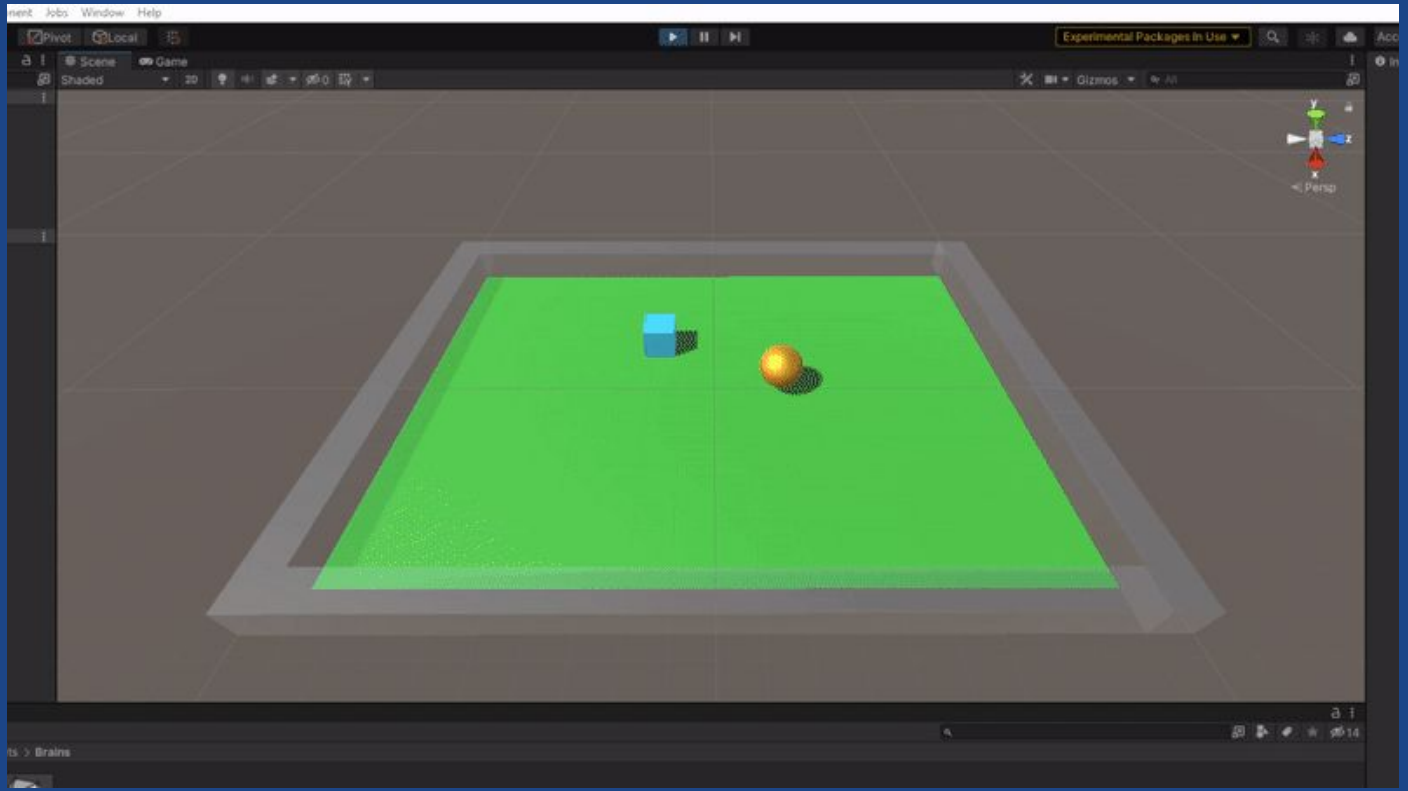
Posición de la recompensa

```
private void OnTriggerEnter(Collider other)
{
  if (other.gameObject.tag == "Reward")
  {
    SetReward(1f);
    floorMeshRenderer.material = winMaterial;
    EndEpisode();
  }
  else if (other.gameObject.tag == "Wall")
  {
    SetReward(-1f);
    floorMeshRenderer.material = loseMaterial;
    EndEpisode();
  }
}
```

ENTRENAMIENTO



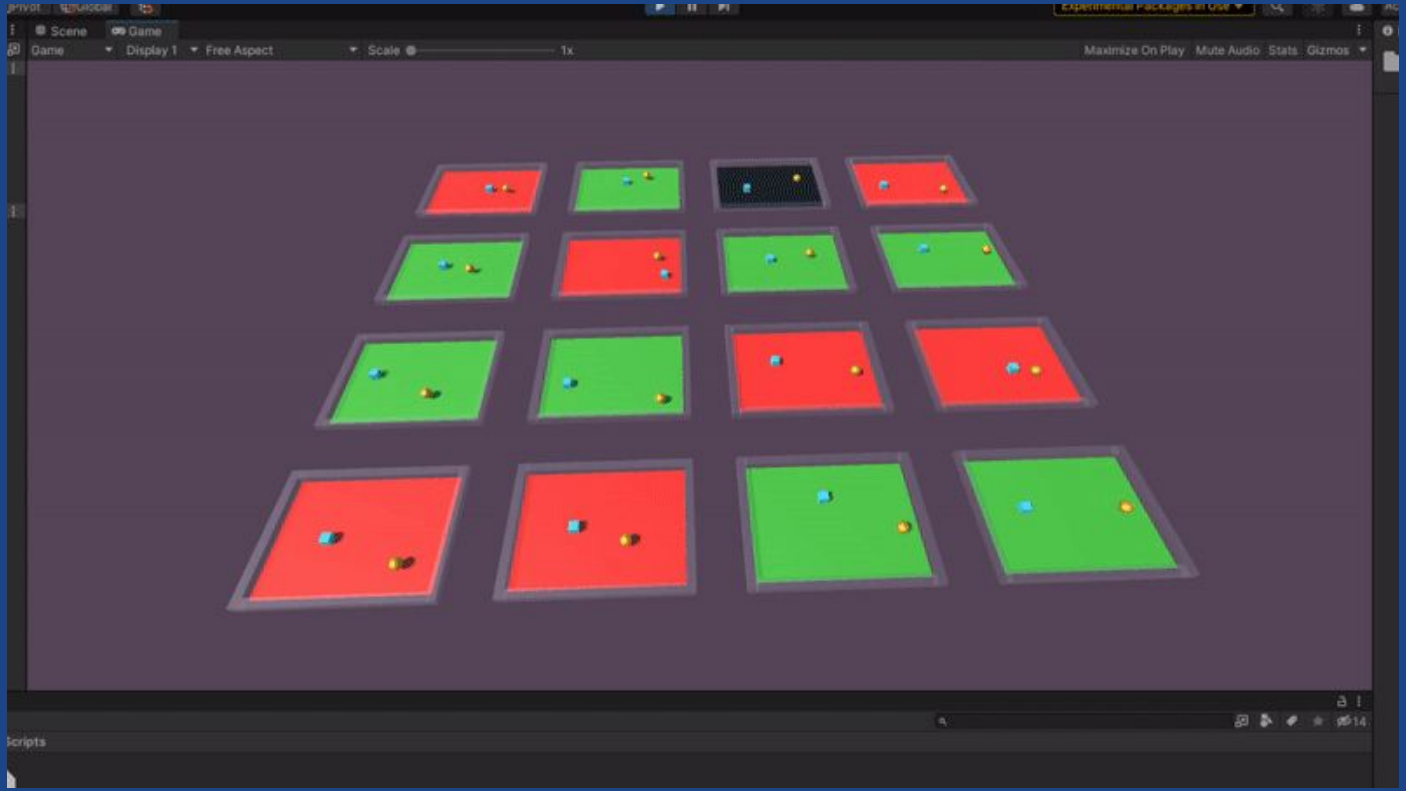
RESULTADOS



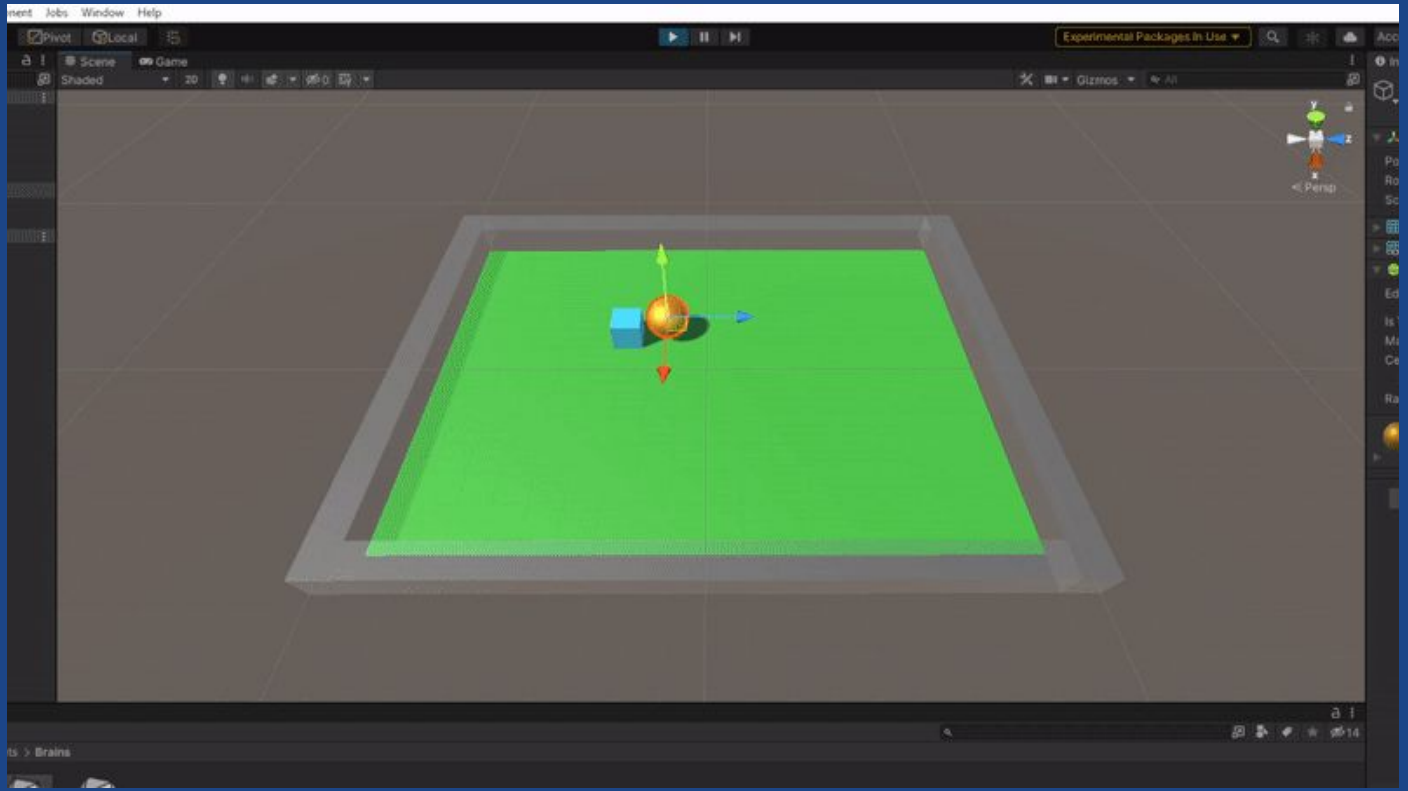
DESARROLLO

```yaml
behaviors:
  CubeBehavior:
    trainer_type: ppo
    hyperparameters:
      batch_size: 10
      buffer_size: 100
      learning_rate: 3.0e-4
      beta: 5.0e-4
      epsilon: 0.2
      lambd: 0.99
      num_epoch: 3
      learning_rate_schedule: linear
      beta_schedule: constant
      epsilon_schedule: linear
    network_settings:
      normalize: false
      hidden_units: 128
      num_layers: 2
    reward_signals:
      extrinsic:
        gamma: 0.99
        strength: 1.0
    max_steps: 500000
    time_horizon: 64
    summary_freq: 10000
```

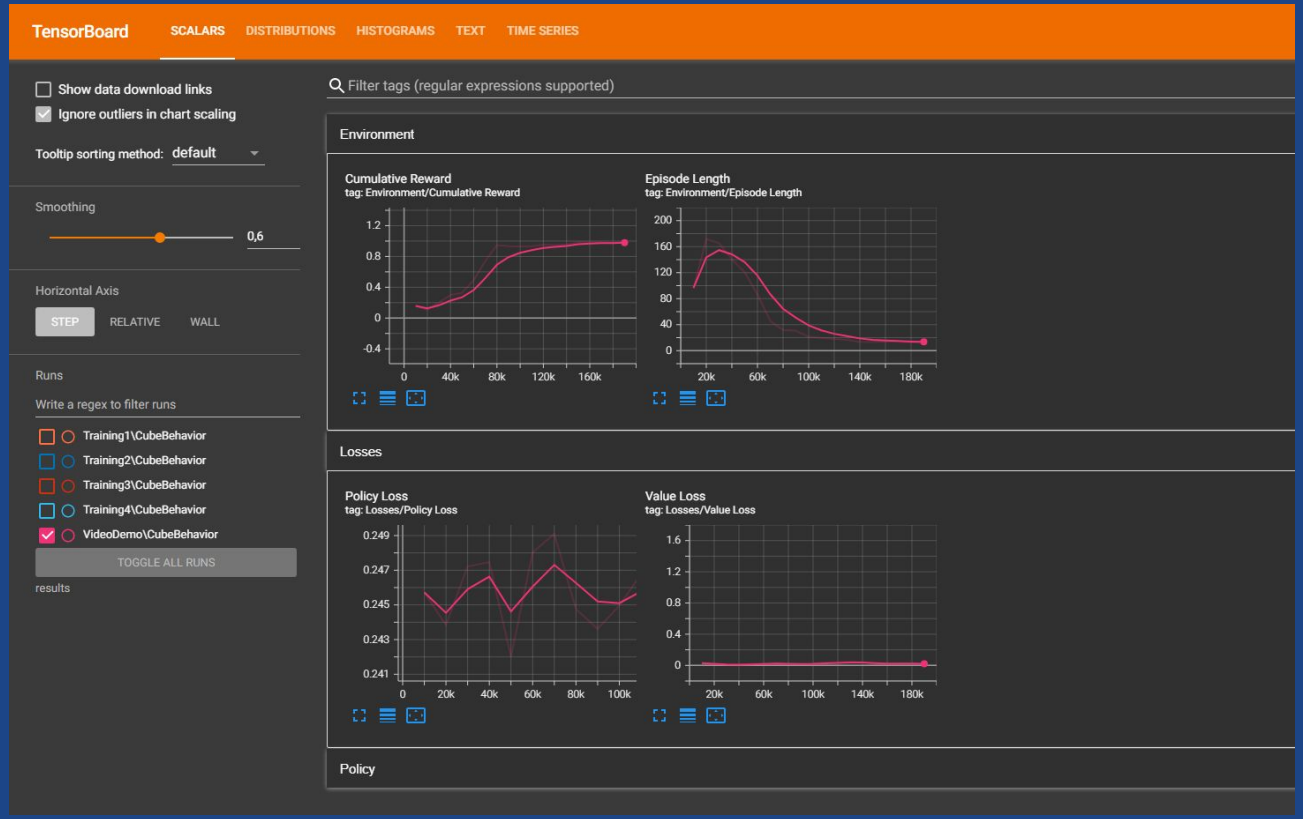
ENTRENAMIENTO



RESULTADOS

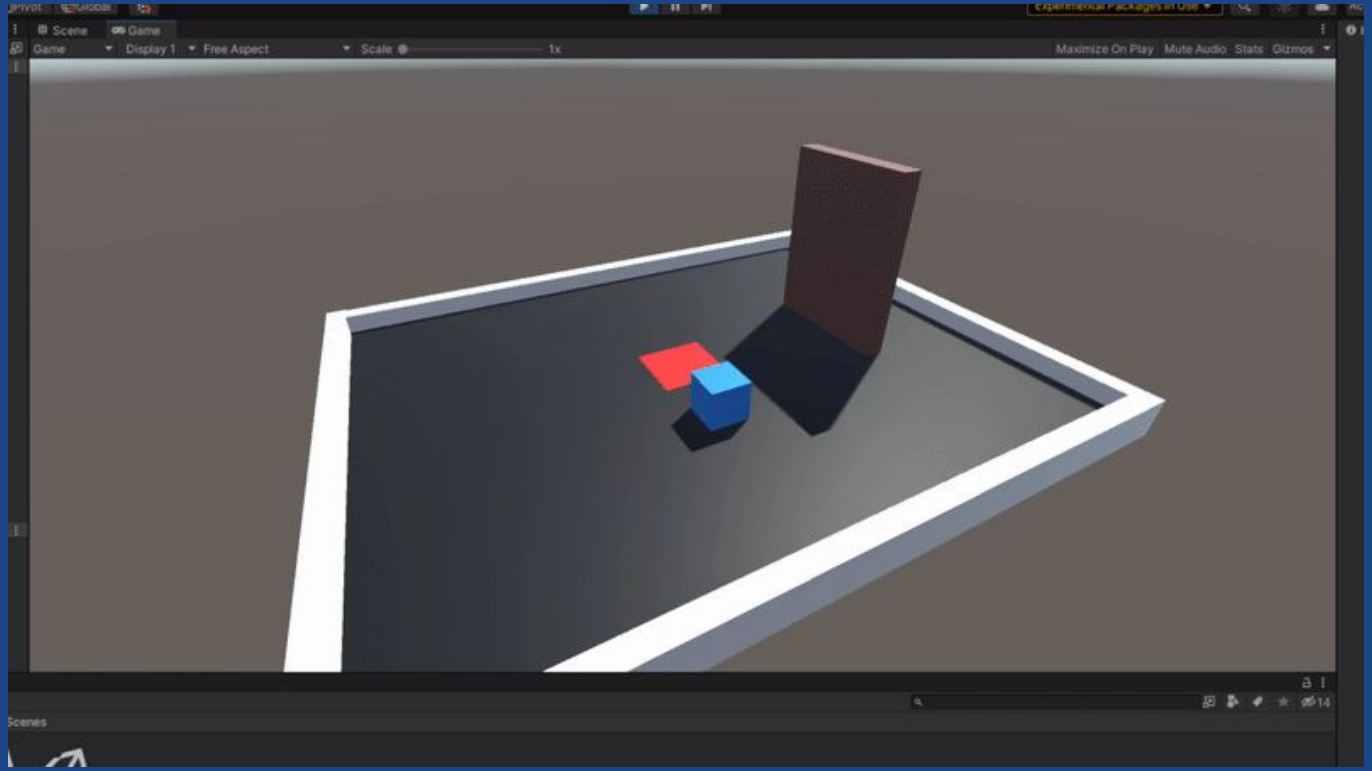


RESULTADOS


TensorBoard
SCALARS
DISTRIBUTIONS
HISTOGRAMS
TEXT
TIME SERIES
Show data download links
Ignore outliers in chart scaling
Tooltip sorting method: default
Smoothing
0,6
Horizontal Axis
STEP
RELATIVE
WALL
Runs
Write a regex to filter runs
Training1\CubeBehavior
Training2\CubeBehavior
Training3\CubeBehavior
Training4\CubeBehavior
VideoDemo\CubeBehavior
TOGGLE ALL RUNS
results
Filter tags (regular expressions supported)
Environment
Cumulative Reward
tag: Environment/Cumulative Reward
Episode Length
tag: Environment/Episode Length
Losses
Policy Loss
tag: Losses/Policy Loss
Value Loss
tag: Losses/Value Loss
Policy

IMITATION LEARNING

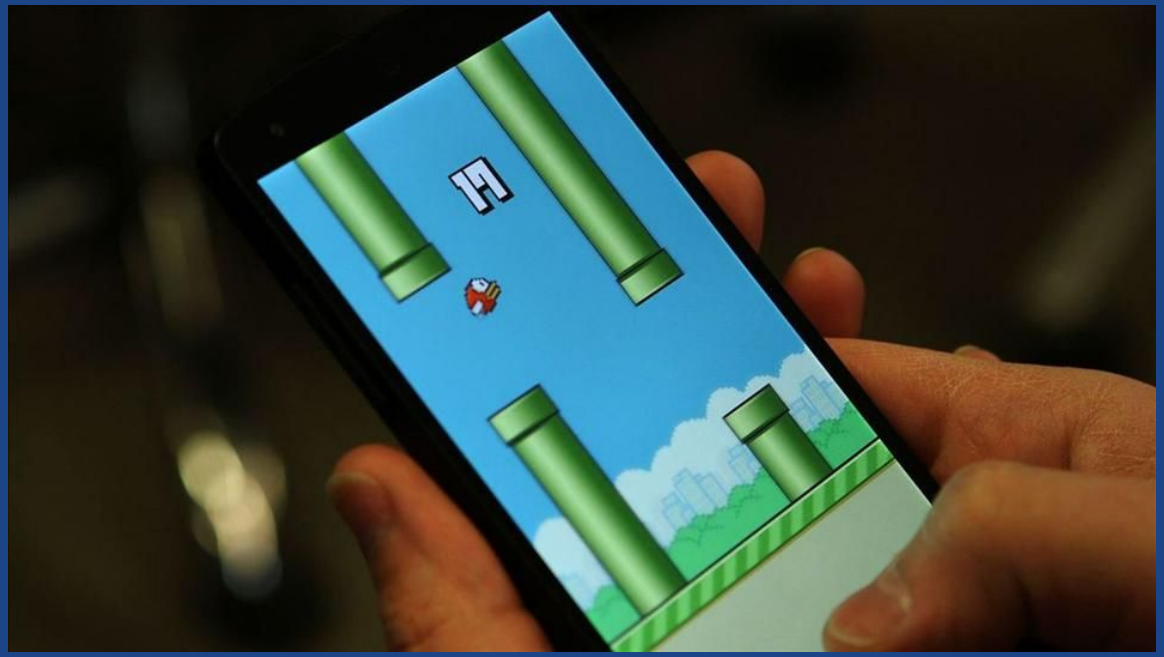
IMITATION LEARNING



03

ESCENARIO CON OBSTÁCULOS

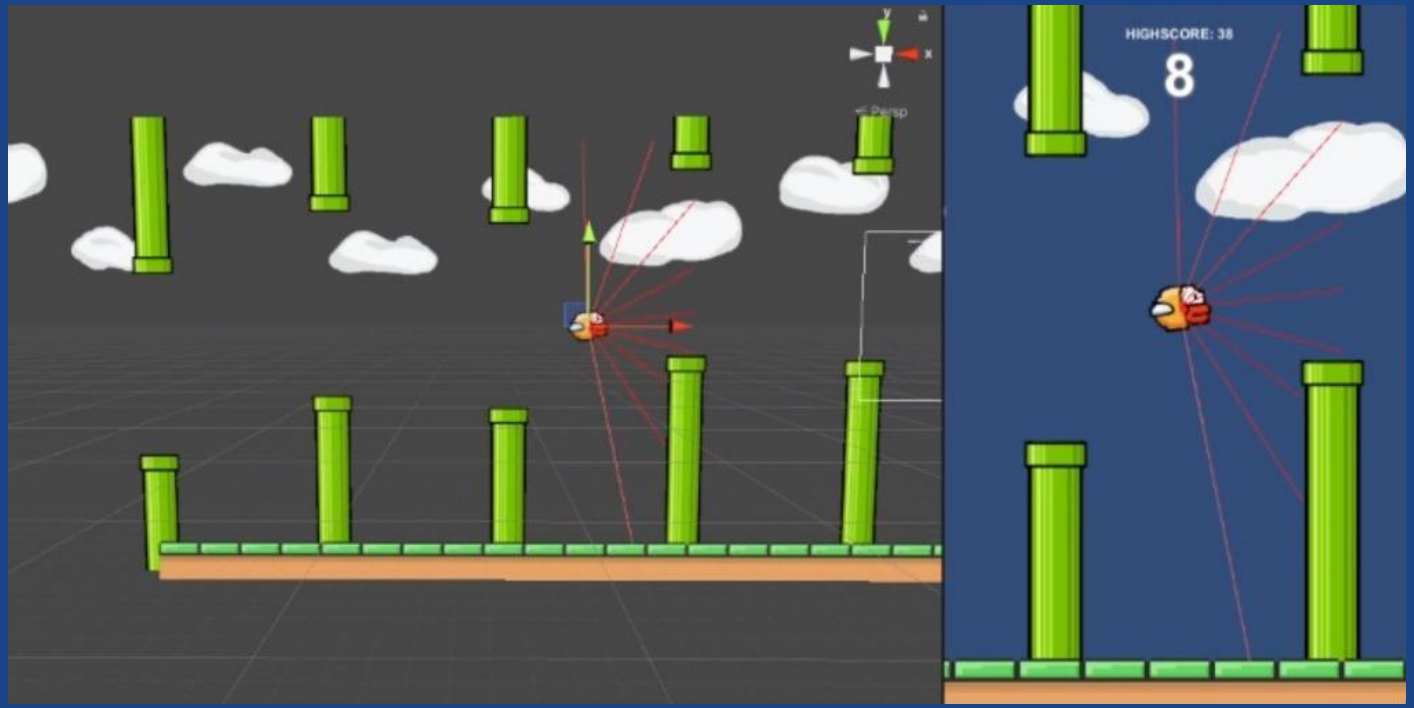
OBJETIVO



DESARROLLO

- Recompensa por tubería superada
- Recompensa por tiempo sobrevivido
- Definir movimientos

DESARROLLO



ENTRENAMIENTO

- Imitation Learning
- Curriculum Learning

ENTRENAMIENTO

ENTRENAMIENTO

```yaml
behaviors:
  FlappyBird:
    trainer_type: ppo
    hyperparameters:
      batch_size: 256
      buffer_size: 10240
      learning_rate: 3.0e-4
      beta: 5.0e-4
      epsilon: 0.2
      lambd: 0.99
      num_epoch: 3
      learning_rate_schedule: linear
    network_settings:
      normalize: false
      hidden_units: 128
      num_layers: 2
```

```yaml
reward_signals:
      extrinsic:
        strength: 0.1
        gamma: 0.8
      gail:
        strength: 0.8
        gamma: 0.8
        demo_path: Demo/FlappyBird.demo
    behavioral_cloning:
        strength: 1.0
        gamma: 0.4
        demo_path: Demo/FlappyBird.demo
    max_steps: 5000000
    time_horizon: 2048
    summary_freq: 10000
```

ENTRENAMIENTO

```yaml
behaviors:
  FlappyBird:
    trainer_type: ppo
    hyperparameters:
      batch_size: 256
      buffer_size: 10240
      learning_rate: 3.0e-4
      beta: 5.0e-4
      epsilon: 0.2
      lambd: 0.99
      num_epoch: 3
      learning_rate_schedule: linear
    network_settings:
      normalize: false
      hidden_units: 128
      num_layers: 2
```

```yaml
reward_signals:
      extrinsic:
        strength: 1.0
        gamma: 0.8
      gail:
        strength: 0.5
        gamma: 0.8
        demo_path: Demo/FlappyBird.demo
    behavioral_cloning:
        strength: 0.4
        gamma: 0.4
        demo_path: Demo/FlappyBird.demo
    max_steps: 5000000
    time_horizon: 2048
    summary_freq: 10000
```

ENTRENAMIENTO

```yaml
behaviors:
  FlappyBird:
    trainer_type: ppo
    hyperparameters:
      batch_size: 256
      buffer_size: 10240
      learning_rate: 3.0e-4
      beta: 5.0e-4
      epsilon: 0.2
      lambd: 0.99
      num_epoch: 3
      learning_rate_schedule: linear
    network_settings:
      normalize: false
      hidden_units: 128
      num_layers: 2
```

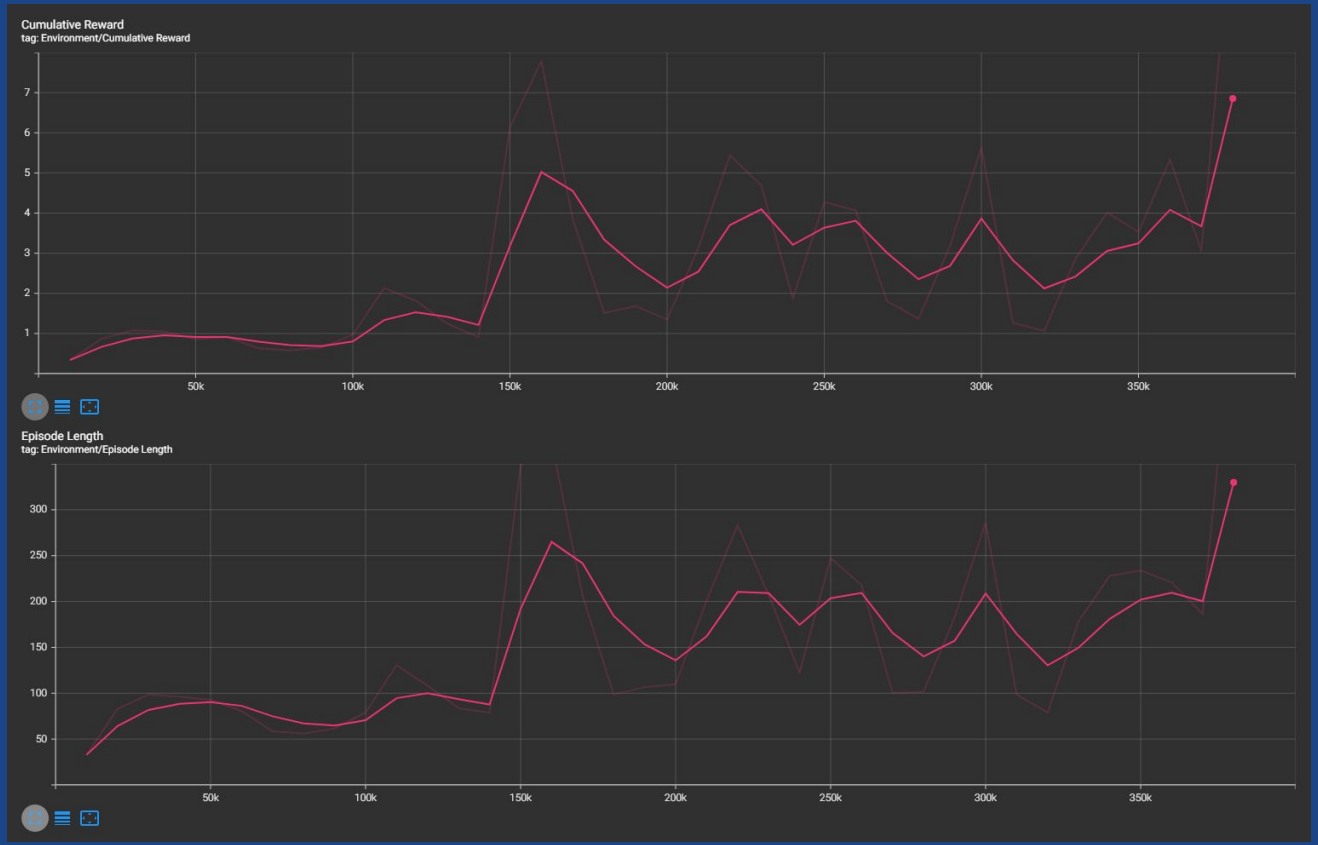
```yaml
reward_signals:
      extrinsic:
        strength: 1.0
        gamma: 0.8
      gail:
        strength: 0.4
        gamma: 0.8
        demo_path: Demo/FlappyBird.demo
    behavioral_cloning:
        strength: 0.1
        gamma: 0.4
        demo_path: Demo/FlappyBird.demo
    max_steps: 5000000
    time_horizon: 2048
    summary_freq: 10000
```

ENTRENAMIENTO

RESULTADOS

TensorBoard
SCALARS
DISTRIBUTIONS
HISTOGRAMS
TEXT
TIME SERIES
INACTIVE
UPLOAD
Show data download links
Ignore outliers in chart scaling
Tooltip sorting method: default
Smoothing
0,6
Horizontal Axis
STEP
RELATIVE
WALL
Runs
Write a regex to filter runs
FlappyTest1\My Behavior
FlappyTest1\FlappyBird
FlappyTestRecto\FlappyBird
FlappyRecto\FlappyBird
FlappyRecto2\FlappyBird
FlappyRectoGabri\FlappyBird
FlappyRectoPositivo - copia\FlappyBird
FlappyRectoPositivo\FlappyBird
FlappyRectoCena\FlappyBird
FlappyRectoCena2\FlappyBird
FlappyRectoCena3 - copia\FlappyBird
FlappyRectoCena3\FlappyBird
TestConPosX2\FlappyBird
TestConPosX3\FlappyBird
TOGGLE ALL RUNS
results
Filter tags (regular expressions supported)
Environment
2
Cumulative Reward
tag: Environment/Cumulative Reward
Episode Length
tag: Environment/Episode Length
Losses
4
GAIL Loss
tag: Losses/GAIL Loss
Policy Loss
tag: Losses/Policy Loss
Pretraining Loss
tag: Losses/Pretraining Loss
Value Loss
tag: Losses/Value Loss
Policy
11

RESULTADOS


Cumulative Reward
tag: Environment/Cumulative Reward
7
6
5
4
3
2
1
50k
100k
150k
200k
250k
300k
350k
Episode Length
tag: Environment/Episode Length
300
250
200
150
100
50
50k
100k
150k
200k
250k
300k
350k

RESULTADOS

RESULTADOS

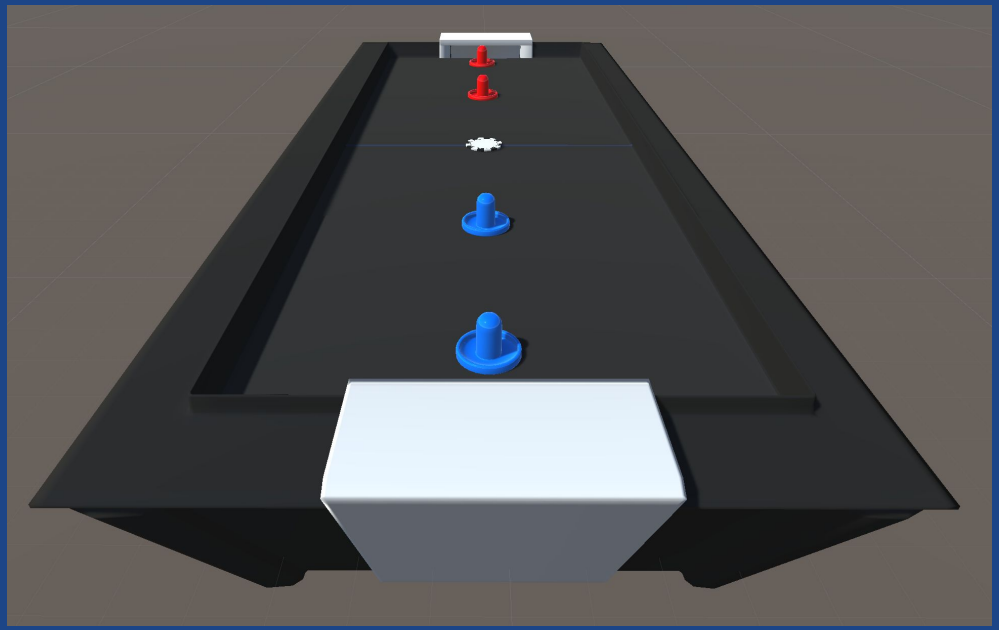
RESULTADOS

04

ESCENARIO COLABORATIVO

OBJETIVO

Aero-jóquey



DESARROLLO

Percepciones

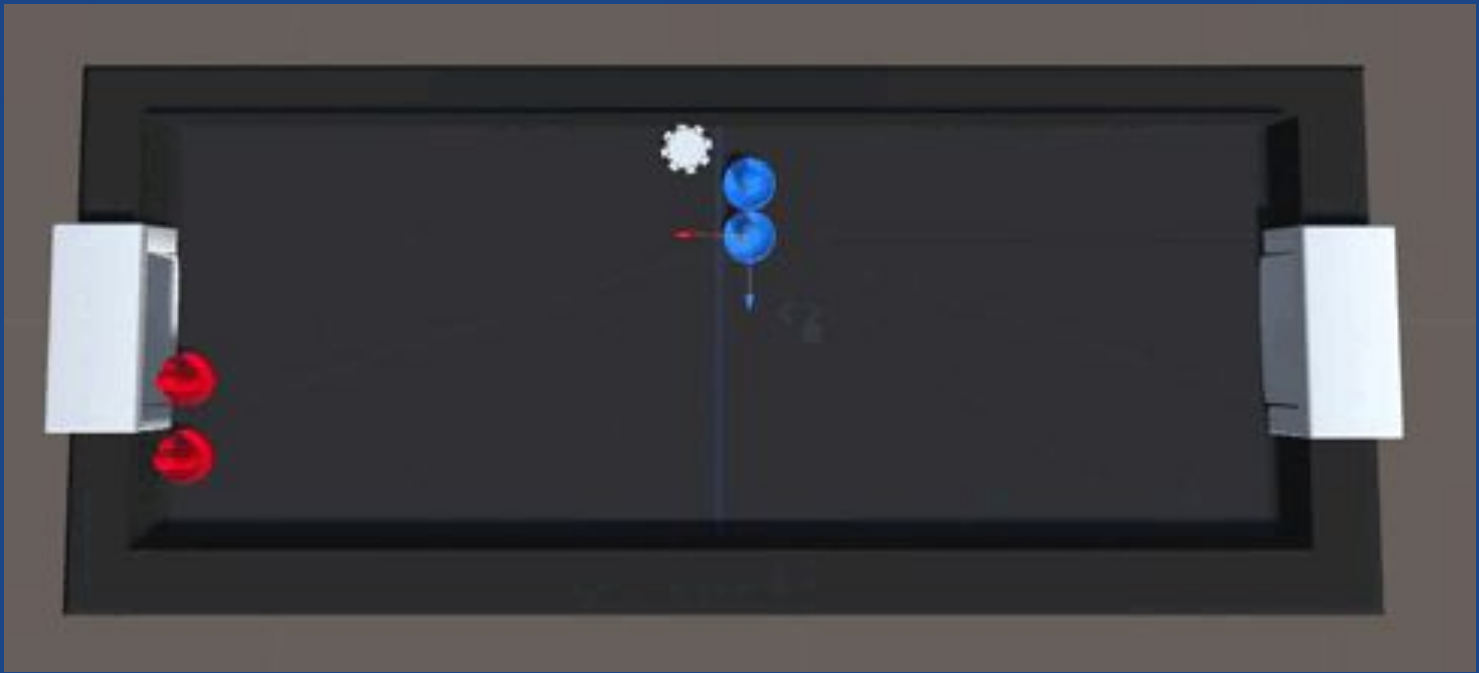
- Disco
- Paredes
- Porterías (azul y roja)
- Otros agentes (azul y rojo)

DESARROLLO

Recompensas

	Individuales	Grupal
+	Tocar el disco	Marcar gol
-	Existencial	Le marcan

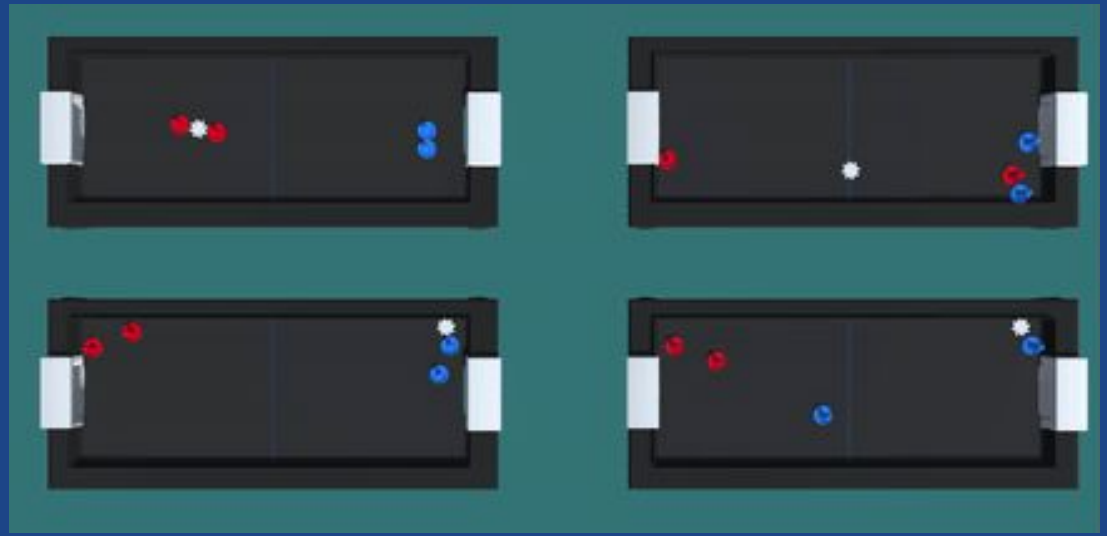
DESARROLLO



ENTRENAMIENTO

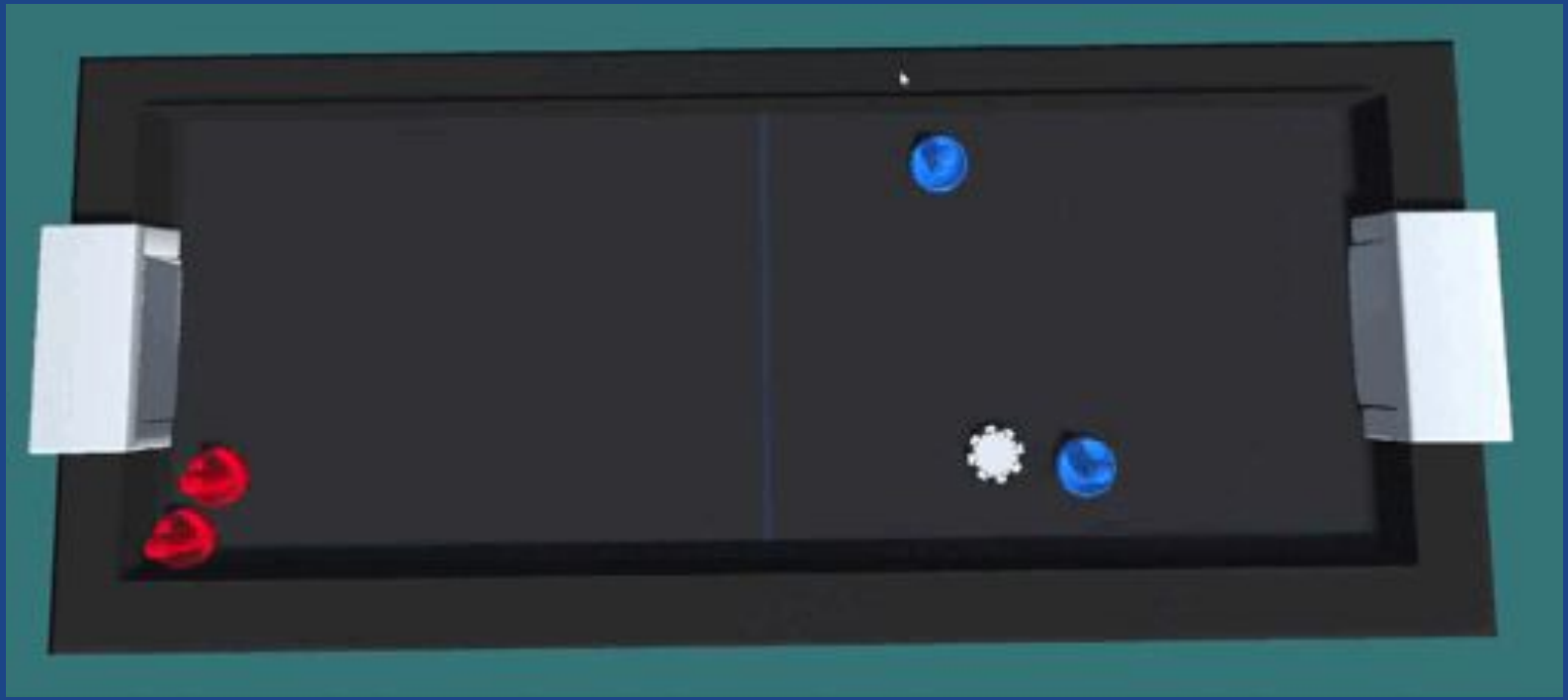
AirHockey

```yaml
behaviors:
  HockeyBehavior:
    trainer_type: poca
    hyperparameters:
      batch_size: 2048
      buffer_size: 20480
      learning_rate: 0.0003
      beta: 0.005
      epsilon: 0.2
      lambd: 0.95
      num_epoch: 3
      learning_rate_schedule: constant
    network_settings:
      normalize: false
      hidden_units: 512
      num_layers: 2
      vis_encode_type: simple
    reward_signals:
      extrinsic:
        gamma: 0.99
        strength: 1.0
    keep_checkpoints: 5
    max_steps: 50000000
    time_horizon: 1000
    summary_freq: 10000
    self_play:
      save_steps: 50000
      team_change: 200000
      swap_steps: 2000
      window: 10
      play_against_latest_model_ratio: 0.5
      initial_elo: 1200.0
```

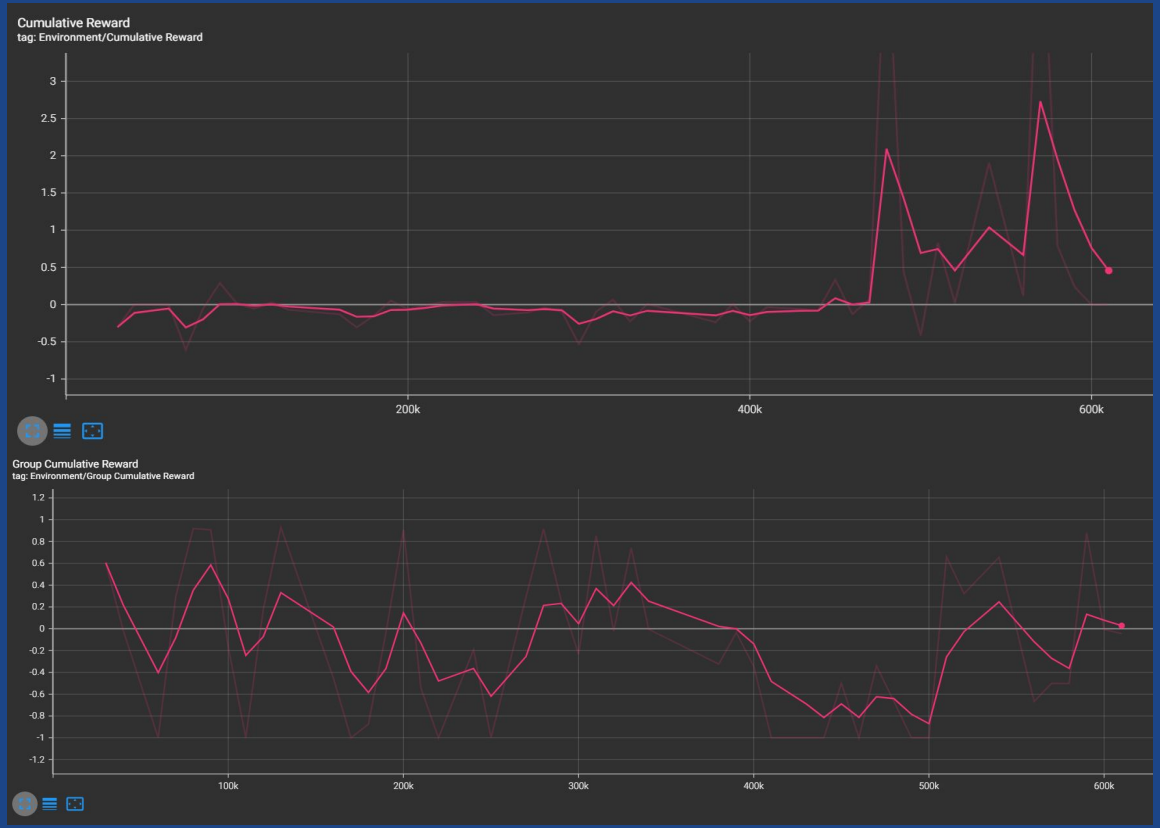


ENTRENAMIENTO

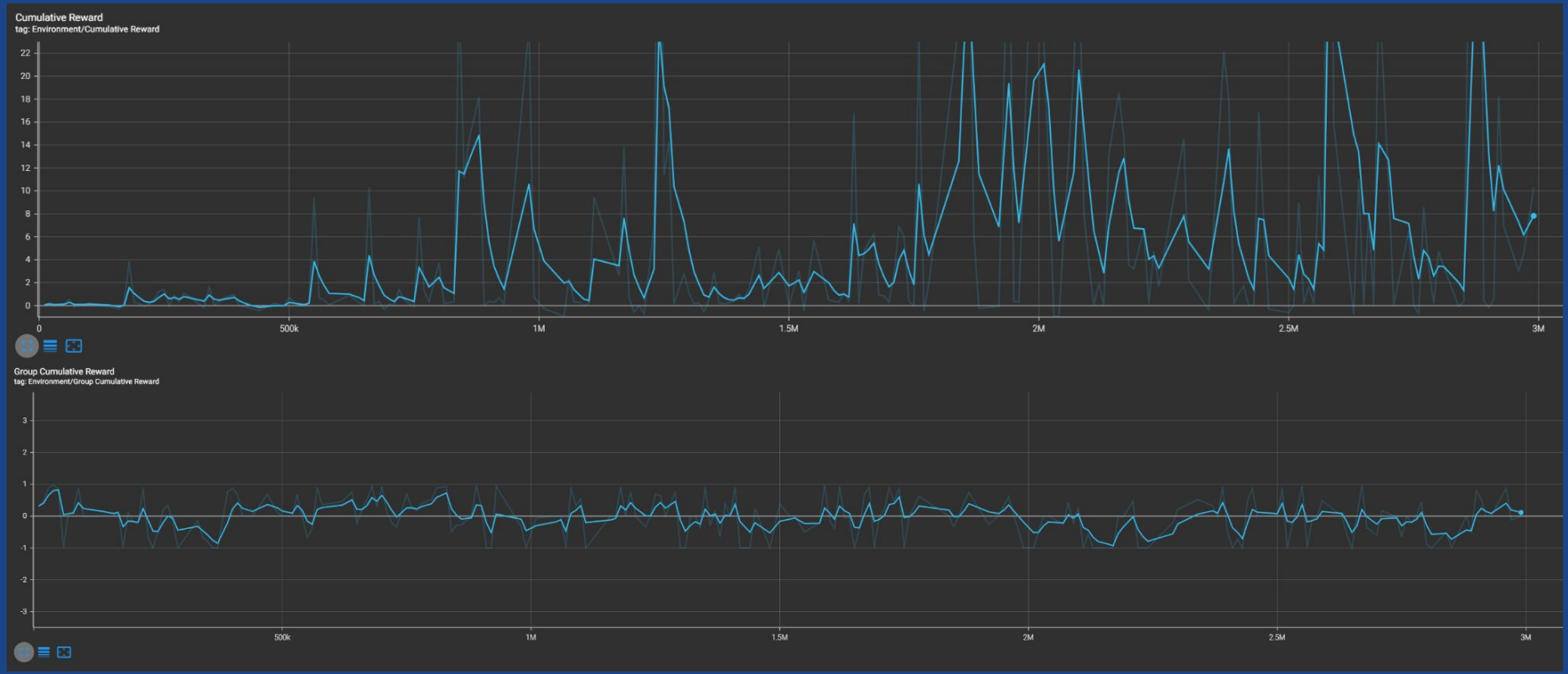
ENTRENAMIENTO



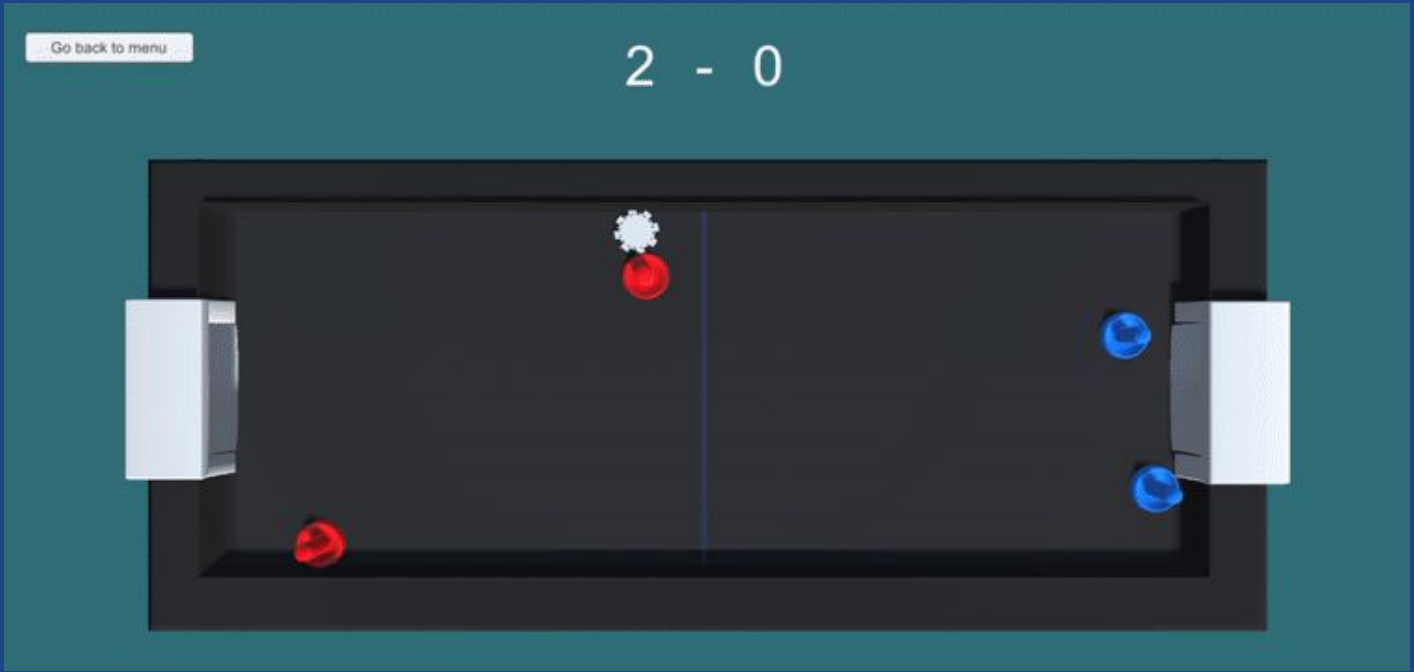
RESULTADOS


Cumulative Reward
tag: Environment/Cumulative Reward
3
2.5
2
1.5
1
0.5
0
-0.5
-1
200k
400k
600k
Group Cumulative Reward
tag: Environment/Group Cumulative Reward
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
100k
200k
300k
400k
500k
600k

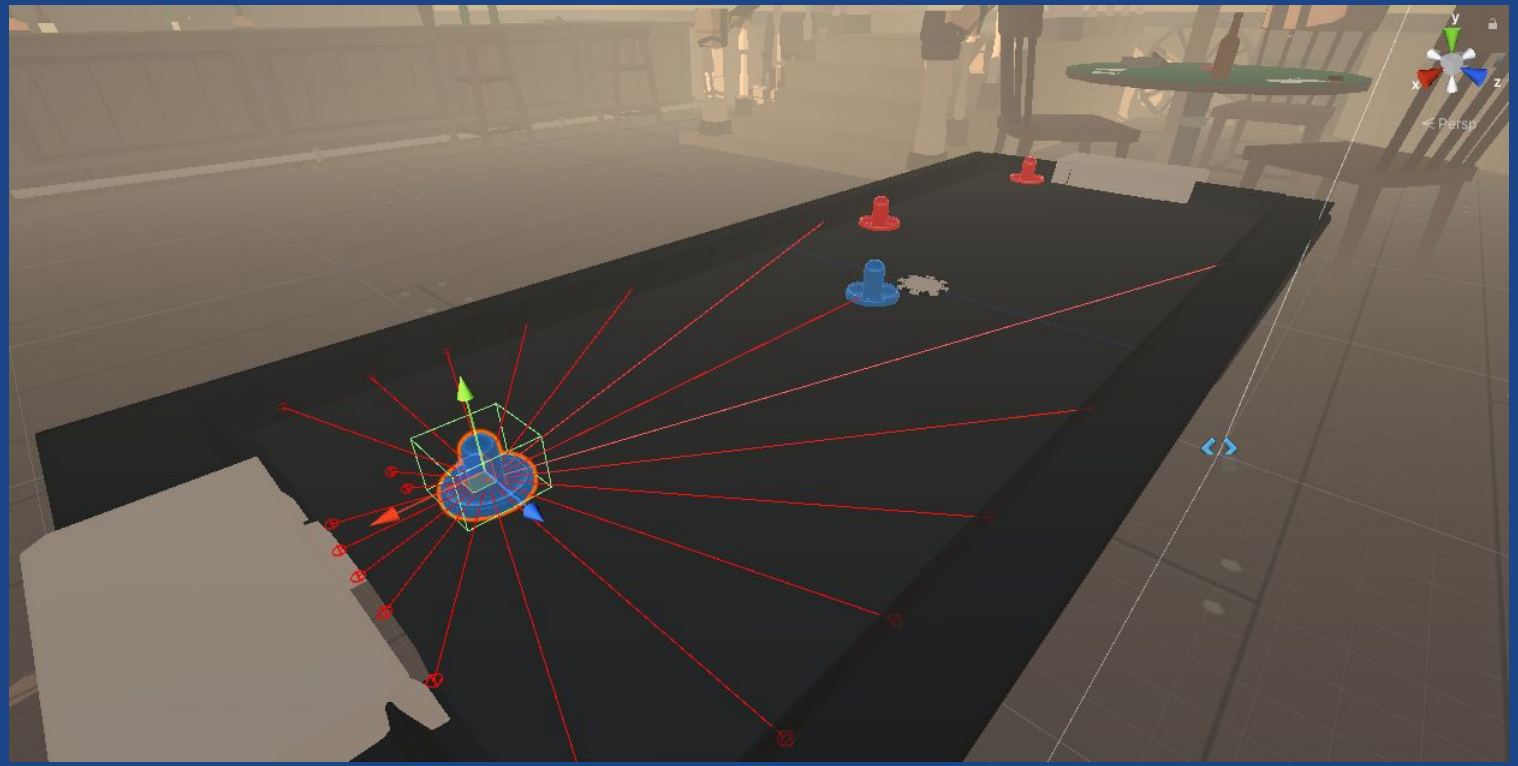
RESULTADOS


Cumulative Reward
tag: Environment/Cumulative Reward
22
20
18
16
14
12
10
8
6
4
2
0
0
500k
1M
1.5M
2M
2.5M
3M
Group Cumulative Reward
tag: Environment/Group Cumulative Reward
3
2
1
0
-1
-2
-3
500k
1M
1.5M
2M
2.5M
3M

RESULTADOS



RESULTADOS



¿PREGUNTAS?