**Wall Breakout Design Notes**

**Prisoner States:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Idle | Walking | PushingCrate | KneelingDown | JumpingUp | PullingUpCollegue | ClimbingUp |

**Actions:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **None** | **StandUpIdle** |  |  |
| ***Movements*** | **RotateLeft** | **RotateRight** | **WalkForward** | **KneelDown** |
| ***Tactical*** | **PushCrate** | **JumpUp** | **ClimbUp** |  |

*Note the Implicit PullUp Animation will Be as a consequence of other Prisoner Performing a Climb Up Action*.

Mask Out Actions Upon the Ability to perform (See below)

**Observations:**

**Ray Cast Spatial Awareness Orientation:**

* Ray Cast Level: Detect**: [ *Crate, Prisoner, Wall, ExitGate, Enclosure*** **]**
* Ray Cast Upper: Detect: **[** ***Crate, Prisoner, Wall, ExitGate, Enclosure* ]**

***Explicit Spatial Awareness:***

* LocalPosition.x // A Sense of Forward Progress
* LocalRotation.Y // A Sense of orientation
* (Crate.LocalPosition.x –Wall.x) // A Sense of Crate Movement Progress
* IsFacingBackward()
* IsFacingFoward()
* IsOnLowerlevel()
* IsOnCrate()
* IsOnUpperLevel()

***One Hot Encoded Binary representation of current Self state:***

* IsWalking
* IsPushing
* IsKneeling
* IsJumpingUp
* IsPullingUp
* IsClimbingUp

**Action Conditional Indicators:**

* IsAbleToMove() ~ **WalkForward, RotateLeft, RotateRight**
* IsAblePush()
* IsAbleKneel()
* IsAbleJump()
* IsAbleClimb()

**Curriculum Learning Profile/ Reward Profile:**

**Decision Penalty:**

Monitor Changes in Proposed Actions (When Able To Action) within Prisoner Agents => Maintain Group DecsionCounter

Final Episode correction=> ObjectiveReward – 0.5f\* DecsionCounter/ MaxNumberDecisions

If DecsionCounter> MaxNumberDecisions => Interrupt and Reset Episode

**Level Objective Reward Profiles**

|  |  |  |  |
| --- | --- | --- | --- |
| **Level** | **Reward** | **Level Objective** | **Objective Reward Profile**  **(And Will End Episode at this Level)** |
| 1 | 0.2 | Both PrisonersInPushingPosition | +0.2f for both Prisoner |
| 2 | 0.4 | Both Prisoners Pushing Crate | +0.2f for bot Prisoner |
| 3 | 1.4 | Crate Moved to Wall | ~~(3.0f-DistanceToWall)/3.0f~~  +1.0f if Crate at Wall |
| 4 | 1.6 | Prisoner Kneeling At Crate | +0.2f if A Prisoner Kneels In Front of Crate |
| 5 | 1.8 | PrisonerIsAbleToJump Onto Crate | +0.2f if A prisoner Achieves Ability to Jump |
| 6 | 2.0 | Prisoner On Crate | +0.2f if A Prisoner is On the Crate |
| 7 | 2.2 | Prisoner Kneeling On Crate | 0.2f if A Prisoner Kneels On Crate Facing Backwards |
| 8 | 2.4 | PrisonerIsAbleToClimb Onto Crate | +0.2f if Prisoner has Ability to Climb Onto Crate |
| 9 | 2.6 | Both Prisoners On Crate | 0.2f if A prisoner On Crate Achieves Ability to Jump |
| 10 | 2.8 | Prisoner Kneeling On Crate | +0.22f i2f A Prisone2r Kneekls On Crate Facing Wall (Both Prisoners on Crate) |
| 11 | 3.0 | PrisonerIsAbleToJump Onto Wall | 0.2f if A prisoner Achieves Ability to Jump (Both Prisoners on Crate) |
| 12 | 3.2 | Prisoner On Upper Level | +0.2f if A Prisoner is On the Crate |
| 13 | 3.4 | Prisoner Kneeling Facing Back On Wall | +0.2f if A Prisoner Kneeling at Wall facing back On Upper Level |
| 14 | 3.6 | PrisonerIsAbleToClimb Onto Wall | +0.2f if Prisoner has Ability to Climb Wall |
| 15 | 3.8 | Both Prisoners on Upper Level | +0.2f if Both Prisoners on Upper level |
| 16 | 4.0 | Both prisoners Facing Gate on Upper Level | Both Prisoners Facing Forward and Local X greater than +5.0f |
| 17 | 5.0 | Prisoners Escaped Gate | +5.0 if Both Prisoners Escaped |
| **18** | **5.0** | **Prisoners Escaped Gate – Final Level Extended No Promotion** | **Ditto** |

**Level Reward**: = Current Level ObjectiveReward + 0.5\*Previous Level ObjectiveReward+ 0.25\*Previous-2 Level ObjectiveReward

**Curriculum Learning Level Progression:**

Episode: Increment ObjectiveCount if Objective Succeed, and Decrement if Objective Not Achieved

Advance Level IF ObjectiveCounts>Level Thresholds (c.f. 25).

**Training Runs**

**Training Run 1**

Gets Stuck on Level 7: Satisfactory Kneeling on Crate. Seems then to fail to Push the crate etc.

**Training Run 2:**

Removed FacingBackwards as an Objective, since can Climb up onto crate so long as Prisoner and crate are aligned for a climb.

Increased Decision Count to 8

**Training Run 3:**

Increased the Sub reward for Crate to the Wall deserves a more significant reward, as a lot of movement effort involved. + 1.-f for having the Crate at the Wall.

Crate Initialised closer to the Wall, and the Prisoners/Crate Proximately thresholds a bit wider. (Easier)

Reaches Level 7: Within ~ 1.5 Million Steps Learns to move the crate, and jump onto the crate (then walks off) to accumulate the Partial +2.0 Reward. So difficulty in Discovering stopping on the Crate and Kneeling down.

Perhaps Decision Boundary (at 8) is now a but too long, keeps walking then off the care.) However note that has time to rotate on the Crate – Which implies also has the Action option and time to Kneel down. Subject to discovery. A “KneelingOnCrate” Objective was achieved across several environments – so then it’s a matter of repeating this consistently.

* Most environments have hit Level 8 at around 3 Million Global Training Steps

So Kneeling on Crate, but Typically facing away (not in proximity) so unable to Climb up. (*Note a few environments did get to Level 9, two to Level 10 and one got to Level 11 !*

**Note The CheckRequestPullUp Is Flawed** for Climbing on the Crate. It Still presumes Prisoner has to IsFacingBackwards. Instead it should only need to be Prisoners facing Each other Check. (This would also Applies to Wall Climb Up as well.)

**Training Run 4:**

* A new Check PrisonersFacingEachOther(), based upon their Fwd vectors.
* Used in Pull Up Request which just needs InProximartyToOther Priosner() + PrisonersFacingEachOther() for the Pull Up Climb Requests to be correct.
* Replace All use of IsFacingBackward() by PrisonersFacingEachOther() including the SensorObservation

Majority of environment reach to level 13 ! At 9 Million Steps => Waiting for Prisoner on Upper level to Kneel Down on top of Wall ***AND facing the other player***.

Maybe worth removing the **PrisonerIsFacingEachOther**() Objective Level Check on At level 13, as the same check will be applied at Level 14 anyway.

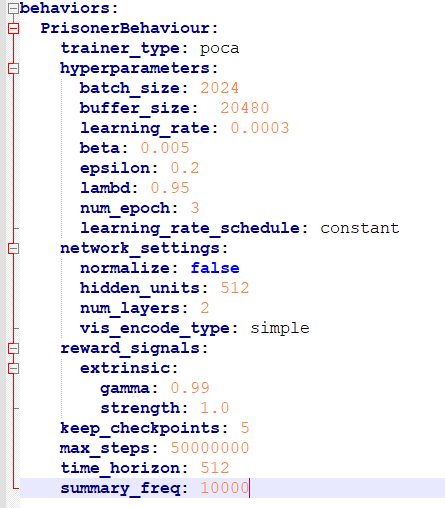
*But Not so obvious, or deliberate moving of Crate at the Level 13 ? And partial Rewards seem to have slumped to remain around 0.4f*

Note Level 13 and above do not have the: && (TheCrateObject.transform.localPosition.x > 2.25f) in their Objective functions: => Suggest Add to ALL Objectives

**Training Run 5 (Baseline):**

Some Objective and Clause Revisions:

* Remove **PrisonersFacingEachOther**() clause in Level 13: KneelingOnWallCheck()
* Add **(TheCrateObject.transform.localPosition.x > 2.25f**) to ALL Checks at level 13 and Above
* MLAgent Line 216: Add Clause isFacingForward() to the Pushing Crate Action – so as to ensure really only pushes the Crate is typically, Only Pushed Forward, closer to the Wall. (This will then be consistent with the corresponding Objective)
* Make reward for Achieving level 16: Prisoners escape = +5.0f to promote Final Movements on Upper Level to Exit gate.



**Figure 1 Baseline Configuration**

**DecsionRequester Period = 8**

Many/Most Environment reach Level 16, but then Performance and behaviours collapse and all Rewards collapse to Zero.

* Loss and Rewards all Collapse around 10 Million, when the Experiment reaches Level 16.
* No Prisoners ever escaped , so no Ultimate Reward
* But No Demotions Occurred either – despite failing to meet the Objectives



**Figure 2 Run 5 Policy Collapse at level 16**

Perhaps Time Horizon to 128 or 256 ? - But Time Horizon is already at 512

But the Rewards Profile Collapses at Level 16.

**Training Run 6**

* Introduce new level 16 – Moving Towards Gate [Level 16 Objective: Both Prisoners Facing Forwards + Local Position.x greater than 5.0f]
* Reduce Time Horizon to 256
* Reduce the Epsilon to 0.1 (from 0.2) ~ presume more policy Clipping helps
* Reduce Learning Rate to 0.00015 (From 0.0003) ~ More conservative Learning

It was noted that was able to Jump from Floor all the way up onto Wall – So need to check Jump Step offset and propose Reduce from 1.0 to 0.8f;

Demotion Count is Not being Incremented upon Max Decision count outs.

**Training Run 7:**

* Demotion Count++ is Missing ! Correct this and set Demotion ~~threshold back to 100~~
* Check Jump Step offset and propose Reduce from 1.0 to 0.8f;
* ~~Reduce Demotion Threshold at 50~~

Can Still Jump to Upper Wall Without Using Crate => which may be warping the Level 14/15 Objectives. Reduce Step further to 0.6f and test.

Climbs to level 15, but then struggles, and still loses ambition to move the Crate.

Seems to be able to get Both on the Wall, but Not meet objective or much Partial rewards. Need to ensure that ALL previous rewards are captured.



**Figure 3 Missing partial and Level Objective** !

Possibly because the Crate is not recognised as being >2.25, even though its sufficiently close for the Prisoners to use.

* + Reduce Crate at Wall threshold to >2.0

This would also explain, why not Pass the **BothOnUpper**() in the above screenshot, as that also include the crate >2.25 ? And can be seen that the carte is not quite at the wall

Demotions do not work, the Demotion Exceeds 50, and yet No Demotion !

Actually they do work, except the Code was Not saved before reducing back down to 50 from 100.

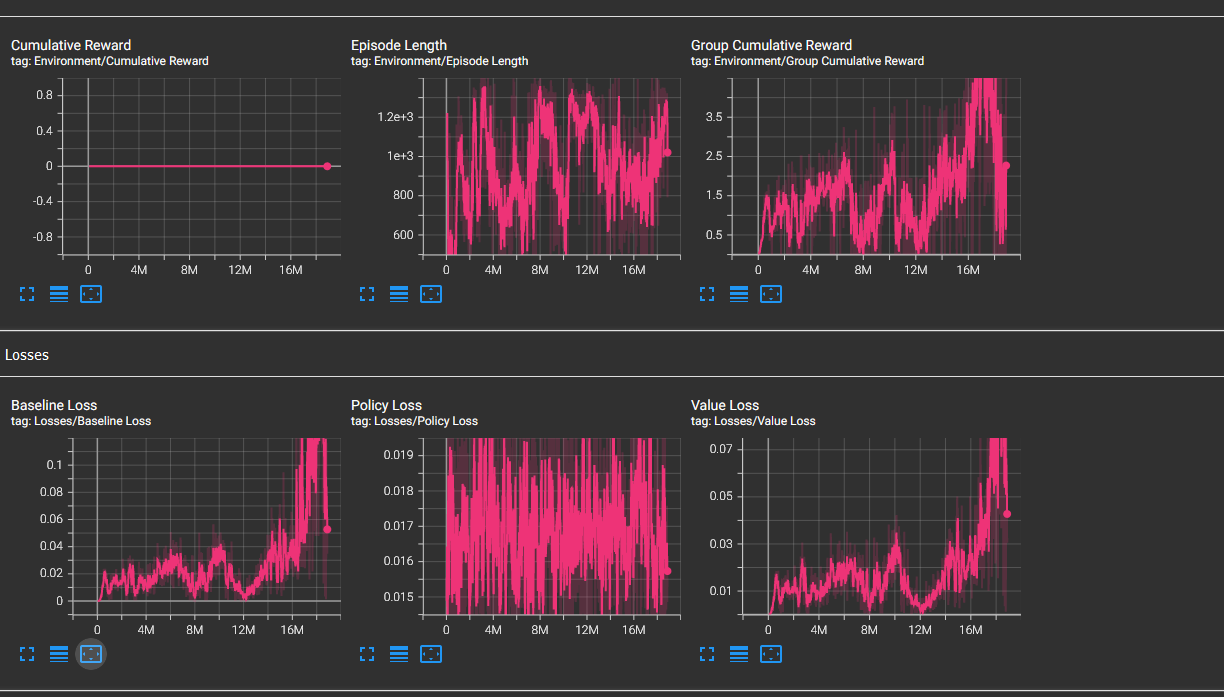
**Training Run 8:**

* + Reduce Crate at Wall threshold to >1.75f – Since the Prisoner Radius, would enable player to still be on the crate, and close to the Wall (Hopefully will then recognise Partial Objectives)
  + Check Reduce further down to 0.6f, and Test to confirm.
  + Confirm Demotion Threshold is set to 50
  + Reduce player Radius to 0.3f All Time, as seems to be standing on Upper Wall, despite not standing on Wall (Will be at 0.3f All the Time)
  + A few other bits of Prisoner fine tuning thresholds

By 18 Million Steps:

**Achieves Level 18, and Level 19 the Prisoner Escape events do occur.**

However not consistent performance and behaviours. But No Policy Loss, is still learning



And at level 18, never Promote beyond 18. – Just Keep Rewards/ Reset cycles.

**Training Run 9:**

***But Need to be careful, the Level 18 Reward Profile Needs to be Equivalent to level 17, so avoid double accounting for the same objective being reached to confuse the estimated reward profile.***

* So only Add +4.8 at Level 18, because 0.2f will also have already double accounted for.
* And Do Not **Promote** Level Beyond Level 18, So As to keep Cycling through the same Rewards/Resets at level 18 – Until end of Training. Hopefully will stay at and reinforce Level 18 behaviour and improve performance.

At 28 Million: Reaches Level 17 But Does not achieve Prisoner breakouts. And does not see rewards above 2.0. See below Run 9 in Green (Previous Run 8 in red)

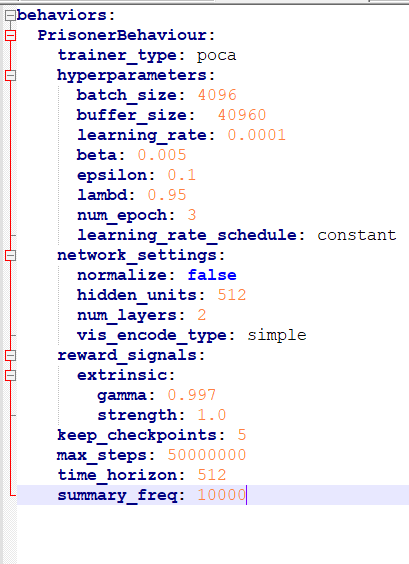


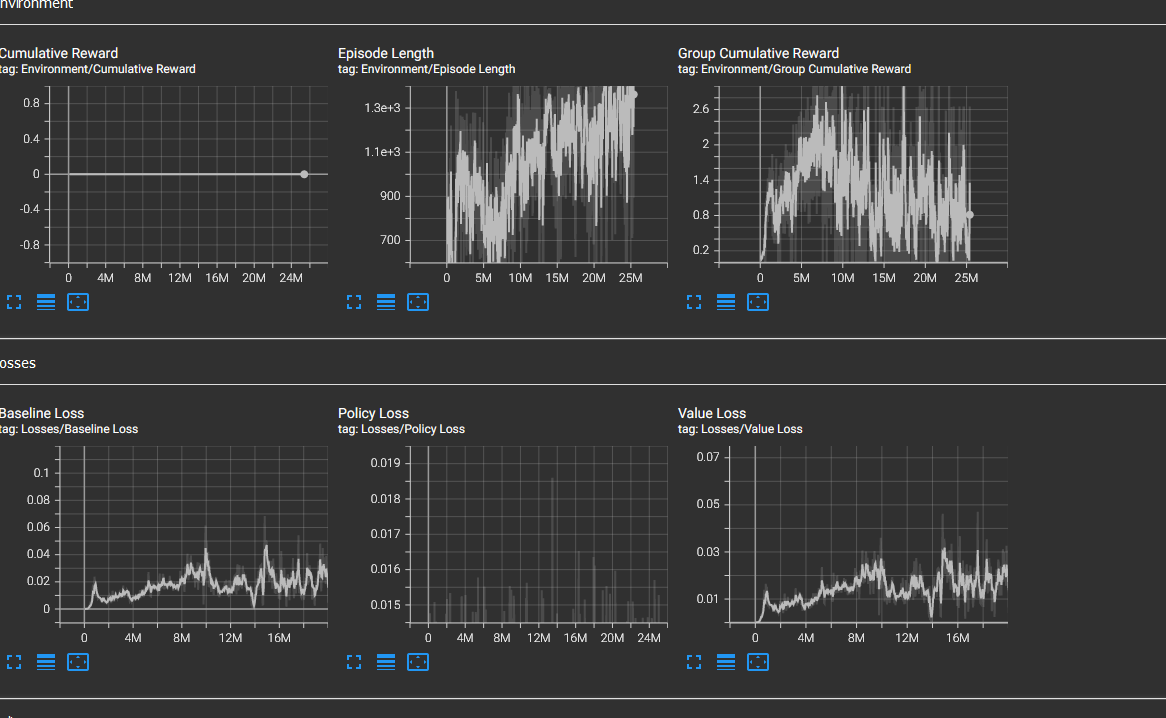
*This could just be random poor luck compared to previous Run 8. ( c.f. Immersive Limit Experience of different run (Slower reward spurts) Experience with same Hyper parameters)*

**Run 10:**

Change some POCA Hyper Parameters:

* Increase Batch size to 4096 (from 2024)
* Reduce learning Rate to 0.0001 (from 0.000015)
* Increase Buffer Size to 40960 (from 20480)
* Increase Time Horizon to 512 (from 256)
* Increase Gamma to 0.997 (from 0.99)





Again struggles On level 17 – Both on the Upper Level, Facing The Gate. But then Difficulty in discovering the Motivation to go towards the Gate. Noting No additional + Reward Signal Until actually stumbles across hitting the Exit gate, within the 50 Demotion Runs.

***Note previous Run 8 Achieved some prisoner escape to get to level 18, but Run 9 and 10 have been struggling at level 17. ~~Perhaps save Run 8 Brain (But extra Observations changed Brain~~)***

Perhaps add a Motivational Reward at Level 16 & 17 (If Both On Upper level:) For distance moved forward. Add another partial Reward ~ Average of both + 1.0f\* (Prisoners’ -Wall.x) /2.0f

**Run 11:**

* Change the ML Agent: Crate Observation to relative to (+**1.75f**-Crate.x) instead of (2.5f-Create.x)
* Add new Observation – InProximatoryToCoPrisoner() AND Increase the Observations Count up to 23 in all ML Agents Prefabs for both Prisoner Agents
* If ( Level >=16 ) && BothOnUpperCheck() ) Then:

PartialReward = PartialReward + (Prisoner1. Local.x-4.0f)/2.5 + (Prisoner2. Local.x-4.0f)/2.5

Still Stuck on Level 17, after 30 Million Runs

**Run 12:**

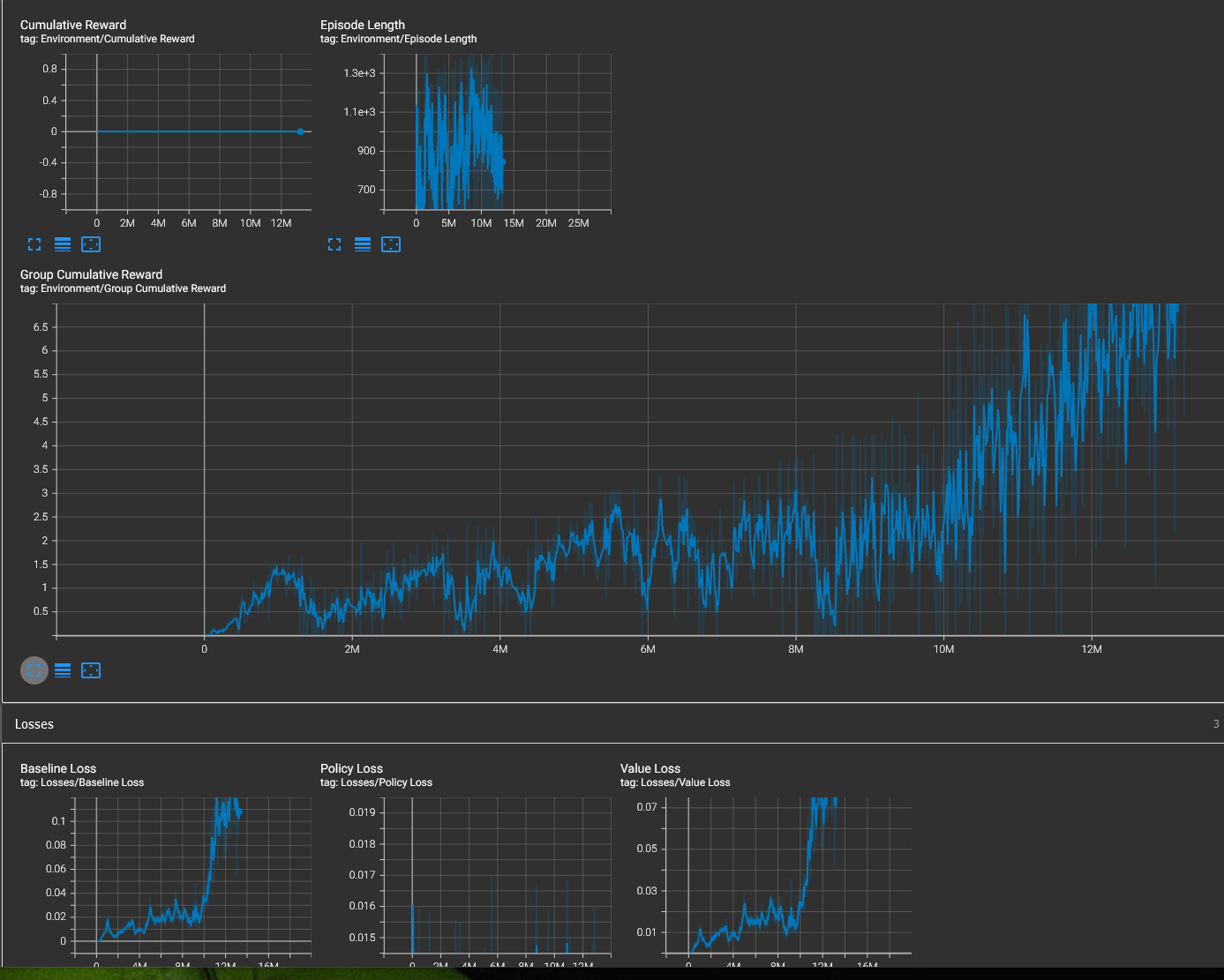
* ifBothOnUpper() Mask Prisoner Agent Actions to None, Walk, Rotate Left, Rotate Right only – No other Kneel, Push, Idle, Jump, Climb Actions

Now Stuck on Level 16 !

**Run 12 again**

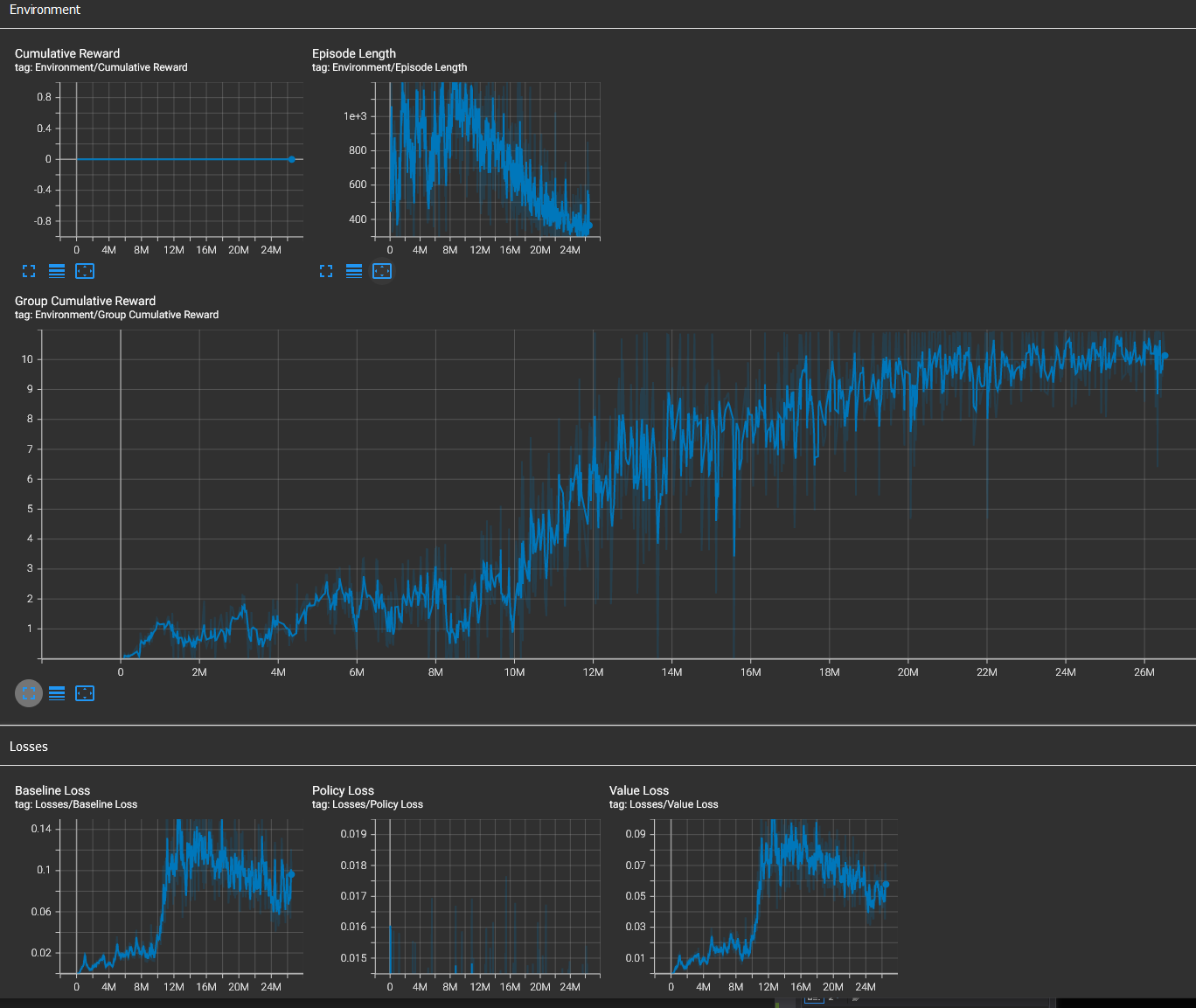
* Allow StandToIdle Action on the Upper Level 16 and 17
* Only Mask the other Actions if >4.0f (So as to ensure Climb, Kneel etc are still supported at Wall Edge)

At 12 Million Runs. Finally getting some prisoner breakouts and All environment Reaching level 18s. With reward signal averaging more than 5.0.

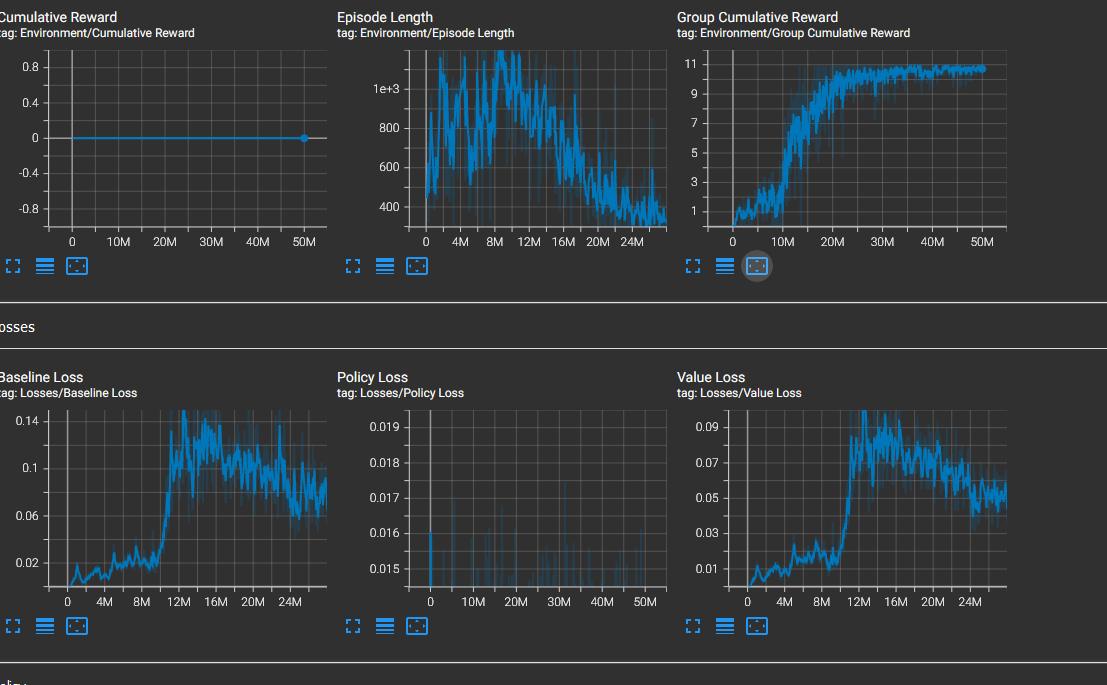


*Keep running to see if develop consistent and more robust behaviours.*

At 26 Million Steps:



And Finally at full 50 Million Steps

****

So learns to Break out but it exhibits a particular style. See the Video Runs:

* Able to push the Crate, even though not behind – Can push when on far side. The Prisoners then slip around to the far side.
* Then Jumps onto Crate from Side, kneel down. Other Prisoner The Climbs Twice from the Side onto the Crate, and then Up onto Wall, whilst the original prisoner remains Kneeling Down.

**Run 13:**

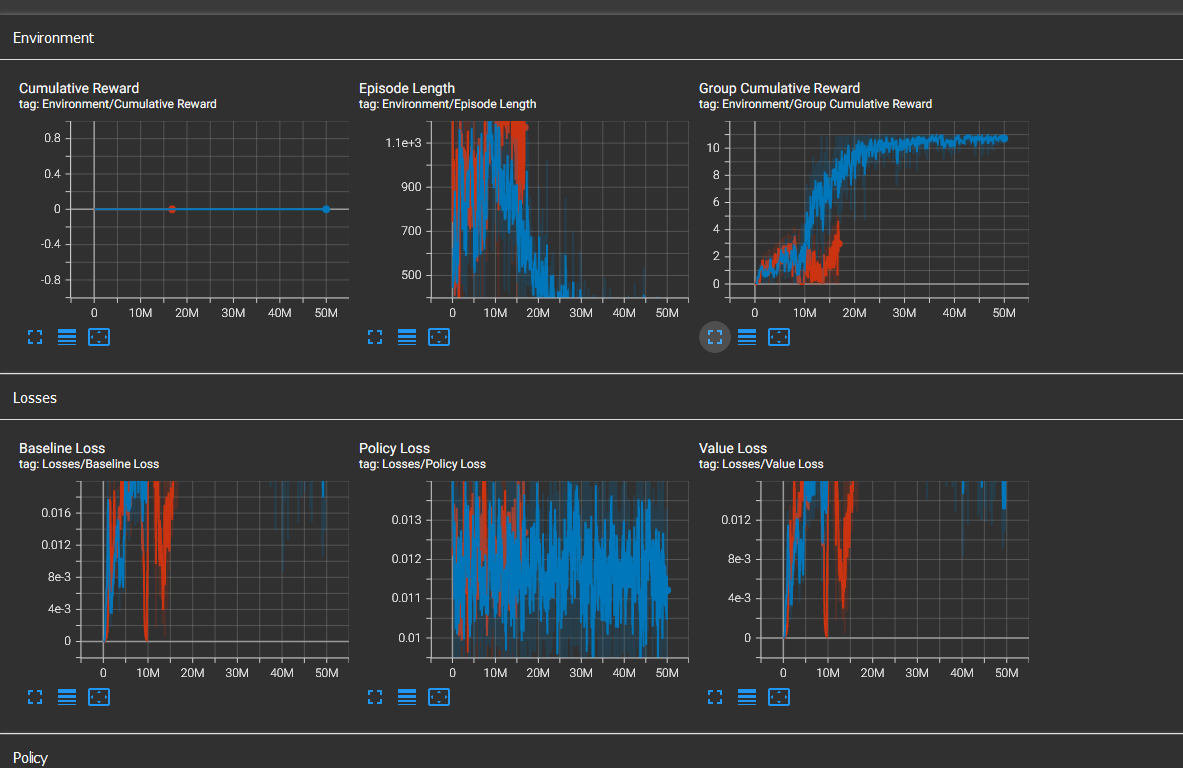
* Can Only Keep Pushing the Crate if in In**PushingZone**() Both Prisoners are abs(Prisoner.z within Crate.z) and Crate.x-Prisoner.x >1.0 Behind the Crate)
* AddObservation In**PushingZone()**
* Only able to Climb IfFacingFoward()
* Change the Crate Reward to an incremental Partial Crate Movement Reward proportional toward 0.5\*1.75f

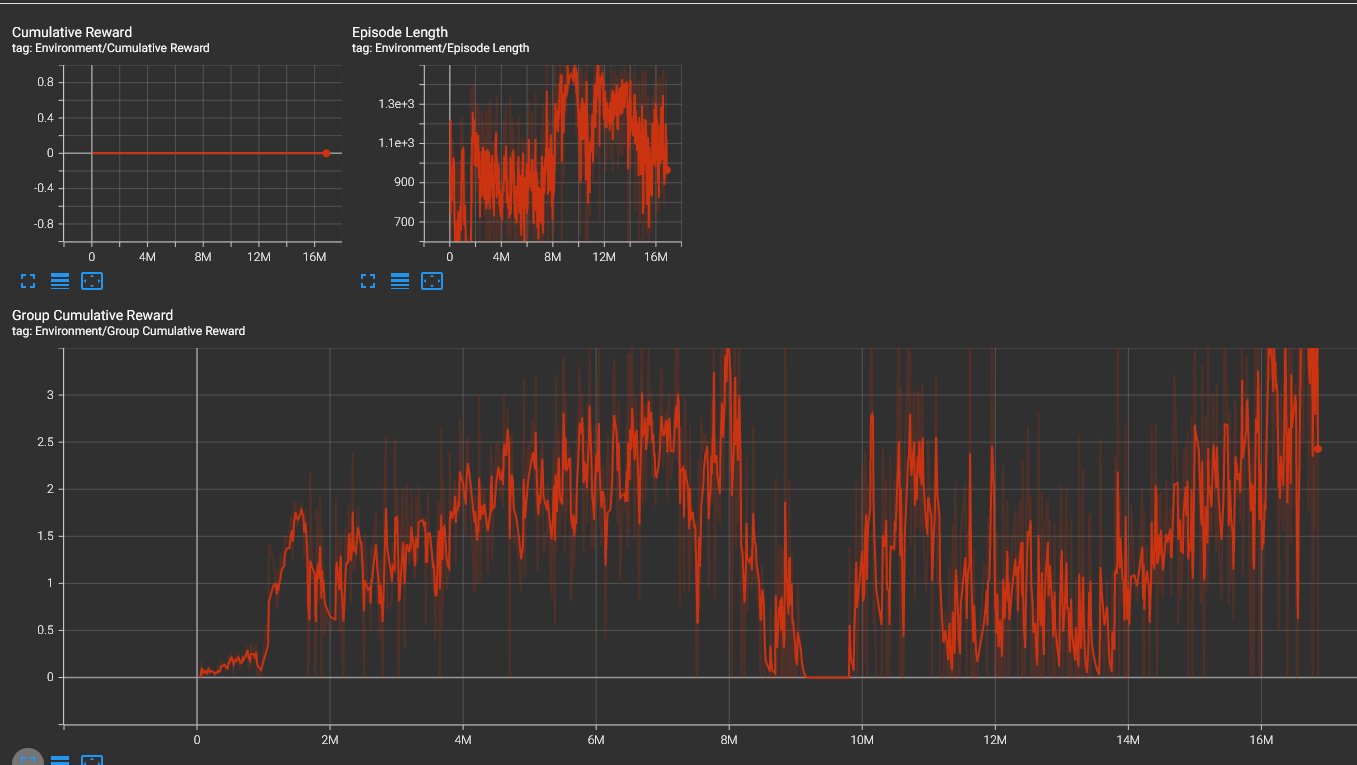
*The Agents are now finding it difficult to discover any advantage to Pushing the Crate, or achieving Level 3.*

***Note crate Movement reward is never achieved, as Level 2 – Resets before Push Applied, and Level 3 Rewards are only assigned upon Meeting the Crate at the Wall. So the partial reward is not achieving any motivation before the crate has actually been pushed to the wall. => may need another 3 Level: Crate has moved - but by how much !***

After Discovering and Finally Getting through Level 3 – Does then Makes progress, but then stumbles at Level 16. At level 16, the Agents Not consistently moving the crate consistently. So the Crate move reward may Not be enough.

Although Looks as though it May recover at 16 Million, and Crate is more consistently Moving: (**Run 13 in orange** compared to Blue Run 12)

******

******

*Some recovery at 16 Million - However will Abort and set Run 14*

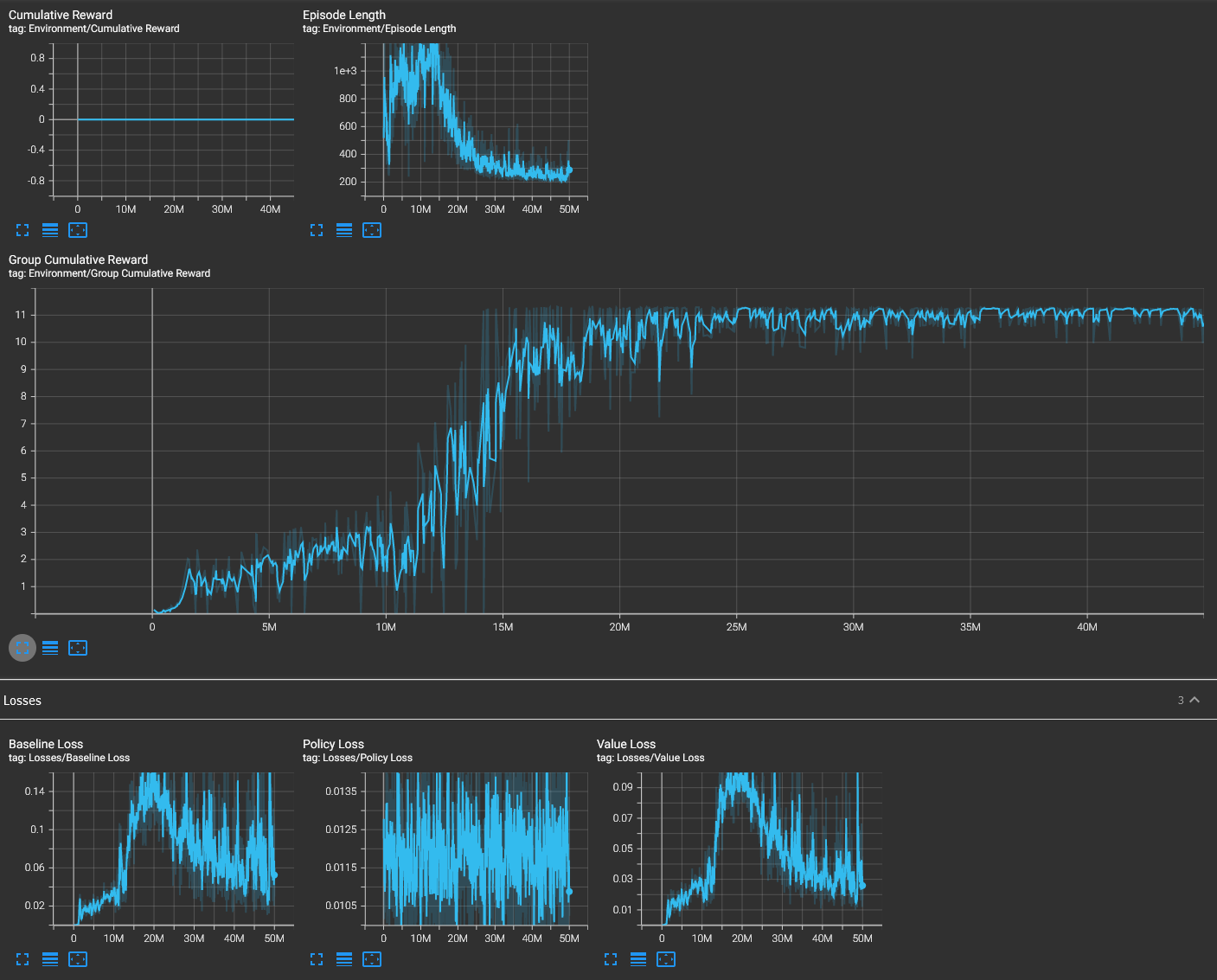
**Run 14:**

* Add new Level 3: Crate Moved, upon Crate.x > 0.5f
* Revert the Crate PartialReward Profile to a Full +1.0f, only if Crate at Wall
* Reduce Demotion Threshold to 40
* Increase Promotion Threshold to 12

|  |  |  |  |
| --- | --- | --- | --- |
| **Level** | **Reward** | **Level Objective** | **Objective Reward Profile**  **(And Will End Episode at this Level)** |
| 1 | 0.2 | Both PrisonersInPushingPosition | Both Prisoners |
| 2 | 0.4 | Both Prisoners Pushing Crate | Both Prisoners |
| 3 | 0.6 | CrateHasMoved | Crate.X > 0.5f |
| 4 | 1.6 | Crate Moved to Wall | Crate.x>1.75 |
| 5 | 1.8 | Prisoner Kneeling At Crate | If A Prisoner Kneels In Front of Crate |
| 6 | 2.0 | PrisonerIsAbleToJump Onto Crate | If A prisoner Achieves Ability to Jump |
| 7 | 2.2 | Prisoner On Crate | If A Prisoner is On the Crate |
| 8 | 2.4 | Prisoner Kneeling On Crate | If A Prisoner Kneels On Crate Facing Other Prisoner |
| 9 | 2.6 | PrisonerIsAbleToClimb Onto Crate | If Prisoner has Ability to Climb Onto Crate |
| 10 | 2.8 | Both Prisoners On Crate | Both Prisoners |
| 11 | 3.0 | Prisoner Kneeling On Crate | If A Prisoner Kneels On Crate Facing Forward && Both Prisoners on Crate |
| 12 | 3.2 | PrisonerIsAbleToJump Onto Wall | If A prisoner Achieves Ability to Jump && Both Prisoners on Crate |
| 13 | 3.4 | Prisoner On Upper Level | If A Prisoner is On the Upper Level |
| 14 | 3.6 | Prisoner Kneeling Facing Back On Wall | If A Prisoner Kneeling at Wall facing other Prisoner && On Upper Level |
| 15 | 3.8 | PrisonerIsAbleToClimb Onto Wall | If Prisoner has Ability to Climb Wall |
| 16 | 4.0 | Both Prisoners on Upper Level | if Both Prisoners on Upper level |
| 17 | 4.2+ delX | Both prisoners Facing Gate on Upper Level | Both Prisoners Facing Forward and Local X greater than +5.0f |
| 18 | 5.0 + delX | Prisoners Escaped Gate | +5.0 if Both Prisoners Escaped |
| **19** | **5.0 + delX** | **Prisoners Escaped Gate – Final Level Extended No Promotion** | **Ditto** |

*Run 14 Seem to Work fine*

*Prisoner breakouts around 18-20 Million Training Steps, and then reinforced optimisation for the remaining 30 steps.*

**

*The Optimised behaviour results in hardly any Kneeling, with just dip downs on Jumps Ups, and little to no time for the Pull Up Animations to play on Climb ups.*

**Further Suggestions**

Increase Decision Count to 20 ?