```
929/1500000: episode: 1, duration: 8.052s, episode steps: 929, steps per second: 115, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
  1753/1500000: episode: 2, duration: 2.604s, episode steps: 824, steps per second: 316, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
  2682/1500000: episode: 3, duration: 2.868s, episode steps: 929, steps per second: 324, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
                                              episode steps: 1042, steps per second: 321, episode reward: -20.000, mean reward: -0.019
  4792/1500000: episode: 5, duration: 3.327s, episode steps: 1068, steps per second: 321, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action:
  5616/1500000: episode: 6, duration: 2.535s, episode steps: 824, steps per second: 325, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
  6458/1500000: episode: 7, duration: 2.641s, episode steps: 842, steps per second: 319, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
                                              episode steps: 1110, steps per second: 325, episode reward: -21.000, mean reward: -0.019 [-1.000, 0.000], mean action
                                      3.418s,
  8593/1500000: episode: 9, duration: 3.149s, episode steps: 1025, steps per second: 325, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
  9530/1500000: episode:
                                         .885s, episode steps: 937, steps per second: 325, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action:
 10542/1500000: episode: 11, duration: 7.427s, episode steps: 1012, steps per second: 136, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action
                                       7.265s, episode steps: 820, steps per second: 113, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
 11362/1500000: episode: 12, duration:
 12258/1500000; episode: 13. duration: 7.951s, episode steps: 896. steps per second: 113. episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
                                         .475s, episode steps: 843, steps per second: 113, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 13986/1500000: episode: 15, duration: 7.678s, episode steps: 885, steps per second: 115, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
 14965/1500000: episode: 16, duration:
                                       8.494s, episode steps: 979, steps per second: 115, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
 15962/1500000: episode: 17, duration: 8.592s, episode steps: 997, steps per second: 116, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
 16861/1500000: episode: 18, duration:
                                         .759s, episode steps: 899, steps per second: 116, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
 17653/1500000; episode: 19. duration: 6.896s, episode steps: 792. steps per second: 115. episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
 18586/1500000: episode: 20, duration:
                                       8.056s, episode steps: 933, steps per second: 116, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
 19488/1500000: episode: 21, duration: 7.815s, episode steps: 902, steps per second: 115, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
                                       9.192s, episode steps: 1066, steps per second: 116,
                                                                                            episode reward: -21.000, mean reward: -0.020 [-1.000, 0.000], mean
 21318/1500000: episode: 23, duration: 7.001s, episode steps: 764, steps per second: 109, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
                                       7.902s, episode steps: 889, steps per second: 113, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
 22207/1500000: episode: 24, duration:
 23166/1500000: episode: 25. duration: 8.301s. episode steps: 959. steps per second: 116. episode reward: -20.000. mean reward: -0.021 [-1.000. 1.000]. mean action:
 24005/1500000: episode: 26, duration:
                                         .405s, episode steps: 839, steps per second: 113, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 24924/1500000: episode: 27, duration: 7.904s, episode steps: 919, steps per second: 116, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
 25855/1500000: episode: 28, duration:
                                       8.023s, episode steps: 931, steps per second: 116, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
 26844/1500000: episode: 29, duration: 8.497s, episode steps: 989, steps per second: 116, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
                                       8.324s, episode steps: 968, steps per second: 116, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
 28730/1500000: episode: 31, duration: 7.977s, episode steps: 918, steps per second: 115, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
                                       7.041s, episode steps: 811, steps per second: 115, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
 29541/1500000: episode: 32, duration:
 30472/1500000; episode: 33, duration: 8.071s, episode steps: 931, steps per second: 115, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
 31584/1500000: episode: 34, duration: 9.559s, episode steps: 1112, steps per second: 116, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action
 32543/1500000: episode: 35, duration: 8.152s, episode steps: 959, steps per second: 118, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
                                       8.254s, episode steps: 972, steps per second: 118, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
 33515/1500000: episode: 36, duration:
 34431/1500000: episode: 37, duration: 7.799s, episode steps: 916, steps per second: 117, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
                                       9.461s, episode steps: 1093, steps per second: 116, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action
 36445/1500000; episode: 39. duration: 7.969s, episode steps: 921. steps per second: 116. episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
                                       8.594s, episode steps: 982, steps per second: 114, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
 37427/1500000: episode: 40, duration:
 38334/1500000; episode: 41. duration: 7.780s, episode steps: 907. steps per second: 117. episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
 39342/1500000: episode: 42, duration: 8.837s, episode steps: 1008, steps
                                                                          per second: 114,
                                                                                            episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action
 40273/1500000: episode: 43, duration: 8.193s, episode steps: 931, steps per second: 114, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
                                       7.131s, episode steps:
                                                              792, steps per second: 111, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
 41065/1500000: episode: 44, duration:
 41985/1500000: episode: 45, duration: 8.315s, episode steps: 920, steps per second: 111, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
 42993/1500000: episode: 46, duration: 9.154s, episode steps: 1008, steps per second: 110, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action
 43757/1500000: episode: 47, duration: 6.904s, episode steps: 764, steps per second: 111, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
                                       7.550s, episode steps: 843, steps per second: 112, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 45565/1500000: episode: 49, duration: 8.533s, episode steps: 965, steps per second: 113, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
 46491/1500000: episode: 50, duration:
                                       7.947s, episode steps: 926, steps per second: 117, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
 47317/1500000: episode: 51, duration: 8.164s, episode steps: 826, steps per second: 101, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
                                       8.540s, episode steps: 880, steps per second: 103, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
 49021/1500000: episode: 53, duration:
                                       7.995s, episode steps: 824, steps per second: 103, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 49966/1500000: episode: 54, duration: 9.098s, episode steps: 945, steps per second: 104, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
 50931/1500000: episode: 55, duration: 9.292s, episode steps: 965, steps per second: 104, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
                                       8.235s, episode steps: 852, steps per second: 103, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 51783/1500000: episode: 56, duration:
 52812/1500000: episode: 57, duration: 9.923s, episode steps: 1029, steps per second: 104, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action
                                       9.127s, episode steps: 957, steps per second: 105, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean acti
 54716/1500000: episode: 59, duration: 9.076s, episode steps: 947, steps per second: 104, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
 55590/1500000: episode: 60, duration: 8.628s, episode steps: 874, steps per second: 101, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
 56414/1500000; episode: 61. duration: 8.201s, episode steps: 824. steps per second: 100. episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 57238/1500000: episode: 62, duration: 7.922s, episode steps: 824, steps per second: 104, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 58062/1500000: episode: 63, duration: 7.799s, episode steps: 824, steps per second: 106, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
                                       8.085s, episode steps: 902, steps per second: 112, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
 58964/1500000: episode: 64, duration:
 59802/1500000: episode: 65, duration: 8.088s, episode steps: 838, steps per second: 104, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
                                        0.008s, episode steps:
                                                               1035, steps per second: 103, episode reward: -20.000, mean reward:
 61629/1500000: episode: 67, duration: 7.630s, episode steps: 792, steps per second: 104, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
 62679/1500000: episode: 68, duration: 10.137s, episode steps: 1050, steps per second: 104, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean actic
 63619/1500000: episode: 69, duration: 8.956s, episode steps: 940, steps per second: 105, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
                episode: 70, duration: 10.311s, episode steps: 1122, steps per second: 109,
                                                                                             episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action
 64741/1500000:
 65686/1500000: episode: 71, duration: 8.412s, episode steps: 945, steps per second: 112, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
                                       8.419s, episode steps: 976, steps per second: 116, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
 66662/1500000: episode: 72, duration:
 67574/1500000: episode: 73, duration: 7.787s, episode steps: 912, steps per second: 117, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
                                       7.061s, episode steps: 824, steps per second: 117, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 69250/1500000: episode: 75, duration: 7.338s, episode steps: 852, steps per second: 116, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
                                       8.345s, episode steps: 977, steps per second: 117, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
 70227/1500000: episode: 76, duration:
 71592/1500000: episode: 77, duration: 11.645s, episode steps: 1365, steps per second: 117, episode reward: -21.000, mean reward: -0.015 [-1.000, 0.000], mean actic
 72782/1500000: episode: 78, duration: 10.195s, episode steps: 1190, steps per second: 117, episode reward: -20.000, mean reward: -0.017 [-1.000, 1.000], mean actic
 73574/1500000: episode: 79, duration: 6.788s, episode steps: 792, steps per second: 117, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
                                       11.912s, episode steps: 1389, steps per second: 117, episode reward: -20.000, mean reward: -0.014 [-1.000, 1.000], mean actic
 75787/1500000: episode: 81, duration:
                                       7.011s, episode steps: 824, steps per second: 118, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 76821/1500000: episode: 82, duration: 8.821s, episode steps: 1034, steps per second: 117, episode reward: -21.000, mean reward: -0.020 [-1.000, 0.000], mean action
 77767/1500000: episode: 83, duration: 8.039s, episode steps: 946, steps per second: 118, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
                                       7.564s, episode steps: 886, steps per second: 117, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
 78653/1500000: episode: 84, duration:
 79599/1500000: episode: 85, duration: 8.021s, episode steps: 946, steps per second: 118, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
                                        .331s, episode steps: 843, steps per second: 115, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
 80442/1500000: episode: 86, duration:
 81263/1500000: episode: 87, duration: 6.936s, episode steps: 821, steps per second: 118, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
 82209/1500000: episode: 88, duration: 8.042s, episode steps: 946, steps per second: 118, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
 83063/1500000: episode: 89, duration:
                                       7.264s, episode steps: 854, steps per second: 118, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 83905/1500000: episode: 90, duration: 7.094s, episode steps: 842, steps per second: 119, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 85106/1500000: episode: 91, duration: 10.086s, episode steps: 1201, steps per second: 119, episode reward: -20.000, mean reward: -0.017 [-1.000, 1.000], mean actic
 86294/1500000: episode: 92, duration: 10.023s, episode steps: 1188, steps per second: 119, episode reward: -21.000, mean reward: -0.018 (-1.000, 0.000), mean actic
 87243/1500000: episode: 93, duration: 8.002s, episode steps: 949, steps per second: 119, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
 88311/1500000: episode: 94, duration: 9.076s, episode steps: 1068, steps per second: 118, episode reward: -21.000, mean reward: -0.020 [-1.000, 0.000], mean action
 89352/1500000: episode: 95, duration: 8.760s, episode steps: 1041, steps per second: 119, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action
 90767/1500000: episode: 96, duration: 12.043s, episode steps: 1415, steps per second: 117, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean actic
 92038/1500000: episode: 97, duration: 10.904s, episode steps: 1271, steps per second: 117, episode reward: -19.000, mean reward: -0.015 [-1.000, 1.000], mean actic
 93376/1500000: episode: 98, duration: 11.224s, episode steps: 1338, steps per second: 119, episode reward: -18.000, mean reward: -0.013 [-1.000, 1.000], mean actic
 94816/1500000: episode: 99, duration: 12.130s, episode steps: 1440, steps per second: 119, episode reward: -18.000, mean reward: -0.013 [-1.000, 1.000], mean actic
 95894/1500000: episode: 100, duration: 8.995s, episode steps: 1078, steps per second: 120, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean actic
 96809/1500000: episode: 101, duration: 7.686s, episode steps: 915, steps per second: 119, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action
 98232/1500000: episode: 102, duration: 11.922s, episode steps: 1423, steps per second: 119, episode reward: -20.000, mean reward: -0.014 [-1.000, 1.000], mean acti
 99451/1500000: episode: 103, duration: 10.290s, episode steps: 1219, steps per second: 118, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean acti
101086/1500000: episode: 104, duration: 13.798s, episode steps: 1635, steps per second: 118, episode reward: -18.000, mean reward: -0.011 [-1.000, 1.000], mean acti
102461/1500000: episode: 105, duration: 11.533s, episode steps: 1375, steps per second: 119, episode reward: -20.000, mean reward: -0.015 [-1.000, 1.000], mean acti
103751/1500000: episode: 106, duration: 10.844s, episode steps: 1290, steps per second: 119, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean acti 105164/1500000: episode: 107, duration: 11.952s, episode steps: 1413, steps per second: 118, episode reward: -20.000, mean reward: -0.014 [-1.000, 1.000], mean acti
106900/1500000: episode: 108, duration: 14.645s, episode steps: 1736, steps per second: 119, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean acti
```

```
108798/1500000: episode: 109, duration: 16.060s, episode steps: 1898, steps per second: 118, episode reward: -15.000, mean reward: -0.008 [-1.000, 1.000], mean acti
110404/1500000: episode: 110, duration: 13.533s, episode steps: 1606, steps per second: 119, episode reward: -17.000, mean reward: -0.011 [-1.000, 1.000], mean acti
                                                                2072, steps per second: 119, episode reward: -18.000, mean reward:
                         111, duration: 17.456s, episode steps:
114112/1500000: episode: 112, duration: 13.825s, episode steps: 1636, steps per second: 118, episode reward: -18.000, mean reward: -0.011
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean acti
116269/1500000: episode: 113, duration: 18.362s, episode steps: 2157, steps per second: 117, episode reward: -18.000, mean reward: -0.008
                                                                                                                                             [-1.000,
                                                                                                                                                     1.000], mean acti
117675/1500000: episode: 114, duration: 11.975s, episode steps: 1406, steps per second: 117, episode reward: -20.000, mean reward: -0.014
                                                                                                                                             [-1.000,
                                                                                                                                                     1.000], mean acti
119883/1500000: episode: 115, duration: 18.754s, episode steps:
                                                                2208, steps per second: 118, episode reward: -15.000, mean reward:
                                                                                                                                                     1.000], mean acti
122585/1500000: episode: 116. duration: 22.972s. episode steps: 2702. steps per second: 118. episode reward: -13.000. mean reward: -0.005
                                                                                                                                             [-1.000.
                                                                                                                                                     1.0001. mean acti
   127/1500000: episode: 117, duration: 21.523s, episode steps:
                                                                 2542, steps per second: 118, episode reward: -16.000, mean reward: -0.006
126751/1500000: episode: 118, duration: 13.780s, episode steps: 1624, steps per second: 118, episode reward: -20.000, mean reward: -0.012 [-1.000, 1.000], mean acti
128369/1500000: episode: 119, duration: 14.061s, episode steps: 1618, steps per second: 115, episode reward: -19.000, mean reward: -0.012
                                                                                                                                                     1.000], mean acti
                                                                                                                                              -1.000,
130136/1500000: episode: 120, duration: 15.006s, episode steps: 1767, steps per second: 118, episode reward: -19.000, mean reward: -0.011
                                                                                                                                             [-1.000,
                                                                                                                                                     1.0001, mean acti
31998/1500000: episode: 121, duration: 15.846s, episode steps:
                                                                 1862, steps per second: 118, episode reward: -16.000, mean reward: -0.009
133534/1500000: episode: 122, duration: 13.151s, episode steps: 1536, steps per second: 117, episode reward: -19.000, mean reward: -0.012
                                                                                                                                            [-1.000.
                                                                                                                                                     1.0001. mean acti
34839/1500000: episode: 123, duration: 11.619s, episode steps:
                                                                 1305, steps per second: 112, episode reward: -20.000, mean reward: -0.015
                                                                                                                                                     1.000], mean acti
136549/1500000: episode: 124, duration: 15.304s, episode steps: 1710, steps per second: 112, episode reward: -17.000, mean reward: -0.010
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean acti
38350/1500000: episode: 125, duration: 16.227s, episode steps:
                                                                 1801, steps per second: 111, episode reward: -17.000, mean reward: -0.009
                                                                                                                                             [-1.000,
                                                                                                                                                     1.000], mean acti
140207/1500000: episode: 126, duration: 16.656s, episode steps: 1857, steps per second: 111, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean acti
                                                                                                                                                     1.000], mean acti
142438/1500000: episode: 127, duration: 20.070s, episode steps: 2231, steps per second: 111, episode reward: -15.000, mean reward: -0.007
                                                                                                                                             -1.000,
144032/1500000: episode: 128, duration: 14.354s, episode steps: 1594, steps per second: 111, episode reward: -19.000, mean reward: -0.012
                                                                                                                                             -1.000,
                                                                                                                                                     1.000], mean acti
146015/1500000: episode: 129, duration: 17.781s, episode steps: 1983, steps per second: 112, episode reward: -15.000, mean reward: -0.008
                                                                                                                                                     1.000], mean acti
147923/1500000: episode: 130, duration: 16.602s, episode steps: 1908, steps per second: 115, episode reward: -16.000, mean reward: -0.008 [-1.000,
                                                                                                                                                     1.000], mean acti
149809/1500000: episode:
                         131, duration: 16.313s, episode steps:
                                                                 1886, steps per second: 116, episode reward: -15.000, mean reward: -0.008
151630/1500000: episode: 132, duration: 15.707s, episode steps: 1821, steps per second: 116, episode reward: -16.000, mean reward: -0.009
                                                                                                                                             [-1.000,
                                                                                                                                                     1.0001, mean acti
. 153943/1500000: episode: 133, duration: 19.749s, episode steps: 2313, steps per second: 117, episode reward: -15.000, mean reward: -0.006
                                                                                                                                                     1.000], mean acti
                                                                                                                                            [-1.000,
155549/1500000: episode: 134, duration: 13.751s, episode steps: 1606, steps per second: 117, episode reward: -17.000, mean reward: -0.011
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean acti
157325/1500000: episode: 135, duration: 15.220s, episode steps:
                                                                 1776, steps per second: 117, episode reward: -16.000, mean reward:
159416/1500000: episode: 136, duration: 18.017s, episode steps: 2091, steps per second: 116, episode reward: -15.000, mean reward: -0.007
                                                                                                                                             [-1.000.
                                                                                                                                                     1.0001. mean acti
61650/1500000: episode: 137, duration: 19.211s, episode steps:
                                                                 2234, steps per second: 116, episode reward: -15.000, mean reward: -0.007
                                                                                                                                                     1.000], mean acti
163819/1500000: episode: 138, duration: 19.159s, episode steps: 2169, steps per second: 113, episode reward: -14.000, mean reward: -0.006
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean acti
165499/1500000: episode: 139, duration: 14.662s, episode steps:
                                                                1680, steps per second: 115, episode reward: -17.000, mean reward: -0.010
167346/1500000: episode: 140, duration: 15.763s, episode steps: 1847, steps per second: 117, episode reward: -16.000, mean reward: -0.009
                                                                                                                                             [-1.000,
                                                                                                                                                     1.0001, mean acti
                                                                 1558, steps per second: 118, episode reward: -19.000, mean reward: -0.012
170405/1500000: episode: 142, duration: 12.694s, episode steps: 1501, steps per second: 118, episode reward: -18.000, mean reward: -0.012 [-1.000,
                                                                                                                                                     1.000], mean acti
172409/1500000: episode: 143, duration: 16.939s, episode steps:
                                                                 2004, steps per second: 118, episode reward: -14.000, mean reward:
                                                                                                                                                     1.000], mean acti
173899/1500000: episode: 144, duration: 12.587s, episode steps: 1490, steps per second: 118, episode reward: -20.000, mean reward: -0.013
                                                                                                                                             [-1.000.
                                                                                                                                                     1.0001, mean acti
176275/1500000: episode: 145, duration: 20.069s, episode steps:
                                                                2376, steps per second: 118, episode reward: -14.000, mean reward: -0.006
178639/1500000: episode: 146, duration: 20.198s, episode steps: 2364, steps per second: 117, episode reward: -13.000, mean reward: -0.005
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean acti
                                                                2358, steps per second: 116, episode reward: -12.000,
180997/1500000: episode: 147, duration: 20.272s, episode steps:
                                                                                                                       mean reward: -0.005
182637/1500000: episode: 148, duration: 14.042s, episode steps: 1640, steps per second: 117, episode reward: -18.000, mean reward: -0.011
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean acti
.84768/1500000: episode: 149, duration: 18.320s, episode steps:
                                                                2131, steps per second: 116, episode reward: -16.000, mean reward: -0.008
186818/1500000: episode: 150. duration: 17.710s. episode steps: 2050. steps per second: 116. episode reward: -17.000. mean reward: -0.008 [-1.000.
                                                                                                                                                     1.0001. mean acti
                         151, duration: 16.310s, episode steps:
188704/1500000: episode:
                                                                 1886, steps per second: 116, episode reward: -15.000, mean reward: -0.008
191061/1500000: episode: 152, duration: 20.282s, episode steps: 2357, steps per second: 116, episode reward: -15.000, mean reward: -0.006
                                                                                                                                             [-1.000.
                                                                                                                                                     1.0001, mean acti
93454/1500000: episode: 153, duration: 20.442s, episode steps:
                                                                2393, steps per second: 117, episode reward: -12.000, mean reward: -0.005
195460/1500000: episode: 154, duration: 17.601s, episode steps: 2006, steps per second: 114, episode reward: -14.000, mean reward: -0.007
                                                                                                                                             -1.000,
                                                                                                                                                     1.000], mean acti
197837/1500000: episode: 155, duration: 21.147s, episode steps:
                                                                2377, steps per second: 112, episode reward: -14.000,
                                                                                                                       mean reward: -0.006
199795/1500000: episode: 156, duration: 17.780s, episode steps: 1958, steps per second: 110, episode reward: -17.000, mean reward: -0.009 [-1.000,
                                                                                                                                                     1.000], mean acti
202068/1500000: episode: 157, duration: 19.849s, episode steps:
                                                                2273, steps per second: 115, episode reward: -15.000, mean reward: -0.007
204311/1500000: episode: 158, duration: 20.303s, episode steps: 2243, steps per second: 110, episode reward: -15.000, mean reward: -0.007
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean acti
206374/1500000: episode: 159, duration: 20.054s, episode steps:
                                                                2063, steps per second: 103, episode reward: -15.000, mean reward: -0.007
208521/1500000: episode: 160, duration: 20.810s, episode steps: 2147, steps per second: 103, episode reward: -16.000, mean reward: -0.007
                                                                                                                                             [-1.000,
                                                                                                                                                     1.0001, mean acti
210162/1500000: episode: 161, duration: 15.698s, episode steps:
                                                                1641, steps per second: 105, episode reward: -17.000, mean reward: -0.010
                                                                                                                                                     1.000], mean acti
                                                                                                                                              -1.000,
212137/1500000: episode: 162, duration: 18.738s, episode steps: 1975, steps per second: 105, episode reward: -15.000, mean reward: -0.008
                                                                                                                                             [-1.000,
                                                                                                                                                     1.000], mean acti
213953/1500000: episode: 163, duration: 17.219s, episode steps:
                                                                 1816, steps per second: 105, episode reward: -15.000, mean reward: -0.008
216631/1500000: episode: 164. duration: 24.506s. episode steps: 2678. steps per second: 109. episode reward: -10.000. mean reward: -0.004
                                                                                                                                             [-1.000.
                                                                                                                                                     1.0001. mean acti
219156/1500000: episode: 165, duration: 23.130s, episode steps:
                                                                 2525, steps per second: 109, episode reward: -13.000, mean reward: -0.005
221503/1500000: episode: 166, duration: 22.432s, episode steps: 2347, steps per second: 105, episode reward: -13.000, mean reward: -0.006 [-1.000,
                                                                                                                                                     1.0001, mean acti
223646/1500000: episode: 167, duration: 19.870s, episode steps:
                                                                2143, steps per second: 108, episode reward: -13.000, mean reward: -0.006
225606/1500000: episode: 168, duration: 17.713s, episode steps: 1960, steps per second: 111, episode reward: -15.000, mean reward: -0.008
                                                                                                                                             [-1.000,
                                                                                                                                                     1.0001, mean acti
                                                                2195, steps per second: 117, episode reward: -15.000, mean reward: -0.007
230038/1500000: episode: 170, duration: 18.676s, episode steps: 2237, steps per second: 120, episode reward: -14.000, mean reward: -0.006 [-1.000,
                                                                                                                                                     1.000], mean acti
231807/1500000: episode:
                         171, duration: 14.816s, episode steps:
                                                                 1769, steps per second: 119, episode reward: -15.000, mean reward: -0.008
233945/1500000: episode: 172, duration: 18.285s, episode steps: 2138, steps per second: 117, episode reward: -13.000, mean reward: -0.006
                                                                                                                                             [-1.000.
                                                                                                                                                     1.0001, mean acti
235574/1500000: episode: 173, duration: 13.626s, episode steps:
                                                                1629, steps per second: 120, episode reward: -19.000, mean reward: -0.012
237638/1500000: episode: 174, duration: 17.302s, episode steps: 2064, steps per second: 119, episode reward: -16.000, mean reward: -0.008 [-1.000,
                                                                                                                                                     1.0001, mean acti
239956/1500000: episode: 175, duration: 19.415s, episode steps: 2318, steps per second: 119, episode reward: -17.000, mean reward: -0.007
242611/1500000: episode: 176, duration: 22.199s, episode steps: 2655, steps per second: 120, episode reward: -13.000, mean reward: -0.005
                                                                                                                                             -1.000,
                                                                                                                                                     1.000], mean acti
245074/1500000: episode: 177, duration: 20.616s, episode steps:
                                                                2463, steps per second: 119, episode reward: -13.000, mean reward: -0.005
247072/1500000: episode: 178, duration: 16.667s, episode steps: 1998, steps per second: 120, episode reward: -16.000, mean reward: -0.008
                                                                                                                                             [-1.000.
                                                                                                                                                     1.0001. mean acti
                         179, duration: 15.830s,
248966/1500000: episode:
                                                 episode steps:
                                                                 1894, steps per second: 120, episode reward: -18.000, mean reward: -0.010
251361/1500000: episode: 180, duration: 19.995s, episode steps: 2395, steps per second: 120, episode reward: -15.000, mean reward: -0.006
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean acti
254016/1500000: episode: 181, duration: 22.245s, episode steps:
                                                                2655, steps per second: 119, episode reward: -14.000, mean reward: -0.005
256244/1500000: episode: 182, duration: 18.872s, episode steps: 2228, steps per second: 118, episode reward: -18.000, mean reward: -0.008
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean acti
258305/1500000: episode: 183, duration: 17.231s, episode steps:
                                                                2061, steps per second: 120, episode reward: -15.000, mean reward:
260178/1500000: episode: 184, duration: 15.647s, episode steps: 1873, steps per second: 120, episode reward: -18.000, mean reward: -0.010
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean acti
262484/1500000: episode: 185, duration: 19.202s, episode steps:
                                                                 2306, steps per second: 120, episode reward: -14.000, mean reward: -0.006
265012/1500000: episode: 186, duration: 21.089s, episode steps: 2528, steps per second: 120, episode reward: -17.000, mean reward: -0.007
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean acti
267621/1500000: episode: 187, duration: 21.716s, episode steps:
                                                                2609, steps per second: 120, episode reward: -13.000, mean reward: -0.005
270075/1500000: episode: 188, duration: 20.681s, episode steps: 2454, steps per second: 119, episode reward: -12.000, mean reward: -0.005
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean acti
                                                                1793, steps per second: 119, episode reward: -17.000, mean reward: -0.009
271868/1500000: episode: 189, duration: 15.112s, episode steps:
274384/1500000: episode: 190, duration: 21.043s, episode steps: 2516, steps per second: 120, episode reward: -15.000, mean reward: -0.006 [-1.000,
                                                                                                                                                     1.000], mean acti
276967/1500000: episode:
                         191, duration: 21.632s, episode steps:
                                                                2583, steps per second: 119, episode reward: -14.000, mean reward:
279919/1500000; episode: 192, duration: 24.633s, episode steps: 2952, steps per second: 120, episode reward: -11.000, mean reward: -0.004
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean acti
282464/1500000: episode: 193, duration: 21.233s, episode steps:
                                                                2545, steps per second: 120, episode reward: -15.000, mean reward: -0.006
                                                                                                                                                     1.000], mean acti
284525/1500000: episode: 194, duration: 17.220s, episode steps: 2061, steps per second: 120, episode reward: -15.000, mean reward: -0.007 [-1.000,
                                                                                                                                                     1.0001, mean acti
287829/1500000: episode: 195, duration: 27.474s, episode steps: 3304, steps per second: 120, episode reward: -6.000, mean reward: -0.002 [
                                                                                                                                             -1.000, 1.000],
290572/1500000: episode: 196, duration: 22.922s, episode steps: 2743, steps per second: 120, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
292647/1500000: episode: 197, duration: 17.470s, episode steps:
                                                                 2075, steps per second: 119, episode reward: -17.000, mean reward: -0.008
294882/1500000: episode: 198, duration: 18.802s, episode steps: 2235, steps per second: 119, episode reward: -15.000, mean reward: -0.007 [-1.000, 1.000], mean acti
297477/1500000: episode: 199, duration: 22.821s, episode steps:
                                                                 2595, steps per second: 114, episode reward: -13.000, mean reward: -0.005
300078/1500000; episode: 200, duration: 22.965s, episode steps: 2601, steps per second: 113, episode reward: -11.000, mean reward: -0.004
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean acti
302051/1500000: episode: 201, duration: 17.375s, episode steps:
                                                                1973, steps per second: 114, episode reward: -16.000, mean reward: -0.008
304296/1500000: episode: 202, duration: 19.847s, episode steps: 2245, steps per second: 113, episode reward: -18.000, mean reward: -0.008
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean acti
                                                                1661, steps per second: 113, episode reward: -18.000, mean reward: -0.011
305957/1500000: episode: 203, duration: 14.722s,
308382/1500000: episode: 204, duration: 21.702s, episode steps: 2425, steps per second: 112, episode reward: -14.000, mean reward: -0.006
                                                                                                                                            [-1.000,
                                                                                                                                                     1.000], mean acti
310998/1500000: episode: 205, duration: 22.388s, episode steps:
                                                                 2616, steps per second: 117, episode reward: -15.000, mean reward: -0.006
313640/1500000: episode: 206, duration: 22.872s, episode steps: 2642, steps per second: 116, episode reward: -13.000, mean reward: -0.005 [-1.000,
                                                                                                                                                     1.0001, mean acti
316396/1500000: episode: 207, duration: 23.772s, episode steps:
                                                                2756, steps per second: 116, episode reward: -11.000, mean reward: -0.004 [-1.000,
319439/1500000; episode: 208, duration: 26.060s, episode steps: 3043, steps per second: 117, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
322073/1500000: episode: 209, duration: 22.711s, episode steps:
                                                                2634, steps per second: 116, episode reward: -13.000, mean reward: -0.005
324780/1500000: episode: 210, duration: 23.537s, episode steps: 2707, steps per second: 115, episode reward: -14.000, mean reward: -0.005 [-1.000,
                                                                                                                                                     1.000], mean acti
                episode: 211, duration: 16.959s,
                                                 episode steps:
                                                                       steps per second: 117, episode reward: -17.000,
                                                                                                                                     -0 009
329750/1500000: episode: 212, duration: 25.490s, episode steps: 2994, steps per second: 117, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
               episode: 213, duration: 29.271s, episode steps:
                                                                 3453, steps per second:
                                                                                         118, episode reward: -4.000, mean reward: -0.001
335410/1500000; episode: 214, duration: 18.716s, episode steps: 2207, steps per second: 118, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
337769/1500000: episode: 215, duration: 20.192s, episode steps: 2359, steps per second: 117, episode reward: -15.000, mean reward: -0.006 [-1.000, 1.000], mean acti
340544/1500000: episode: 216, duration: 23.724s, episode steps: 2775, steps per second: 117, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic 342905/1500000: episode: 217, duration: 20.494s, episode steps: 2361, steps per second: 115, episode reward: -14.000, mean reward: -0.006 [-1.000, 1.000], mean actic
```

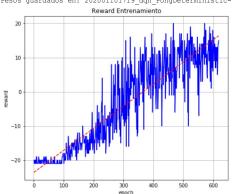
```
346141/1500000: episode: 218, duration: 27.698s, episode steps: 3236, steps per second: 117, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
348700/1500000: episode: 219, duration: 21.833s, episode steps: 2559, steps per second: 117, episode reward: -9.000, mean reward: -0.004 [-1.000, 1.000], mean actic 350849/1500000: episode: 220, duration: 18.424s, episode steps: 2149, steps per second: 117, episode reward: -16.000, mean reward: -0.007 [-1.000, 1.000], mean actic
353570/1500000: episode: 221, duration: 23.875s, episode steps: 2721, steps per second: 114, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
356365/1500000: episode: 222, duration: 24.778s, episode steps: 2795, steps per second: 113, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
358709/1500000: episode: 223, duration: 20.555s, episode steps: 2344, steps per second: 114, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
361791/1500000: episode: 224, duration: 26.614s, episode steps: 3082, steps per second: 116, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
364208/1500000: episode: 225, duration: 22.891s, episode steps: 2417, steps per second: 106, episode reward: -14.000, mean reward: -0.006 [-1.000, 1.000], mean acti
367295/1500000: episode: 226, duration: 29.867s, episode steps:
                                                                  3087, steps per second: 103, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
370708/1500000: episode: 227, duration: 32.575s, episode steps: 3413, steps per second: 105, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
373662/1500000: episode: 228, duration: 27.636s, episode steps: 2954, steps per second: 107, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
376795/1500000: episode: 229, duration: 28.439s, episode steps: 3133, steps per second: 110, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
                                                                 1938, steps per second: 111, episode reward: -15.000, mean reward: -0.008 [-1.000, 1.000], mean acti
378733/1500000: episode: 230, duration: 17.438s, episode steps:
381476/1500000: episode: 231, duration: 25.896s, episode steps: 2743, steps per second: 106, episode reward: -7.000, mean reward: -0.003 [-1.000, 1.000], mean actic
383976/1500000: episode: 232, duration: 22.993s, episode steps:
                                                                 2500, steps per second: 109, episode reward:
                                                                                                               -7.000, mean reward: -0.003 [-1.000, 1.000], mean actic
386003/1500000; episode: 233, duration: 18.171s, episode steps: 2027, steps per second: 112, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
388297/1500000: episode: 234, duration: 20.181s, episode steps:
                                                                 2294, steps per second: 114, episode reward: -15.000, mean reward: -0.007 [-1.000, 1.000], mean acti
390906/1500000: episode: 235, duration: 21.923s, episode steps: 2609, steps per second: 119, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
393503/1500000: episode: 236, duration: 21.803s, episode steps: 2597, steps per second: 119, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
396065/1500000: episode: 237, duration: 21.634s, episode steps: 2562, steps per second: 118, episode reward: -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
399555/1500000: episode: 238, duration: 29.391s, episode steps: 3490, steps per second: 119, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
402294/1500000: episode: 239, duration: 22.982s, episode steps: 2739, steps per second: 119, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
405122/1500000: episode: 240, duration: 23.686s, episode steps:
                                                                                          119, episode reward: -12.000, mean reward: -0.004 [-1.000,
                                                                 2828, steps per second:
407360/1500000: episode: 241, duration: 18.725s, episode steps: 2238, steps per second: 120, episode reward: -18.000, mean reward: -0.008 [-1.000, 1.000], mean acti
409693/1500000: episode: 242, duration: 19.597s, episode steps: 2333, steps per second: 119, episode reward: -16.000, mean reward: -0.007
                                                                                                                                             [-1.000, 1.000], mean acti
412713/1500000: episode: 243, duration: 25.450s, episode steps: 3020, steps per second: 119, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
416256/1500000: episode: 244, duration: 29.886s, episode steps:
                                                                 3543, steps per second: 119, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
419680/1500000: episode: 245, duration: 28.672s, episode steps: 3424, steps per second: 119, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
422398/1500000: episode: 246, duration: 22.7578, episode steps: 2718, steps per second: 119, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
425869/1500000: episode: 247, duration: 29.022s, episode steps: 3471, steps per second: 120, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
429127/1500000: episode: 248, duration: 27.182s, episode steps:
                                                                 3258, steps per second: 120, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
432661/1500000: episode: 249, duration: 29.726s, episode steps: 3534, steps per second: 119, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
435665/1500000: episode: 250, duration: 25.193s, episode steps:
                                                                 3004, steps per second: 119, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean action
438514/1500000: episode: 251, duration: 23.786s, episode steps: 2849, steps per second: 120, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
                                                                 2581, steps per second: 119, episode reward: -16.000, mean reward: -0.006 [-1.000, 1.000], mean acti
441095/1500000: episode: 252, duration: 21.608s, episode steps:
443503/1500000; episode: 253, duration: 20.122s, episode steps: 2408, steps per second: 120, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
446814/1500000: episode: 254, duration: 27.719s, episode steps:
                                                                 3311, steps per second: 119, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean action
448998/1500000: episode: 255, duration: 18.316s, episode steps: 2184, steps per second: 119, episode reward: -15.000, mean reward: -0.007 [-1.000, 1.000], mean acti
451884/1500000: episode: 256, duration: 24.461s, episode steps:
                                                                 2886, steps per second: 118, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
454537/1500000: episode: 257, duration: 22.307s, episode steps: 2653, steps per second: 119, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
                                                                 2952, steps per second: 115, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
457489/1500000: episode: 258, duration: 25.620s, episode steps:
461237/1500000: episode: 259, duration: 33.172s, episode steps: 3748, steps per second: 113, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
                                                                 3257, steps per second: 113, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
464494/1500000: episode: 260, duration: 28.817s, episode steps:
468036/1500000; episode: 261, duration: 31.370s, episode steps: 3542, steps per second: 113, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
471697/1500000: episode: 262, duration: 31.425s, episode steps:
                                                                 3661, steps per second: 117, episode reward:
                                                                                                               -9.000, mean reward: -0.002 [-1.000, 1.000], mean actio
475300/1500000: episode: 263, duration: 31.009s, episode steps: 3603, steps per second: 116, episode reward: -5.000, mean reward: -0.001 [-1.000, 1.000], mean actic
                                                                                                               -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
477865/1500000: episode: 264, duration: 22.139s, episode steps:
                                                                 2565, steps per second: 116, episode reward:
480873/1500000: episode: 265, duration: 25.902s, episode steps: 3008, steps per second: 116, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
484618/1500000: episode: 266, duration: 32.441s, episode steps:
                                                                  3745, steps per second: 115, episode reward:
                                                                                                                -2.000, mean reward: -0.001 [-1.000, 1.000], mean actic
487866/1500000; episode: 267, duration: 28.301s, episode steps: 3248, steps per second: 115, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
                                                                                                               -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
490888/1500000: episode: 268, duration: 25.550s, episode steps:
                                                                 3022, steps per second: 118, episode reward:
493193/1500000: episode: 269, duration: 19.483s, episode steps: 2305, steps per second: 118, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
496592/1500000: episode: 270, duration: 28.828s, episode steps:
                                                                 3399, steps per second: 118, episode reward:
                                                                                                                2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
500059/1500000: episode: 271, duration: 29.631s, episode steps: 3467, steps per second: 117, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
503575/1500000: episode: 272, duration: 30.259s, episode steps:
                                                                 3516, steps per second: 116, episode reward:
                                                                                                                -1.000, mean reward: -0.000 [-1.000, 1.000], mean action
506507/1500000: episode: 273. duration: 25.311s. episode steps: 2932. steps per second: 116. episode reward: 8.000. mean reward: 0.003 [-1.000. 1.000]. mean action:
509200/1500000: episode: 274, duration: 23.228s, episode steps:
                                                                 2693, steps per second: 116, episode reward:
                                                                                                                -11.000, mean reward: -0.004 [-1.000, 1.0
512079/1500000; episode: 275, duration: 24.914s, episode steps: 2879, steps per second: 116, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
                                                                                                                -5.000, mean reward: -0.002 [-1.000, 1.000], mean actic
515081/1500000: episode: 276, duration: 26.777s, episode steps:
                                                                 3002, steps per second: 112, episode reward:
517896/1500000: episode: 277, duration: 24.603s, episode steps: 2815, steps per second: 114, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
520751/1500000: episode: 278, duration: 27.850s, episode steps:
                                                                 2855, steps per second: 103, episode reward:
                                                                                                                -9.000, mean reward: -0.003 [-1.000, 1.000], mean action
523826/1500000: episode: 279, duration: 29.962s, episode steps: 3075, steps per second: 103, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
527079/1500000: episode: 280, duration: 31.022s, episode steps:
                                                                 3253, steps per second: 105, episode reward:
                                                                                                               7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
529720/1500000: episode: 281, duration: 24.563s, episode steps: 2641, steps per second: 108, episode reward:
                                                                                                               9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
532979/1500000: episode: 282, duration: 30.741s, episode steps:
                                                                 3259, steps per second: 106, episode reward:
                                                                                                               3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
536193/1500000: episode: 283, duration: 27.515s, episode steps: 3214, steps per second: 117, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
539035/1500000: episode: 284, duration: 24.206s, episode steps: 2842, steps per second: 117, episode reward:
                                                                                                               8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
541472/1500000: episode: 285, duration: 20.405s, episode steps: 2437, steps per second: 119, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
544301/1500000: episode: 286, duration: 23.796s, episode steps: 2829, steps per second: 119, episode reward:
                                                                                                               -8.000, mean reward: -0.003 [-1.000, 1.000], mean action
546956/1500000: episode: 287. duration: 22.334s, episode steps: 2655. steps per second: 119. episode reward: -11.000. mean reward: -0.004 [-1.000. 1.000], mean acti
550733/1500000: episode: 288, duration: 31.800s, episode steps:
                                                                                                               1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
                                                                 3777, steps per second: 119, episode reward:
553408/1500000; episode: 289, duration: 22.511s, episode steps: 2675, steps per second: 119, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
                                                                                                               -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
556073/1500000: episode: 290, duration: 22.707s, episode steps:
                                                                 2665, steps per second: 117, episode reward:
558530/1500000: episode: 291, duration: 20.906s, episode steps: 2457, steps per second: 118, episode reward: -15.000, mean reward: -0.006 [-1.000, 1.000], mean acti
                                                                                                               -7.000, mean reward: -0.002 [-1.000, 1.000], mean actio
561425/1500000: episode: 292, duration: 28.625s, episode steps:
                                                                 2895, steps per second: 101, episode reward:
564550/1500000: episode: 293, duration: 26.385s, episode steps: 3125, steps per second: 118, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
567613/1500000: episode: 294, duration: 25.910s, episode steps:
                                                                 3063, steps per second: 118, episode reward:
                                                                                                               9.000, mean reward: 0.003 [-1.000, 1.000],
571143/1500000; episode: 295, duration: 29.729s, episode steps: 3530, steps per second: 119, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
                                                                 3002, steps per second: 119, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean action
574145/1500000: episode: 296, duration: 25.313s, episode steps:
577339/1500000: episode: 297, duration: 26.756s, episode steps: 3194, steps per second: 119, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
                                                                 3927, steps per second: 119, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000],
581266/1500000: episode: 298, duration: 33.031s, episode steps:
584277/1500000: episode: 299, duration: 25.453s, episode steps: 3011, steps per second: 118, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
587647/1500000: episode: 300, duration: 28.739s, episode steps:
                                                                 3370, steps per second: 117, episode reward:
                                                                                                               -1.000, mean reward: -0.000 [-1.000, 1.000], mean action
591688/1500000; episode: 301, duration: 34.490s, episode steps: 4041, steps per second: 117, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
594708/1500000: episode: 302, duration: 26.943s, episode steps: 3020, steps per second: 112, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
597902/1500000: episode: 303, duration: 28.778s, episode steps: 3194, steps per second: 111, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
601469/1500000: episode: 304, duration: 32.253s, episode steps: 3567, steps per second: 111, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
604553/1500000: episode: 305, duration: 27.719s, episode steps: 3084, steps per second: 111, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
607640/1500000: episode: 306, duration: 27.023s, episode steps:
                                                                 3087, steps per second: 114, episode reward:
                                                                                                               -8.000, mean reward: -0.003 [-1.000, 1.000], mean action
610737/1500000: episode: 307, duration: 27.408s, episode steps: 3097, steps per second: 113, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
613075/1500000: episode: 308, duration: 20.466s, episode steps:
                                                                 2338, steps per second: 114, episode reward: 15.000, mean reward: 0.006 [-1.000, 1.000], mean action
615671/1500000; episode: 309, duration: 22.811s, episode steps: 2596, steps per second: 114, episode reward: -16.000, mean reward: -0.006 [-1.000, 1.000], mean acti
619630/1500000: episode: 310, duration: 34.897s, episode steps:
                                                                 3959, steps per second: 113, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
623234/1500000: episode: 311, duration: 31.675s, episode steps: 3604, steps per second: 114, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
                                                                 3527, steps per second: 116, episode reward:
                                                                                                               -5.000, mean reward: -0.001 [-1.000, 1.000], mean actio
626761/1500000: episode: 312, duration: 30.425s, episode steps:
630912/1500000: episode: 313, duration: 35.965s, episode steps: 4151, steps per second: 115, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
 33913/1500000: episode: 314, duration: 26.052s, episode steps:
                                                                 3001, steps per second: 115, episode reward:
                                                                                                                8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
636941/1500000; episode: 315, duration: 26.283s, episode steps: 3028, steps per second: 115, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
640149/1500000: episode: 316, duration: 27.909s, episode steps:
                                                                 3208, steps per second: 115, episode reward:
                                                                                                               -7.000, mean reward: -0.002 [-1.000, 1.000], mean action
643683/1500000; episode: 317, duration: 30.943s, episode steps: 3534, steps per second: 114, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
645690/1500000: episode: 318, duration: 17.531s, episode steps:
                                                                 2007, steps per second: 114, episode reward:
                                                                                                                -16.000, mean reward: -0.008 [-1.000, 1.000],
650228/1500000: episode: 319, duration: 40.743s, episode steps: 4538, steps per second: 111, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
                                                                                                                                     -0.001 [-1.000, 1.000], mean actio
                         320, duration: 31.764s,
                                                  episode steps:
                                                                       steps per second: 111, episode reward:
                                                                                                                -5.000, mean reward:
657695/1500000: episode: 321, duration: 38.532s, episode steps: 3953, steps per second: 103, episode reward:
                                                                                                                -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
                         322, duration: 33.796s, episode steps:
                                                                 3495, steps per second: 103, episode reward:
                                                                                                                  .000, mean reward: 0.000 [-1.000, 1.000], mean action:
663480/1500000; episode: 323, duration: 21.532s, episode steps: 2290, steps per second: 106, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean action
667380/1500000: episode: 324, duration: 37.100s, episode steps:
                                                                 3900, steps per second: 105, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
670032/1500000: episode: 325, duration: 23.962s, episode steps: 2652, steps per second: 111, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action: 673872/1500000: episode: 326, duration: 33.042s, episode steps: 3840, steps per second: 116, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
```

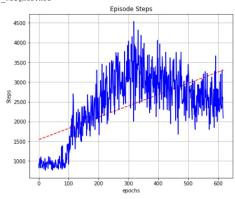
```
677145/1500000: episode: 327, duration: 27.911s, episode steps: 3273, steps per second: 117, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
680235/1500000: episode: 328, duration: 26.081s, episode steps: 3090, steps per second: 118, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
                                                                  2861, steps per second: 118, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 683096/1500000: episode: 329, duration: 24.295s, episode steps:
686560/1500000: episode: 330, duration: 29.433s, episode steps: 3464, steps per second: 118, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 689928/1500000: episode: 331, duration: 28.801s, episode steps: 3368, steps per second: 117, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
693239/1500000: episode: 332, duration: 28.830s, episode steps: 3311, steps per second: 115, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 697557/1500000: episode: 333, duration: 38.005s, episode steps:
                                                                  4318, steps per second: 114, episode reward:
                                                                                                                -2.000, mean reward: -0.000 [-1.000, 1.000], mean actic
 701102/1500000: episode: 334, duration: 30.808s, episode steps: 3545, steps per second: 115, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
                                                                                                                4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
         .500000: episode: 335, duration: 29.792s, episode steps:
                                                                  3429, steps per second: 115, episode reward:
 708207/1500000: episode: 336, duration: 31.983s, episode steps: 3676, steps per second: 115, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 711402/1500000: episode: 337, duration: 27.708s, episode steps:
                                                                  3195, steps per second: 115, episode reward:
                                                                                                                7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 713867/1500000: episode: 338, duration: 21.230s, episode steps: 2465, steps per second: 116, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
 717146/1500000: episode: 339, duration: 28.410s, episode steps:
                                                                  3279, steps per second: 115, episode reward:
                                                                                                                  6.000, mean reward: -0.002 [-1.000, 1.000], mean actio
 720339/1500000: episode: 340, duration: 27.637s, episode steps: 3193, steps per second: 116, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
                          341, duration: 31.699s, episode steps:
 723992/1500000: episode:
                                                                  3653, steps per second: 115, episode reward:
                                                                                                                  4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
 727482/1500000; episode: 342, duration: 30.257s, episode steps: 3490, steps per second: 115, episode reward: -3.000, mean reward: -0.001 (-1.000, 1.000), mean actic
 31668/1500000: episode: 343, duration: 37.679s, episode steps:
                                                                  4186, steps per second: 111, episode reward:
                                                                                                                -2.000, mean reward: -0.000 [-1.000, 1.000], mean actic
734251/1500000: episode: 344, duration: 23.898s, episode steps: 2583, steps per second: 108, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
 737875/1500000: episode: 345, duration: 32.988s, episode steps: 3624, steps per second: 110, episode reward:
                                                                                                                4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 740381/1500000: episode: 346, duration: 22.654s, episode steps: 2506, steps per second: 111, episode reward:
                                                                                                                -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
 743477/1500000: episode: 347, duration: 27.231s, episode steps:
                                                                  3096, steps per second: 114, episode reward:
                                                                                                                -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
746341/1500000: episode: 348, duration: 25.537s, episode steps: 2864, steps per second: 112, episode reward: 750189/1500000: episode: 349, duration: 34.168s, episode steps: 3848, steps per second: 113, episode reward:
                                                                                                                -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
                                                                                                                  .000, mean reward: 0.001 [-1.000, 1.000], mean action:
753266/1500000: episode: 350, duration: 27.333s, episode steps: 3077, steps per second: 113, episode reward:
                                                                                                                5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 756112/1500000: episode: 351, duration: 25.298s, episode steps: 2846, steps per second: 113, episode reward:
                                                                                                                9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
 759900/1500000: episode: 352, duration: 33.608s, episode steps: 3788, steps per second: 113, episode reward:
                                                                                                                -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
                                                                                                                -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
 763881/1500000: episode:
                          353, duration: 34.546s, episode steps:
                                                                  3981, steps per second: 115, episode reward:
 767285/1500000: episode: 354, duration: 29.846s, episode steps: 3404, steps per second: 114, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 70627/1500000: episode: 355, duration: 29.403s, episode steps:
                                                                  3342, steps per second: 114, episode reward:
                                                                                                                 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 773333/1500000: episode: 356, duration: 23.808s, episode steps: 2706, steps per second: 114, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
 776153/1500000: episode: 357, duration: 24.824s, episode steps:
                                                                  2820, steps per second: 114, episode reward:
                                                                                                                10.000, mean reward: 0.004 [-1.000, 1.000], mean action
 779182/1500000: episode: 358, duration: 26.705s, episode steps: 3029, steps per second: 113, episode reward:
                                                                                                                7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
                                                                  2943, steps per second: 112, episode reward:
 785416/1500000: episode: 360, duration: 29.398s, episode steps: 3291, steps per second: 112, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 788156/1500000: episode: 361, duration: 24.916s, episode steps:
                                                                  2740, steps per second: 110, episode reward:
                                                                                                                 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
 791422/1500000: episode: 362, duration: 30.275s, episode steps: 3266, steps per second: 108, episode reward:
                                                                                                                3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 793782/1500000: episode: 363, duration: 23.145s, episode steps:
                                                                  2360, steps per second: 102, episode reward:
                                                                                                                11.000, mean reward: 0.005 [-1.000, 1.000], mean action
 796973/1500000: episode: 364, duration: 31.114s, episode steps: 3191, steps per second: 103, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 800651/1500000: episode: 365, duration: 35.012s, episode steps:
                                                                  3678, steps per second: 105, episode reward:
                                                                                                                 -3.000, mean reward: -0.001 [-1.000, 1.000], mean actio
803238/1500000: episode: 366, duration: 25.527s, episode steps: 2587, steps per second: 101, episode reward:
                                                                                                                11.000, mean reward: 0.004 [-1.000, 1.000], mean action
 306380/1500000: episode: 367, duration: 29.165s, episode steps:
                                                                  3142, steps per second: 108, episode reward:
                                                                                                                 6.000, mean reward: 0.002 [-1.000, 1.000],
809461/1500000: episode: 368. duration: 26.654s. episode steps: 3081. steps per second: 116. episode reward:
                                                                                                                -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
 812209/1500000: episode: 369, duration: 23.761s, episode steps:
                                                                                                                  .000, mean reward: 0.003 [-1.000, 1.000], mean action:
                                                                  2748, steps per second: 116, episode reward:
815903/1500000: episode: 370, duration: 32.409s, episode steps: 3694, steps per second: 114, episode reward:
                                                                                                                1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
 818312/1500000: episode: 371, duration: 20.924s, episode steps:
                                                                  2409, steps per second: 115, episode reward:
                                                                                                                 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
820532/1500000: episode: 372, duration: 19.362s, episode steps: 2220, steps per second: 115, episode reward:
                                                                                                                16.000, mean reward: 0.007 [-1.000, 1.000], mean action
                          373, duration: 27.992s,
 823742/1500000: episode:
                                                  episode steps:
                                                                  3210, steps per second: 115, episode reward:
                                                                                                                 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
826153/1500000: episode: 374, duration: 20.781s, episode steps: 2411, steps per second: 116, episode reward:
                                                                                                                11.000, mean reward: 0.005 [-1.000, 1.000], mean action
 828831/1500000: episode: 375, duration: 22.793s, episode steps:
                                                                  2678, steps per second: 117, episode reward:
                                                                                                                   .000, mean reward: 0.004 [-1.000, 1.000], mean action
831136/1500000; episode: 376, duration: 19.608s, episode steps: 2305, steps per second: 118, episode reward: 13.000, mean reward: 0.006 [-1.000, 1.000], mean action
 834511/1500000: episode: 377, duration: 28.623s, episode steps:
                                                                  3375, steps per second: 118, episode reward:
                                                                                                                4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
837600/1500000: episode: 378, duration: 26.579s, episode steps: 3089, steps per second: 116, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 841162/1500000: episode: 379, duration: 30.317s, episode steps:
                                                                  3562, steps per second: 117, episode reward:
                                                                                                                1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
844647/1500000: episode: 380, duration: 29.500s, episode steps: 3485, steps per second: 118, episode reward:
                                                                                                                3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 848294/1500000: episode: 381, duration: 30.780s, episode steps:
                                                                  3647, steps per second: 118, episode reward:
                                                                                                                 5.000, mean reward: 0.001
851233/1500000: episode: 382. duration: 25.167s. episode steps: 2939. steps per second: 117. episode reward:
                                                                                                                9.000, mean reward: 0.003 [=1.000, 1.000], mean action:
 54115/1500000: episode: 383, duration: 24.441s, episode steps:
                                                                  2882, steps per second: 118, episode reward:
                                                                                                                 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
856848/1500000; episode: 384, duration: 23.212s, episode steps: 2733, steps per second: 118, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 860065/1500000: episode: 385, duration: 27.394s, episode steps:
                                                                  3217, steps per second: 117, episode reward:
                                                                                                                 -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
863428/1500000: episode: 386, duration: 28.509s, episode steps: 3363, steps per second: 118, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
                                                                  2859, steps per second: 115, episode reward:
869550/1500000: episode: 388, duration: 29.230s, episode steps: 3263, steps per second: 112, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 872001/1500000: episode: 389, duration: 22.075s, episode steps:
                                                                  2451, steps per second: 111, episode reward:
                                                                                                                 10.000, mean reward: 0.004
874970/1500000: episode: 390, duration: 26.884s, episode steps: 2969, steps per second: 110, episode reward:
                                                                                                                7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 878369/1500000: episode: 391, duration: 30.137s, episode steps:
                                                                  3399, steps per second: 113, episode reward:
                                                                                                                 -2.000, mean reward: -0.001 [-1.000, 1.000], mean actic
881043/1500000: episode: 392, duration: 23.282s, episode steps: 2674, steps per second: 115, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
                                                                                                                6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 884212/1500000: episode: 393, duration: 27.966s, episode steps:
                                                                  3169, steps per second: 113, episode reward:
886334/1500000: episode: 394, duration: 18.668s, episode steps: 2122, steps per second: 114, episode reward:
                                                                                                                17.000, mean reward: 0.008 [-1.000, 1.000], mean action
 389507/1500000: episode: 395, duration: 27.986s, episode steps:
                                                                  3173, steps per second: 113, episode reward:
                                                                                                                  .000, mean reward: 0.002 [-1.000, 1.000], mean action:
891977/1500000: episode: 396. duration: 21.693s. episode steps: 2470. steps per second: 114. episode reward: 13.000. mean reward: 0.005 [-1.000, 1.000]. mean action
 895145/1500000: episode: 397, duration: 27.860s,
                                                                                                                 -6.000, mean reward: -0.002 [-1.000, 1.000], mean action
                                                  episode steps:
                                                                  3168, steps per second: 114, episode reward:
899205/1500000: episode: 398, duration: 34.661s, episode steps: 4060, steps per second: 117, episode reward:
                                                                                                                -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
 02138/1500000: episode: 399, duration: 25.076s, episode steps:
                                                                  2933, steps per second: 117, episode reward:
                                                                                                                11.000, mean reward: 0.004 [-1.000, 1.000], mean action
 905746/1500000: episode: 400, duration: 31.014s, episode steps: 3608, steps per second: 116, episode reward:
                                                                                                                -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
 908584/1500000: episode: 401, duration: 25.097s, episode steps:
                                                                  2838, steps per second: 113, episode reward:
                                                                                                                7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 911567/1500000: episode: 402, duration: 25.820s, episode steps: 2983, steps per second: 116, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 914557/1500000: episode: 403, duration: 25.922s, episode steps:
                                                                  2990, steps per second: 115, episode reward:
                                                                                                                 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
 917184/1500000; episode: 404, duration: 22.768s, episode steps: 2627, steps per second: 115, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
 919505/1500000: episode: 405, duration: 20.406s, episode steps:
                                                                  2321, steps per second: 114, episode reward:
                                                                                                                13.000, mean reward: 0.006 [-1.000, 1.000], mean action
 923473/1500000: episode: 406, duration: 35.155s, episode steps: 3968, steps per second: 113, episode reward:
                                                                                                                1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
                                                                                                                7.000, mean reward: 0.003 [-1.000, 1.000],
 926215/1500000: episode: 407, duration: 24.263s, episode steps:
                                                                  2742, steps per second: 113, episode reward:
 929608/1500000: episode: 408, duration: 31.705s, episode steps: 3393, steps per second: 107, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
 032689/1500000: episode: 409, duration: 29.855s, episode steps:
                                                                  3081, steps per second: 103, episode reward:
                                                                                                                 -5.000, mean reward: -0.002 [-1.000, 1.000], mean action
 935984/1500000; episode: 410, duration: 31.609s, episode steps: 3295, steps per second: 104, episode reward:
                                                                                                                1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
                                                                                                                7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 939080/1500000: episode: 411, duration: 28.244s, episode steps:
                                                                  3096, steps per second: 110, episode reward:
 941637/1500000: episode: 412, duration: 24.666s, episode steps: 2557, steps per second: 104, episode reward:
                                                                                                                11.000, mean reward: 0.004 [-1.000, 1.000], mean action
 944770/1500000: episode: 413, duration: 27.909s,
                                                   episode steps: 3133, steps per second: 112, episode reward:
                                                                                                                6.000, mean reward: 0.002 [-1.000, 1.000],
 948416/1500000: episode: 414, duration: 31.023s, episode steps: 3646, steps per second: 118, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
 951862/1500000: episode: 415, duration: 29.327s, episode steps:
                                                                  3446, steps per second: 118, episode reward:
                                                                                                                 -5.000, mean reward: -0.001 [-1.000, 1.000], mean action
 955128/1500000: episode: 416, duration: 28.137s, episode steps: 3266, steps per second: 116, episode reward:
                                                                                                                -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
 958281/1500000: episode: 417, duration: 26.932s, episode steps:
                                                                        steps per second: 117, episode reward:
                                                                                                                2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 960580/1500000: episode: 418, duration: 19.648s, episode steps: 2299, steps per second: 117, episode reward:
                                                                                                                12.000, mean reward: 0.005 [-1.000, 1.000], mean action
 963480/1500000: episode: 419, duration: 24.701s, episode steps:
                                                                  2900, steps per second: 117, episode reward:
                                                                                                                7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 966784/1500000: episode: 420, duration: 28.154s, episode steps: 3304, steps per second: 117, episode reward:
                                                                                                                1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
 969400/1500000: episode: 421, duration: 22.332s,
                                                                                                                12.000, mean reward: 0.005 [-1.000, 1.000], mean action
                                                                  2616, steps per second: 117, episode reward:
 972785/1500000: episode: 422, duration: 28.846s, episode steps: 3385, steps per second: 117, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 975415/1500000: episode: 423, duration: 22.353s, episode steps:
                                                                  2630, steps per second: 118, episode reward:
 978500/1500000: episode: 424. duration: 26.735s, episode steps: 3085, steps per second: 115, episode reward:
                                                                                                                9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
 981451/1500000: episode: 425, duration: 25.069s, episode steps:
                                                                  2951, steps per second: 118, episode reward:
                                                                                                                7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 985208/1500000; episode: 426, duration: 31.996s, episode steps: 3757, steps per second: 117, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
 088350/1500000: episode: 427, duration: 26.852s, episode steps:
                                                                  3142, steps per second: 117, episode reward:
                                                                                                                2.000, mean reward: 0.001 [-1.000, 1.000],
 990946/1500000: episode: 428, duration: 22.027s, episode steps: 2596, steps per second: 118, episode reward:
                                                                                                                -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
                 episode: 429, duration: 27.672s,
                                                                                                                 -3.000, mean reward: -0.001 [-1.000, 1.000], mean actio
                                                   episode steps:
                                                                  3253,
                                                                        steps per second: 118, episode reward:
 997144/1500000: episode: 430, duration: 24.597s, episode steps: 2945, steps per second: 120, episode reward:
                                                                                                                8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
                episode: 431, duration: 19.102s, episode steps:
                                                                  2300, steps per second:
1002449/1500000: episode: 432, duration: 25.121s, episode steps: 3005, steps per second: 120, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1005601/1500000: episode: 433, duration: 27.608s, episode steps:
                                                                  3152, steps per second: 114, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1008422/1500000: episode: 434, duration: 25.138s, episode steps: 2821, steps per second: 112, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action: 1011608/1500000: episode: 435, duration: 28.298s, episode steps: 3186, steps per second: 113, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
                                                                                                                7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
```

```
1014124/1500000: episode: 436, duration: 22.639s, episode steps: 2516, steps per second: 111, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1016402/1500000: episode: 437, duration: 20.251s, episode steps: 2278, steps per second: 112, episode reward: 15.000, mean reward: 0.007 [-1.000, 1.000], mean action
                                                                 2705, steps per second: 117, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
1019107/1500000: episode: 438, duration: 23.124s, episode steps:
1022149/1500000: episode: 439, duration: 26.122s, episode steps: 3042, steps per second: 116, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1025766/1500000: episode: 440, duration: 31.146s, episode steps: 3617, steps per second: 116, episode reward:
                                                                                                                -2.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1029187/1500000: episode: 441, duration: 29.380s, episode steps: 3421, steps per second: 116, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1031131/1500000: episode: 442, duration: 16.992s, episode steps:
                                                                  1944, steps per second: 114, episode reward:
                                                                                                                17.000, mean reward: 0.009 [-1.000, 1.000], mean action
1034415/1500000: episode: 443. duration: 27.843s. episode steps: 3284. steps per second: 118. episode reward: 3.000. mean reward: 0.001 [-1.000. 1.000], mean action:
.037103/1500000: episode: 444, duration: 22.871s, episode steps:
                                                                  2688, steps per second: 118, episode reward:
                                                                                                                10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1039985/1500000: episode: 445, duration: 24.504s, episode steps: 2882, steps per second: 118, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1042563/1500000: episode: 446, duration: 22.125s, episode steps:
                                                                 2578, steps per second: 117, episode reward:
                                                                                                                9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1044968/1500000: episode: 447, duration: 20.715s, episode steps: 2405, steps per second: 116, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
1047760/1500000: episode: 448, duration: 24.188s, episode steps:
                                                                  2792, steps per second: 115, episode reward:
                                                                                                                10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1051062/1500000: episode: 449, duration: 28.872s, episode steps: 3302, steps per second: 114, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1054602/1500000: episode: 450, duration: 30.481s, episode steps:
                                                                  3540, steps per second: 116, episode reward:
                                                                                                                7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1057976/1500000: episode: 451, duration: 28.973s, episode steps: 3374, steps per second: 116, episode reward:
                                                                                                                4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1060983/1500000: episode: 452, duration: 26.591s, episode steps:
                                                                  3007, steps per second: 113, episode reward:
                                                                                                                5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1064569/1500000: episode: 453, duration: 31.402s, episode steps: 3586, steps per second: 114, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1067382/1500000: episode: 454, duration: 26.710s, episode steps: 2813, steps per second: 105, episode reward:
                                                                                                                7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1070368/1500000: episode: 455, duration: 28.695s, episode steps: 2986, steps per second: 104, episode reward:
                                                                                                                11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1073428/1500000: episode: 456, duration: 28.841s, episode steps:
                                                                  3060, steps per second: 106, episode reward:
                                                                                                                8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1076473/1500000: episode: 457, duration: 28.016s, episode steps: 3045, steps per second: 109, episode reward: 1078954/1500000: episode: 458, duration: 23.466s, episode steps: 2481, steps per second: 106, episode reward:
                                                                                                                3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
                                                                  2481, steps per second:
                                                                                                                14.000, mean reward: 0.006
1080978/1500000: episode: 459, duration: 18.447s, episode steps: 2024, steps per second: 110, episode reward:
                                                                                                                16.000, mean reward: 0.008 [-1.000, 1.000], mean action
1083641/1500000: episode: 460, duration: 22.375s, episode steps: 2663, steps per second: 119, episode reward:
                                                                                                                8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1087394/1500000: episode: 461, duration: 31.629s, episode steps: 3753, steps per second: 119, episode reward:
                                                                                                                -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1090450/1500000: episode: 462, duration: 25.474s, episode steps:
                                                                  3056, steps per second: 120, episode reward:
1093620/1500000: episode: 463, duration: 26.641s, episode steps: 3170, steps per second: 119, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1096574/1500000: episode: 464, duration: 24.725s, episode steps:
                                                                  2954, steps per second: 119, episode reward:
                                                                                                                8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1098670/1500000: episode: 465, duration: 17.484s, episode steps: 2096, steps per second: 120, episode reward: 15.000, mean reward: 0.007 [-1.000, 1.000], mean action
                                                                                                                6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1101592/1500000: episode: 466, duration: 24.450s, episode steps:
                                                                  2922, steps per second: 120, episode reward:
1103269/1500000: episode: 467, duration: 14.051s, episode steps: 1677, steps per second: 119, episode reward: 20.000, mean reward: 0.012 [-1.000, 1.000], mean action
                                                                  3269, steps per second: 119, episode reward:
1108857/1500000: episode: 469, duration: 19.557s, episode steps: 2319, steps per second: 119, episode reward: 15.000, mean reward: 0.006 [-1.000, 1.000], mean action
1111705/1500000: episode: 470, duration: 23.855s, episode steps:
                                                                  2848, steps per second: 119, episode reward:
                                                                                                                3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1114054/1500000: episode: 471, duration: 19.661s, episode steps: 2349, steps per second: 119, episode reward: 13,000, mean reward: 0.006 [-1.000, 1.000], mean action
1116852/1500000: episode: 472, duration: 23.427s, episode steps:
                                                                 2798, steps per second: 119, episode reward:
                                                                                                                -6.000, mean reward: -0.002 [-1.000, 1.000], mean actio
1120416/1500000: episode: 473, duration: 29.903s, episode steps: 3564, steps per second: 119, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1123639/1500000: episode: 474, duration: 27.399s,
                                                                  3223, steps per second: 118, episode reward:
                                                                                                                7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1126512/1500000: episode: 475, duration: 24.095s, episode steps: 2873, steps per second: 119, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1129803/1500000: episode: 476, duration: 27.799s, episode steps:
                                                                  3291, steps per second: 118, episode reward:
                                                                                                                1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1133567/1500000: episode: 477, duration: 31.600s, episode steps: 3764, steps per second: 119, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1136377/1500000: episode: 478, duration: 23.524s, episode steps:
                                                                  2810, steps per second: 119, episode reward:
                                                                                                                6.000, mean reward: 0.002 [-1.000,
1139531/1500000: episode: 479, duration: 26.527s, episode steps: 3154, steps per second: 119, episode reward:
                                                                                                                6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1142043/1500000: episode: 480, duration: 21.384s, episode steps:
                                                                  2512, steps per second: 117, episode reward:
                                                                                                                12.000, mean reward: 0.005
1144501/1500000: episode: 481, duration: 21.716s, episode steps: 2458, steps per second: 113, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
1147009/1500000: episode: 482, duration: 22.366s, episode steps:
                                                                  2508, steps per second: 112, episode reward:
                                                                                                                12.000, mean reward: 0.005
1149464/1500000: episode: 483, duration: 21.892s, episode steps: 2455, steps per second: 112, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
 52030/1500000: episode: 484, duration: 23.002s, episode steps:
                                                                  2566, steps per second: 112, episode reward:
                                                                                                                   .000, mean reward: 0.004
1154534/1500000: episode: 485, duration: 22.271s, episode steps: 2504, steps per second: 112, episode reward: 13,000, mean reward: 0.005 [-1.000, 1.000], mean action
1157523/1500000: episode: 486, duration: 26.023s, episode steps:
                                                                 2989, steps per second: 115, episode reward:
1160552/1500000: episode: 487, duration: 26.125s, episode steps: 3029, steps per second: 116, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1162706/1500000: episode: 488, duration: 18.431s, episode steps:
                                                                 2154, steps per second: 117, episode reward:
                                                                                                                17.000, mean reward: 0.008
1165180/1500000: episode: 489, duration: 21.443s, episode steps: 2474, steps per second: 115, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean action
1168073/1500000: episode: 490, duration: 24.897s, episode steps:
                                                                  2893, steps per second: 116, episode reward:
1170853/1500000: episode: 491, duration: 24.036s, episode steps: 2780, steps per second: 116, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
 73246/1500000: episode: 492, duration: 20.661s, episode steps:
                                                                  2393, steps per second: 116, episode reward:
                                                                                                                   .000, mean reward: 0.006
1175265/1500000: episode: 493, duration: 17.316s, episode steps: 2019, steps per second: 117, episode reward: 17.000, mean reward: 0.008 [-1.000, 1.000], mean action
1177689/1500000: episode: 494, duration: 20.784s, episode steps:
                                                                  2424, steps per second: 117, episode reward:
                                                                                                                13.000, mean reward: 0.005 [-1.000, 1.000], mean action
1180836/1500000: episode: 495, duration: 26.977s, episode steps: 3147, steps per second: 117, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1183130/1500000: episode: 496, duration: 19.823s, episode steps:
                                                                  2294, steps per second: 116, episode reward:
                                                                                                                13.000, mean reward: 0.006
1185409/1500000: episode: 497, duration: 19.698s, episode steps: 2279, steps per second: 116, episode reward: 15.000, mean reward: 0.007 [-1.000, 1.000], mean action
1188596/1500000: episode: 498, duration: 27.609s, episode steps:
                                                                  3187, steps per second: 115, episode reward:
                                                                                                                  .000, mean reward: 0.001 [
1191599/1500000; episode: 499, duration: 26.015s, episode steps: 3003, steps per second: 115, episode reward:
                                                                                                                5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1194738/1500000: episode: 500, duration: 27.499s, episode steps:
                                                                  3139, steps per second: 114, episode reward:
                                                                                                                5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1198189/1500000: episode: 501, duration: 29.862s, episode steps: 3451, steps per second: 116, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1201402/1500000: episode: 502, duration: 28.656s, episode steps:
                                                                 3213, steps per second: 112, episode reward:
                                                                                                                4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1204494/1500000: episode: 503, duration: 27.393s, episode steps: 3092, steps per second: 113, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1206589/1500000: episode: 504, duration: 20.324s, episode steps:
                                                                  2095, steps per second: 103, episode reward:
                                                                                                                15.000, mean reward: 0.007
1209491/1500000: episode: 505. duration: 28.259s. episode steps: 2902. steps per second: 103. episode reward: -3.000. mean reward: -0.001 [-1.000. 1.0001. mean actic
1212272/1500000: episode: 506, duration: 26.587s,
                                                  episode steps:
                                                                  2781, steps per second: 105, episode reward:
                                                                                                                12.000, mean reward: 0.004
                                                                                                                                              -1.000, 1.000], mean action
1215020/1500000: episode: 507, duration: 25.311s, episode steps: 2748, steps per second: 109, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
1217839/1500000: episode: 508, duration: 27.013s, episode steps:
                                                                 2819, steps per second: 104, episode reward:
                                                                                                                10.000, mean reward: 0.004
1220631/1500000: episode: 509, duration: 25.024s, episode steps: 2792, steps per second: 112, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
1223226/1500000: episode: 510, duration: 21.832s, episode steps:
                                                                 2595, steps per second: 119, episode reward:
                                                                                                                13.000, mean reward:
                                                                                                                                            [-1.000, 1.000], mean action
1226145/1500000: episode: 511, duration: 24.517s, episode steps: 2919, steps per second: 119, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
                          512, duration: 20.422s, episode steps:
                                                                  2394, steps per second: 117, episode reward: 14.000, mean reward: 0.006
1230747/1500000: episode: 513, duration: 18.633s, episode steps: 2208, steps per second: 119, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
1232798/1500000: episode: 514, duration: 17.262s, episode steps:
                                                                 2051, steps per second: 119, episode reward: 17.000, mean reward: 0.008 [-1.000, 1.000], mean action
1235864/1500000: episode: 515, duration: 25.933s, episode steps: 3066, steps per second: 118, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
1238424/1500000: episode: 516, duration: 21.418s, episode steps:
                                                                  2560, steps per second: 120, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000],
1241395/1500000: episode: 517, duration: 24.943s, episode steps: 2971, steps per second: 119, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
1243738/1500000: episode:
                          518, duration: 19.735s, episode steps:
                                                                  2343, steps per second: 119, episode reward:
                                                                                                                15.000, mean reward: 0.006
1246534/1500000: episode: 519, duration: 23.508s, episode steps: 2796, steps per second: 119, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1248996/1500000: episode: 520, duration: 20.753s, episode steps:
                                                                 2462, steps per second: 119, episode reward:
                                                                                                                12.000, mean reward: 0.005 [-1.000, 1.000], mean action
1251169/1500000: episode: 521, duration: 18.250s, episode steps: 2173, steps per second: 119, episode reward: 15.000, mean reward: 0.007 [-1.000, 1.000], mean action
1254192/1500000: episode: 522, duration: 25.419s,
                                                   episode steps: 3023, steps per second: 119, episode reward:
                                                                                                                4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1256127/1500000: episode: 523, duration: 16.271s, episode steps: 1935, steps per second: 119, episode reward: 17.000, mean reward: 0.009 [-1.000, 1.000], mean action
                                                                                                                19.000, mean reward: 0.011 [-1.000, 1.000], mean action
1257926/1500000: episode: 524, duration: 15.133s, episode steps:
                                                                  1799, steps per second: 119, episode reward:
1260327/1500000; episode: 525, duration: 20.252s, episode steps: 2401, steps per second: 119, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
1262445/1500000: episode: 526, duration: 17.779s,
                                                  episode steps:
                                                                  2118, steps per second: 119, episode reward:
                                                                                                                16.000, mean reward: 0.008
1264773/1500000; episode: 527, duration: 19.895s, episode steps: 2328, steps per second: 117, episode reward:
                                                                                                                9.000, mean reward: 0.004 [-1.000, 1.000], mean action:
1267029/1500000: episode: 528, duration: 19.062s, episode steps:
                                                                  2256, steps per second: 118, episode reward:
                                                                                                                16.000, mean reward: 0.007
1269675/1500000: episode: 529, duration: 22.453s, episode steps: 2646, steps per second: 118, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
                                                                 1986, steps per second: 119, episode reward:
1271661/1500000: episode: 530, duration: 16.680s,
                                                                                                                16.000, mean reward: 0.008 [-1.000, 1.000], mean action
1274337/1500000: episode: 531, duration: 22.431s, episode steps: 2676, steps per second: 119, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1276856/1500000: episode: 532, duration: 21.204s, episode steps:
                                                                  2519, steps per second: 119, episode reward:
1279797/1500000: episode: 533, duration: 24.882s, episode steps: 2941, steps per second: 118, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1282322/1500000: episode: 534, duration: 21.963s, episode steps:
                                                                 2525, steps per second: 115, episode reward:
                                                                                                                11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1284490/1500000: episode: 535, duration: 19.439s, episode steps: 2168, steps per second: 112, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
1287421/1500000: episode: 536, duration: 26.219s, episode steps:
                                                                 2931, steps per second: 112, episode reward:
                                                                                                                10.000, mean reward: 0.003 [-1.000, 1.000], mean action
1289895/1500000: episode: 537, duration: 22.136s, episode steps: 2474, steps per second: 112, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean action
                          538, duration: 25.179s,
                                                   episode steps:
                                                                        steps per second: 112, episode reward:
                                                                                                                7.000, mean reward: 0.002 [
1295085/1500000: episode: 539, duration: 20.494s, episode steps: 2370, steps per second: 116, episode reward:
                                                                                                                13.000, mean reward: 0.005
                                                                                                                                            [-1.000, 1.000], mean action
                          540, duration: 23.677s, episode steps:
                                                                  2734, steps per second:
                                                                                                                   .000, mean reward: 0.004 [-1.000, 1.000], mean
1300657/1500000: episode: 541, duration: 24.908s, episode steps: 2838, steps per second: 114, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1302988/1500000: episode: 542, duration: 20.212s, episode steps: 2331, steps per second: 115, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean action
1305755/1500000: episode: 543, duration: 24.170s, episode steps: 2767, steps per second: 114, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action: 1308441/1500000: episode: 544, duration: 23.426s, episode steps: 2686, steps per second: 115, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
```

```
1311459/1500000: episode: 545, duration: 26.293s, episode steps: 3018, steps per second: 115, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1314176/1500000: episode: 546, duration: 23.403s, episode steps: 2717, steps per second: 116, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1316322/1500000: episode: 547, duration: 18.384s, episode steps:
                                                                  2146, steps per second: 117, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean action
1319392/1500000: episode: 548, duration: 26.355s, episode steps: 3070, steps per second: 116, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
                                                                                                                 -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1322446/1500000: episode: 549, duration: 26.562s, episode steps: 3054, steps per second: 115, episode reward:
1324863/1500000: episode: 550, duration: 21.015s, episode steps: 2417, steps per second: 115, episode reward: 11.000, mean reward: 0.005 [-1.000, 1.000], mean action
 1327387/1500000: episode:
                           551, duration: 21.964s, episode steps:
                                                                   2524, steps per second: 115, episode reward:
                                                                                                                 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1330593/1500000: episode: 552, duration: 27.911s, episode steps: 3206, steps per second: 115, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
  32864/1500000: episode:
                           553, duration: 19.849s, episode steps:
                                                                   2271, steps per second: 114, episode reward:
                                                                                                                 13.000, mean reward: 0.006 [-1.000, 1.000], mean action
1335429/1500000: episode: 554, duration: 22.776s, episode steps: 2565, steps per second: 113, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1337717/1500000: episode: 555, duration: 20.913s, episode steps: 2288, steps per second: 109, episode reward:
                                                                                                                 14.000, mean reward: 0.006
                                                                                                                                            [-1.000, 1.000], mean action
1340406/1500000: episode: 556, duration: 23.735s, episode steps: 2689, steps per second: 113, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
 1343320/1500000: episode: 557, duration: 28.390s, episode steps:
                                                                   2914, steps per second: 103, episode reward:
                                                                                                                 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1346586/1500000: episode: 558, duration: 31.718s, episode steps: 3266, steps per second: 103, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
                           559, duration: 25.239s, episode steps:
 1349192/1500000: episode:
                                                                         steps per second: 103, episode reward:
                                                                                                                 9.000, mean reward: 0.003
                                                                                                                                            [-1.000, 1.000], mean action:
1351990/1500000: episode: 560, duration: 25.626s, episode steps: 2798, steps per second: 109, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1354333/1500000: episode: 561, duration: 22.682s, episode steps:
                                                                   2343, steps per second: 103, episode reward:
                                                                                                                 14.000, mean reward: 0.006
1356756/1500000: episode: 562, duration: 21.992s, episode steps: 2423, steps per second: 110, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1358909/1500000: episode: 563, duration: 18.382s, episode steps:
                                                                  2153, steps per second: 117, episode reward:
                                                                                                                 14.000, mean reward: 0.007
                                                                                                                                             [-1.000, 1.000], mean action
1362133/1500000: episode: 564, duration: 27.266s, episode steps:
                                                                  3224, steps per second: 118, episode reward:
                                                                                                                 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 1364669/1500000: episode: 565, duration: 21.488s, episode steps:
                                                                   2536, steps per second: 118, episode reward:
                                                                                                                 10.000, mean reward: 0.004
                                                                                                                                                     1.000], mean action
1367247/1500000: episode: 566, duration: 21.889s, episode steps:
                                                                  2578, steps per second: 118, episode reward: 14.000, mean reward: 0.005
                                                                                                                                            [-1.000.
                           567, duration: 23.041s,
                                                   episode steps:
                 episode:
                                                                         steps per second:
                                                                                           116, episode reward:
                                                                                                                 11.000, mean reward: 0.004
1372284/1500000: episode: 568, duration: 20.146s, episode steps: 2357, steps per second: 117, episode reward: 16.000, mean reward: 0.007
                                                                                                                                            [-1.000,
                                                                                                                                                     1.0001, mean action
                                                                  2345, steps per second: 117, episode reward:
1374629/1500000: episode: 569, duration: 20.059s, episode steps:
                                                                                                                 16.000, mean reward: 0.007
                                                                                                                                             [-1.000, 1.000], mean action
1377247/1500000: episode: 570, duration: 22.109s, episode steps:
                                                                  2618, steps per second: 118, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
 1379044/1500000: episode:
                           571, duration: 15.182s,
                                                   episode steps:
                                                                   1797, steps per second: 118, episode reward:
                                                                                                                 20.000, mean reward:
1381259/1500000: episode: 572, duration: 18.774s, episode steps: 2215, steps per second: 118, episode reward: 15.000, mean reward: 0.007
                                                                                                                                            [-1.000, 1.000], mean action
 1383317/1500000: episode: 573, duration: 17.494s, episode steps:
                                                                   2058, steps per second: 118, episode reward:
                                                                                                                 18.000, mean reward: 0.009
1386171/1500000: episode: 574, duration: 24.208s, episode steps: 2854, steps per second: 118, episode reward: 10.000, mean reward: 0.004 [-1.000,
                                                                                                                                                     1.0001, mean action
1388107/1500000: episode: 575, duration: 16.399s, episode steps:
                                                                  1936, steps per second: 118, episode reward:
                                                                                                                 18.000, mean reward: 0.009
1390763/1500000: episode: 576, duration: 22.460s, episode steps: 2656, steps per second: 118, episode reward:
                                                                                                                 11.000, mean reward: 0.004 [-1.000,
                                                                                                                                                     1.0001, mean action
                                                                   2998, steps per second: 118, episode reward:
1396188/1500000: episode: 578, duration: 20.544s, episode steps: 2427, steps per second: 118, episode reward: 11.000, mean reward: 0.005 [-1.000, 1.000], mean action
 1398879/1500000: episode:
                           579, duration: 22.762s, episode steps:
                                                                   2691, steps per second: 118, episode reward:
                                                                                                                 5.000, mean reward: 0.002 [
1402010/1500000: episode: 580, duration: 26.456s, episode steps: 3131, steps per second: 118, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 1404792/1500000: episode: 581, duration: 23.484s, episode steps:
                                                                   2782, steps per second: 118, episode reward:
                                                                                                                 9.000, mean reward: 0.003
1407370/1500000: episode: 582, duration: 22.296s, episode steps: 2578, steps per second: 116, episode reward: 12.000, mean reward: 0.005
                                                                                                                                            [-1.000, 1.000], mean action
1409599/1500000: episode: 583, duration: 18.988s,
                                                                  2229, steps per second: 117, episode reward:
                                                   episode steps:
                                                                                                                 12.000, mean reward: 0.005
                                                                                                                                             [-1.000,
1411723/1500000: episode: 584, duration: 18.051s, episode steps: 2124, steps per second: 118, episode reward: 15.000, mean reward: 0.007
                                                                                                                                            [-1.000, 1.000], mean action
 1414263/1500000: episode: 585, duration: 21.543s, episode steps:
                                                                   2540, steps per second: 118, episode reward:
                                                                                                                 15.000, mean reward: 0.006
1417080/1500000: episode: 586, duration: 24.139s, episode steps: 2817, steps per second: 117, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action: 1419223/1500000: episode: 587, duration: 19.307s, episode steps: 2143, steps per second: 111, episode reward: 14.000, mean reward: 0.007 [-1.000, 1.000], mean action
1421071/1500000: episode: 588, duration: 16.700s, episode steps: 1848, steps per second: 111, episode reward: 19.000, mean reward: 0.010 [-1.000, 1.000], mean action
 1423887/1500000: episode: 589, duration: 25.305s, episode steps:
                                                                  2816, steps per second: 111, episode reward:
                                                                                                                 7.000, mean reward: 0.002 [
1427161/1500000: episode: 590, duration: 29.541s, episode steps: 3274, steps per second: 111, episode reward:
                                                                                                                 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1429550/1500000: episode:
                           591, duration: 20.842s,
                                                   episode steps:
                                                                   2389,
                                                                         steps per second: 115, episode reward:
                                                                                                                 17.000, mean reward: 0.007
1432302/1500000: episode: 592, duration: 23.856s, episode steps: 2752, steps per second: 115, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
 1435967/1500000: episode: 593, duration: 31.742s, episode steps:
                                                                   3665, steps per second: 115, episode reward:
                                                                                                                   000, mean reward: 0.000 [-1.000, 1.000], mean action:
1438095/1500000: episode: 594, duration: 18.457s, episode steps: 2128, steps per second: 115, episode reward: 15.000, mean reward: 0.007 [-1.000, 1.000], mean action
1440774/1500000: episode: 595, duration: 23.244s, episode steps:
                                                                  2679, steps per second: 115, episode reward:
                                                                                                                 10.000, mean reward: 0.004
                                                                                                                                            [-1.000, 1.000], mean action
1443606/1500000: episode: 596, duration: 25.020s, episode steps: 2832, steps per second: 113, episode reward:
                                                                                                                 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 1446196/1500000: episode: 597, duration: 23.206s, episode steps:
                                                                                                                 11.000, mean reward: 0.004
                                                                  2590, steps per second: 112, episode reward:
1448198/1500000: episode: 598, duration: 17.297s, episode steps: 2002, steps per second: 116, episode reward:
                                                                                                                 18.000, mean reward: 0.009 [-1.000, 1.000], mean action
                                                                         steps per second: 115, episode reward:
                 episode:
                           599, duration: 24.485s, episode steps:
                                                                                                                 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1453288/1500000: episode: 600. duration: 19.537s. episode steps: 2264. steps per second: 116. episode reward:
                                                                                                                 13.000, mean reward: 0.006 [-1.000, 1.000], mean action
 456098/1500000: episode: 601, duration: 24.260s, episode steps:
                                                                   2810, steps per second: 116, episode reward:
                                                                                                                   .000, mean reward: 0.004 [-1.000, 1.000], mean action
1458391/1500000: episode: 602, duration: 19.953s, episode steps: 2293, steps per second: 115, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean action
1460987/1500000: episode: 603, duration: 22.564s, episode steps:
                                                                  2596, steps per second: 115, episode reward:
                                                                                                                 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1463295/1500000: episode: 604, duration: 20.018s, episode steps: 2308, steps per second: 115, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
 1465534/1500000: episode: 605, duration: 19.486s, episode steps:
                                                                   2239, steps per second: 115, episode reward:
1467812/1500000: episode: 606, duration: 19.855s, episode steps: 2278, steps per second: 115, episode reward: 13.000, mean reward: 0.006 [-1.000, 1.000], mean action
1470815/1500000: episode:
                           607, duration: 26.175s, episode steps:
                                                                   3003, steps per second: 115, episode reward:
                                                                                                                 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1473242/1500000; episode: 608, duration: 21.201s, episode steps: 2427, steps per second: 114, episode reward:
                                                                                                                 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
 1476443/1500000: episode: 609, duration: 28.648s, episode steps:
                                                                  3201, steps per second: 112, episode reward:
                                                                                                                 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1478870/1500000: episode: 610, duration: 21.984s, episode steps: 2427, steps per second: 110, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean action
1481190/1500000: episode: 611, duration: 21.242s, episode steps:
                                                                  2320, steps per second: 109, episode reward:
                                                                                                                 13.000, mean reward: 0.006
                                                                                                                                            [-1.000, 1.000], mean action
1483608/1500000: episode: 612, duration: 23.772s, episode steps: 2418, steps per second: 102, episode reward:
                                                                                                                 9.000, mean reward: 0.004 [-1.000, 1.000], mean action:
 1485951/1500000: episode: 613, duration: 22.901s, episode steps:
                                                                   2343, steps per second: 102, episode reward:
                                                                                                                 13.000, mean reward: 0.006
1489250/1500000: episode: 614, duration: 31.782s, episode steps: 3299, steps per second: 104, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
                           615, duration: 23.028s, episode steps:
                 episode:
                                                                         steps per second: 110, episode reward:
                                                                                                                 9.000, mean reward: 0.004
1494399/1500000: episode: 616, duration: 24.940s, episode steps: 2619, steps per second: 105, episode reward: 14.000, mean reward: 0.005 [-1.000, 1.000], mean action
 1496660/1500000: episode: 617, duration: 21.256s, episode steps:
                                                                  2261, steps per second: 106, episode reward:
                                                                                                                 14.000, mean reward: 0.006
                                                                                                                                            [-1.000, 1.000], mean action
1498748/1500000: episode: 618, duration: 18.592s, episode steps: 2088, steps per second: 112, episode reward: 15.000, mean reward: 0.007 [-1.000, 1.000], mean action
done, took 13087.561 seconds
```

Pesos guardados en: 202001101719_dqn_PongDeterministic-v0_weights.h5f





Testing for 10 episodes Episode 1: reward: 18.000, steps: 2097 Episode 2: reward: 13.000, steps: 2524 Episode 3: reward: 16.000, steps: 2123 Episode 4: reward: 14.000, steps: 2272 Episode 5: reward: 13.000, steps: 2279 Episode 6: reward: 15.000, steps: 2435 7: reward: 8.000, steps: 2669 Episode 8: reward: 13.000, steps: 2309 Episode 9: reward: 12.000, steps: 2665 Episode 10: reward: 12.000, steps: 2484 mean reward: 13.4

mean_steps: 2385.7 (13.4, 2385.7)

7 of 7