```
822/2500000: episode: 1, duration: 7.890s, episode steps: 822, steps per second: 104, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
  1646/2500000: episode: 2, duration: 2.550s, episode steps: 824, steps per second: 323, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
  2526/2500000: episode: 3, duration: 2.808s, episode steps: 880, steps per second: 313, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
                            4, duration: 2.945s, episode steps: 921, steps per second: 313, episode reward: -20.000, mean reward: -0.022
                                                                                                                                                         [-1.000,
                                                                                                                                                                   1.000], mean action:
  4271/2500000: episode: 5, duration: 2.620s, episode steps: 824, steps per second: 315, episode reward: -21.000, mean reward: -0.025
                                                                                                                                                         [-1.000, 0.000], mean action:
  5035/2500000: episode: 6, duration: 2.441s, episode steps: 764, steps per second: 313, episode reward: -21.000, mean reward: -0.027
  5859/2500000: episode: 7, duration: 2.586s, episode steps: 824, steps per second: 319, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
                                                    episode steps: 1011, steps per second: 311, episode reward: -20.000, mean reward: -0.020
                            8, duration: 3.249s,
                                                                                                                                                         [-1.000, 1.000], mean action
  7670/2500000: episode: 9, duration: 2.571s, episode steps: 800, steps per second: 311, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
                                            2.527s, episode steps: 792, steps per second: 313, episode reward: -21.000, mean reward: -0.027
  9347/2500000: episode: 11, duration: 2.781s, episode steps: 885, steps per second: 318, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
 10171/2500000: episode: 12, duration: 4.949s, episode steps: 824, steps per second: 166, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 11130/2500000; episode: 13. duration: 8.645s, episode steps: 959. steps per second: 111. episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
                                             .183s, episode steps: 794, steps per second: 111, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
 12891/2500000: episode: 15, duration: 8.804s, episode steps: 967, steps per second: 110, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action:
 13743/2500000: episode: 16, duration:
                                            7.664s, episode steps: 852, steps per second: 111, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 14550/2500000: episode: 17, duration: 7.334s, episode steps: 807, steps per second: 110, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
  15774/2500000: episode: 18, duration:
                                               .042s, episode steps:
                                                                        1224, steps per second: 111, episode reward: -18.000, mean reward:
                                                                                                                                                    -0.015 [-1.000, 1.000], mean actic
 16690/2500000: episode: 19, duration: 8.511s, episode steps: 916, steps per second: 108, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
                                            6.947s, episode steps: 764, steps per second: 110, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
 17454/2500000: episode: 20, duration:
 18237/2500000: episode: 21, duration:
                                            7.384s, episode steps: 783, steps per second: 106, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
                                            8.585s, episode steps: 931, steps per second: 108, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
 20039/2500000: episode: 23, duration: 7.901s, episode steps: 871, steps per second: 110, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
                                            7.977s, episode steps: 868, steps per second: 109, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
 20907/2500000: episode: 24, duration:
 21827/2500000: episode: 25. duration: 8.269s. episode steps: 920. steps per second: 111. episode reward: -20.000. mean reward: -0.022 [-1.000. 1.000]. mean action:
                                            9.198s, episode steps: 1012, steps per second: 110, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action
 23777/2500000: episode: 27, duration: 8.443s, episode steps: 938, steps per second: 111, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
                                            8.141s, episode steps: 900, steps per second: 111, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
 25617/2500000: episode: 29, duration: 8.485s, episode steps: 940, steps per second: 111, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
                                             .728s, episode steps: 848, steps per second: 110, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000],
 27558/2500000: episode: 31, duration: 10.050s, episode steps: 1093, steps per second: 109, episode reward: -17.000, mean reward: -0.016 [-1.000, 1.000], mean actic
 28531/2500000: episode: 32, duration:
                                            8.782s, episode steps: 973, steps per second: 111, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
 29491/2500000: episode: 33, duration: 8.586s, episode steps: 960, steps per second: 112, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
                                            7.972s, episode steps: 884, steps per second: 111, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
 31207/2500000: episode: 35, duration: 7.487s, episode steps: 832, steps per second: 111, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
                                            7.441s, episode steps: 825, steps per second: 111, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 32032/2500000: episode: 36, duration:
 32903/2500000: episode: 37, duration: 7.879s, episode steps: 871, steps per second: 111, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
                                              .064s, episode steps: 783, steps per second: 111, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
 34587/2500000; episode: 39. duration: 8.183s, episode steps: 901. steps per second: 110. episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
                                            7.775s, episode steps: 853, steps per second: 110, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 35440/2500000: episode: 40, duration:
 36461/2500000; episode: 41. duration: 9.304s, episode steps: 1021, steps per second: 110. episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action
 37419/2500000: episode: 42, duration: 8.705s, episode steps: 958, steps per second: 110, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000],
 38620/2500000: episode: 43, duration: 10.885s, episode steps: 1201, steps per second: 110, episode reward: -17.000, mean reward: -0.014 [-1.000, 1.000], mean actic
 39579/2500000: episode: 44, duration: 8.708s, episode steps: 959, steps per second: 110, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action: 40362/2500000: episode: 45, duration: 7.077s, episode steps: 783, steps per second: 111, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
 41427/2500000: episode: 46, duration: 9.632s, episode steps: 1065, steps per second: 111, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action
 42293/2500000: episode: 47, duration: 7.877s, episode steps: 866, steps per second: 110, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
                                            7.665s, episode steps: 856, steps per second: 112, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
 43149/2500000: episode: 48, duration:
 43913/2500000: episode: 49, duration: 6.931s, episode steps: 764, steps per second: 110, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
 44863/2500000: episode: 50, duration: 8.528s, episode steps: 950, steps per second: 111, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
 45861/2500000: episode: 51, duration: 9.063s, episode steps: 998, steps per second: 110, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean action:
                                            8.908s, episode steps: 990, steps per second: 111, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
 47772/2500000; episode: 53, duration: 8.337s, episode steps: 921, steps per second: 110, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
 48681/2500000: episode: 54, duration: 8.235s, episode steps: 909, steps per second: 110, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
 49581/2500000; episode: 55, duration: 8.203s, episode steps: 900, steps per second: 110, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
                                            9.010s, episode steps: 950, steps
                                                                                   per second: 105, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
 50531/2500000: episode: 56, duration:
 51314/2500000: episode: 57, duration: 7.087s, episode steps: 783, steps per second: 110, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
                                            8.159s, episode steps: 888, steps per second: 109, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
 53100/2500000: episode: 59, duration: 8.089s, episode steps: 898, steps per second: 111, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
 53922/2500000: episode: 60, duration: 7.509s, episode steps: 822, steps per second: 109, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
 54948/2500000; episode: 61. duration: 9.293s, episode steps: 1026, steps per second: 110. episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action
 55731/2500000: episode: 62, duration: 7.060s, episode steps: 783, steps per second: 111, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.0001, mean action: 56801/2500000: episode: 63, duration: 9.620s, episode steps: 1070, steps per second: 111, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action: 56801/2500000: episode: 63, duration: 9.620s, episode steps: 1070, steps per second: 111, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action: 56801/2500000: episode: 63, duration: 9.620s, episode steps: 1070, steps per second: 111, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action: 56801/2500000: episode: 63, duration: 9.620s, episode steps: 1070, steps per second: 111, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action: 56801/2500000: episode: 63, duration: 9.620s, episode steps: 1070, steps per second: 111, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action: 56801/2500000: episode: 63, duration: 9.620s, episode steps: 1070, steps per second: 111, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action: 56801/2500000: episode: 63, duration: 9.620s, episode steps: 1070, steps per second: 111, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action: 56801/2500000: episode: 63, duration: 9.620s, episode steps: 1070, steps per second: 111, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action: 56801/2500000: episode: 63, duration: 9.620s, episode steps: 1070, steps per second: 111, episode reward: -19.000, mean reward: -10.000 [-1.000], mean action: 9.000 [-1.000], mean episode: 63, duration: 9.620s, episode steps: 1070, steps per second: 111, episode reward: -19.000, mean reward: -10.000 [-1.000], mean episode: 9.000 [-1.000], mean episod
                                            8.885s, episode steps: 990, steps per second: 111, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
 57791/2500000: episode: 64, duration:
 58741/2500000: episode: 65, duration: 8.638s, episode steps: 950, steps per second: 110, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
                                             .760s, episode steps: 1090, steps per second: 112, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean action
 60698/2500000: episode: 67, duration:
                                            7.902s, episode steps: 867, steps per second: 110, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
 61874/2500000: episode: 68, duration: 10.511s, episode steps: 1176, steps per second: 112, episode reward: -18.000, mean reward: -0.015 [-1.000, 1.000], mean actic
 62980/2500000: episode: 69, duration: 10.000s, episode steps: 1106, steps per second: 111, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean actic
                   episode: 70, duration: 8.979s, episode steps: 993, steps per second: 111, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action:
 63973/2500000:
 64918/2500000: episode: 71, duration: 8.483s, episode steps: 945, steps per second: 111, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
                                            7.868s, episode steps: 881, steps per second: 112, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
 65799/2500000: episode: 72, duration:
 66704/2500000: episode: 73, duration: 8.109s, episode steps: 905, steps per second: 112, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
  57858/2500000: episode: 74, duration: 10.512s, episode steps: 1154, steps per second: 110, episode reward: -19.000, mean reward: -0.016 [-1.000, 1.00]
 69116/2500000: episode: 75, duration: 11.465s, episode steps: 1258, steps per second: 110, episode reward: -18.000, mean reward: -0.014 [-1.000, 1.000], mean actic
                                            11.466s, episode steps: 1273, steps per second: 111, episode reward: -19.000, mean reward: -0.015 [-1.000, 1.000], mean actic
 70389/2500000: episode: 76, duration:
 71480/2500000: episode: 77, duration:
                                            9.881s, episode steps: 1091, steps per second: 110, episode reward: -18.000, mean reward: -0.016 [-1.000, 1.000], mean action
                                            9.910s, episode steps: 1099, steps per second: 111, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean action
 73697/2500000: episode: 79, duration: 10.191s, episode steps: 1118, steps per second: 110, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean actic
                                            9.168s, episode steps: 1022, steps per second: 111, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action
 75825/2500000: episode: 81, duration: 9.943s, episode steps: 1106, steps per second: 111, episode reward: -18.000, mean reward: -0.016 [-1.000, 1.000], mean action
 77026/2500000: episode: 82, duration: 10.789s, episode steps: 1201, steps per second: 111, episode reward: -20.000, mean reward: -0.017 [-1.000, 1.000], mean action
 78124/2500000: episode: 83, duration: 9.912s, episode steps: 1098, steps per second: 111, episode reward: -21.000, mean reward: -0.019 [-1.000, 0.000], mean action
 79027/2500000: episode: 84, duration: 8.133s, episode steps: 903, steps per second: 111, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
 79959/2500000: episode: 85, duration: 8.309s, episode steps: 932, steps per second: 112, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
                                            6.925s, episode steps: 764, steps per second: 110, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
 80723/2500000: episode: 86, duration:
 81549/2500000: episode: 87, duration: 7.471s, episode steps: 826, steps per second: 111, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
 82448/2500000: episode: 88, duration: 8.044s, episode steps: 899, steps per second: 112, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
 83212/2500000; episode: 89. duration: 6.893s, episode steps: 764. steps per second: 111. episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
                                                                                                       episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action
 84282/2500000: episode: 90, duration: 9.822s, episode steps: 1070, steps per second: 109,
 85410/2500000: episode: 91, duration: 10.620s, episode steps: 1128, steps per second: 106, episode reward: -18.000, mean reward: -0.016 [-1.000, 1.000], mean actic
 86624/2500000: episode: 92, duration: 10.953s, episode steps: 1214, steps per second: 111, episode reward: -19.000, mean reward: -0.016 (-1.000, 1.000), mean actic
 87827/2500000: episode: 93, duration: 10.819s, episode steps: 1203, steps per second: 111, episode reward: -17.000, mean reward: -0.014 [-1.000, 1.000], mean actic
 89221/2500000: episode: 94, duration: 12.577s, episode steps: 1394, steps per second: 111, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean actic
 90406/2500000; episode: 95, duration: 10.734s, episode steps: 1185, steps per second: 110, episode reward: -19.000, mean reward: -0.016 [-1.000, 1.000], mean actic
 91534/2500000: episode: 96, duration: 10.154s, episode steps: 1128, steps per second: 111, episode reward: -21.000, mean reward: -0.019 [-1.000, 0.000], mean actic
 92774/2500000: episode: 97, duration: 11.048s, episode steps: 1240, steps per second: 112, episode reward: -21.000, mean reward: -0.017 [-1.000, 0.000], mean actic
 94223/2500000: episode: 98, duration: 13.073s, episode steps: 1449, steps per second: 111, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean actic
 95883/2500000: episode: 99, duration: 14.929s, episode steps: 1660, steps per second: 111, episode reward: -17.000, mean reward: -0.010 [-1.000, 1.000], mean actic
 96903/2500000: episode: 100, duration: 9.201s, episode steps: 1020, steps per second: 111, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000]
 98704/2500000: episode: 101, duration: 16.122s, episode steps: 1801, steps per second: 112, episode reward: -17.000, mean reward: -0.009 [-1.000, 1.000], mean acti
100191/2500000: episode: 102, duration: 13.343s, episode steps: 1487, steps per second: 111, episode reward: -21.000, mean reward: -0.014 [-1.000, 0.000], mean acti
101490/2500000; episode: 103, duration: 11.563s, episode steps: 1299, steps per second: 112, episode reward: -21.000, mean reward: -0.016 [-1.000, 0.000], mean acti
102786/2500000: episode: 104, duration: 11.501s, episode steps: 1296, steps per second: 113, episode reward: -20.000, mean reward: -0.015 [-1.000, 1.000], mean acti
104334/2500000: episode: 105, duration: 13.835s, episode steps: 1548, steps per second: 112, episode reward: -18.000, mean reward: -0.012 [-1.000, 1.000], mean acti
106140/2500000: episode: 106, duration: 16.090s, episode steps: 1806, steps per second: 112, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean acti 107303/2500000: episode: 107, duration: 10.402s, episode steps: 1163, steps per second: 112, episode reward: -20.000, mean reward: -0.017 [-1.000, 1.000], mean acti
108755/2500000: episode: 108, duration: 12.979s, episode steps: 1452, steps per second: 112, episode reward: -21.000, mean reward: -0.014 [-1.000, 0.000], mean acti
```

```
110406/2500000: episode: 109, duration: 14.809s, episode steps: 1651, steps per second: 111, episode reward: -16.000, mean reward: -0.010 [-1.000, 1.000], mean acti
111863/2500000: episode: 110, duration: 13.002s, episode steps: 1457, steps per second: 112, episode reward: -18.000, mean reward: -0.012 [-1.000, 1.000], mean acti
113302/2500000: episode: 111, duration: 12.780s, episode steps: 1439, steps per second: 113, episode reward: -20.000, mean reward: -0.014 [-1.000,
114754/2500000: episode: 112, duration: 12.984s, episode steps: 1452, steps per second: 112, episode reward: -21.000, mean reward: -0.014 [-1.000,
                                                                                                                                                         0.0001, mean acti
116338/2500000: episode: 113, duration: 14.067s, episode steps: 1584, steps per second: 113, episode reward: -21.000, mean reward: -0.013
                                                                                                                                                [-1.000,
                                                                                                                                                         0.000], mean acti
117852/2500000: episode: 114, duration: 13.555s, episode steps: 1514, steps per second: 112, episode reward: -20.000, mean reward: -0.013 [-1.000,
                                                                                                                                                         1.000], mean acti
119087/2500000: episode: 115, duration: 11.621s, episode steps: 1235, steps per second: 106, episode reward: -20.000, mean reward: -0.016
                                                                                                                                                         1.000], mean acti
120980/2500000: episode: 116, duration: 17.086s, episode steps: 1893, steps per second: 111, episode reward: -17.000, mean reward: -0.009 [-1.000, 122269/2500000: episode: 117, duration: 11.615s, episode steps: 1289, steps per second: 111, episode reward: -20.000, mean reward: -0.016 [-1.000,
                                                                                                                                                         1.0001. mean acti
123764/2500000: episode: 118, duration: 13.359s, episode steps: 1495, steps per second: 112, episode reward: -19.000, mean reward: -0.013 [-1.000,
                                                                                                                                                         1.0001, mean acti
124742/2500000: episode: 119, duration: 8.717s, episode steps: 978, steps per second: 112, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action
125831/2500000: episode: 120, duration: 9.695s, episode steps: 1089, steps per second: 112, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean actic
127009/2500000: episode: 121, duration: 10.571s, episode steps: 1178, steps per second: 111, episode reward: -19.000, mean reward: -0.016
128678/2500000: episode: 122, duration: 15.004s, episode steps: 1669, steps per second: 111, episode reward: -17.000, mean reward: -0.010 [-1.000, 1.000], mean acti
 30708/2500000: episode: 123, duration: 18.086s, episode steps:
                                                                  2030, steps per second: 112, episode reward: -18.000, mean reward: -0.009
                                                                                                                                                         1.000], mean acti
132726/2500000: episode: 124, duration: 18.084s, episode steps: 2018, steps per second: 112, episode reward: -16.000, mean reward: -0.008 [-1.000,
                                                                                                                                                         1.0001, mean acti
 34843/2500000: episode: 125, duration: 18.875s, episode steps: 2117, steps per second: 112, episode reward: -14.000, mean reward: -0.007
                                                                                                                                                [-1.000,
136985/2500000: episode: 126, duration: 19.190s, episode steps: 2142, steps per second: 112, episode reward: -14.000, mean reward: -0.007 [-1.000, 1.000], mean acti
138373/2500000: episode: 127, duration: 12.449s, episode steps: 1388, steps per second: 111, episode reward: -20.000, mean reward: -0.014
                                                                                                                                                [-1.000, 1.000], mean acti
140257/2500000: episode: 128, duration: 16.823s, episode steps: 1884, steps per second: 112, episode reward: -18.000, mean reward: -0.010
                                                                                                                                                [-1.000, 1.000], mean acti
141728/2500000: episode: 129, duration: 13.153s, episode steps: 1471, steps per second: 112, episode reward: -18.000, mean reward: -0.012
                                                                                                                                                [-1.000, 1.000], mean acti
143608/2500000: episode: 130, duration: 16.886s, episode steps: 1880, steps per second: 111, episode reward: -20.000, mean reward: -0.011 144985/2500000: episode: 131, duration: 12.328s, episode steps: 1377, steps per second: 112, episode reward: -19.000, mean reward: -0.014
                                                                                                                                                [-1.000, 1.000], mean acti
146745/2500000: episode: 132, duration: 15.771s, episode steps: 1760, steps per second: 112, episode reward: -19.000, mean reward: -0.011
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
48212/2500000: episode: 133, duration: 13.116s, episode steps: 1467, steps per second: 112, episode reward: -18.000, mean reward: -0.012
                                                                                                                                                [-1.000, 1.000], mean acti
150159/2500000: episode: 134, duration: 17.351s, episode steps: 1947, steps per second: 112, episode reward: -15.000, mean reward: -0.008
                                                                                                                                                         1.0001, mean acti
                                                                                                                                                [-1.000,
151820/2500000: episode: 135, duration: 14.944s, episode steps:
                                                                  1661, steps per second: 111, episode reward: -19.000, mean reward: -0.011
153628/2500000: episode: 136, duration: 16.779s, episode steps: 1808, steps per second: 108, episode reward: -18.000, mean reward: -0.010 [-1.000,
                                                                                                                                                         1.0001. mean acti
 55633/2500000: episode: 137, duration: 18.111s, episode steps:
                                                                  2005, steps per second: 111, episode reward: -18.000, mean reward: -0.009
                                                                                                                                                [-1.000,
                                                                                                                                                         1.000], mean acti
157473/2500000: episode: 138, duration: 16.487s, episode steps: 1840, steps per second: 112, episode reward: -19.000, mean reward: -0.010 [-1.000,
                                                                                                                                                         1.0001, mean acti
158975/2500000: episode: 139, duration: 13.405s, episode steps:
                                                                  1502, steps per second: 112, episode reward: -20.000, mean reward: -0.013
160589/2500000: episode: 140, duration: 14.425s, episode steps: 1614, steps per second: 112, episode reward: -18.000, mean reward: -0.011
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
162697/2500000: episode: 141, duration: 18.904s, episode steps:
                                                                  2108, steps per second: 112, episode reward: -15.000, mean reward: -0.007
164565/2500000: episode: 142, duration: 16.641s, episode steps: 1868, steps per second: 112, episode reward: -17.000, mean reward: -0.009 [-1.000,
                                                                                                                                                         1.000], mean acti
166184/2500000: episode: 143, duration: 14.467s, episode steps:
                                                                  1619, steps per second: 112, episode reward: -19.000, mean reward:
                                                                                                                                                         1.000], mean acti
168264/2500000: episode: 144, duration: 18.614s, episode steps: 2080, steps per second: 112, episode reward: -18.000, mean reward: -0.009 [-1.000,
                                                                                                                                                         1.0001, mean acti
170551/2500000: episode: 145, duration: 20.409s, episode steps:
                                                                  2287, steps per second: 112, episode reward: -18.000, mean reward: -0.008
                                                                                                                                                         1.000], mean acti
172950/2500000: episode: 146, duration: 21.418s, episode steps: 2399, steps per second: 112, episode reward: -14.000, mean reward: -0.006
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
174803/2500000: episode: 147, duration: 16.570s, episode steps:
                                                                  1853, steps per second: 112, episode reward: -18.000, mean reward: -0.010
177476/2500000: episode: 148, duration: 23.748s, episode steps: 2673, steps per second: 113, episode reward: -14.000, mean reward: -0.005 [-1.000,
                                                                                                                                                         1.000], mean acti
178907/2500000: episode: 149, duration: 12.810s, episode steps:
                                                                  1431, steps per second: 112, episode reward: -20.000, mean reward: -0.014
180922/2500000: episode: 150, duration: 17.930s, episode steps: 2015, steps per second: 112, episode reward: -19.000, mean reward: -0.009 [-1.000,
                                                                                                                                                         1.0001. mean acti
                          151, duration: 10.993s, episode steps:
                                                                  1222, steps per second: 111, episode reward: -21.000, mean reward: -0.017
182144/2500000: episode:
183390/2500000; episode: 152, duration: 11.171s, episode steps: 1246, steps per second: 112, episode reward: -20.000, mean reward: -0.016
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
 85499/2500000: episode: 153, duration: 18.853s, episode steps:
                                                                  2109, steps per second: 112, episode reward: -18.000, mean reward: -0.009
187425/2500000: episode: 154, duration: 17.550s, episode steps: 1926, steps per second: 110, episode reward: -18.000, mean reward: -0.009 [-1.000,
                                                                                                                                                         1.000], mean acti
188448/2500000: episode: 155, duration: 9.572s, episode steps: 1023, steps per second: 107, episode reward: -20.000, mean reward: -0.020 [
190777/2500000: episode: 156, duration: 20.791s, episode steps: 2329, steps per second: 112, episode reward: -16.000, mean reward: -0.007 [-1.000, 1.000], mean acti
 93007/2500000: episode: 157, duration: 19.939s, episode steps: 2230, steps per second: 112, episode reward: -16.000, mean reward: -0.007
194829/2500000: episode: 158, duration: 16.355s, episode steps: 1822, steps per second: 111, episode reward: -16.000, mean reward: -0.009 [-1.000,
                                                                                                                                                         1.0001, mean acti
196770/2500000: episode: 159, duration: 17.306s, episode steps:
                                                                  1941, steps per second: 112, episode reward: -17.000, mean reward: -0.009
199265/2500000: episode: 160, duration: 22.435s, episode steps: 2495, steps per second: 111, episode reward: -15.000, mean reward: -0.006
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
201561/2500000: episode: 161, duration: 20.509s, episode steps: 2296, steps per second: 112, episode reward: -17.000, mean reward: -0.007
                                                                                                                                                         1.000], mean acti
                                                                                                                                                 -1.000,
203536/2500000: episode: 162, duration: 17.774s, episode steps: 1975, steps per second: 111, episode reward: -15.000, mean reward: -0.008 [-1.000,
                                                                                                                                                         1.000], mean acti
205494/2500000: episode: 163, duration: 17.575s, episode steps:
                                                                  1958, steps per second: 111, episode reward: -17.000, mean reward: -0.009
207997/2500000: episode: 164. duration: 22,479s. episode steps: 2503. steps per second: 111. episode reward: -15,000. mean reward: -0.006 [-1,000.
                                                                                                                                                         1.0001. mean acti
210180/2500000: episode: 165, duration: 19.511s, episode steps:
                                                                  2183, steps per second: 112, episode reward: -17.000, mean reward: -0.008
212433/2500000; episode: 166, duration: 20.282s, episode steps: 2253, steps per second: 111, episode reward: -16.000, mean reward: -0.007 [-1.000, 1.000], mean acti
214972/2500000: episode: 167, duration: 22.592s, episode steps:
                                                                  2539, steps per second: 112, episode reward: -14.000, mean reward: -0.006
217765/2500000: episode: 168, duration: 24.800s, episode steps: 2793, steps per second: 113, episode reward: -14.000, mean reward: -0.005 220062/2500000: episode: 169, duration: 20.465s, episode steps: 2297, steps per second: 112, episode reward: -16.000, mean reward: -0.007
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
222589/2500000: episode: 170, duration: 23.113s, episode steps: 2527, steps per second: 109, episode reward: -16.000, mean reward: -0.006 [-1.000,
                                                                                                                                                         1.000], mean acti
224658/2500000: episode:
                          171, duration: 18.509s, episode steps:
                                                                  2069, steps per second: 112, episode reward: -19.000, mean reward: -0.009
                                                                                                                                                           .000], mean acti
227119/2500000: episode: 172, duration: 21.957s, episode steps: 2461, steps per second: 112, episode reward: -18.000, mean reward: -0.007
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
229525/2500000: episode: 173, duration: 21.461s, episode steps:
                                                                  2406, steps per second: 112, episode reward: -16.000, mean reward: -0.007
231602/2500000: episode: 174, duration: 18.497s, episode steps: 2077, steps per second: 112, episode reward: -20.000, mean reward: -0.010 [-1.000,
                                                                                                                                                         1.0001, mean acti
233807/2500000: episode: 175, duration: 19.625s, episode steps: 2205, steps per second: 112, episode reward: -19.000, mean reward: -0.009
                                                                                                                                                         1.000], mean acti
                                                                                                                                                 -1.000,
236684/2500000: episode: 176, duration: 25.604s, episode steps: 2877, steps per second: 112, episode reward: -19.000, mean reward: -0.007
                                                                                                                                                 -1.000,
                                                                                                                                                         1.000], mean acti
238881/2500000: episode: 177, duration: 19.594s, episode steps:
                                                                  2197, steps per second: 112, episode reward: -17.000, mean reward: -0.008
240636/2500000: episode: 178, duration: 15.613s, episode steps: 1755, steps per second: 112, episode reward: -19.000, mean reward: -0.011 [-1.000,
                                                                                                                                                         1.0001. mean acti
                          179, duration: 22.193s, episode steps:
243119/2500000: episode:
                                                                  2483, steps per second: 112, episode reward: -17.000, mean reward: -0.007
245648/2500000; episode: 180, duration: 22.652s, episode steps: 2529, steps per second: 112, episode reward: -16.000, mean reward: -0.006
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
247989/2500000: episode: 181, duration: 20.844s, episode steps:
                                                                  2341, steps per second: 112, episode reward: -18.000, mean reward: -0.008
250939/2500000: episode: 182, duration: 26.345s, episode steps: 2950, steps per second: 112, episode reward: -13.000, mean reward: -0.004 [-1.000,
                                                                                                                                                         1.000], mean acti
                                                                  1850, steps per second: 112, episode reward: -18.000, mean reward: -0.010
252789/2500000: episode: 183, duration: 16.574s, episode steps:
254775/2500000: episode: 184, duration: 17.795s, episode steps: 1986, steps per second: 112, episode reward: -19.000, mean reward: -0.010 [-1.000,
                                                                                                                                                         1.000], mean acti
256768/2500000: episode: 185, duration: 18.109s, episode steps:
                                                                  1993, steps per second: 110, episode reward: -17.000, mean reward: -0.009
259288/2500000: episode: 186, duration: 22.857s, episode steps: 2520, steps per second: 110, episode reward: -16.000, mean reward: -0.006 [-1.000,
                                                                                                                                                         1.0001, mean acti
261067/2500000: episode: 187, duration: 15.771s, episode steps:
                                                                  1779, steps per second: 113, episode reward: -17.000, mean reward: -0.010
263374/2500000: episode: 188, duration: 20.612s, episode steps: 2307, steps per second: 112, episode reward: -16.000, mean reward: -0.007
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
                                                                  1979, steps per second: 112, episode reward: -16.000, mean reward: -0.008
265353/2500000: episode: 189, duration: 17.648s, episode steps:
267883/2500000: episode: 190, duration: 22.568s, episode steps: 2530, steps per second: 112, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
270447/2500000: episode: 191, duration: 22.779s, episode steps:
                                                                  2564, steps per second: 113, episode reward: -14.000, mean reward:
273575/2500000: episode: 192, duration: 27.880s, episode steps: 3128, steps per second: 112, episode reward: -15.000, mean reward: -0.005
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
275910/2500000: episode: 193, duration: 20.869s, episode steps: 2335, steps per second: 112, episode reward: -15.000, mean reward: -0.006
                                                                                                                                                 -1.000,
                                                                                                                                                         1.000], mean acti
278199/2500000: episode: 194, duration: 20.257s, episode steps: 2289, steps per second: 113, episode reward: -17.000, mean reward: -0.007
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
280512/2500000: episode: 195, duration: 20.553s, episode steps: 2313, steps per second: 113, episode reward: -11.000, mean reward: -0.005
                                                                                                                                                [-1.000,
                                                                                                                                                         1.000], mean acti
283417/2500000: episode: 196, duration: 25.940s, episode steps: 2905, steps per second: 112, episode reward: -11.000, mean reward: -0.004
                                                                                                                                                [-1.000,
                                                                                                                                                         1.000], mean acti
285739/2500000: episode: 197, duration: 20.623s, episode steps:
                                                                  2322, steps per second: 113, episode reward: -14.000, mean reward: -0.006
288289/2500000: episode: 198, duration: 22.638s, episode steps: 2550, steps per second: 113, episode reward: -15.000, mean reward: -0.006 [-1.000,
                                                                                                                                                         1.000], mean acti
290633/2500000: episode: 199, duration: 21.111s, episode steps:
                                                                  2344,
                                                                        steps per second: 111, episode reward: -17.000, mean reward: -0.007
291893/2500000: episode: 200, duration: 11.507s, episode steps: 1260, steps per second: 110, episode reward: -21.000, mean reward: -0.017
                                                                                                                                                [-1.000.
                                                                                                                                                         0.0001, mean acti
294630/2500000: episode: 201, duration: 24.473s, episode steps:
                                                                  2737, steps per second: 112, episode reward: -11.000, mean reward: -0.004
297200/2500000: episode: 202, duration: 22.961s, episode steps: 2570, steps per second: 112, episode reward: -13.000, mean reward: -0.005
                                                                                                                                                [-1.000,
                                                                                                                                                         1.000], mean acti
299489/2500000: episode: 203, duration: 20.318s,
                                                                  2289, steps per second: 113, episode reward: -14.000, mean reward: -0.006
                                                                                                                                                         1.0001, mean acti
301356/2500000: episode: 204, duration: 16.526s, episode steps: 1867, steps per second: 113, episode reward: -15.000, mean reward: -0.008
                                                                                                                                                [-1.000,
                                                                                                                                                         1.000], mean acti
303917/2500000: episode: 205, duration: 22.663s, episode steps:
                                                                  2561, steps per second: 113, episode reward: -16.000, mean reward: -0.006
306910/2500000: episode: 206, duration: 26.629s, episode steps: 2993, steps per second: 112, episode reward: -12.000, mean reward: -0.004 [-1.000,
                                                                                                                                                         1.0001, mean acti
309687/2500000: episode: 207, duration: 24.796s, episode steps:
                                                                  2777, steps per second: 112, episode reward: -13.000, mean reward: -0.005
311503/2500000: episode: 208, duration: 16.144s, episode steps: 1816, steps per second: 112, episode reward: -15.000, mean reward: -0.008
                                                                                                                                                [-1.000,
                                                                                                                                                         1.0001, mean acti
314014/2500000: episode: 209, duration: 22.397s, episode steps:
                                                                  2511, steps per second: 112, episode reward: -15.000, mean reward: -0.006
316746/2500000: episode: 210, duration: 24.316s, episode steps: 2732, steps per second: 112, episode reward: -15.000, mean reward: -0.005
                                                                                                                                                [-1.000,
                                                                                                                                                         1.000], mean acti
318984/2500000:
                episode: 211, duration: 20.003s,
                                                  episode steps:
                                                                  2238,
                                                                         steps per second: 112, episode reward: -17.000,
                                                                                                                                        -0 008
321723/2500000: episode: 212, duration: 24.261s, episode steps: 2739, steps per second: 113, episode reward: -11.000, mean reward: -0.004
                                                                                                                                                [-1.000,
                                                                                                                                                         1.000], mean acti
                episode: 213, duration: 22.365s, episode steps:
                                                                  2481, steps per second:
                                                                                           111, episode reward: -12.000, mean reward:
                                                                                                                                                [-1.000,
                                                                                                                                                           .000], mean act:
327082/2500000: episode: 214, duration: 25.994s, episode steps: 2878, steps per second: 111, episode reward: -13.000, mean reward: -0.005 [-1.000,
                                                                                                                                                         1.0001, mean acti
329800/2500000: episode: 215, duration: 24.290s, episode steps: 2718, steps per second: 112, episode reward: -17.000, mean reward: -0.006
                                                                                                                                                         1.000], mean acti
332677/2500000: episode: 216, duration: 25.923s, episode steps: 2877, steps per second: 111, episode reward: -18.000, mean reward: -0.006 [-1.000, 335437/2500000: episode: 217, duration: 24.462s, episode steps: 2760, steps per second: 113, episode reward: -13.000, mean reward: -0.005 [-1.000,
                                                                                                                                                         1.0001, mean acti
                                                                                                                                                         1.0001, mean acti
```

```
338291/2500000: episode: 218, duration: 25.429s, episode steps: 2854, steps per second: 112, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
340803/2500000: episode: 219, duration: 22.323s, episode steps: 2512, steps per second: 113, episode reward: -16.000, mean reward: -0.006 [-1.000, 1.000], mean acti
343493/2500000: episode: 220, duration: 24.026s, episode steps: 2690, steps per second: 112, episode reward: -15.000, mean reward: -0.006 [-1.000,
346521/2500000: episode: 221, duration: 27.024s, episode steps: 3028, steps per second: 112, episode reward: -13.000, mean reward: -0.004 [-1.000,
                                                                                                                                                      1.0001, mean acti
349413/2500000: episode: 222, duration: 25.824s, episode steps: 2892, steps per second: 112, episode reward: -15.000, mean reward: -0.005
                                                                                                                                             [-1.000,
                                                                                                                                                      1.000], mean acti
352245/2500000: episode: 223, duration: 25.263s, episode steps: 2832, steps per second: 112, episode reward: -12.000, mean reward: -0.004
                                                                                                                                             [-1.000,
                                                                                                                                                      1.000], mean acti
355276/2500000: episode: 224, duration: 27.015s, episode steps:
                                                                 3031, steps per second: 112, episode reward: -15.000, mean reward:
                                                                                                                                                      1.000], mean acti
358422/2500000: episode: 225. duration: 28.343s. episode steps: 3146. steps per second: 111. episode reward: -11.000. mean reward: -0.003
                                                                                                                                              [-1.000.
                                                                                                                                                      1.0001. mean acti
361173/2500000: episode: 226, duration: 24.931s, episode steps:
                                                                  2751, steps per second: 110, episode reward: -13.000, mean reward: -0.005
364553/2500000: episode: 227, duration: 30.435s, episode steps: 3380, steps per second: 111, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
367831/2500000: episode: 228, duration: 29.240s, episode steps: 3278, steps per second: 112, episode reward: -12.000, mean reward: -0.004
                                                                                                                                                      1.000], mean acti
                                                                                                                                             [-1.000,
371280/2500000: episode: 229, duration: 30.874s, episode steps: 3449, steps per second: 112, episode reward: -13.000, mean reward: -0.004
                                                                                                                                             [-1.000,
                                                                                                                                                      1.0001, mean acti
                                                                 2535, steps per second: 111, episode reward: -16.000, mean reward: -0.006
373815/2500000: episode: 230, duration: 22.751s, episode steps:
376309/2500000: episode: 231, duration: 22.406s, episode steps: 2494, steps per second: 111, episode reward: -17.000, mean reward: -0.007 [-1.000,
                                                                                                                                                      1.0001. mean acti
378831/2500000: episode: 232, duration: 22.669s, episode steps:
                                                                 2522, steps per second: 111, episode reward: -15.000, mean reward: -0.006
                                                                                                                                                       1.000], mean acti
381223/2500000: episode: 233, duration: 21.454s, episode steps: 2392, steps per second: 111, episode reward: -12.000, mean reward: -0.005 [-1.000,
                                                                                                                                                      1.0001, mean acti
384196/2500000: episode: 234, duration: 26.687s, episode steps:
                                                                 2973, steps per second: 111, episode reward: -11.000, mean reward: -0.004 [-1.000,
                                                                                                                                                      1.000], mean acti
388173/2500000: episode: 235, duration: 35.689s, episode steps: 3977, steps per second: 111, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
390943/2500000: episode: 236, duration: 24.867s, episode steps: 2770, steps per second: 111, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
394461/2500000: episode: 237, duration: 31.947s, episode steps: 3518, steps per second: 110, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
397636/2500000: episode: 238, duration: 28.937s, episode steps: 3175, steps per second: 110, episode reward: -10.000, mean reward: -0.003
                                                                                                                                             [-1.000, 1.000], mean acti
400083/2500000: episode: 239, duration: 21.796s, episode steps: 2447, steps per second: 112, episode reward: -14.000, mean reward: -0.006 [-1.000, 403261/2500000: episode: 240, duration: 28.795s, episode steps: 3178, steps per second: 110, episode reward: -10.000, mean reward: -0.003 [-1.000,
                                                                                                                                                      1.000], mean acti
405810/2500000: episode: 241, duration: 22.910s, episode steps: 2549, steps per second: 111, episode reward: -12.000, mean reward: -0.005 [-1.000,
                                                                                                                                                      1.0001, mean acti
408559/2500000: episode: 242, duration: 24.611s, episode steps: 2749, steps per second: 112, episode reward: -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
412780/2500000: episode: 243, duration: 37.822s, episode steps: 4221, steps per second: 112, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
                                                                 3332, steps per second: 112, episode reward: -11.000, mean reward: -0.003
416112/2500000: episode: 244, duration: 29.770s, episode steps:
419816/2500000: episode: 245, duration: 33.304s, episode steps: 3704, steps per second: 111, episode reward: -9.000, mean reward: -0.002 [-1.000, 1.000], mean actic
423611/2500000: episode: 246, duration: 34.002s, episode steps:
                                                                 3795, steps per second: 112, episode reward: -10.000, mean reward: -0.003
                                                                                                                                                    00,
427096/2500000: episode: 247, duration: 31.568s, episode steps: 3485, steps per second: 110, episode reward: -13.000, mean reward: -0.004 [-1.000,
                                                                                                                                                      1.0001, mean acti
429695/2500000: episode: 248, duration: 23.361s, episode steps:
                                                                 2599, steps per second: 111, episode reward: -13.000, mean reward: -0.005 [-1.000,
433026/2500000: episode: 249, duration: 30.257s, episode steps: 3331, steps per second: 110, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
                                                                 3190, steps per second: 111, episode reward: -11.000, mean reward: -0.003
436216/2500000: episode: 250, duration: 28.614s, episode steps:
438720/2500000: episode: 251, duration: 22.408s, episode steps: 2504, steps per second: 112, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
442856/2500000: episode: 252, duration: 36.979s, episode steps:
                                                                 4136, steps per second: 112, episode reward: -9.000, mean reward: -0.002 [-1.000, 1.000], mean actic
445620/2500000; episode: 253, duration: 24.776s, episode steps: 2764, steps per second: 112, episode reward: -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
450238/2500000: episode: 254, duration: 41.346s, episode steps:
                                                                 4618, steps per second: 112, episode reward: -6.000, mean reward: -0.001 [-1.000, 1.000], mean actic
454872/2500000: episode: 255, duration: 41.484s, episode steps: 4634, steps per second: 112, episode reward: -6.000, mean reward: -0.001 [-1.000, 1.000], mean actic
458815/2500000: episode: 256, duration: 35.403s, episode steps: 3943, steps per second: 111, episode reward: -9.000, mean reward: -0.002
                                                                                                                                            [-1.000,
463420/2500000: episode: 257, duration: 41.493s, episode steps: 4605, steps per second: 111, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
466007/2500000: episode: 258, duration: 23.463s, episode steps:
                                                                 2587, steps per second: 110, episode reward: -15.000, mean reward: -0.006
469152/2500000: episode: 259, duration: 28.230s, episode steps: 3145, steps per second: 111, episode reward: -13.000, mean reward: -0.004 [-1.000, 1.000], mean acti
471587/2500000: episode: 260, duration: 21.742s, episode steps:
                                                                 2435, steps per second: 112, episode reward: -16.000, mean reward: -0.007
474265/2500000; episode: 261, duration: 23.930s, episode steps: 2678, steps per second: 112, episode reward: -17.000, mean reward: -0.006
                                                                                                                                             [-1.000,
                                                                                                                                                      1.0001, mean acti
476815/2500000: episode: 262, duration: 22.721s, episode steps:
                                                                 2550, steps per second: 112, episode reward: -14.000, mean reward: -0.005
480249/2500000: episode: 263, duration: 30.781s, episode steps: 3434, steps per second: 112, episode reward: -14.000, mean reward: -0.004
                                                                                                                                             [-1.000,
                                                                                                                                                      1.000], mean acti
                                                                 3187, steps per second: 111, episode reward: -14.000, mean reward: -0.004
483436/2500000: episode: 264, duration: 28.681s, episode steps:
486419/2500000: episode: 265, duration: 26.610s, episode steps: 2983, steps per second: 112, episode reward: -11.000, mean reward: -0.004 [-1.000,
                                                                                                                                                      1.000], mean acti
489943/2500000: episode: 266, duration: 31.473s, episode steps:
                                                                 3524, steps per second: 112, episode reward: -10.000, mean reward: -0.003
493883/2500000: episode: 267, duration: 35.435s, episode steps: 3940, steps per second: 111, episode reward: -14.000, mean reward: -0.004 [-1.000,
                                                                                                                                                      1.0001, mean acti
496488/2500000: episode: 268, duration: 23.259s, episode steps:
                                                                 2605, steps per second: 112, episode reward: -17.000, mean reward: -0.007
499372/2500000: episode: 269, duration: 26.259s, episode steps: 2884, steps per second: 110, episode reward: -20.000, mean reward: -0.007
                                                                                                                                             [-1.000,
                                                                                                                                                      1.0001, mean acti
 02067/2500000: episode: 270, duration: 24.088s, episode steps:
                                                                 2695, steps per second: 112, episode reward: -13.000, mean reward: -0.005
                                                                                                                                               -1.000,
504844/2500000: episode: 271, duration: 24.918s, episode steps: 2777, steps per second: 111, episode reward: -14.000, mean reward: -0.005 [-1.000,
                                                                                                                                                      1.000], mean acti
508331/2500000: episode: 272, duration: 31.170s, episode steps:
                                                                 3487, steps per second: 112, episode reward: -15.000, mean reward: -0.004
509511/2500000: episode: 273. duration: 10.638s. episode steps: 1180. steps per second: 111. episode reward: -19.000. mean reward: -0.016 [-1.000.
                                                                                                                                                      1.0001. mean acti
512573/2500000: episode: 274, duration: 27.268s, episode steps:
                                                                 3062, steps per second: 112, episode reward: -15.000, mean reward: -0.005
516519/2500000; episode: 275, duration: 35.297s, episode steps: 3946, steps per second: 112, episode reward: -9.000, mean reward: -0.002 [-1.000, 1.000], mean actic
519459/2500000: episode: 276, duration: 26.267s, episode steps:
                                                                 2940, steps per second: 112, episode reward: -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
522707/2500000: episode: 277, duration: 28.942s, episode steps: 3248, steps per second: 112, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
                                                                 3067, steps per second: 112, episode reward: -10.000, mean reward: -0.003
525774/2500000: episode: 278, duration: 27.288s, episode steps:
528232/2500000: episode: 279, duration: 22.269s, episode steps: 2458, steps per second: 110, episode reward: -17.000, mean reward: -0.007 [-1.000, 1.000], mean acti
530645/2500000: episode: 280, duration: 21.726s, episode steps:
                                                                 2413, steps per second: 111, episode reward: -14.000, mean reward: -0.006 [-1.000,
533496/2500000; episode: 281, duration: 25.504s, episode steps: 2851, steps per second: 112, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
537093/2500000: episode: 282, duration: 32.419s, episode steps:
                                                                 3597, steps per second: 111, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
541241/2500000: episode: 283, duration: 37.138s, episode steps: 4148, steps per second: 112, episode reward: -5.000, mean reward: -0.001 [-1.000, 1.000], mean actic
545171/2500000: episode: 284, duration: 35.042s, episode steps: 3930, steps per second: 112, episode reward: -5.000, mean reward: -0.001 [-1.000, 1.000], mean actic
549047/2500000: episode: 285, duration: 34.674s, episode steps: 3876, steps per second: 112, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
                                                                 3890, steps per second: 113, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean action
552937/2500000: episode: 286, duration: 34.554s, episode steps:
555493/2500000: episode: 287, duration: 22.809s, episode steps: 2556, steps per second: 112, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
558050/2500000: episode: 288, duration: 22.775s, episode steps:
                                                                 2557, steps per second: 112, episode reward: -12.000, mean reward: -0.005 [-1.000,
561620/2500000; episode: 289, duration: 31.992s, episode steps: 3570, steps per second: 112, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
664727/2500000: episode: 290, duration: 27.877s, episode steps:
                                                                 3107, steps per second: 111, episode reward: -16.000, mean reward: -0.005
569934/2500000: episode: 291, duration: 47.119s, episode steps: 5207, steps per second: 111, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
574596/2500000: episode: 292, duration: 41.588s, episode steps:
                                                                 4662, steps per second: 112, episode reward:
                                                                                                               -5.000, mean reward: -0.001 [-1.000, 1.000], mean action
578585/2500000: episode: 293, duration: 35.753s, episode steps: 3989, steps per second: 112, episode reward: -9.000, mean reward: -0.002 [-1.000, 1.000], mean actic
580898/2500000: episode: 294, duration: 20.667s, episode steps:
                                                                 2313, steps per second: 112, episode reward: -13.000, mean reward: -0.006
583666/2500000; episode: 295, duration: 24.763s, episode steps; 2768, steps per second: 112, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
586438/2500000: episode: 296, duration: 24.737s, episode steps:
                                                                 2772, steps per second: 112, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
590395/2500000: episode: 297, duration: 35.387s, episode steps: 3957, steps per second: 112, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
593490/2500000: episode: 298, duration: 27.714s, episode steps:
                                                                 3095, steps per second: 112, episode reward: -15.000, mean reward: -0.005 [-1.000, 1.000], mean acti
597598/2500000: episode: 299, duration: 36.979s, episode steps: 4108, steps per second: 111, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
                                                                 3667, steps per second: 112, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean action
601265/2500000: episode: 300, duration: 32.809s, episode steps:
603480/2500000; episode: 301, duration: 20.218s, episode steps: 2215, steps per second: 110, episode reward: -16.000, mean reward: -0.007 [-1.000, 1.000], mean acti
605024/2500000: episode: 302, duration: 13.867s, episode steps: 1544, steps per second: 111, episode reward: -18.000, mean reward: -0.012 [-1.000, 1.000], mean acti
607486/2500000: episode: 303, duration: 22.017s, episode steps: 2462, steps per second: 112, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
611209/2500000: episode: 304, duration: 33.298s, episode steps: 3723, steps per second: 112, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000],
614949/2500000: episode: 305, duration: 33.013s, episode steps: 3740, steps per second: 113, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
                                                                 2980, steps per second: 112, episode reward: -15.000, mean reward: -0.005 [-1.000,
617929/2500000: episode: 306, duration: 26.571s, episode steps:
620342/2500000: episode: 307, duration: 21.470s, episode steps: 2413, steps per second: 112, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
624318/2500000: episode: 308, duration: 35.491s, episode steps:
                                                                       steps per second: 112, episode reward: -11.000, mean reward: -0.003 [-1.000,
628617/2500000; episode: 309, duration: 38.444s, episode steps; 4299, steps per second: 112, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
632869/2500000: episode: 310, duration: 38.173s, episode steps:
                                                                 4252, steps per second: 111, episode reward: -10.000, mean reward: -0.002
635911/2500000: episode: 311, duration: 27.143s, episode steps: 3042, steps per second: 112, episode reward: -16.000, mean reward: -0.005 [-1.000, 1.000], mean acti
639385/2500000: episode: 312, duration: 31.374s, episode steps:
                                                                 3474, steps per second: 111, episode reward: -15.000, mean reward: -0.004 [-1.000,
643224/2500000: episode: 313, duration: 34.242s, episode steps: 3839, steps per second: 112, episode reward: -11.000, mean reward: -0.003 [-1.000, 1.000], mean acti
547039/2500000: episode: 314, duration: 34.076s, episode steps:
                                                                 3815, steps per second: 112, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean action
649857/2500000; episode: 315, duration: 25.319s, episode steps: 2818, steps per second: 111, episode reward: -18.000, mean reward: -0.006 [-1.000, 1.000], mean acti
654354/2500000: episode: 316, duration: 40.123s, episode steps: 4497, steps per second: 112, episode reward: -9.000, mean reward: -0.002 [-1.000, 1.000], mean actio
659008/2500000: episode: 317, duration: 41.609s, episode steps: 4654, steps per second: 112, episode reward: -11.000, mean reward: -0.002 [-1.000, 1.000], mean acti
662763/2500000: episode: 318, duration: 33.328s, episode steps:
                                                                 3755, steps per second: 113, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean action
666978/2500000: episode: 319, duration: 37.773s, episode steps: 4215, steps per second: 112, episode reward: -6.000, mean reward: -0.001 [-1.000, 1.000], mean actic
                         320, duration: 31.179s,
670471/2500000: episode:
                                                  episode steps:
                                                                       steps per second: 112, episode reward: -9.000, mean reward: -0.003 [-1.000,
673809/2500000: episode: 321, duration: 30.092s, episode steps: 3338, steps per second: 111, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
                         322, duration: 36.979s, episode steps:
                                                                  4131, steps per second: 112, episode reward:
                                                                                                               -9.000, mean reward: -0.002 [-1.000, 1.000], mean action
681231/2500000; episode: 323, duration: 29.365s, episode steps; 3291, steps per second: 112, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
684321/2500000: episode: 324, duration: 27.702s, episode steps:
                                                                 3090, steps per second: 112, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
687452/2500000: episode: 325, duration: 27.971s, episode steps: 3131, steps per second: 112, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean actic 691096/2500000: episode: 326, duration: 32.544s, episode steps: 3644, steps per second: 112, episode reward: -9.000, mean reward: -0.002 [-1.000, 1.000], mean actic
```

```
695017/2500000: episode: 327, duration: 35.035s, episode steps: 3921, steps per second: 112, episode reward: -12.000, mean reward: -0.003 [-1.000, 1.000], mean acti
 698916/2500000: episode: 328, duration: 34.912s, episode steps: 3899, steps per second: 112, episode reward: -9.000, mean reward: -0.002 [-1.000, 1.000], mean actic
                                                                         2611, steps per second: 112, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
 701527/2500000: episode: 329, duration: 23.291s, episode steps:
 704735/2500000: episode: 330, duration: 28.657s, episode steps: 3208, steps per second: 112, episode reward: -13.000, mean reward: -0.004 [-1.000, 1.000], mean acti
 708806/2500000: episode: 331, duration: 36.798s, episode steps: 4071, steps per second: 111, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
 712187/2500000: episode: 332, duration: 30.191s, episode steps: 3381, steps per second: 112, episode reward: -14.000, mean reward: -0.004 [-1.000, 1.000], mean acti
 715168/2500000: episode: 333, duration: 26.618s, episode steps:
                                                                         2981, steps per second: 112, episode reward: -13.000, mean reward: -0.004 [-1.000,
 718524/2500000: episode: 334, duration: 29.898s, episode steps: 3356, steps per second: 112, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean active reward: -0.001 [-1.000, 1.000],
 727491/2500000: episode: 336, duration: 28.642s, episode steps: 3198, steps per second: 112, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
 730535/2500000: episode: 337, duration: 27.063s, episode steps:
                                                                         3044, steps per second: 112, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
 734254/2500000: episode: 338, duration: 33.473s, episode steps: 3719, steps per second: 111, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
                                                                         2201, steps per second: 111, episode reward: -16.000, mean reward: -0.007
 736455/2500000: episode: 339, duration: 19.742s, episode steps:
 739785/2500000: episode: 340, duration: 29.614s, episode steps: 3330, steps per second: 112, episode reward: -11.000, mean reward: -0.003 [-1.000, 1.000], mean acti
                             341, duration: 21.877s, episode steps:
 742197/2500000: episode:
                                                                         2412, steps per second: 110, episode reward: -13.000, mean reward: -0.005
                                                                                                                                                                       1.000], mean acti
 745694/2500000; episode: 342, duration: 31.388s, episode steps: 3497, steps per second: 111, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
 748398/2500000: episode: 343, duration: 24.248s, episode steps:
                                                                         2704, steps per second: 112, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
 752083/2500000: episode: 344, duration: 32.941s, episode steps: 3685, steps per second: 112, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
 755923/2500000: episode: 345, duration: 34.219s, episode steps: 3840, steps per second: 112, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
 759867/2500000: episode: 346, duration: 35.154s, episode steps: 3944, steps per second: 112, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
 762917/2500000: episode: 347, duration: 27.316s, episode steps:
                                                                         3050, steps per second: 112, episode reward: -12.000, mean reward: -0.004
                                                                                                                                                             [-1.000, 1.000], mean acti
 765667/2500000: episode: 348, duration: 24.761s, episode steps: 2750, steps per second: 111, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti 769791/2500000: episode: 349, duration: 37.101s, episode steps: 4124, steps per second: 111, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
 773546/2500000: episode: 350, duration: 33.597s, episode steps: 3755, steps per second: 112, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
  76505/2500000: episode: 351, duration: 26.955s, episode steps: 2959, steps per second: 110, episode reward: -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
 780071/2500000: episode: 352, duration: 31.975s, episode steps: 3566, steps per second: 112, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
                                                                         3026, steps per second: 112, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
 783097/2500000: episode: 353, duration: 26.923s, episode steps:
 786478/2500000: episode: 354, duration: 30.266s, episode steps: 3381, steps per second: 112, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
 789358/2500000: episode: 355, duration: 25.797s, episode steps:
                                                                         2880, steps per second: 112, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
 792580/2500000: episode: 356, duration: 28.810s, episode steps: 3222, steps per second: 112, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
                                                                         2761, steps per second: 112, episode reward: -11.000, mean reward:
 795341/2500000: episode: 357, duration: 24.645s, episode steps:
                                                                                                                                                             [-1.000, 1.000], mean acti
 798293/2500000: episode: 358, duration: 26.392s, episode steps: 2952, steps per second: 112, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
 801262/2500000: episode: 359, duration: 26.596s, episode steps:
                                                                         2969, steps per second: 112, episode reward: -11.000, mean reward: -0.004 [-1.000,
 804583/2500000: episode: 360, duration: 29.645s, episode steps: 3321, steps per second: 112, episode reward: -11.000, mean reward: -0.003 [-1.000, 1.000], mean acti
 807018/2500000: episode: 361, duration: 21.821s, episode steps:
                                                                         2435, steps per second: 112, episode reward: -16.000, mean reward: -0.007
 809718/2500000; episode: 362, duration: 24.025s, episode steps: 2700, steps per second: 112, episode reward: -15.000, mean reward: -0.006 [-1.000, 1.000], mean acti
 812980/2500000: episode: 363, duration: 29.707s, episode steps:
                                                                         3262, steps per second: 110, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
 816305/2500000: episode: 364, duration: 29.787s, episode steps: 3325, steps per second: 112, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
 819522/2500000: episode: 365, duration: 28.806s, episode steps:
                                                                         3217, steps per second: 112, episode reward: -10.000, mean reward: -0.003 [-1.000,
                                                                                                                                                                       1.0001, mean acti
 823353/2500000: episode: 366, duration: 34.229s, episode steps: 3831, steps per second: 112, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
 827582/2500000: episode: 367, duration: 37.710s, episode steps:
                                                                         4229, steps per second: 112, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 831708/2500000: episode: 368, duration: 36.963s, episode steps: 4126, steps per second: 112, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
                                                                         4050, steps per second: 112, episode reward:
                                                                                                                            -6.000, mean reward: -0.001 [-1.000, 1.000], mean actic
 835758/2500000: episode: 369, duration: 36.284s, episode steps:
 839286/2500000; episode: 370, duration: 31.536s, episode steps: 3528, steps per second: 112, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
 343646/2500000: episode: 371, duration: 38.840s, episode steps:
                                                                         4360, steps per second: 112, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 847098/2500000: episode: 372, duration: 31.283s, episode steps: 3452, steps per second: 110, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
                                                                                                                            -2.000, mean reward: -0.001 [-1.000, 1.000], mean action
 850840/2500000: episode: 373, duration: 33.511s, episode steps:
                                                                         3742, steps per second: 112, episode reward:
 853750/2500000: episode: 374, duration: 26.087s, episode steps: 2910, steps per second: 112, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
 356798/2500000: episode: 375, duration: 27.283s, episode steps:
                                                                         3048, steps per second: 112, episode reward:
                                                                                                                            -10.000, mean reward: -0.003 [-1.000, 1.00
 860705/2500000: episode: 376, duration: 34.917s, episode steps: 3907, steps per second: 112, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
 864386/2500000: episode: 377, duration: 33.110s, episode steps:
                                                                         3681, steps per second: 111, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean action
 867811/2500000: episode: 378, duration: 30.855s, episode steps: 3425, steps per second: 111, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
 872418/2500000: episode: 379, duration: 41.586s, episode steps:
                                                                         4607, steps per second: 111, episode reward:
                                                                                                                             -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
 875483/2500000: episode: 380, duration: 27.555s, episode steps: 3065, steps per second: 111, episode reward: 14.000, mean reward: 0.005 [-1.000, 1.000], mean action
 878984/2500000: episode: 381, duration: 31.334s, episode steps:
                                                                         3501, steps per second: 112, episode reward:
                                                                                                                             9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
 881518/2500000: episode: 382. duration: 23.210s. episode steps: 2534. steps per second: 109. episode reward: 15.000. mean reward: 0.006 [-1.000, 1.000]. mean action
  85161/2500000: episode: 383, duration: 32.592s, episode steps:
                                                                         3643, steps per second: 112, episode reward:
                                                                                                                             -12.000, mean reward: -0.003 [-1.000, 1.00
 888567/2500000; episode: 384, duration: 30.512s, episode steps: 3406, steps per second: 112, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
 891881/2500000: episode: 385, duration: 29.806s, episode steps:
                                                                         3314, steps per second: 111, episode reward:
                                                                                                                             -5.000, mean reward: -0.002 [-1.000, 1.000], mean actic
 895040/2500000: episode: 386, duration: 28.245s, episode steps: 3159, steps per second: 112, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
 898658/2500000: episode: 387, duration: 32.501s, episode steps:
                                                                         3618, steps per second: 111, episode reward:
                                                                                                                             -5.000, mean reward: -0.001 [-1.000, 1.000], mean action
 901027/2500000: episode: 388, duration: 21.206s, episode steps: 2369, steps per second: 112, episode reward: 17.000, mean reward: 0.007 [-1.000, 1.000], mean action
 904479/2500000: episode: 389, duration: 31.251s, episode steps:
                                                                         3452, steps per second: 110, episode reward:
                                                                                                                               000, mean reward: 0.001 [-1.000, 1.000], mean action:
 907458/2500000: episode: 390, duration: 26.595s, episode steps: 2979, steps per second: 112, episode reward: -7.000, mean reward: -0.002 (-1.000, 1.000), mean actic
 911159/2500000: episode: 391, duration: 33.266s, episode steps:
                                                                         3701, steps per second: 111, episode reward:
                                                                                                                            -2.000, mean reward: -0.001 [-1.000, 1.000], mean actic
 914572/2500000: episode: 392, duration: 30.959s, episode steps: 3413, steps per second: 110, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 917805/2500000: episode: 393, duration: 29.049s, episode steps:
                                                                         3233, steps per second: 111, episode reward:
                                                                                                                            10.000, mean reward: 0.003 [-1.000, 1.000], mean action
 921332/2500000: episode: 394, duration: 31.616s, episode steps: 3527, steps per second: 112, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 924853/2500000: episode: 395, duration: 31.487s, episode steps:
                                                                         3521, steps per second: 112, episode reward:
                                                                                                                            4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 928620/2500000: episode: 396, duration: 33.736s, episode steps: 3767, steps per second: 112, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 933574/2500000: episode: 397, duration: 44.267s, episode steps:
                                                                         4954, steps per second: 112, episode reward:
                                                                                                                             3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 937291/2500000: episode: 398, duration: 33.298s, episode steps: 3717, steps per second: 112, episode reward:
                                                                                                                            5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 339670/2500000: episode: 399, duration: 21.228s, episode steps:
                                                                         2379, steps per second: 112, episode reward:
                                                                                                                            18.000, mean reward: 0.008 [-1.000, 1.000], mean action
 943033/2500000: episode: 400, duration: 30.113s, episode steps: 3363, steps per second: 112, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
                                                                                                                            18.000, mean reward: 0.008 [-1.000, 1.000], mean action
 945356/2500000: episode: 401, duration: 20.725s, episode steps:
                                                                         2323, steps per second: 112, episode reward:
 947712/2500000: episode: 402, duration: 21.024s, episode steps: 2356, steps per second: 112, episode reward: 17.000, mean reward: 0.007 [-1.000, 1.000], mean action
 950972/2500000: episode: 403, duration: 29.613s, episode steps:
                                                                         3260, steps per second: 110, episode reward:
                                                                                                                             4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 954417/2500000; episode: 404, duration: 30.833s, episode steps: 3445, steps per second: 112, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 958461/2500000: episode: 405, duration: 36.194s, episode steps:
                                                                         4044, steps per second: 112, episode reward:
                                                                                                                            -4.000, mean reward: -0.001 [-1.000, 1.000], mean action
 961437/2500000: episode: 406, duration: 26.686s, episode steps: 2976, steps per second: 112, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
                                                                         3948, steps per second: 112, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 965385/2500000: episode: 407, duration: 35.168s, episode steps:
 968956/2500000: episode: 408, duration: 31.902s, episode steps: 3571, steps per second: 112, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 973279/2500000: episode: 409, duration: 38.805s, episode steps:
                                                                         4323, steps per second: 111, episode reward:
                                                                                                                             6.000, mean reward: 0.001
 977251/2500000: episode: 410, duration: 35.341s, episode steps: 3972, steps per second: 112, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
 981037/2500000: episode: 411, duration: 33.809s, episode steps:
                                                                         3786, steps per second: 112, episode reward:
                                                                                                                            7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
 984053/2500000: episode: 412, duration: 27.520s, episode steps: 3016, steps per second: 110, episode reward:
                                                                                                                            -5.000, mean reward: -0.002 [-1.000, 1.000], mean actic
 986473/2500000: episode: 413, duration: 21.576s, episode steps: 2420, steps per second: 112, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
 98885/2500000: episode: 414, duration: 21.657s, episode steps: 2412, steps per second: 111, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
 992217/2500000: episode: 415, duration: 29.844s, episode steps:
                                                                         3332, steps per second: 112, episode reward:
                                                                                                                              .000, mean reward: 0.002 [-1.000, 1.000], mean action:
 995716/2500000: episode: 416, duration: 31.343s, episode steps: 3499, steps per second: 112, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
 998031/2500000: episode: 417, duration: 20.788s, episode steps:
                                                                         2315, steps per second: 111, episode reward:
                                                                                                                             13.000, mean reward: 0.006
                                                                                                                                                           [-1.000,
1001687/2500000; episode: 418, duration: 32.603s, episode steps: 3656, steps per second: 112, episode reward:
                                                                                                                            3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1004707/2500000: episode: 419, duration: 27.269s, episode steps:
                                                                         3020, steps per second: 111, episode reward:
                                                                                                                             6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1007574/2500000: episode: 420, duration: 25.722s, episode steps: 2867, steps per second: 111, 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 action
1011175/2500000: episode: 421, duration: 32.336s,
                                                                         3601, steps per second: 111, episode reward:
1014057/2500000: episode: 422, duration: 25.709s, episode steps: 2882, steps per second: 112, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1017488/2500000: episode: 423, duration: 30.944s, episode steps:
                                                                         3431, steps per second: 111, episode reward:
                                                                                                                             8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1020378/2500000: episode: 424, duration: 26.213s, episode steps: 2890, steps per second: 110, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1023499/2500000: episode: 425, duration: 28.044s, episode steps:
                                                                         3121, steps per second: 111, episode reward:
                                                                                                                            9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1026334/2500000: episode: 426, duration: 25.377s, episode steps: 2835, steps per second: 112, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1029139/2500000: episode: 427, duration: 25.055s, episode steps:
                                                                         2805, steps per second: 112, episode reward:
                                                                                                                            14.000, mean reward: 0.005 [-1.000, 1.000], mean action
1031980/2500000: episode: 428, duration: 25.417s, episode steps: 2841, steps per second: 112, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
                                                                         3714,
                                                                                                                                                    -0.000 [-1.000, 1.000], mean action
1035694/2500000:
                   episode: 429, duration: 33.257s,
                                                        episode steps:
                                                                                steps per second: 112, episode reward:
                                                                                                                             -1.000, mean reward:
1039320/2500000: episode: 430, duration: 32.610s, episode steps: 3626, steps per second: 111, episode reward:
                                                                                                                            -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
                  episode: 431, duration: 34.568s, episode steps:
                                                                         3888, steps per second:
                                                                                                    112, episode reward:
                                                                                                                            2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1047116/2500000: episode: 432, duration: 34.947s, episode steps: 3908, steps per second: 112, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1050670/2500000: episode: 433, duration: 31.647s, episode steps:
                                                                         3554, steps per second: 112, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1053687/2500000: episode: 434, duration: 27.423s, episode steps: 3017, steps per second: 110, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action: 1056524/2500000: episode: 435, duration: 25.327s, episode steps: 2837, steps per second: 112, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
```

```
1059793/2500000: episode: 436, duration: 29.278s, episode steps: 3269, steps per second: 112, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1062147/2500000: episode: 437, duration: 20.957s, episode steps: 2354, steps per second: 112, episode reward: 17.000, mean reward: 0.007 [-1.000, 1.000], mean action
1065524/2500000: episode: 438, duration: 30.130s, episode steps: 3377, steps per second: 112, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1068387/2500000: episode: 439, duration: 25.551s, episode steps: 2863, steps per second: 112, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
1071283/2500000: episode: 440, duration: 25.870s, episode steps: 2896, steps per second: 112, episode reward:
                                                                                                                 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1073806/2500000: episode: 441, duration: 22.647s, episode steps: 2523, steps per second: 111, episode reward: 15.000, mean reward: 0.006 [-1.000, 1.000], mean action
1076808/2500000: episode: 442, duration: 26.795s, episode steps:
                                                                   3002, steps per second: 112, episode reward:
                                                                                                                  8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1080552/2500000: episode: 443, duration: 33.436s, episode steps: 3744, steps per second: 112, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
                                                                   2966, steps per second: 112, episode reward:
1083518/2500000: episode: 444, duration: 26.457s, episode steps:
                                                                                                                  6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1086348/2500000: episode: 445, duration: 25.098s, episode steps: 2830, steps per second: 113, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
1089890/2500000: episode: 446, duration: 31.953s, episode steps: 3542, steps per second: 111, episode reward:
                                                                                                                 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1093285/2500000: episode: 447, duration: 30.224s, episode steps: 3395, steps per second: 112, episode reward:
                                                                                                                 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1095964/2500000: episode: 448, duration: 24.171s, episode steps:
                                                                   2679, steps per second: 111, episode reward:
                                                                                                                  11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1099165/2500000: episode: 449, duration: 28.498s, episode steps: 3201, steps per second: 112, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
                                                                                                                   3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1102945/2500000: episode: 450, duration: 33.540s, episode steps:
                                                                   3780, steps per second: 113, episode reward:
1106143/2500000: episode: 451, duration: 28.780s, episode steps: 3198, steps per second: 111, episode reward:
                                                                                                                 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
                                                                   3549, steps per second: 113, episode reward:
                                                                                                                   .000, mean reward: 0.002 [-1.000, 1.000], mean action:
1109692/2500000: episode: 452, duration: 31.535s, episode steps:
1112999/2500000: episode: 453, duration: 29.506s, episode steps: 3307, steps per second: 112, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
1116752/2500000: episode: 454, duration: 33.562s, episode steps:
                                                                  3753, steps per second: 112, episode reward:
                                                                                                                  -2.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1120218/2500000: episode: 455, duration: 31.022s, episode steps: 3466, steps per second: 112, episode reward:
                                                                                                                 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1123217/2500000: episode: 456, duration: 27.352s, episode steps:
                                                                   2999, steps per second: 110, episode reward:
                                                                                                                  9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1125289/2500000: episode: 457, duration: 18.494s, episode steps: 2072, steps per second: 112, episode reward: 19.000, mean reward: 0.009 [-1.000, 1.000], mean action: 1128694/2500000: episode: 458, duration: 30.505s, episode steps: 3405, steps per second: 112, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
                                                                                                                                              -1.000, 1.000], mean action:
                                                                   3405, steps per second:
1132139/2500000: episode: 459, duration: 30.764s, episode steps: 3445, steps per second: 112, episode reward:
                                                                                                                 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1135808/2500000: episode: 460, duration: 32.729s, episode steps:
                                                                  3669, steps per second: 112, episode reward:
                                                                                                                  6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1138009/2500000: episode: 461, duration: 19.789s, episode steps: 2201, steps per second: 111, episode reward: 15.000, mean reward: 0.007 [-1.000, 1.000], mean action
1141144/2500000: episode: 462, duration: 28.191s, episode steps:
                                                                   3135, steps per second: 111, episode reward:
1144683/2500000: episode: 463, duration: 31.655s, episode steps: 3539, steps per second: 112, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1148839/2500000: episode: 464, duration: 37.189s, episode steps:
                                                                   4156, steps per second: 112, episode reward:
                                                                                                                  2.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1152315/2500000: episode: 465, duration: 31.043s, episode steps: 3476, steps per second: 112, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1156297/2500000: episode: 466, duration: 35.930s, episode steps:
                                                                   3982, steps per second: 111, episode reward:
                                                                                                                  4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1160306/2500000: episode: 467, duration: 35.833s, episode steps: 4009, steps per second: 112, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
                                                                                                                  -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1164375/2500000: episode: 468, duration: 36.303s, episode steps:
                                                                   4069, steps per second: 112, episode reward:
1167824/2500000: episode: 469, duration: 30.888s, episode steps: 3449, steps per second: 112, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1171197/2500000: episode: 470, duration: 30.050s, episode steps:
                                                                   3373, steps per second: 112, episode reward:
                                                                                                                  8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1174790/2500000: episode: 471, duration: 32.127s, episode steps: 3593, steps per second: 112, episode reward:
                                                                                                                 -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1178069/2500000: episode: 472, duration: 29.091s, episode steps:
                                                                   3279, steps per second: 113, episode reward:
                                                                                                                 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1182245/2500000: episode: 473, duration: 37.170s, episode steps: 4176, steps per second: 112, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1186396/2500000: episode: 474, duration: 36.859s, episode steps:
                                                                   4151, steps per second: 113, episode reward:
                                                                                                                  -1.000, mean reward: -0.000 [-1.000, 1.000], mean actio
1189187/2500000: episode: 475, duration: 24.670s, episode steps: 2791, steps per second: 113, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1191745/2500000: episode: 476, duration: 23.127s, episode steps:
                                                                   2558, steps per second: 111, episode reward:
                                                                                                                  13.000, mean reward: 0.005 [-1.000, 1.000], mean action
1194898/2500000: episode: 477, duration: 28.126s, episode steps: 3153, steps per second: 112, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1197800/2500000: episode: 478, duration: 25.985s, episode steps:
                                                                   2902, steps per second: 112, episode reward:
                                                                                                                  13.000, mean reward: 0.004 [-1.000, 1.000], mean action
1201635/2500000: episode: 479, duration: 34.117s, episode steps: 3835, steps per second: 112, episode reward:
                                                                                                                 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1205203/2500000: episode: 480, duration: 31.828s, episode steps:
                                                                   3568, steps per second: 112, episode reward:
                                                                                                                  7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1208539/2500000: episode: 481, duration: 30.287s, episode steps: 3336, steps per second: 110, episode reward:
                                                                                                                 13.000, mean reward: 0.004 [-1.000, 1.000], mean action
1212208/2500000: episode: 482, duration: 32.793s, episode steps:
                                                                   3669, steps per second: 112, episode reward:
                                                                                                                   .000, mean reward: 0.002 [-1.000, 1.000], mean action:
1215701/2500000: episode: 483, duration: 31.207s, episode steps: 3493, steps per second: 112, episode reward:
                                                                                                                 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1219076/2500000: episode: 484, duration: 29.933s, episode steps:
                                                                   3375, steps per second: 113, episode reward:
                                                                                                                  10.000, mean reward: 0.003 [-1.000, 1.000], mean action
1223026/2500000: episode: 485, duration: 35.231s, episode steps: 3950, steps per second: 112, episode reward: -3.000, mean reward: -0.001 (-1.000, 1.000), mean actic
1225696/2500000: episode: 486, duration: 24.161s, episode steps:
                                                                   2670, steps per second: 111, episode reward:
                                                                                                                 15.000, mean reward: 0.006 [-1.000, 1.000], mean action
1229065/2500000: episode: 487, duration: 30.360s, episode steps: 3369, steps per second: 111, episode reward:
                                                                                                                 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1233109/2500000: episode: 488, duration: 36.064s, episode steps:
                                                                   4044, steps per second: 112, episode reward:
                                                                                                                  3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1236291/2500000: episode: 489, duration: 28.372s, episode steps: 3182, steps per second: 112, episode reward:
                                                                                                                 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
1238944/2500000: episode: 490, duration: 23.512s, episode steps:
                                                                   2653, steps per second: 113, episode reward:
                                                                                                                  16.000, mean reward: 0.006 [-1.000, 1.000], mean action
1242366/2500000: episode: 491. duration: 30.846s. episode steps: 3422. steps per second: 111. episode reward:
                                                                                                                 -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
1246217/2500000: episode: 492, duration: 34.476s, episode steps:
                                                                   3851, steps per second: 112, episode reward:
                                                                                                                   .000, mean reward: 0.001 [-1.000, 1.000], mean action:
1249500/2500000: episode: 493, duration: 29.250s, episode steps: 3283, steps per second: 112, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1253380/2500000: episode: 494, duration: 34.583s, episode steps:
                                                                   3880, steps per second: 112, episode reward:
                                                                                                                  -2.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1257049/2500000: episode: 495, duration: 32.776s, episode steps: 3669, steps per second: 112, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action: 1260409/2500000: episode: 496, duration: 30.219s, episode steps: 3360, steps per second: 111, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1263168/2500000: episode: 497, duration: 24.570s, episode steps: 2759, steps per second: 112, episode reward: 15.000, mean reward: 0.005 [-1.000, 1.000], mean action
1266020/2500000: episode: 498, duration: 25.409s, episode steps:
                                                                   2852, steps per second: 112, episode reward:
                                                                                                                  10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1269052/2500000: episode: 499, duration: 27.007s, episode steps: 3032, steps per second: 112, episode reward:
                                                                                                                 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1272366/2500000: episode: 500, duration: 29.343s, episode steps:
                                                                   3314, steps per second: 113, episode reward:
                                                                                                                  9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1275656/2500000: episode: 501, duration: 29.276s, episode steps: 3290, steps per second: 112, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
1278210/2500000: episode: 502, duration: 22.962s, episode steps:
                                                                  2554, steps per second: 111, episode reward:
                                                                                                                 14.000, mean reward: 0.005 [-1.000, 1.000], mean action
1281849/2500000: episode: 503, duration: 32.653s, episode steps: 3639, steps per second: 111, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1285472/2500000: episode: 504, duration: 32.382s, episode steps:
                                                                   3623, steps per second: 112, episode reward:
                                                                                                                  8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1289545/2500000: episode: 505, duration: 36.391s, episode steps: 4073, steps per second: 112, episode reward: 2.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1292520/2500000: episode: 506, duration: 26.508s, episode steps:
                                                                   2975, steps per second: 112, episode reward:
                                                                                                                  9.000, mean reward: 0.003
1295896/2500000: episode: 507, duration: 30.447s, episode steps: 3376, steps per second: 111, episode reward:
                                                                                                                 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1298893/2500000: episode: 508, duration: 26.804s, episode steps:
                                                                   2997, steps per second: 112, episode reward:
                                                                                                                  9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1301550/2500000: episode: 509, duration: 23.738s, episode steps: 2657, steps per second: 112, episode reward: 14.000, mean reward: 0.005 [-1.000, 1.000], mean action
1304153/2500000: episode: 510, duration: 23.256s, episode steps:
                                                                  2603, steps per second: 112, episode reward:
                                                                                                                  15.000, mean reward: 0.006 [-1.000,
1307443/2500000: episode: 511, duration: 29.379s, episode steps: 3290, steps per second: 112, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean action
1310184/2500000: episode: 512, duration: 24.656s, episode steps:
                                                                   2741, steps per second: 111, episode reward:
                                                                                                                  15.000, mean reward: 0.005 [-1.000, 1.000], mean action
1313625/2500000: episode: 513, duration: 31.059s, episode steps: 3441, steps per second: 111, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
                                                                                                                 15.000, mean reward: 0.006 [-1.000, 1.000], mean action
1316117/2500000: episode: 514, duration: 22.438s, episode steps:
                                                                  2492, steps per second: 111, episode reward:
1319834/2500000: episode: 515, duration: 33.140s, episode steps: 3717, steps per second: 112, episode reward:
                                                                                                                 -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
                                                                                                                 12.000, mean reward: 0.004 [-1.000, 1.000], mean a
                                                                   2724, steps per second: 112, episode reward:
1326397/2500000: episode: 517, duration: 34.200s, episode steps: 3839, steps per second: 112, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
                          518, duration: 29.966s, episode steps:
1329736/2500000: episode:
                                                                   3339, steps per second: 111, episode reward:
                                                                                                                  5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1333050/2500000; episode: 519, duration: 29.650s, episode steps: 3314, steps per second: 112, episode reward:
                                                                                                                 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1335493/2500000: episode: 520, duration: 21.749s, episode steps:
                                                                  2443, steps per second: 112, episode reward:
                                                                                                                  15.000, mean reward: 0.006 [-1.000, 1.000], mean action
1338349/2500000: episode: 521, duration: 25.541s, episode steps: 2856, steps per second: 112, episode reward: 14.000, mean reward: 0.005 [-1.000, 1.000], mean action
1341251/2500000: episode: 522, duration: 25.812s,
                                                   episode steps: 2902, steps per second: 112, episode reward:
                                                                                                                 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1344342/2500000: episode: 523, duration: 27.715s, episode steps: 3091, steps per second: 112, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1347850/2500000: episode: 524, duration: 31.400s, episode steps:
                                                                   3508, steps per second: 112, episode reward:
                                                                                                                    .000, mean reward: 0.001 [-1.000, 1.000], mean action:
1350946/2500000: episode: 525, duration: 27.607s, episode steps: 3096, steps per second: 112, episode reward: 13.000, mean reward: 0.004 [-1.000, 1.000], mean action
1353956/2500000: episode: 526, duration: 26.926s, episode steps:
                                                                   3010, steps per second: 112, episode reward:
                                                                                                                  7.000, mean reward: 0.002 [
1356872/2500000: episode: 527, duration: 26.198s, episode steps:
                                                                  2916, steps per second: 111, episode reward:
                                                                                                                 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1360066/2500000: episode: 528, duration: 28.588s, episode steps:
                                                                   3194, steps per second: 112, episode reward:
                                                                                                                  -13.000, mean reward:
                                                                                                                                        -0.004 [-1.000, 1.000], mean acti
1363185/2500000: episode: 529, duration: 28.224s, episode steps: 3119, steps per second: 111, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
                                                                                                                  4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1366642/2500000: episode: 530, duration: 31.480s,
                                                                   3457, steps per second: 110, episode reward:
1370815/2500000: episode: 531, duration: 37.310s, episode steps: 4173, steps per second: 112, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1374477/2500000: episode: 532, duration: 32.832s, episode steps:
                                                                   3662, steps per second: 112, episode reward:
                                                                                                                   .000, mean reward: 0.001 [-1.000,
1377510/2500000: episode: 533, duration: 27.135s, episode steps: 3033, steps per second: 112, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
1379763/2500000: episode: 534, duration: 20.506s, episode steps:
                                                                  2253, steps per second: 110, episode reward: 17.000, mean reward: 0.008 [-1.000, 1.000], mean action
1384132/2500000: episode: 535, duration: 39.036s, episode steps: 4369, steps per second: 112, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
                                                                                                                 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1386940/2500000: episode: 536, duration: 25.104s, episode steps:
                                                                   2808, steps per second: 112, episode reward:
1389599/2500000: episode: 537, duration: 23.749s, episode steps: 2659, steps per second: 112, episode reward: 15.000, mean reward: 0.006 [-1.000, 1.000], mean action
1393049/2500000: episode:
                          538, duration: 30.776s,
                                                   episode steps:
                                                                   3450,
                                                                         steps per second: 112, episode reward:
                                                                                                                  8.000,
                                                                                                                         mean reward: 0.002 [-1.000, 1.000],
1396867/2500000: episode: 539, duration: 34.150s, episode steps: 3818, steps per second: 112, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
                                                                   3904, steps per second:
                                                                                                                                        -0.000 [-1.000, 1.000], mean actio
1403936/2500000: episode: 541, duration: 28.410s, episode steps: 3165, steps per second: 111, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1407604/2500000: episode: 542, duration: 32.898s, episode steps:
                                                                   3668, steps per second: 111, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean action
1410419/2500000: episode: 543, duration: 25.194s, episode steps: 2815, steps per second: 112, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action 1412919/2500000: episode: 544, duration: 22.647s, episode steps: 2500, steps per second: 110, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
```

```
1416398/2500000: episode: 545, duration: 31.065s, episode steps: 3479, steps per second: 112, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1419051/2500000: episode: 546, duration: 23.654s, episode steps: 2653, steps per second: 112, episode reward: 14.000, mean reward: 0.005 [-1.000, 1.000], mean action
1421559/2500000: episode: 547, duration: 22.492s, episode steps: 2508, steps per second: 112, episode reward: 16.000, mean reward: 0.006 [-1.000, 1.000], mean action
1424754/2500000: episode: 548, duration: 28.566s, episode steps: 3195, steps per second: 112, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1427469/2500000: episode: 549, duration: 24.155s, episode steps: 2715, steps per second: 112, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
1430431/2500000: episode: 550, duration: 26.299s, episode steps: 2962, steps per second: 113, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean action
1433575/2500000: episode: 551, duration: 28.392s, episode steps: 3144, steps per second: 111, episode reward:
                                                                                                               9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1436214/2500000: episode: 552. duration: 23.763s. episode steps: 2639. steps per second: 111. episode reward: 16.000. mean reward: 0.006 [-1.000, 1.000]. mean action
439592/2500000: episode:
                          553, duration: 30.198s, episode steps: 3378, steps per second: 112, episode reward: 3.000, mean reward: 0.001 [-1.000, 1
1443256/2500000: episode: 554, duration: 32.554s, episode steps: 3664, steps per second: 113, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1446489/2500000: episode: 555, duration: 28.855s, episode steps: 3233, steps per second: 112, episode reward:
                                                                                                               11.000, mean reward: 0.003 [-1.000, 1.000], mean action
1449387/2500000: episode: 556, duration: 25.918s, episode steps: 2898, steps per second: 112, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1452282/2500000: episode: 557, duration: 25.902s, episode steps: 2895, steps per second: 112, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
1455647/2500000: episode: 558, duration: 30.165s, episode steps: 3365, steps per second: 112, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
                                                                                                                13.000, mean reward: 0.004 [-1.000, 1.000], mean action
1458661/2500000: episode: 559, duration: 26.848s, episode steps:
                                                                  3014, steps per second: 112, episode reward:
1462014/2500000: episode: 560, duration: 29.913s, episode steps: 3353, steps per second: 112, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1464438/2500000: episode: 561, duration: 21.699s, episode steps: 2424, steps per second: 112, episode reward:
                                                                                                               16.000, mean reward: 0.007 [-1.000, 1.000], mean action
1467247/2500000: episode: 562, duration: 24.872s, episode steps: 2809, steps per second: 113, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1470707/2500000: episode: 563, duration: 31.385s, episode steps: 3460, steps per second: 110, episode reward:
                                                                                                               7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1474137/2500000: episode: 564, duration: 30.857s, episode steps: 3430, steps per second: 111, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1476469/2500000: episode: 565, duration: 20.815s, episode steps: 2332, steps per second: 112, episode reward: 15.000, mean reward: 0.006 [-1.000, 1.000], mean action
1480538/2500000: episode: 566, duration: 36.411s, episode steps: 4069, steps per second: 112, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action: 1484464/2500000: episode: 567, duration: 35.095s, episode steps: 3926, steps per second: 112, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1488347/2500000: episode: 568, duration: 34.735s, episode steps: 3883, steps per second: 112, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1490790/2500000: episode: 569, duration: 21.929s, episode steps: 2443, steps per second: 111, episode reward: 15.000, mean reward: 0.006 [-1.000, 1.000], mean action
1494574/2500000: episode: 570, duration: 33.756s, episode steps: 3784, steps per second: 112, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1497918/2500000: episode: 571, duration: 29.776s, episode steps: 3344, steps per second: 112, episode reward: 11.000, mean reward: 0.003 [-1.000, 1.000], mean action
1501081/2500000: episode: 572, duration: 28.212s, episode steps: 3163, steps per second: 112, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
504253/2500000: episode: 573, duration: 28.652s, episode steps: 3172, steps per second: 111, episode reward:
                                                                                                                9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1506573/2500000: episode: 574, duration: 20.790s, episode steps: 2320, steps per second: 112, episode reward: 17.000, mean reward: 0.007 [-1.000, 1.000], mean action
1510019/2500000: episode: 575, duration: 30.785s, episode steps: 3446, steps per second: 112, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1513133/2500000: episode: 576, duration: 27.656s, episode steps: 3114, steps per second: 113, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1515967/2500000: episode: 577, duration: 25.573s, episode steps: 2834, steps per second: 111, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1519191/2500000: episode: 578, duration: 28.721s, episode steps: 3224, steps per second: 112, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1523362/2500000: episode: 579, duration: 37.221s, episode steps: 4171, steps per second: 112, episode reward: 2.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1526410/2500000: episode: 580, duration: 27.122s, episode steps: 3048, steps per second: 112, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean action
1529307/2500000: episode: 581, duration: 25.831s, episode steps: 2897, steps per second: 112, episode reward: -5.000, mean reward: -0.002 [-1.000, 1.000], mean actic
1532435/2500000: episode: 582, duration: 27.767s, episode steps: 3128, steps per second: 113, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1535177/2500000: episode: 583, duration: 24.697s, episode steps: 2742, steps per second: 111, episode reward: 12.000, mean reward: 0.004 [-1.000,
1537980/2500000: episode: 584, duration: 25.337s, episode steps: 2803, steps per second: 111, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1540963/2500000: episode: 585, duration: 26.683s, episode steps: 2983, steps per second: 112, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1544084/2500000: episode: 586, duration: 27.821s, episode steps: 3121, steps per second: 112, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1546810/2500000: episode: 587, duration: 24.397s, episode steps: 2726, steps per second: 112, episode reward: 15.000, mean reward: 0.006 [-1.000, 1.000], mean action
1550427/2500000: episode: 588, duration: 32.488s, episode steps: 3617, steps per second: 111, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1553719/2500000: episode: 589, duration: 29.460s, episode steps: 3292, steps per second: 112, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1557175/2500000: episode: 590, duration: 30.863s, episode steps: 3456, steps per second: 112, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1560302/2500000: episode: 591, duration: 28.121s, episode steps: 3127, steps per second: 111, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000],
1564223/2500000: episode: 592, duration: 35.067s, episode steps: 3921, steps per second: 112, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
.566890/2500000: episode: 593, duration: 23.890s, episode steps: 2667, steps per second: 112, episode reward: -9.000, mean reward: -0.003
                                                                                                                                            [-1.000,
1569797/2500000: episode: 594, duration: 26.090s, episode steps: 2907, steps per second: 111, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1572313/2500000: episode: 595, duration: 22.875s, episode steps: 2516, steps per second: 110, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
1575369/2500000: episode: 596, duration: 27.381s, episode steps: 3056, steps per second: 112, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean action
1578366/2500000: episode: 597, duration: 26.723s, episode steps: 2997, steps per second: 112, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
1581000/2500000: episode: 598, duration: 23.657s, episode steps: 2634, steps per second: 111, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean action
1584865/2500000: episode: 599, duration: 34.714s, episode steps: 3865, steps per second: 111, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1587594/2500000: episode: 600. duration: 24.352s. episode steps: 2729. steps per second: 112. episode reward: 13.000. mean reward: 0.005 [-1.000. 1.000]. mean action
590714/2500000: episode: 601, duration: 27.807s, episode steps: 3120, steps per second: 112, episode reward: 12.000, mean reward: 0.004 [-1.000, 1
1593611/2500000: episode: 602, duration: 25.846s, episode steps: 2897, steps per second: 112, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean action
1596285/2500000: episode: 603, duration: 23.855s, episode steps: 2674, steps per second: 112, episode reward: 14.000, mean reward: 0.005 [-1.000, 1.000], mean action
1599273/2500000: episode: 604, duration: 26.654s, episode steps: 2988, steps per second: 112, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1601723/2500000: episode: 605, duration: 21.905s, episode steps: 2450, steps per second: 112, episode reward: 15.000, mean reward: 0.006
1604638/2500000: episode: 606, duration: 26.022s, episode steps: 2915, steps per second: 112, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean action
1607454/2500000: episode: 607, duration: 25.486s, episode steps: 2816, steps per second: 110, episode reward:
                                                                                                               6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1609987/2500000: episode: 608, duration: 22.634s, episode steps: 2533, steps per second: 112, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
1612969/2500000: episode: 609, duration: 26.662s, episode steps: 2982, steps per second: 112, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000],
1614174/2500000: episode: 610, duration: 10.774s, episode steps: 1205, steps per second: 112, episode reward: -18.000, mean reward: -0.015 [-1.000, 1.000], mean acti
1617167/2500000: episode: 611, duration: 27.001s, episode steps: 2993, steps per second: 111, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1619062/2500000; episode: 612, duration: 17.001s, episode steps: 1895, steps per second: 111, episode reward: -15.000, mean reward: -0.008 [-1.000, 1.000], mean acti
1622825/2500000: episode: 613, duration: 33.614s, episode steps: 3763, steps per second: 112, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1625544/2500000: episode: 614, duration: 24.231s, episode steps: 2719, steps per second: 112, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
                          615, duration: 24.978s, episode steps: 2788,
                                                                       steps per second: 112, episode reward: 12.000, mean reward: 0.004 [-1.000,
1630799/2500000: episode: 616, duration: 21.907s, episode steps: 2467, steps per second: 113, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
1634388/2500000: episode: 617, duration: 31.999s, episode steps: 3589, steps per second: 112, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1638172/2500000: episode: 618, duration: 33.943s, episode steps: 3784, steps per second: 111, episode reward: -5.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1641262/2500000: episode: 619, duration: 28.041s, episode steps: 3090, steps per second: 110, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
```

6 of 6