Default 🡪 Blue

* max\_steps: 3.0e4
* epsilon: 0.2
* beta: 1.0e-2
* hidden\_units: 256
* learning\_rate: 3.0e-4

Epsilon 🡪 Red

* max\_steps: 3.0e4
* **epsilon: 0.5**
* beta: 1.0e-2
* hidden\_units: 256
* learning\_rate: 3.0e-4

Beta 🡪 Grey

* max\_steps: 3.0e4
* epsilon: 0.5
* **beta: 1.0e-1**
* hidden\_units: 256
* learning\_rate: 3.0e-4

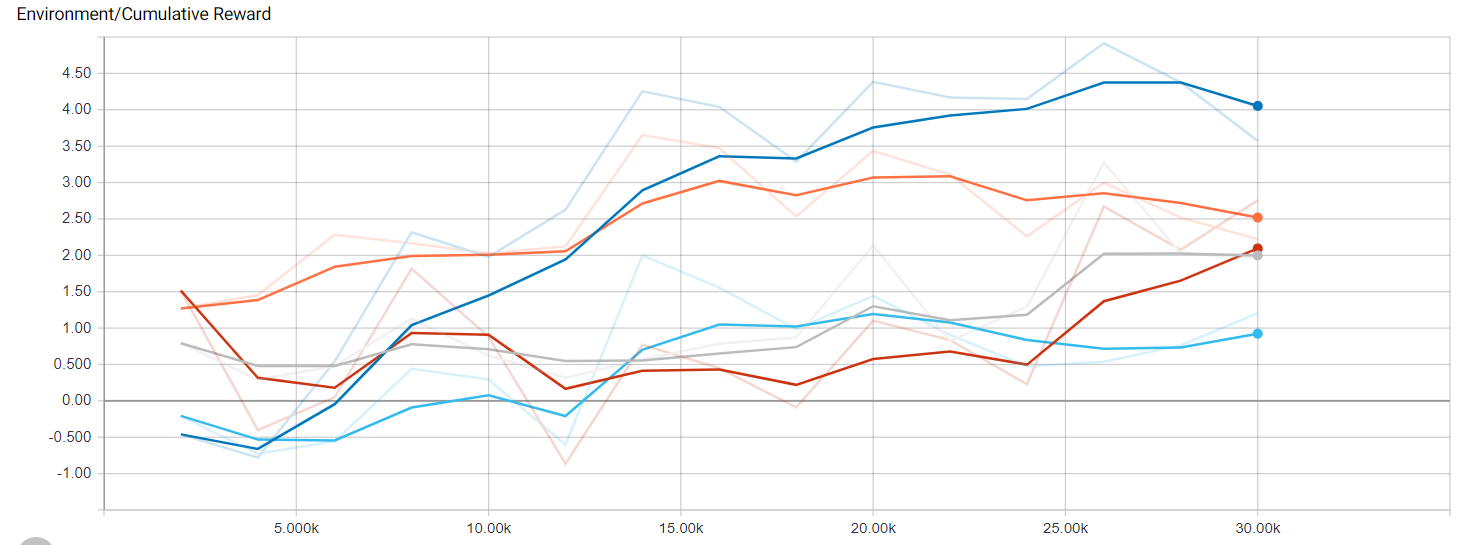
Hidden 🡪 Orange

* max\_steps: 3.0e4
* epsilon: 0.5
* beta: 1.0e-1
* **hidden\_units: 512**
* learning\_rate: 3.0e-4

Learning 🡪 Green/Blue

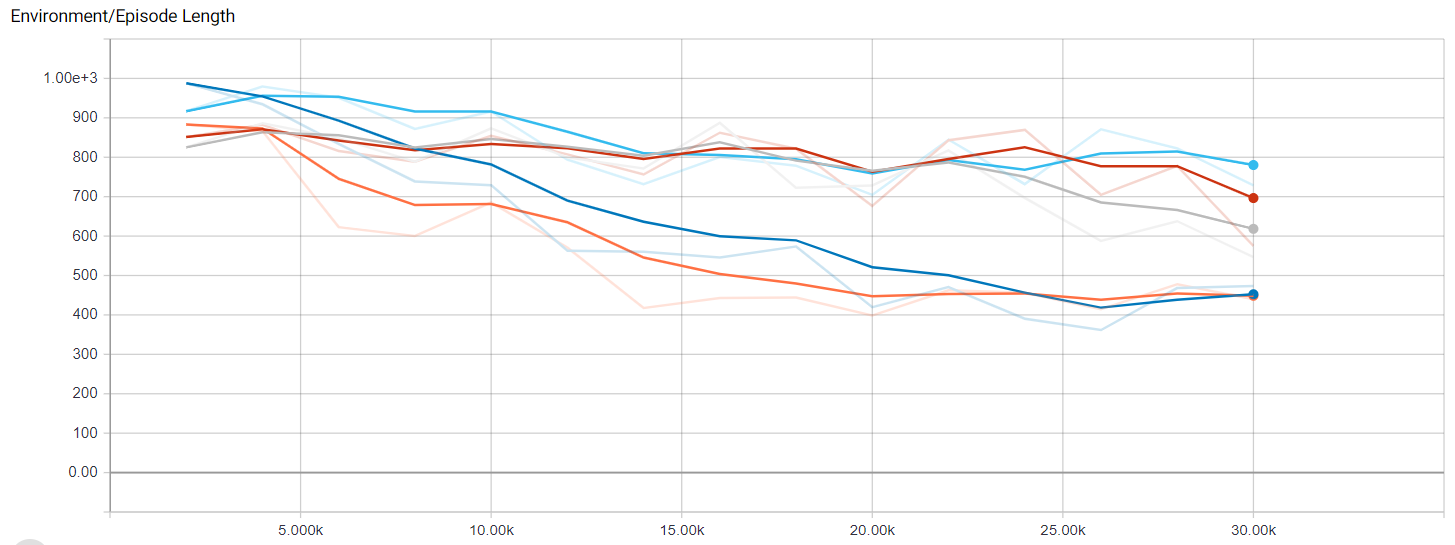
* max\_steps: 3.0e4
* epsilon: 0.5
* beta: 1.0e-1
* hidden\_units: 512
* **learning\_rate: 3.0e-5**

Cumulative reward



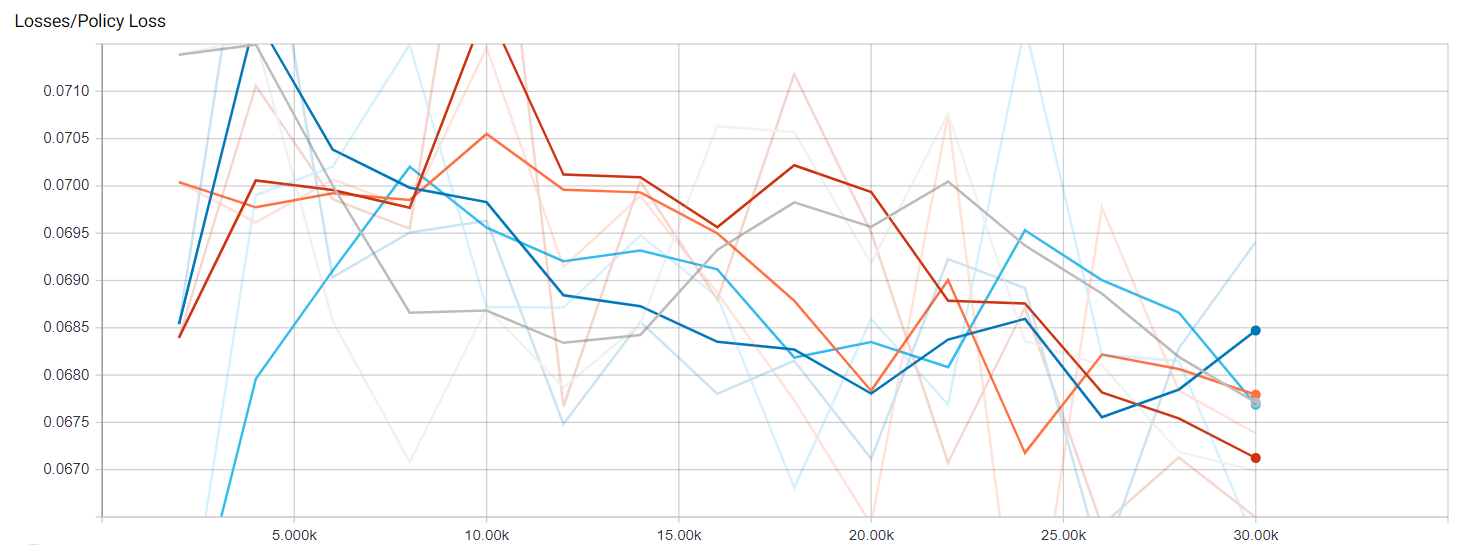
The best one is the Default and the worst are the Learning and the Epsilon

Episode Lenght



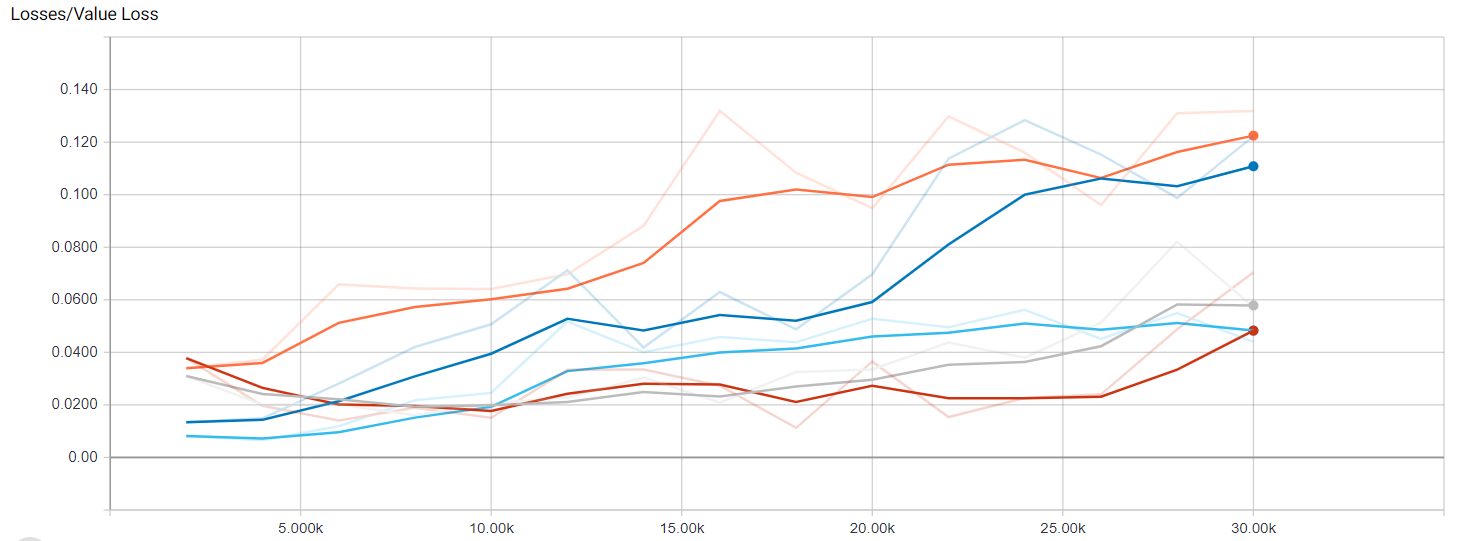
The better ones are the Default and the Hidden

Policy Loss



The better one Epsilon and the worst, the Default

Value Loss



The better ones are the Default and the Hidden, and the worst ones are the Epsilon and the Learning