



PROYECTO JAVIVI

Grupo 6

QUIÉNES SOMOS

JAVIER CORREA MARICHAL

VIREN SAJJU DHANWANI DHANWANI

GABRIEL GARCÍA JAUBERT



▮ TABLA DE CONTENIDOS ▮

01

Propuesta

03

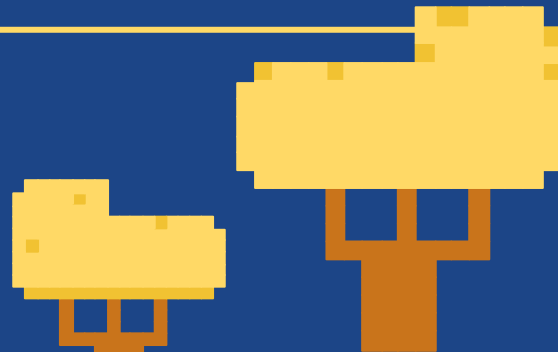
Escenario con obstáculos

02

Escenario sencillo

04

Escenario colaborativo





01

PROPUESTA



ESCENARIOS



3 escenarios

- Llegar a un objetivo simple
- Esquivar obstáculos
- Colaborar





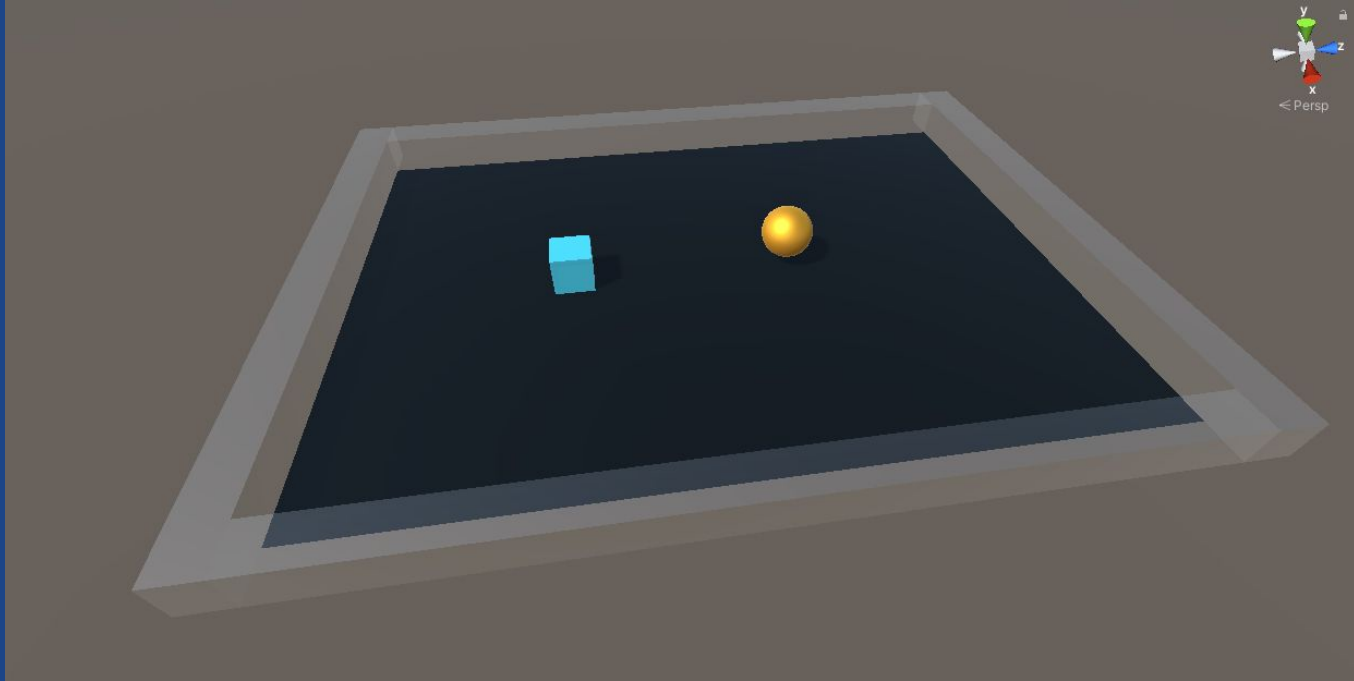
02

ESCENARIO SENCILLO





OBJETIVO





DESARROLLO



Posición del agente

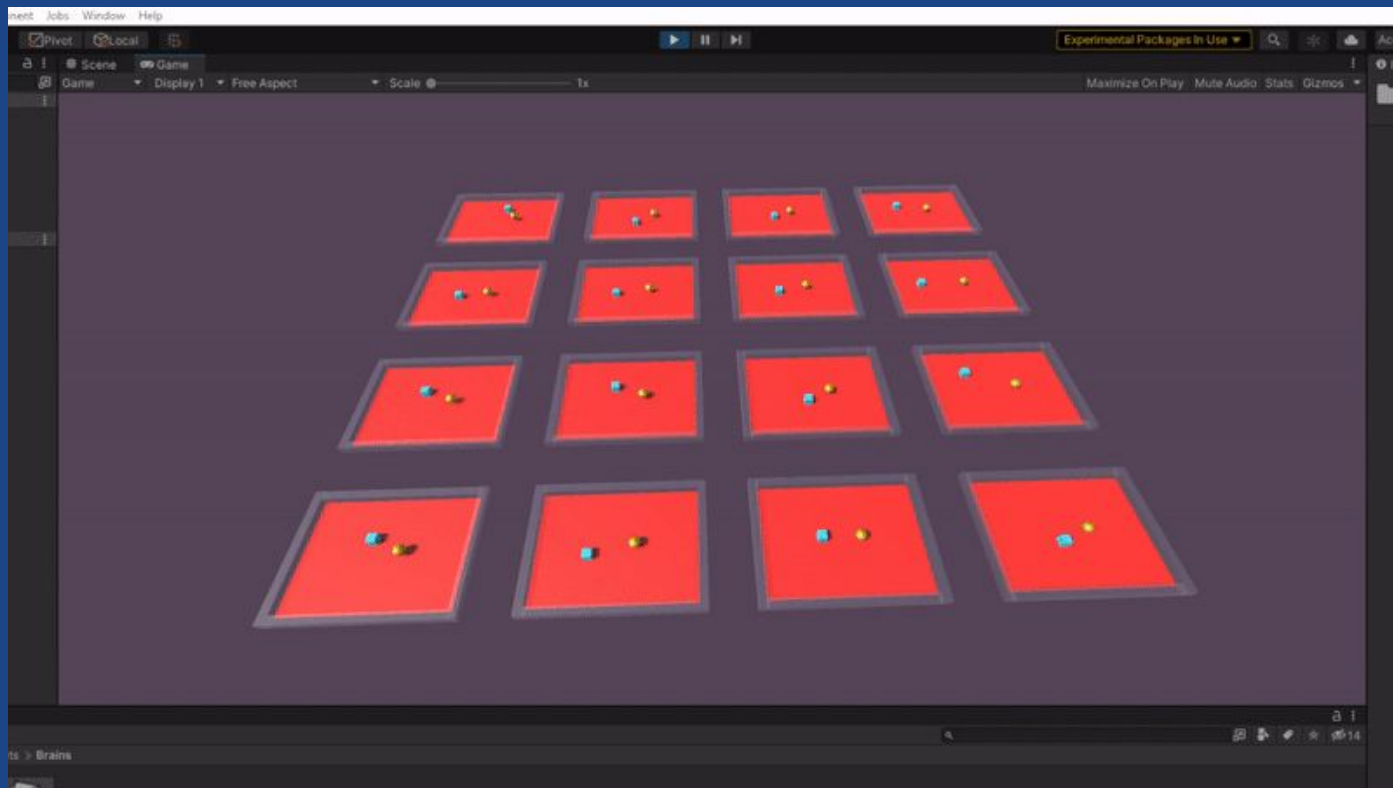
Posición de la recompensa



```
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();
    }
}
```

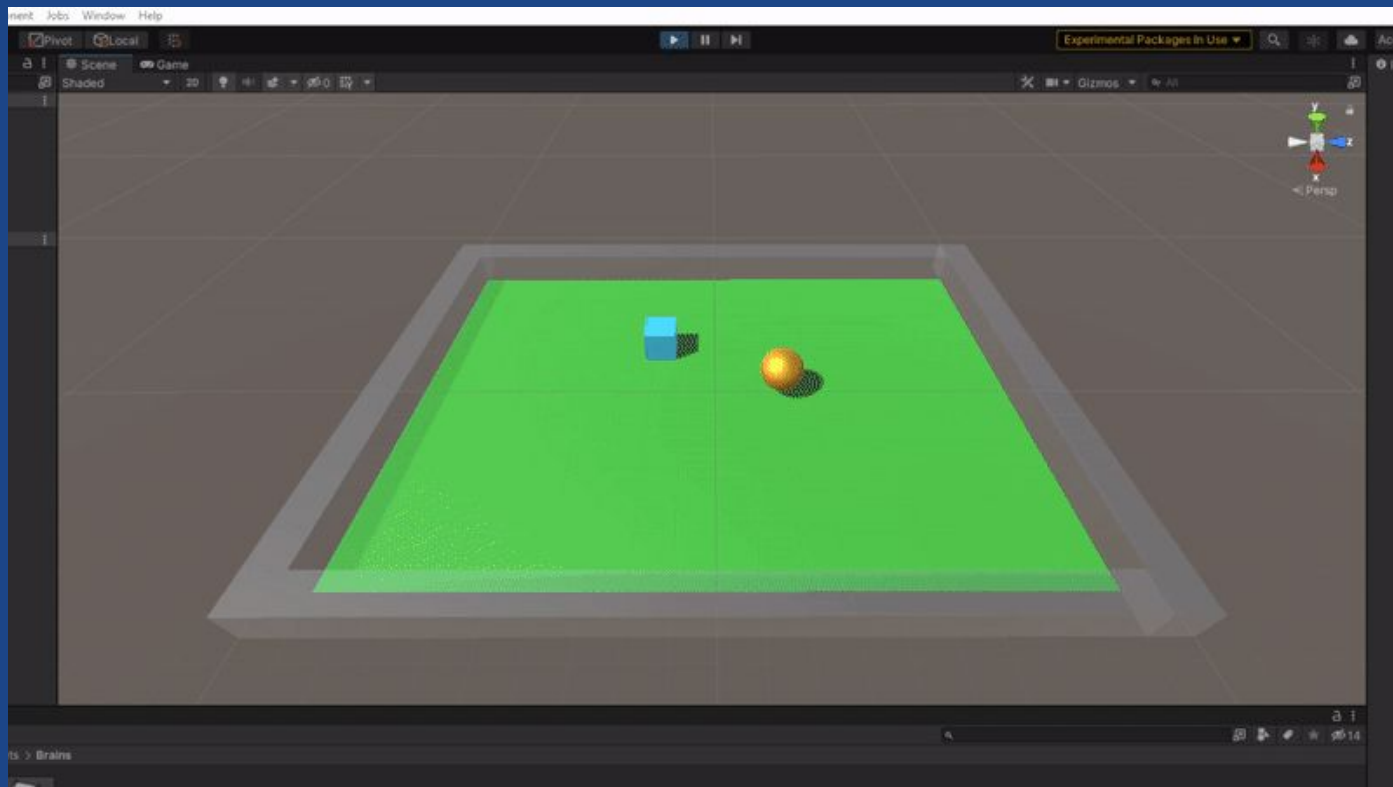



ENTRENAMIENTO





RESULTADOS





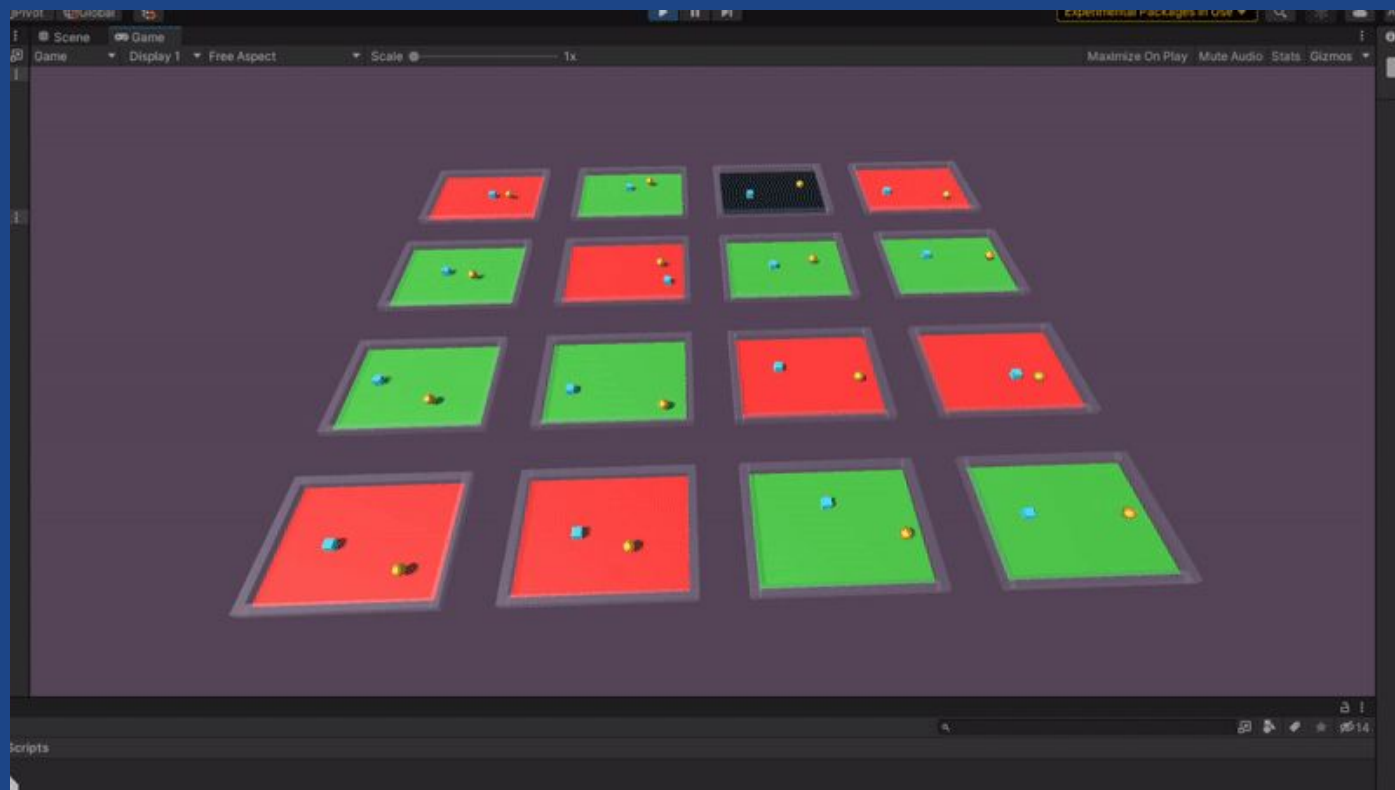
DESARROLLO



```
behaviors:
  CubeBehavior:
    trainer_type: ppo
    hyperparameters:
      batch_size: 10
      buffer_size: 100
      learning_rate: 3.0e-4
      beta: 5.0e-4
      epsilon: 0.2
      lambda: 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: 1000
```

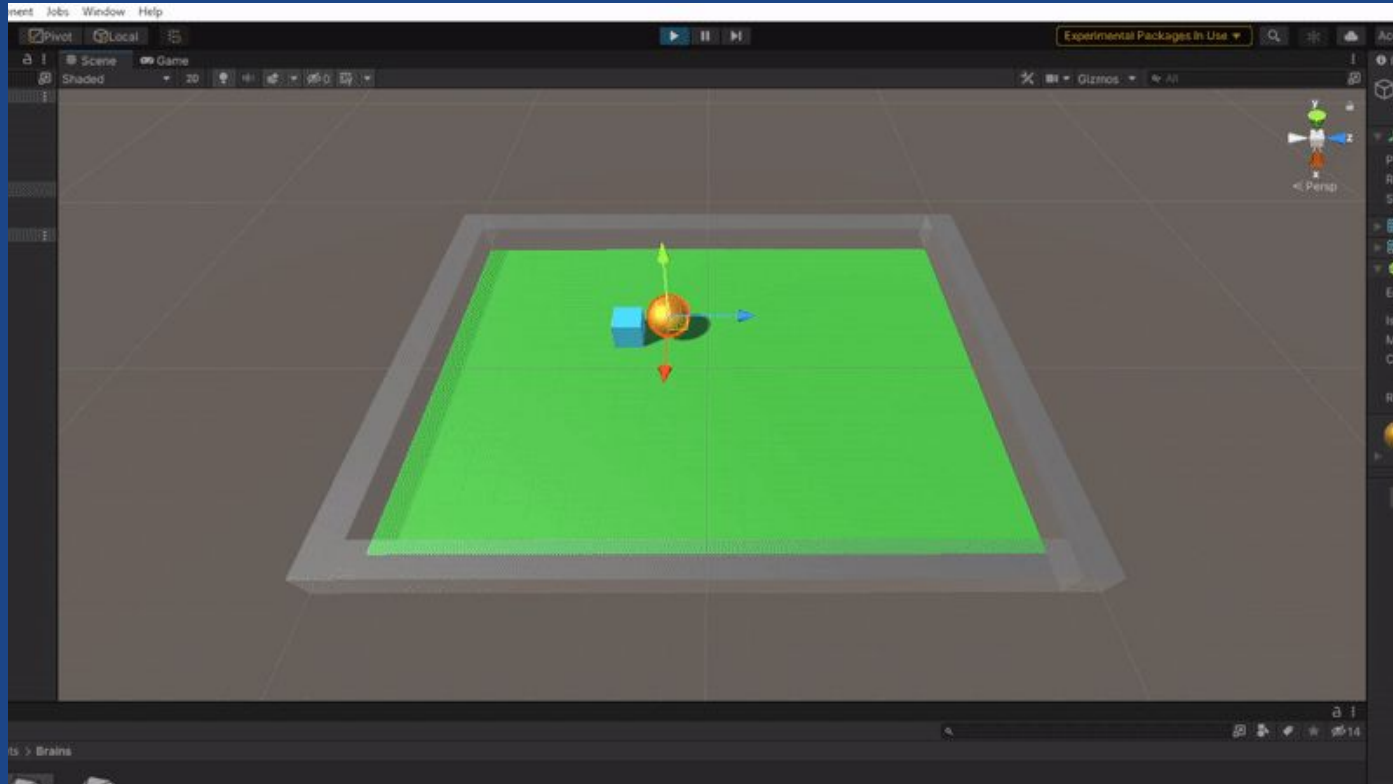


ENTRENAMIENTO



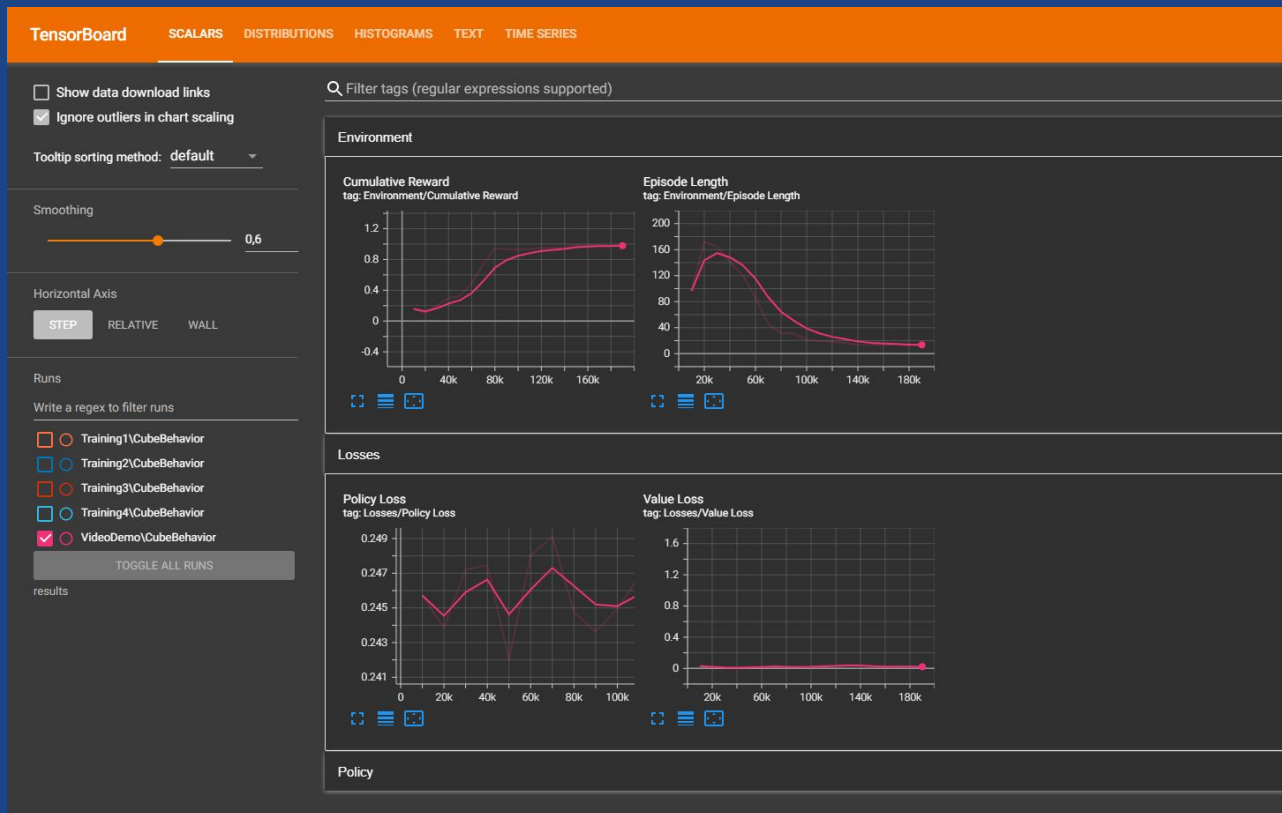


RESULTADOS





RESULTADOS

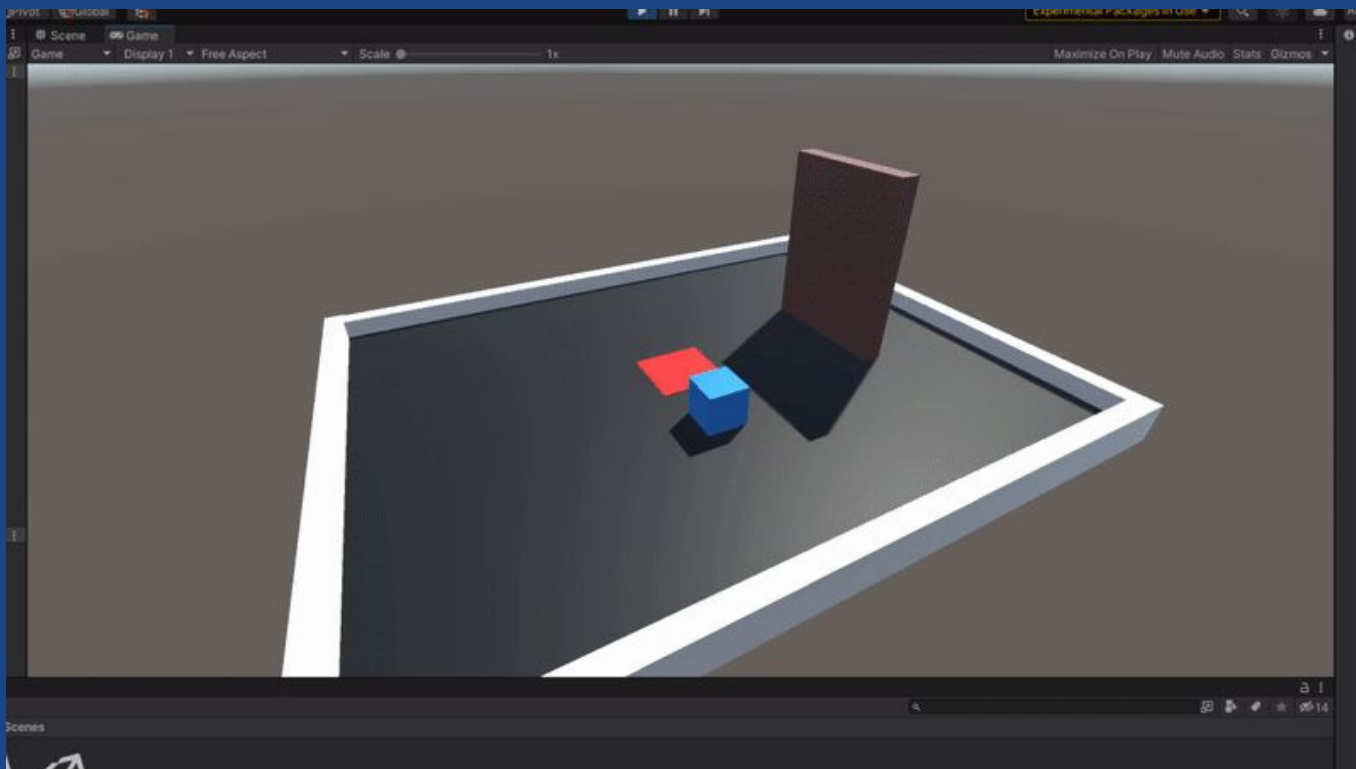


IMITATION LEARNING

```
reward_signals:
  extrinsic:
    strength: 0.2
    gamma: 0.8
  gail:
    strength: 0.8
    gamma: 0.8
  demo_path: Demos/SimpleDemo.demo
behavioral_cloning:
  strength: 1.0
  gamma: 0.8
  demo_path: Demos/SimpleDemo.demo
max_steps: 5000000
time_horizon: 2048
summary_freq: 20000
```

```
reward_signals:
  extrinsic:
    strength: 1.0
    gamma: 0.8
  gail:
    strength: 0.1
    gamma: 0.8
  demo_path: Demos/SimpleDemo.demo
behavioral_cloning:
  strength: 0.1
  gamma: 0.8
  demo_path: Demos/SimpleDemo.demo
max_steps: 5000000
time_horizon: 2048
summary_freq: 20000
```

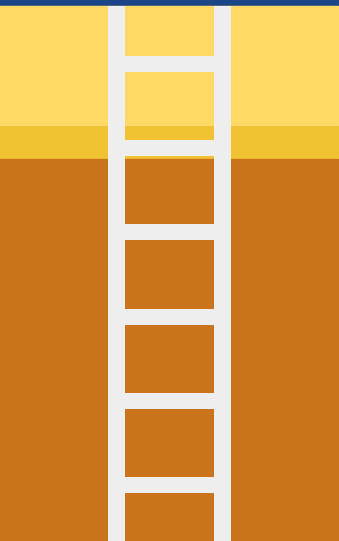
IMITATION LEARNING





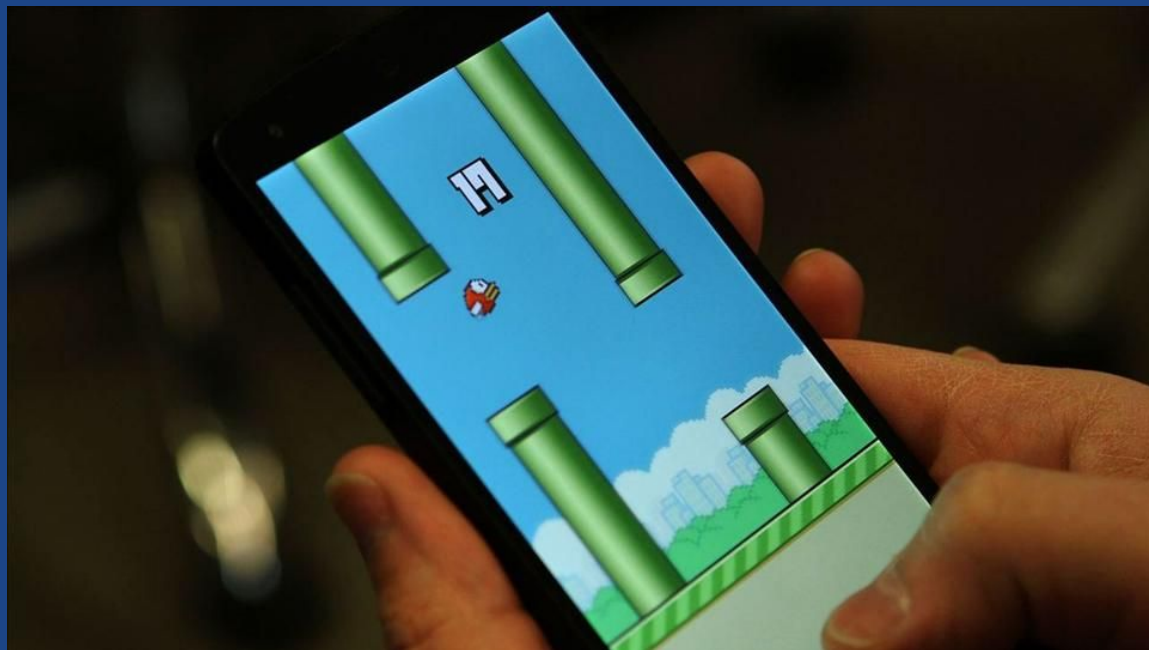
03

ESCENARIO CON OBSTÁCULOS





OBJETIVO





DESARROLLO

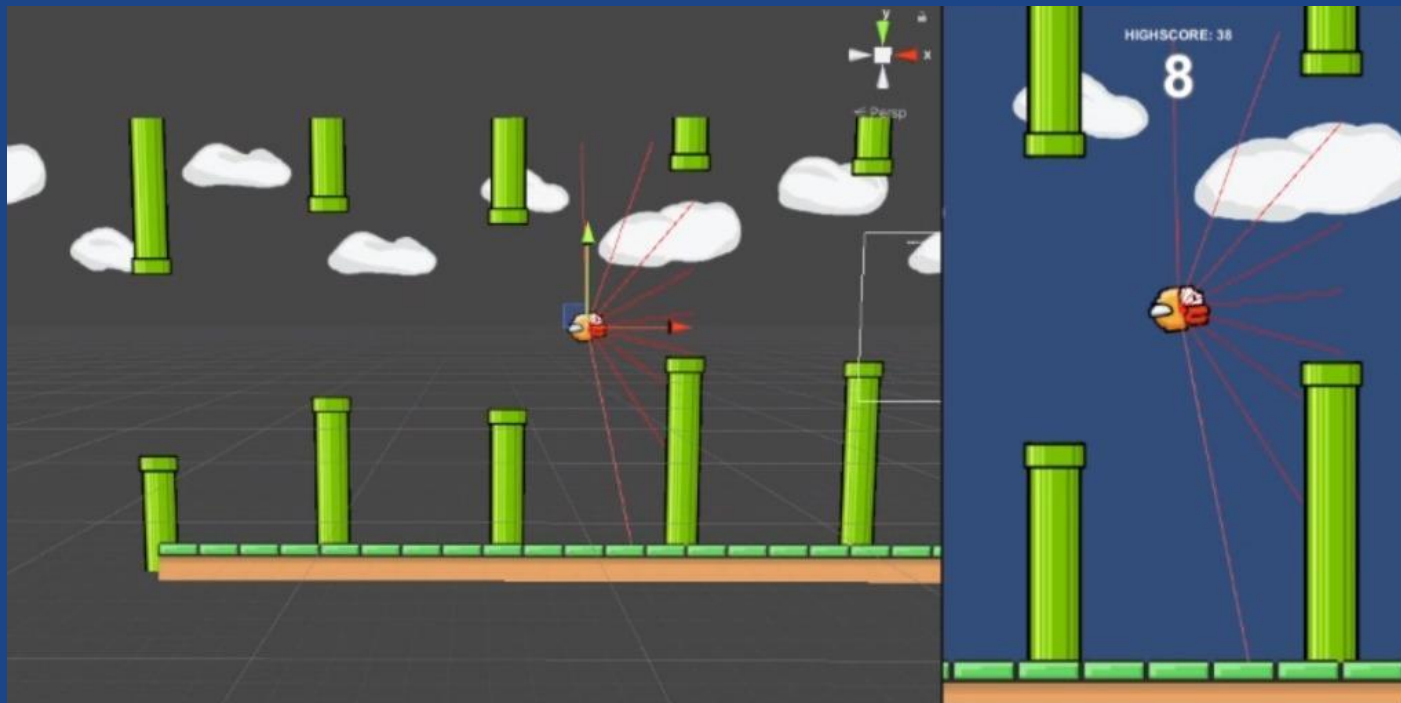


- Recompensa por tubería superada
- Recompensa por tiempo sobrevivido
- Definir movimientos





DESARROLLO





ENTRENAMIENTO



- Imitation Learning
- Curriculum Learning





ENTRENAMIENTO





ENTRENAMIENTO



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

```
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: 1000
```



ENTRENAMIENTO



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

```
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



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

```
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



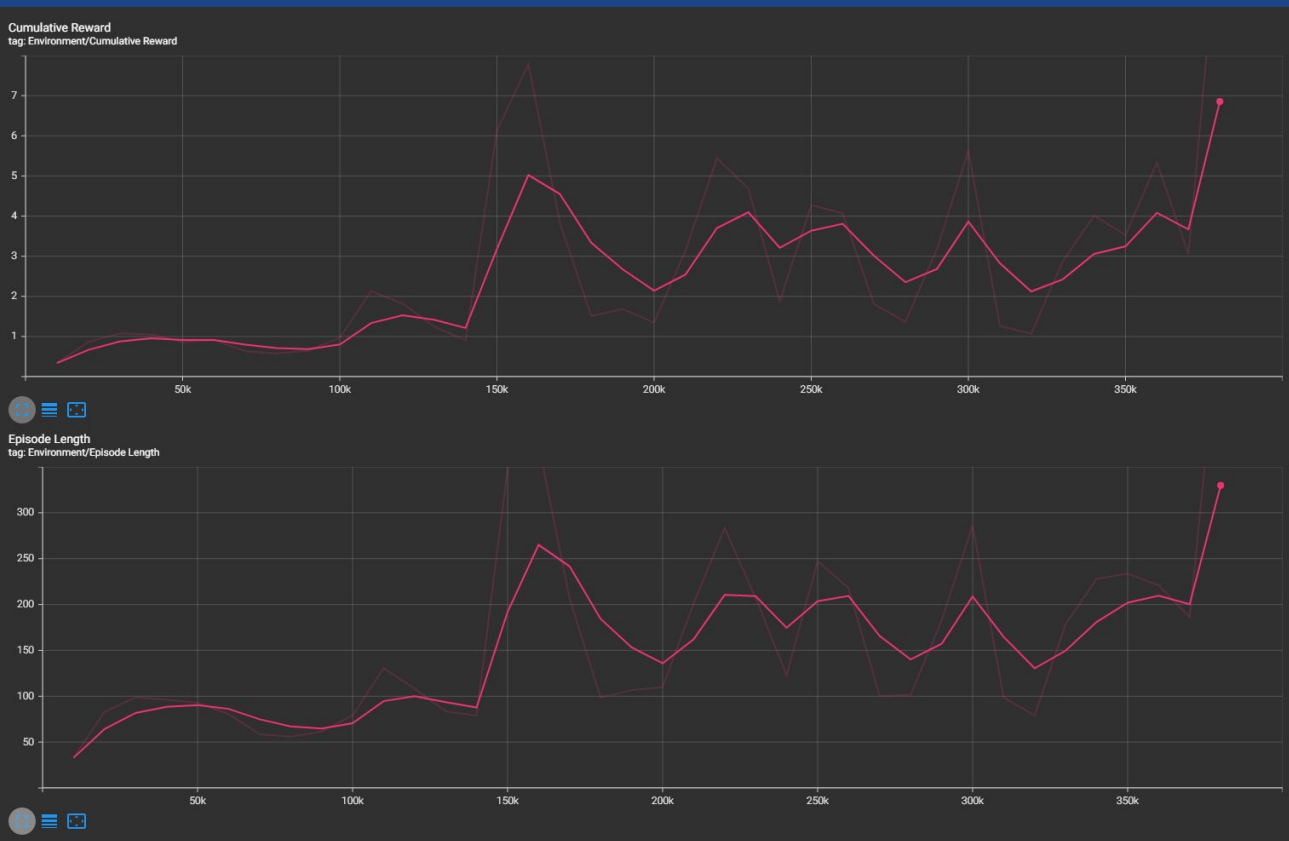


RESULTADOS



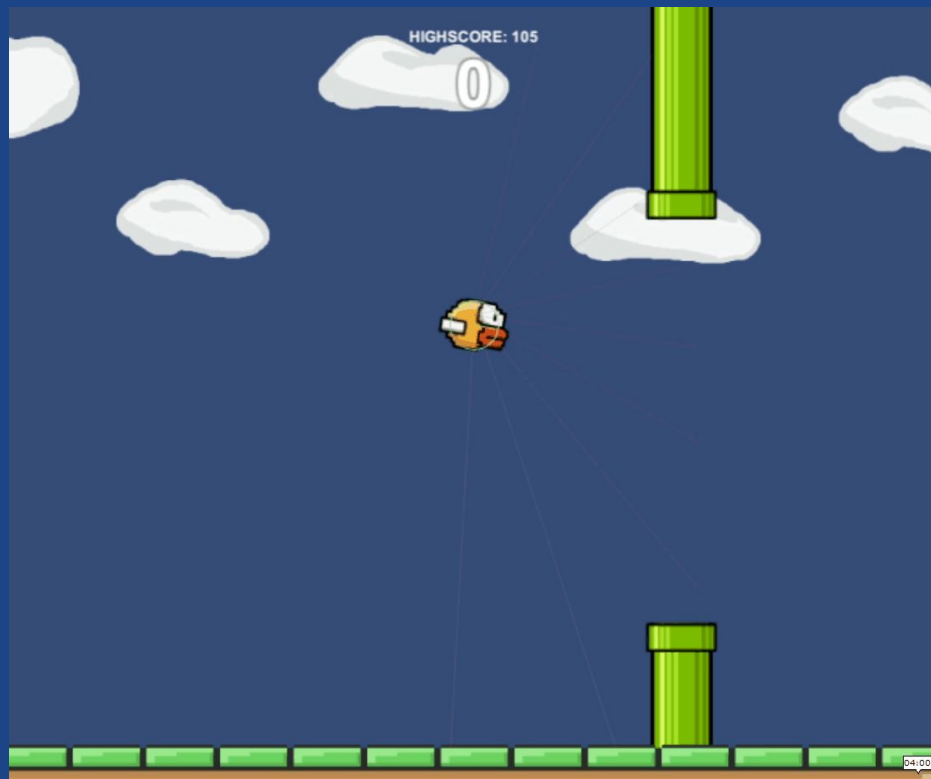


RESULTADOS





RESULTADOS





RESULTADOS





RESULTADOS





04

ESCENARIO COLABORATIVO





OBJETIVO



Aero-jóquey





DESARROLLO



Percepciones

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





DESARROLLO

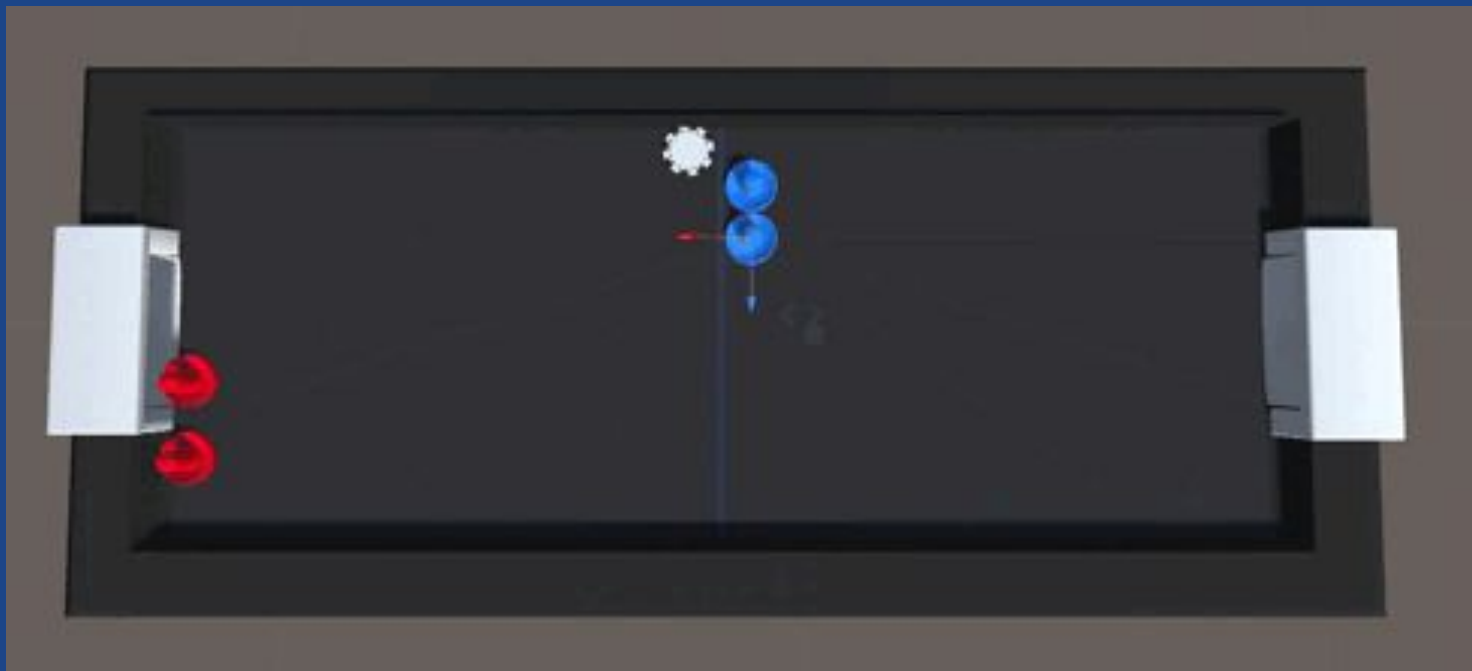


Recompensas

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



DESARROLLO





ENTRENAMIENTO



```
AirHockey

behaviors:
  HockeyBehavior:
    trainer_type: poca
    hyperparameters:
      batch_size: 2048
      buffer_size: 20480
      learning_rate: 0.0003
      beta: 0.005
      epsilon: 0.2
      lambda: 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: 5000000
    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

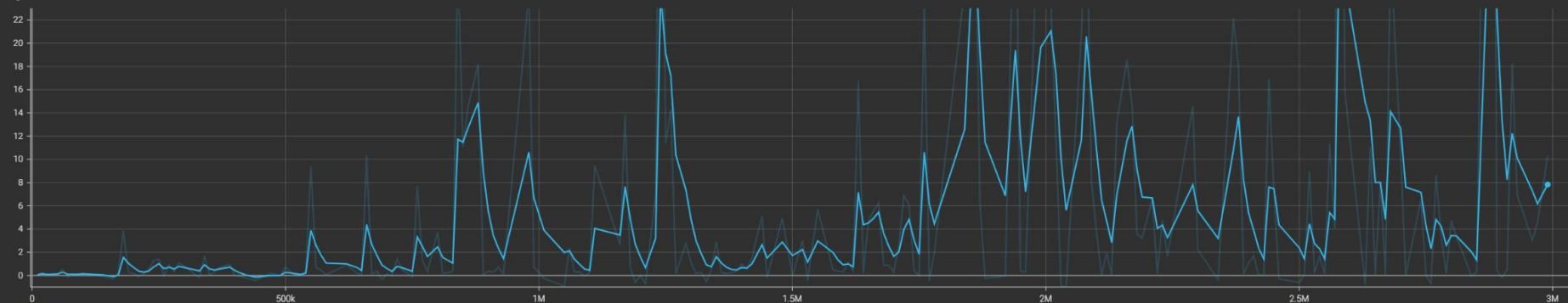




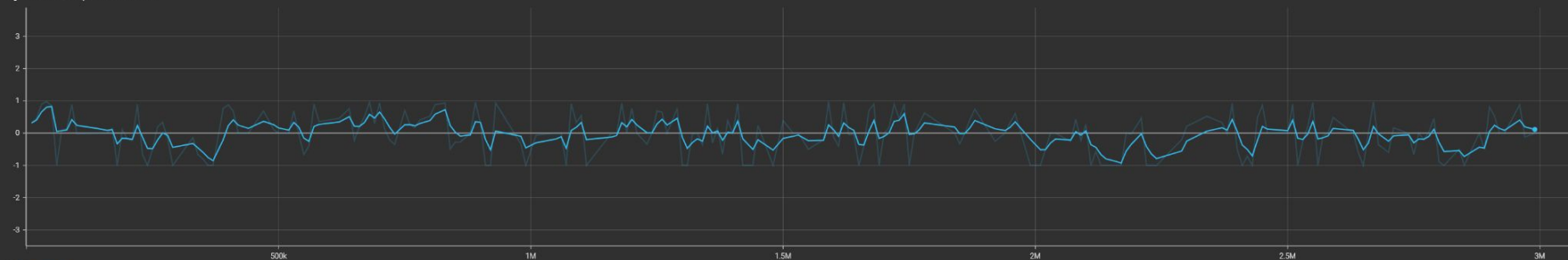
RESULTADOS



Cumulative Reward
tag: Environment/Cumulative Reward

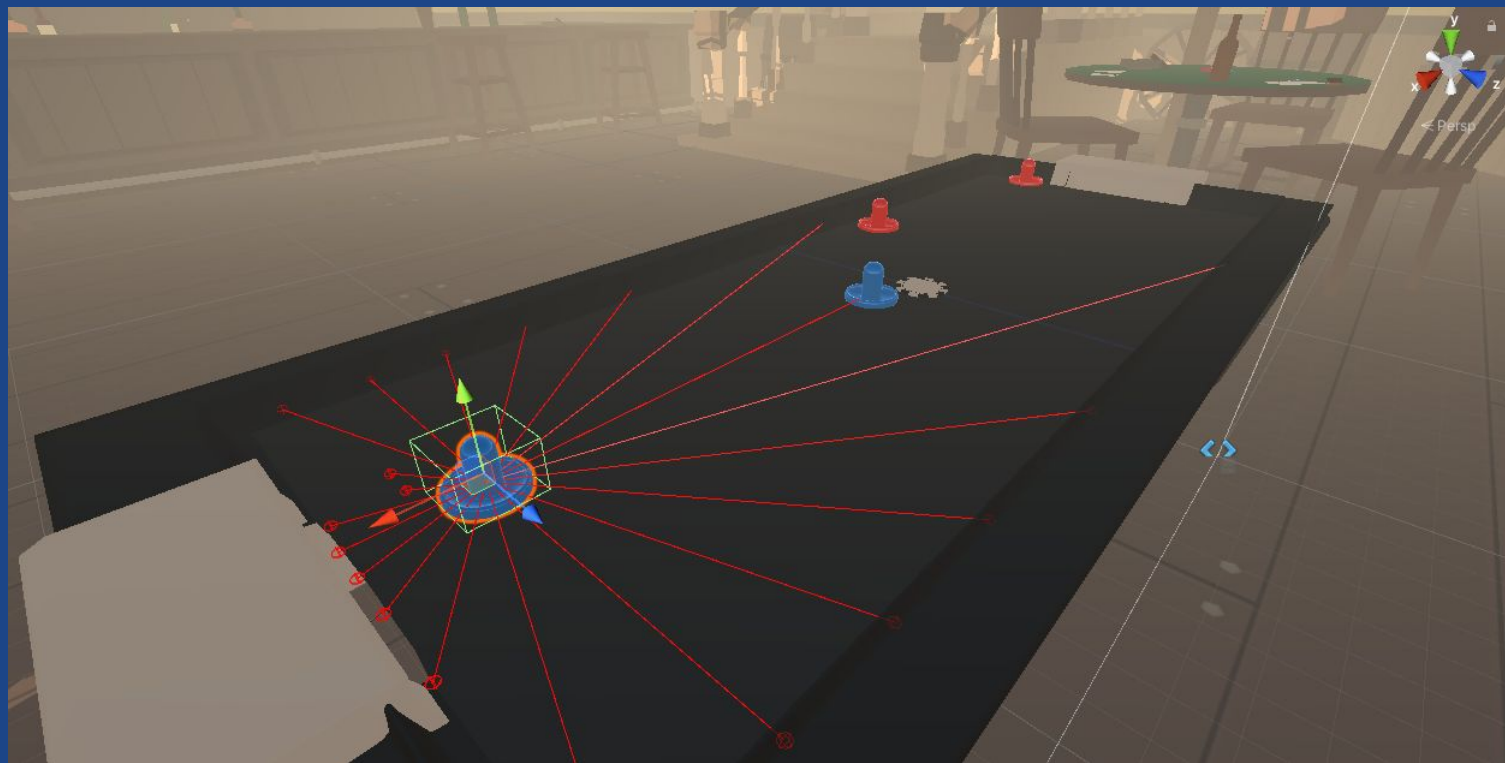


Group Cumulative Reward
tag: Environment/Group Cumulative Reward





RESULTADOS



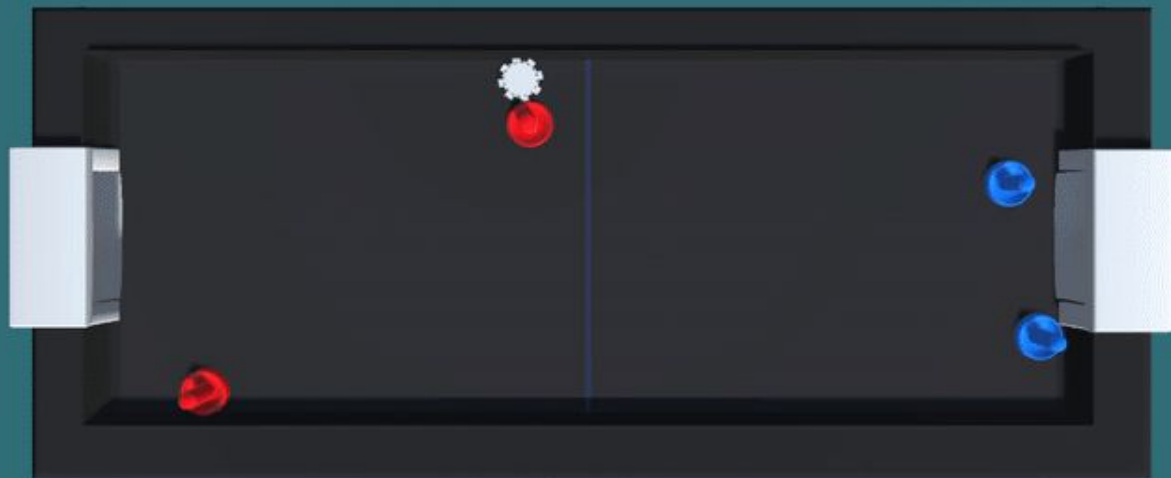


RESULTADOS



[Go back to menu](#)

2 - 0





¿PREGUNTAS?