

# HW2

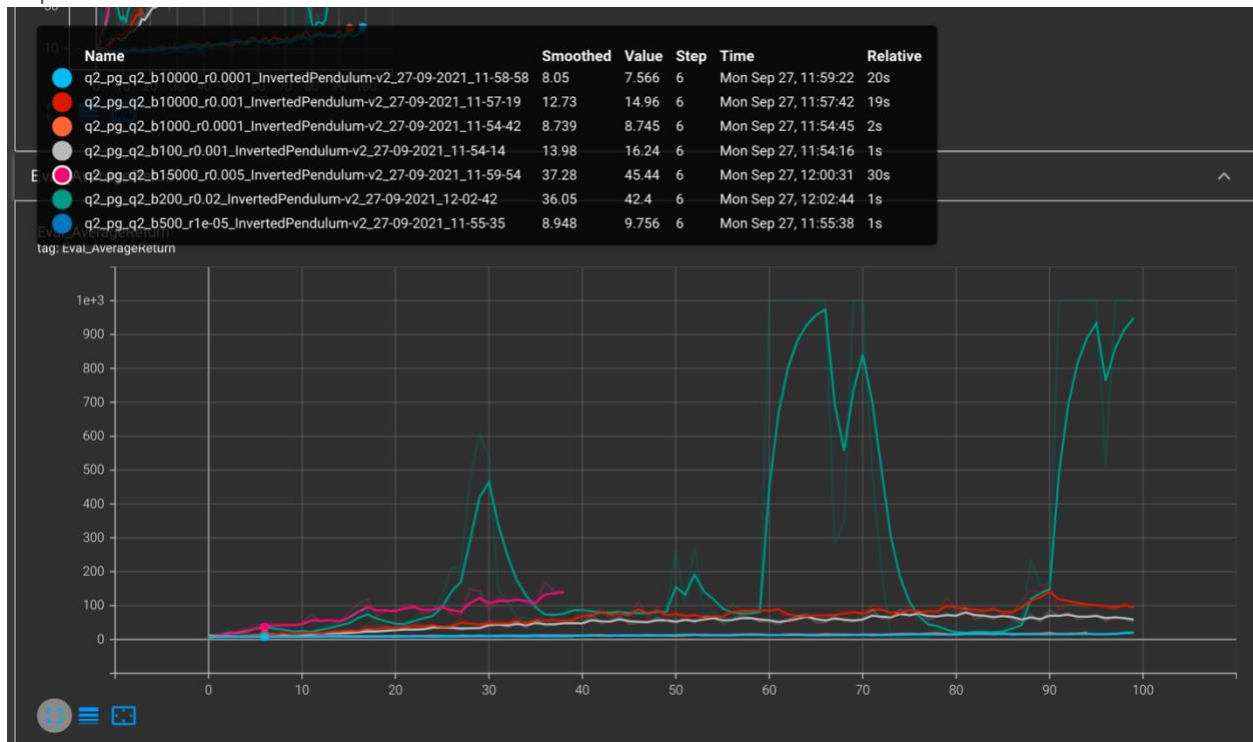
Alejandro Escontrela

## Experiment 1

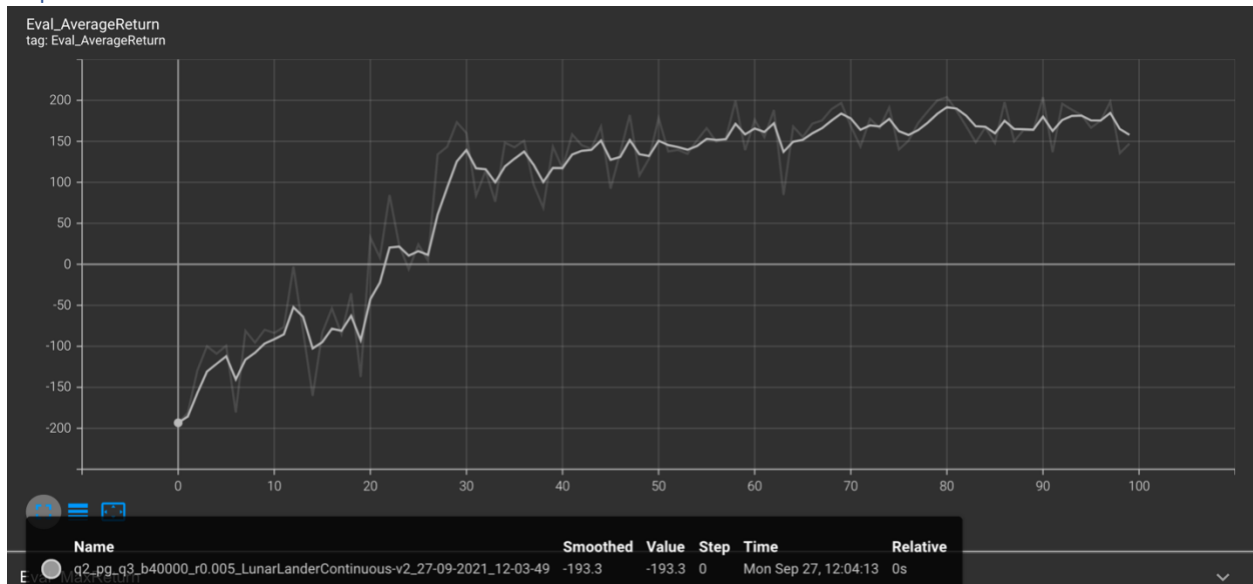


- Reward to go performs significantly better
- Normalizing advantages leads to more stable learning
- Batch size reduced the amount of training steps required to achieve the max reward

## Experiment 2

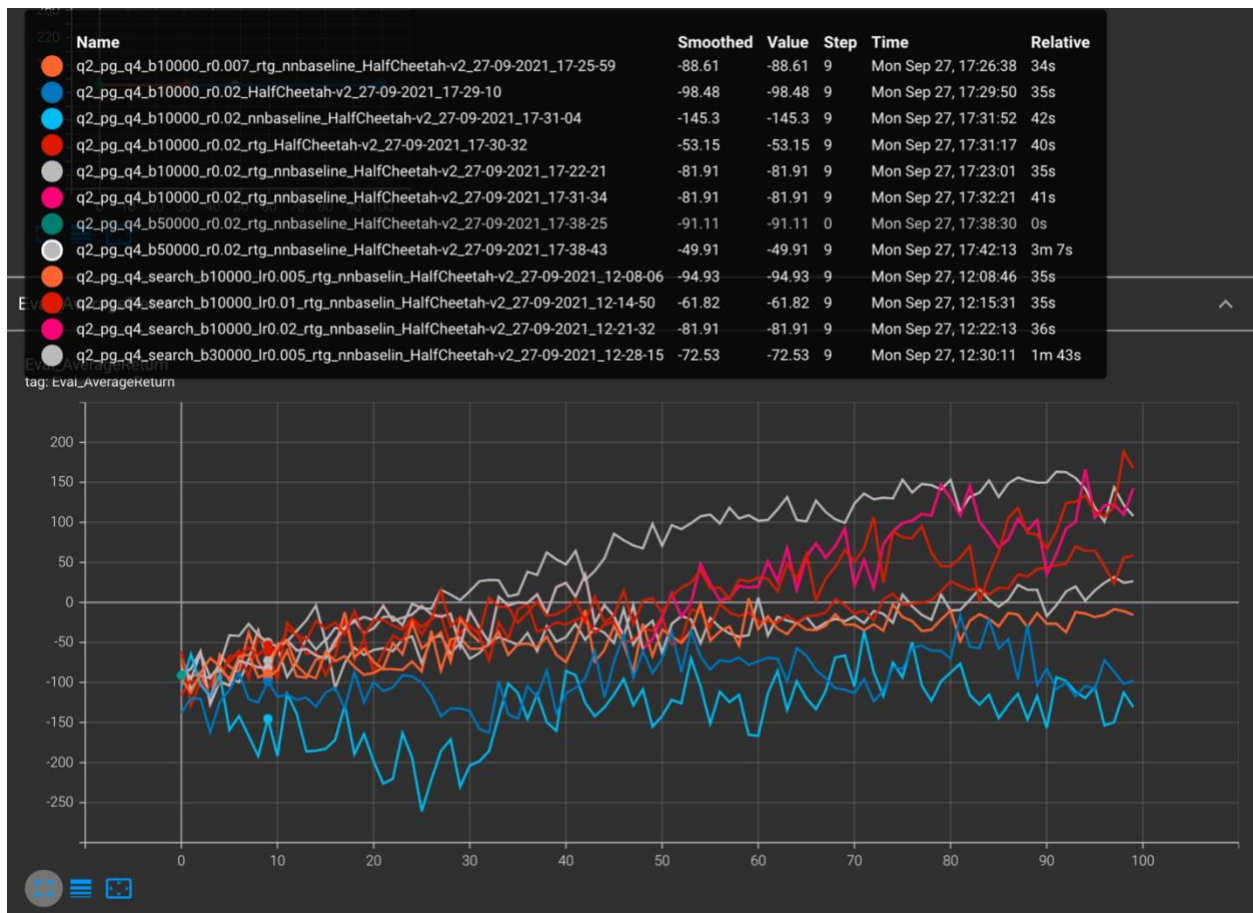


## Experiment 3



## Experiment 4





## Experiment 5

