```
Training for 2500000 steps ...
     822/2500000: episode: 1, duration: 4.845s, episode steps: 822, steps per second: 170, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
    1646/2500000: episode: 2, duration: 2.732s, episode steps: 824, steps per second: 302, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
    2526/2500000: episode: 3, duration: 2.869s, episode steps: 880, steps per second: 307, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
    3447/2500000: episode: 4, duration: 2.898s, episode steps: 921, steps per second: 318, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
    4271/2500000: episode: 5, duration: 2.676s, episode steps: 824, steps per second: 308, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
    5035/2500000: episode: 6, duration: 2.525s, episode steps: 764, steps per second: 303, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
    5859/2500000: episode: 7, duration: 2.653s, episode steps: 824, steps per second: 311, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
    6870/2500000: episode: 8, duration: 3.250s, episode steps: 1011, steps per second: 311, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
    7670/2500000: episode: 9, duration: 2.617s, episode steps: 800, steps per second: 306, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
    8462/2500000: episode: 10, duration: 2.639s, episode steps: 792, steps per second: 300, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
    9347/2500000: episode: 11, duration: 2.925s, episode steps: 885, steps per second: 303, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
   10139/2500000: episode: 12, duration: 7.292s, episode steps: 792, steps per second: 109, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   11009/2500000: episode: 13, duration: 6.458s, episode steps: 870, steps per second: 135, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
   11863/2500000: episode: 14, duration: 6.415s, episode steps: 854, steps per second: 133, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   12704/2500000: episode: 15, duration: 6.109s, episode steps: 841, steps per second: 138, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   13544/2500000: episode: 16, duration: 6.358s, episode steps: 840, steps per second: 132, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   14356/2500000: episode: 17, duration: 6.107s, episode steps: 812, steps per second: 133, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
   15283/2500000: episode: 18, duration: 6.713s, episode steps: 927, steps per second: 138, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
   16136/2500000: episode: 19, duration: 6.394s, episode steps: 853, steps per second: 133, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   17039/2500000: episode: 20, duration: 6.798s, episode steps: 903, steps per second: 133, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
   17831/2500000: episode: 21, duration: 5.719s, episode steps: 792, steps per second: 138, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   18714/2500000: episode: 22, duration: 6.472s, episode steps: 883, steps per second: 136, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
   19585/2500000: episode: 23, duration: 6.738s, episode steps: 871, steps per second: 129, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
   20524/2500000: episode: 24, duration: 7.249s, episode steps: 939, steps per second: 130, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action:
   21348/2500000: episode: 25, duration: 6.186s, episode steps: 824, steps per second: 133, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   22419/2500000: episode: 26, duration: 7.735s, episode steps: 1071, steps per second: 138, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action
   23309/2500000: episode: 27, duration: 6.376s, episode steps: 890, steps per second: 140, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
   24092/2500000: episode: 28, duration: 5.803s, episode steps: 783, steps per second: 135, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   24937/2500000: episode: 29, duration: 6.284s, episode steps: 845, steps per second: 134, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   25870/2500000: episode: 30, duration: 6.882s, episode steps: 933, steps per second: 136, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   26714/2500000: episode: 31, duration: 6.314s, episode steps: 844, steps per second: 134, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
   27645/2500000: episode: 32, duration: 6.773s, episode steps: 931, steps per second: 137, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   28456/2500000: episode: 33, duration: 5.789s, episode steps: 811, steps per second: 140, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
   29299/2500000: episode: 34, duration: 6.207s, episode steps: 843, steps per second: 136, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
   30361/2500000: episode: 35, duration: 7.538s, episode steps: 1062, steps per second: 141, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action
   31144/2500000: episode: 36, duration: 5.754s, episode steps: 783, steps per second: 136, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   32071/2500000: episode: 37, duration: 6.796s, episode steps: 927, steps per second: 136, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
   33091/2500000: episode: 38, duration: 7.269s, episode steps: 1020, steps per second: 140, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action
   33975/2500000: episode: 39, duration: 6.496s, episode steps: 884, steps per second: 136, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
   34818/2500000: episode: 40, duration: 6.278s, episode steps: 843, steps per second: 134, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
   35838/2500000: episode: 41, duration: 7.297s, episode steps: 1020, steps per second: 140, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
   36858/2500000: episode: 42, duration: 7.424s, episode steps: 1020, steps per second: 137, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action
   37701/2500000: episode: 43, duration: 6.307s, episode steps: 843, steps per second: 134, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   38495/2500000: episode: 44, duration: 5.638s, episode steps: 794, steps per second: 141, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
   39305/2500000: episode: 45, duration: 5.807s, episode steps: 810, steps per second: 139, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
   40264/2500000: episode: 46, duration: 6.994s, episode steps: 959, steps per second: 137, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   41248/2500000: episode: 47, duration: 7.283s, episode steps: 984, steps per second: 135, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
   42031/2500000: episode: 48, duration: 5.749s, episode steps: 783, steps per second: 136, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   43091/2500000: episode: 49, duration: 7.759s, episode steps: 1060, steps per second: 137, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean action
   44042/2500000: episode: 50, duration: 6.949s, episode steps: 951, steps per second: 137, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   45179/2500000: episode: 51, duration: 8.403s, episode steps: 1137, steps per second: 135, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean action
   46232/2500000: episode: 52, duration: 7.857s, episode steps: 1053, steps per second: 134, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action
   47191/2500000: episode: 53, duration: 7.054s, episode steps: 959, steps per second: 136, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   48094/2500000: episode: 54, duration: 6.538s, episode steps: 903, steps per second: 138, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
   48914/2500000: episode: 55, duration: 5.973s, episode steps: 820, steps per second: 137, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
   49768/2500000: episode: 56, duration: 6.249s, episode steps: 854, steps per second: 137, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   50736/2500000: episode: 57, duration: 7.202s, episode steps: 968, steps per second: 134, episode reward: -21.000, mean reward: -0.022 [-1.000, 1.000], mean action:
   51500/2500000: episode: 58, duration: 5.511s, episode steps: 764, steps per second: 139, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   52322/2500000: episode: 59, duration: 6.073s, episode steps: 822, steps per second: 135, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
   53164/2500000: episode: 60, duration: 6.172s, episode steps: 842, steps per second: 136, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
   54235/2500000: episode: 61, duration: 7.952s, episode steps: 1071, steps per second: 135, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean action
   54999/2500000: episode: 62, duration: 5.583s, episode steps: 764, steps per second: 137, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   55782/2500000: episode: 63, duration: 5.750s, episode steps: 783, steps per second: 136, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   56663/2500000: episode: 64, duration: 6.443s, episode steps: 881, steps per second: 137, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
   57877/2500000: episode: 65, duration: 8.860s, episode steps: 1214, steps per second: 137, episode reward: -18.000, mean reward: -0.015 [-1.000, 1.000], mean action
   58703/2500000: episode: 66, duration: 6.131s, episode steps: 826, steps per second: 135, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   59467/2500000: episode: 67, duration: 5.515s, episode steps: 764, steps per second: 139, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   60292/2500000: episode: 68, duration: 6.291s, episode steps: 825, steps per second: 131, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   61084/2500000: episode: 69, duration: 5.868s, episode steps: 792, steps per second: 135, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   61876/2500000: episode: 70, duration: 6.127s, episode steps: 792, steps per second: 129, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   62843/2500000: episode: 71, duration: 7.439s, episode steps: 967, steps per second: 130, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   63667/2500000: episode: 72, duration: 5.949s, episode steps: 824, steps per second: 139, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   64625/2500000: episode: 73, duration: 7.097s, episode steps: 958, steps per second: 135, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   65594/2500000: episode: 74, duration: 7.216s, episode steps: 969, steps per second: 134, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   66576/2500000: episode: 75, duration: 7.301s, episode steps: 982, steps per second: 135, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action:
   67581/2500000: episode: 76, duration: 7.614s, episode steps: 1005, steps per second: 132, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action
   68808/2500000: episode: 77, duration: 8.840s, episode steps: 1227, steps per second: 139, episode reward: -17.000, mean reward: -0.014 [-1.000, 1.000], mean action
   69762/2500000: episode: 78, duration: 7.110s, episode steps: 954, steps per second: 134, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   70633/2500000: episode: 79, duration: 6.438s, episode steps: 871, steps per second: 135, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
   71565/2500000: episode: 80, duration: 6.853s, episode steps: 932, steps per second: 136, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   72466/2500000: episode: 81, duration: 6.557s, episode steps: 901, steps per second: 137, episode reward: -21.000, mean reward: -0.023 [-1.000, 0.000], mean action:
   73308/2500000: episode: 82, duration: 6.258s, episode steps: 842, steps per second: 135, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean action:
   74072/2500000: episode: 83, duration: 5.595s, episode steps: 764, steps per second: 137, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   75038/2500000: episode: 84, duration: 7.225s, episode steps: 966, steps per second: 135, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   76130/2500000: episode: 85, duration: 8.013s, episode steps: 1092, steps per second: 136, episode reward: -18.000, mean reward: -0.016 [-1.000, 1.000], mean action
   77001/2500000: episode: 86, duration: 6.544s, episode steps: 871, steps per second: 133, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean action:
   78087/2500000: episode: 87, duration: 7.995s, episode steps: 1086, steps per second: 136, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean action
   78974/2500000: episode: 88, duration: 6.504s, episode steps: 887, steps per second: 136, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
   79758/2500000: episode: 89, duration: 5.604s, episode steps: 784, steps per second: 140, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   80603/2500000: episode: 90, duration: 6.276s, episode steps: 845, steps per second: 135, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   81533/2500000: episode: 91, duration: 6.869s, episode steps: 930, steps per second: 135, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
   82403/2500000: episode: 92, duration: 6.334s, episode steps: 870, steps per second: 137, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean action:
   83195/2500000: episode: 93, duration: 6.044s, episode steps: 792, steps per second: 131, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   83978/2500000: episode: 94, duration: 5.735s, episode steps: 783, steps per second: 137, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean action:
   84937/2500000: episode: 95, duration: 7.031s, episode steps: 959, steps per second: 136, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action:
   85834/2500000: episode: 96, duration: 6.706s, episode steps: 897, steps per second: 134, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action:
   86654/2500000: episode: 97, duration: 6.029s, episode steps: 820, steps per second: 136, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action:
   87629/2500000: episode: 98, duration: 7.124s, episode steps: 975, steps per second: 137, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean action:
   88412/2500000: episode: 99, duration: 5.794s, episode steps: 783, steps per second: 135, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action:
   89312/2500000: episode: 100, duration: 6.651s, episode steps: 900, steps per second: 135, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action
   90138/2500000: episode: 101, duration: 6.245s, episode steps: 826, steps per second: 132, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action
   91038/2500000: episode: 102, duration: 6.880s, episode steps: 900, steps per second: 131, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action
   92008/2500000: episode: 103, duration: 7.008s, episode steps: 970, steps per second: 138, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean action
   92819/2500000: episode: 104, duration: 6.266s, episode steps: 811, steps per second: 129, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean action
   93814/2500000: episode: 105, duration: 7.382s, episode steps: 995, steps per second: 135, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean action
   94739/2500000: episode: 106, duration: 6.854s, episode steps: 925, steps per second: 135, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean action
   95670/2500000: episode: 107, duration: 6.883s, episode steps: 931, steps per second: 135, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action
   96494/2500000: episode: 108, duration: 6.037s, episode steps: 824, steps per second: 136, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean action
```

```
 97492/2500000: episode: 109, duration: 7.281s, episode steps: 998, steps per second: 137, episode reward: -19.000, mean reward: -0.019 [-1.000, 1.000], mean actio
 98304/2500000: episode: 110, duration: 6.035s, episode steps: 812, steps per second: 135, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean actio
 99124/2500000: episode: 111, duration: 6.033s, episode steps: 820, steps per second: 136, episode reward: -21.000, mean reward: -0.026 [-1.000, 0.000], mean actio
100069/2500000: episode: 112, duration: 6.892s, episode steps: 945, steps per second: 137, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean actio
101036/2500000: episode: 113, duration: 7.223s, episode steps: 967, steps per second: 134, episode reward: -19.000, mean reward: -0.020 [-1.000, 1.000], mean actio
101941/2500000: episode: 114, duration: 6.776s, episode steps: 905, steps per second: 134, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean actio
102825/2500000: episode: 115, duration: 6.813s, episode steps: 884, steps per second: 130, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean actio
103698/2500000: episode: 116, duration: 6.721s, episode steps: 873, steps per second: 130, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean actio
104845/2500000: episode: 117, duration: 8.501s, episode steps: 1147, steps per second: 135, episode reward: -20.000, mean reward: -0.017 [-1.000, 1.000], mean actio
105989/2500000: episode: 118, duration: 8.555s, episode steps: 1144, steps per second: 134, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
106922/2500000: episode: 119, duration: 6.801s, episode steps: 933, steps per second: 137, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
107940/2500000: episode: 120, duration: 7.423s, episode steps: 1018, steps per second: 137, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean actio
109016/2500000: episode: 121, duration: 7.989s, episode steps: 1076, steps per second: 135, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean actio
110235/2500000: episode: 122, duration: 9.007s, episode steps: 1219, steps per second: 135, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean actio
111460/2500000: episode: 123, duration: 9.210s, episode steps: 1225, steps per second: 133, episode reward: -19.000, mean reward: -0.016 [-1.000, 1.000], mean actio
112408/2500000: episode: 124, duration: 6.965s, episode steps: 948, steps per second: 136, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
113492/2500000: episode: 125, duration: 8.006s, episode steps: 1084, steps per second: 135, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean actio
114501/2500000: episode: 126, duration: 7.635s, episode steps: 1009, steps per second: 132, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean actio
115363/2500000: episode: 127, duration: 6.327s, episode steps: 862, steps per second: 136, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean actio
116319/2500000: episode: 128, duration: 7.055s, episode steps: 956, steps per second: 136, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
117307/2500000: episode: 129, duration: 7.254s, episode steps: 988, steps per second: 136, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean actio
118281/2500000: episode: 130, duration: 7.427s, episode steps: 974, steps per second: 131, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
119225/2500000: episode: 131, duration: 7.002s, episode steps: 944, steps per second: 135, episode reward: -21.000, mean reward: -0.022 [-1.000, 0.000], mean actio
120188/2500000: episode: 132, duration: 6.997s, episode steps: 963, steps per second: 138, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
121377/2500000: episode: 133, duration: 8.838s, episode steps: 1189, steps per second: 135, episode reward: -21.000, mean reward: -0.018 [-1.000, 0.000], mean actio
122717/2500000: episode: 134, duration: 9.995s, episode steps: 1340, steps per second: 134, episode reward: -19.000, mean reward: -0.014 [-1.000, 1.000], mean actio
123826/2500000: episode: 135, duration: 8.297s, episode steps: 1109, steps per second: 134, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean actio
125013/2500000: episode: 136, duration: 8.767s, episode steps: 1187, steps per second: 135, episode reward: -20.000, mean reward: -0.017 [-1.000, 1.000], mean actio
125837/2500000: episode: 137, duration: 6.098s, episode steps: 824, steps per second: 135, episode reward: -21.000, mean reward: -0.025 [-1.000, 0.000], mean actio
126882/2500000: episode: 138, duration: 7.891s, episode steps: 1045, steps per second: 132, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean actio
128391/2500000: episode: 139, duration: 10.945s, episode steps: 1509, steps per second: 138, episode reward: -14.000, mean reward: -0.009 [-1.000, 1.000], mean acti
129325/2500000: episode: 140, duration: 6.982s, episode steps: 934, steps per second: 134, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
130265/2500000: episode: 141, duration: 6.967s, episode steps: 940, steps per second: 135, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
131029/2500000: episode: 142, duration: 5.758s, episode steps: 764, steps per second: 133, episode reward: -21.000, mean reward: -0.027 [-1.000, 0.000], mean actio
131928/2500000: episode: 143, duration: 6.653s, episode steps: 899, steps per second: 135, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean actio
132799/2500000: episode: 144, duration: 6.370s, episode steps: 871, steps per second: 137, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean actio
134020/2500000: episode: 145, duration: 9.243s, episode steps: 1221, steps per second: 132, episode reward: -19.000, mean reward: -0.016 [-1.000, 1.000], mean actio
134961/2500000: episode: 146, duration: 6.891s, episode steps: 941, steps per second: 137, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
135886/2500000: episode: 147, duration: 6.775s, episode steps: 925, steps per second: 137, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean actio
136877/2500000: episode: 148, duration: 7.254s, episode steps: 991, steps per second: 137, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean actio
137764/2500000: episode: 149, duration: 6.499s, episode steps: 887, steps per second: 136, episode reward: -20.000, mean reward: -0.023 [-1.000, 1.000], mean actio
138928/2500000: episode: 150, duration: 8.708s, episode steps: 1164, steps per second: 134, episode reward: -19.000, mean reward: -0.016 [-1.000, 1.000], mean actio
139865/2500000: episode: 151, duration: 6.938s, episode steps: 937, steps per second: 135, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
140741/2500000: episode: 152, duration: 6.439s, episode steps: 876, steps per second: 136, episode reward: -21.000, mean reward: -0.024 [-1.000, 0.000], mean actio
141993/2500000: episode: 153, duration: 9.277s, episode steps: 1252, steps per second: 135, episode reward: -19.000, mean reward: -0.015 [-1.000, 1.000], mean actio
142953/2500000: episode: 154, duration: 7.092s, episode steps: 960, steps per second: 135, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
144177/2500000: episode: 155, duration: 9.184s, episode steps: 1224, steps per second: 133, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean actio
145569/2500000: episode: 156, duration: 10.510s, episode steps: 1392, steps per second: 132, episode reward: -18.000, mean reward: -0.013 [-1.000, 1.000], mean acti
146677/2500000: episode: 157, duration: 8.308s, episode steps: 1108, steps per second: 133, episode reward: -21.000, mean reward: -0.019 [-1.000, 0.000], mean actio
148106/2500000: episode: 158, duration: 10.589s, episode steps: 1429, steps per second: 135, episode reward: -15.000, mean reward: -0.010 [-1.000, 1.000], mean acti
149282/2500000: episode: 159, duration: 8.666s, episode steps: 1176, steps per second: 136, episode reward: -21.000, mean reward: -0.018 [-1.000, 0.000], mean actio
150925/2500000: episode: 160, duration: 12.006s, episode steps: 1643, steps per second: 137, episode reward: -15.000, mean reward: -0.009 [-1.000, 1.000], mean acti
151916/2500000: episode: 161, duration: 7.437s, episode steps: 991, steps per second: 133, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean actio
153505/2500000: episode: 162, duration: 11.858s, episode steps: 1589, steps per second: 134, episode reward: -18.000, mean reward: -0.011 [-1.000, 1.000], mean acti
154432/2500000: episode: 163, duration: 6.917s, episode steps: 927, steps per second: 134, episode reward: -20.000, mean reward: -0.022 [-1.000, 1.000], mean actio
155585/2500000: episode: 164, duration: 8.582s, episode steps: 1153, steps per second: 134, episode reward: -18.000, mean reward: -0.016 [-1.000, 1.000], mean actio
156427/2500000: episode: 165, duration: 6.131s, episode steps: 842, steps per second: 137, episode reward: -20.000, mean reward: -0.024 [-1.000, 1.000], mean actio
157470/2500000: episode: 166, duration: 7.853s, episode steps: 1043, steps per second: 133, episode reward: -19.000, mean reward: -0.018 [-1.000, 1.000], mean actio
158949/2500000: episode: 167, duration: 10.922s, episode steps: 1479, steps per second: 135, episode reward: -17.000, mean reward: -0.011 [-1.000, 1.000], mean acti
159975/2500000: episode: 168, duration: 7.507s, episode steps: 1026, steps per second: 137, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean actio
161167/2500000: episode: 169, duration: 8.693s, episode steps: 1192, steps per second: 137, episode reward: -18.000, mean reward: -0.015 [-1.000, 1.000], mean actio
162589/2500000: episode: 170, duration: 10.556s, episode steps: 1422, steps per second: 135, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean acti
164441/2500000: episode: 171, duration: 13.564s, episode steps: 1852, steps per second: 137, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean acti
165533/2500000: episode: 172, duration: 7.923s, episode steps: 1092, steps per second: 138, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean actio
166871/2500000: episode: 173, duration: 9.685s, episode steps: 1338, steps per second: 138, episode reward: -18.000, mean reward: -0.013 [-1.000, 1.000], mean actio
167993/2500000: episode: 174, duration: 8.170s, episode steps: 1122, steps per second: 137, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean actio
169483/2500000: episode: 175, duration: 10.922s, episode steps: 1490, steps per second: 136, episode reward: -15.000, mean reward: -0.010 [-1.000, 1.000], mean acti
171065/2500000: episode: 176, duration: 11.748s, episode steps: 1582, steps per second: 135, episode reward: -15.000, mean reward: -0.009 [-1.000, 1.000], mean acti
172298/2500000: episode: 177, duration: 9.036s, episode steps: 1233, steps per second: 136, episode reward: -21.000, mean reward: -0.017 [-1.000, 0.000], mean actio
173752/2500000: episode: 178, duration: 10.761s, episode steps: 1454, steps per second: 135, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean acti
174910/2500000: episode: 179, duration: 8.567s, episode steps: 1158, steps per second: 135, episode reward: -19.000, mean reward: -0.016 [-1.000, 1.000], mean actio
176084/2500000: episode: 180, duration: 8.700s, episode steps: 1174, steps per second: 135, episode reward: -20.000, mean reward: -0.017 [-1.000, 1.000], mean actio
177094/2500000: episode: 181, duration: 7.421s, episode steps: 1010, steps per second: 136, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean actio
178223/2500000: episode: 182, duration: 8.277s, episode steps: 1129, steps per second: 136, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean actio
179502/2500000: episode: 183, duration: 9.367s, episode steps: 1279, steps per second: 137, episode reward: -17.000, mean reward: -0.013 [-1.000, 1.000], mean actio
180973/2500000: episode: 184, duration: 10.813s, episode steps: 1471, steps per second: 136, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean acti
182231/2500000: episode: 185, duration: 9.333s, episode steps: 1258, steps per second: 135, episode reward: -17.000, mean reward: -0.014 [-1.000, 1.000], mean actio
183199/2500000: episode: 186, duration: 7.157s, episode steps: 968, steps per second: 135, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean actio
184791/2500000: episode: 187, duration: 11.802s, episode steps: 1592, steps per second: 135, episode reward: -15.000, mean reward: -0.009 [-1.000, 1.000], mean acti
185993/2500000: episode: 188, duration: 8.969s, episode steps: 1202, steps per second: 134, episode reward: -18.000, mean reward: -0.015 [-1.000, 1.000], mean actio
187282/2500000: episode: 189, duration: 9.695s, episode steps: 1289, steps per second: 133, episode reward: -19.000, mean reward: -0.015 [-1.000, 1.000], mean actio
188634/2500000: episode: 190, duration: 10.020s, episode steps: 1352, steps per second: 135, episode reward: -18.000, mean reward: -0.013 [-1.000, 1.000], mean acti
190003/2500000: episode: 191, duration: 10.007s, episode steps: 1369, steps per second: 137, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean acti
191523/2500000: episode: 192, duration: 11.181s, episode steps: 1520, steps per second: 136, episode reward: -16.000, mean reward: -0.011 [-1.000, 1.000], mean acti
192625/2500000: episode: 193, duration: 8.160s, episode steps: 1102, steps per second: 135, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean actio
193625/2500000: episode: 194, duration: 7.297s, episode steps: 1000, steps per second: 137, episode reward: -21.000, mean reward: -0.021 [-1.000, 0.000], mean actio
194711/2500000: episode: 195, duration: 7.945s, episode steps: 1086, steps per second: 137, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean actio
195822/2500000: episode: 196, duration: 8.102s, episode steps: 1111, steps per second: 137, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean actio
197326/2500000: episode: 197, duration: 11.096s, episode steps: 1504, steps per second: 136, episode reward: -16.000, mean reward: -0.011 [-1.000, 1.000], mean acti
198742/2500000: episode: 198, duration: 10.467s, episode steps: 1416, steps per second: 135, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean acti
200079/2500000: episode: 199, duration: 9.853s, episode steps: 1337, steps per second: 136, episode reward: -18.000, mean reward: -0.013 [-1.000, 1.000], mean actio
201543/2500000: episode: 200, duration: 10.671s, episode steps: 1464, steps per second: 137, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean acti
202840/2500000: episode: 201, duration: 9.657s, episode steps: 1297, steps per second: 134, episode reward: -19.000, mean reward: -0.015 [-1.000, 1.000], mean actio
204110/2500000: episode: 202, duration: 9.359s, episode steps: 1270, steps per second: 136, episode reward: -17.000, mean reward: -0.013 [-1.000, 1.000], mean actio
205355/2500000: episode: 203, duration: 9.209s, episode steps: 1245, steps per second: 135, episode reward: -19.000, mean reward: -0.015 [-1.000, 1.000], mean actio
206581/2500000: episode: 204, duration: 9.053s, episode steps: 1226, steps per second: 135, episode reward: -19.000, mean reward: -0.015 [-1.000, 1.000], mean actio
207678/2500000: episode: 205, duration: 8.008s, episode steps: 1097, steps per second: 137, episode reward: -19.000, mean reward: -0.017 [-1.000, 1.000], mean actio
208900/2500000: episode: 206, duration: 9.015s, episode steps: 1222, steps per second: 136, episode reward: -19.000, mean reward: -0.016 [-1.000, 1.000], mean actio
210607/2500000: episode: 207, duration: 12.478s, episode steps: 1707, steps per second: 137, episode reward: -14.000, mean reward: -0.008 [-1.000, 1.000], mean acti
212043/2500000: episode: 208, duration: 10.555s, episode steps: 1436, steps per second: 136, episode reward: -18.000, mean reward: -0.013 [-1.000, 1.000], mean acti
213320/2500000: episode: 209, duration: 9.481s, episode steps: 1277, steps per second: 135, episode reward: -18.000, mean reward: -0.014 [-1.000, 1.000], mean acti
215131/2500000: episode: 210, duration: 13.125s, episode steps: 1811, steps per second: 138, episode reward: -17.000, mean reward: -0.009 [-1.000, 1.000], mean acti
216894/2500000: episode: 211, duration: 13.006s, episode steps: 1763, steps per second: 136, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean acti
218475/2500000: episode: 212, duration: 11.517s, episode steps: 1581, steps per second: 137, episode reward: -18.000, mean reward: -0.011 [-1.000, 1.000], mean acti
219466/2500000: episode: 213, duration: 7.350s, episode steps: 991, steps per second: 135, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean actio
221055/2500000: episode: 214, duration: 11.751s, episode steps: 1589, steps per second: 135, episode reward: -16.000, mean reward: -0.010 [-1.000, 1.000], mean acti
222582/2500000: episode: 215, duration: 11.172s, episode steps: 1527, steps per second: 137, episode reward: -16.000, mean reward: -0.010 [-1.000, 1.000], mean acti
223767/2500000: episode: 216, duration: 8.625s, episode steps: 1185, steps per second: 137, episode reward: -20.000, mean reward: -0.017 [-1.000, 1.000], mean actio
225393/2500000: episode: 217, duration: 11.613s, episode steps: 1626, steps per second: 140, episode reward: -17.000, mean reward: -0.010 [-1.000, 1.000], mean acti
```

```
226933/2500000: episode: 218, duration: 11.350s, episode steps: 1540, steps per second: 136, episode reward: -17.000, mean reward: -0.011 [-1.000, 1.000], mean acti
228383/2500000: episode: 219, duration: 10.774s, episode steps: 1450, steps per second: 135, episode reward: -15.000, mean reward: -0.010 [-1.000, 1.000], mean acti
229647/2500000: episode: 220, duration: 9.604s, episode steps: 1264, steps per second: 132, episode reward: -19.000, mean reward: -0.015 [-1.000, 1.000], mean actio
230609/2500000: episode: 221, duration: 7.189s, episode steps: 962, steps per second: 134, episode reward: -20.000, mean reward: -0.021 [-1.000, 1.000], mean action
232147/2500000: episode: 222, duration: 11.127s, episode steps: 1538, steps per second: 138, episode reward: -17.000, mean reward: -0.011 [-1.000, 1.000], mean acti
233816/2500000: episode: 223, duration: 12.453s, episode steps: 1669, steps per second: 134, episode reward: -10.000, mean reward: -0.010 [-1.000, 1.000], mean acti
234899/2500000: episode: 224, duration: 7.675s, episode steps: 1083, steps per second: 141, episode reward: -20.000, mean reward: -0.018 [-1.000, 1.000], mean actic
236426/2500000: episode: 225, duration: 11.179s, episode steps: 1527, steps per second: 137, episode reward: -16.000, mean reward: -0.010 [-1.000, 1.000], mean acti
237781/2500000: episode: 226, duration: 9.930s, episode steps: 1355, steps per second: 136, episode reward: -18.000, mean reward: -0.013 [-1.000, 1.000], mean acti
239383/2500000: episode: 227, duration: 11.518s, episode steps: 1602, steps per second: 139, episode reward: -13.000, mean reward: -0.008 [-1.000, 1.000], mean acti
240897/2500000: episode: 228, duration: 11.093s, episode steps: 1514, steps per second: 136, episode reward: -18.000, mean reward: -0.012 [-1.000, 1.000], mean acti
242657/2500000: episode: 229, duration: 12.484s, episode steps: 1760, steps per second: 141, episode reward: -16.000, mean reward: -0.009 [-1.000, 1.000], mean acti
244327/2500000: episode: 230, duration: 12.174s, episode steps: 1670, steps per second: 137, episode reward: -18.000, mean reward: -0.011 [-1.000, 1.000], mean acti
246147/2500000: episode: 231, duration: 12.944s, episode steps: 1820, steps per second: 141, episode reward: -15.000, mean reward: -0.008 [-1.000, 1.000], mean acti
247665/2500000: episode: 232, duration: 10.873s, episode steps: 1518, steps per second: 140, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean acti
248898/2500000: episode: 233, duration: 8.963s, episode steps: 1233, steps per second: 138, episode reward: -18.000, mean reward: -0.015 [-1.000, 1.000], mean actic
250063/2500000: episode: 234, duration: 8.116s, episode steps: 1165, steps per second: 144, episode reward: -19.000, mean reward: -0.016 [-1.000, 1.000], mean actic
251436/2500000: episode: 235, duration: 10.047s, episode steps: 1373, steps per second: 137, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean acti
253099/2500000: episode: 236, duration: 12.118s, episode steps: 1663, steps per second: 137, episode reward: -16.000, mean reward: -0.010 [-1.000, 1.000], mean acti
254349/2500000: episode: 237, duration: 9.039s, episode steps: 1250, steps per second: 138, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean actic
255641/2500000: episode: 238, duration: 9.108s, episode steps: 1292, steps per second: 142, episode reward: -18.000, mean reward: -0.014 [-1.000, 1.000], mean actic
257395/2500000: episode: 239, duration: 12.561s, episode steps: 1754, steps per second: 140, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean acti
259012/2500000: episode: 240, duration: 11.315s, episode steps: 1617, steps per second: 143, episode reward: -19.000, mean reward: -0.012 [-1.000, 1.000], mean acti
260004/2500000: episode: 241, duration: 7.170s, episode steps: 992, steps per second: 138, episode reward: -20.000, mean reward: -0.020 [-1.000, 1.000], mean action
261655/2500000: episode: 242, duration: 11.955s, episode steps: 1651, steps per second: 138, episode reward: -17.000, mean reward: -0.010 [-1.000, 1.000], mean acti
263475/2500000: episode: 243, duration: 12.978s, episode steps: 1820, steps per second: 140, episode reward: -16.000, mean reward: -0.009 [-1.000, 1.000], mean acti
265269/2500000: episode: 244, duration: 13.094s, episode steps: 1794, steps per second: 137, episode reward: -19.000, mean reward: -0.011 [-1.000, 1.000], mean acti
266295/2500000: episode: 245, duration: 7.144s, episode steps: 1026, steps per second: 144, episode reward: -20.000, mean reward: -0.019 [-1.000, 1.000], mean actic
267635/2500000: episode: 246, duration: 9.781s, episode steps: 1340, steps per second: 137, episode reward: -20.000, mean reward: -0.015 [-1.000, 1.000], mean actic
269177/2500000: episode: 247, duration: 11.089s, episode steps: 1542, steps per second: 139, episode reward: -20.000, mean reward: -0.013 [-1.000, 1.000], mean acti
270665/2500000: episode: 248, duration: 10.782s, episode steps: 1488, steps per second: 138, episode reward: -16.000, mean reward: -0.011 [-1.000, 1.000], mean acti
271923/2500000: episode: 249, duration: 9.035s, episode steps: 1258, steps per second: 139, episode reward: -19.000, mean reward: -0.015 [-1.000, 1.000], mean actic
273386/2500000: episode: 250, duration: 10.874s, episode steps: 1463, steps per second: 135, episode reward: -18.000, mean reward: -0.012 [-1.000, 1.000], mean acti
274754/2500000: episode: 251, duration: 9.613s, episode steps: 1368, steps per second: 142, episode reward: -19.000, mean reward: -0.014 [-1.000, 1.000], mean actic
276060/2500000: episode: 252, duration: 9.286s, episode steps: 1306, steps per second: 141, episode reward: -17.000, mean reward: -0.013 [-1.000, 1.000], mean actic
277735/2500000: episode: 253, duration: 11.967s, episode steps: 1675, steps per second: 140, episode reward: -13.000, mean reward: -0.008 [-1.000, 1.000], mean acti
279435/2500000: episode: 254, duration: 12.179s, episode steps: 1700, steps per second: 140, episode reward: -15.000, mean reward: -0.009 [-1.000, 1.000], mean acti
281005/2500000: episode: 255, duration: 11.309s, episode steps: 1570, steps per second: 139, episode reward: -17.000, mean reward: -0.011 [-1.000, 1.000], mean acti
282864/2500000: episode: 256, duration: 13.372s, episode steps: 1859, steps per second: 139, episode reward: -15.000, mean reward: -0.008 [-1.000, 1.000], mean acti
284183/2500000: episode: 257, duration: 9.589s, episode steps: 1319, steps per second: 138, episode reward: -16.000, mean reward: -0.012 [-1.000, 1.000], mean actic
286052/2500000: episode: 258, duration: 13.370s, episode steps: 1869, steps per second: 140, episode reward: -13.000, mean reward: -0.007 [-1.000, 1.000], mean acti
287419/2500000: episode: 259, duration: 9.955s, episode steps: 1367, steps per second: 137, episode reward: -19.000, mean reward: -0.014 [-1.000, 1.000], mean acti
288692/2500000: episode: 260, duration: 9.264s, episode steps: 1273, steps per second: 137, episode reward: -17.000, mean reward: -0.013 [-1.000, 1.000], mean actic
290525/2500000: episode: 261, duration: 13.253s, episode steps: 1833, steps per second: 138, episode reward: -17.000, mean reward: -0.009 [-1.000, 1.000], mean acti
292259/2500000: episode: 262, duration: 12.451s, episode steps: 1734, steps per second: 139, episode reward: -16.000, mean reward: -0.009 [-1.000, 1.000], mean acti
294140/2500000: episode: 263, duration: 13.798s, episode steps: 1881, steps per second: 136, episode reward: -13.000, mean reward: -0.007 [-1.000, 1.000], mean acti
295594/2500000: episode: 264, duration: 10.427s, episode steps: 1454, steps per second: 139, episode reward: -19.000, mean reward: -0.013 [-1.000, 1.000], mean acti
296930/2500000: episode: 265, duration: 9.624s, episode steps: 1336, steps per second: 139, episode reward: -19.000, mean reward: -0.014 [-1.000, 1.000], mean actic
298377/2500000: episode: 266, duration: 10.507s, episode steps: 1447, steps per second: 138, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean acti
300149/2500000: episode: 267, duration: 12.811s, episode steps: 1772, steps per second: 138, episode reward: -15.000, mean reward: -0.008 [-1.000, 1.000], mean acti
301566/2500000: episode: 268, duration: 10.340s, episode steps: 1417, steps per second: 137, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean acti
302886/2500000: episode: 269, duration: 9.628s, episode steps: 1320, steps per second: 137, episode reward: -18.000, mean reward: -0.014 [-1.000, 1.000], mean actic
304278/2500000: episode: 270, duration: 10.039s, episode steps: 1392, steps per second: 139, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean acti
305511/2500000: episode: 271, duration: 8.999s, episode steps: 1233, steps per second: 137, episode reward: -20.000, mean reward: -0.016 [-1.000, 1.000], mean acti
306882/2500000: episode: 272, duration: 10.023s, episode steps: 1371, steps per second: 137, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean acti
309112/2500000: episode: 273, duration: 15.975s, episode steps: 2230, steps per second: 140, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
311033/2500000: episode: 274, duration: 14.190s, episode steps: 1921, steps per second: 135, episode reward: -12.000, mean reward: -0.006 [-1.000, 1.000], mean acti
312691/2500000: episode: 275, duration: 11.996s, episode steps: 1658, steps per second: 138, episode reward: -16.000, mean reward: -0.010 [-1.000, 1.000], mean acti
314682/2500000: episode: 276, duration: 14.490s, episode steps: 1991, steps per second: 137, episode reward: -13.000, mean reward: -0.007 [-1.000, 1.000], mean acti
316190/2500000: episode: 277, duration: 11.076s, episode steps: 1508, steps per second: 136, episode reward: -15.000, mean reward: -0.010 [-1.000, 1.000], mean acti
317714/2500000: episode: 278, duration: 11.098s, episode steps: 1524, steps per second: 137, episode reward: -16.000, mean reward: -0.010 [-1.000, 1.000], mean acti
319170/2500000: episode: 279, duration: 10.578s, episode steps: 1456, steps per second: 138, episode reward: -16.000, mean reward: -0.011 [-1.000, 1.000], mean acti
320919/2500000: episode: 280, duration: 12.655s, episode steps: 1749, steps per second: 138, episode reward: -14.000, mean reward: -0.008 [-1.000, 1.000], mean acti
322760/2500000: episode: 281, duration: 13.444s, episode steps: 1841, steps per second: 137, episode reward: -14.000, mean reward: -0.008 [-1.000, 1.000], mean acti
324302/2500000: episode: 282, duration: 11.147s, episode steps: 1542, steps per second: 138, episode reward: -16.000, mean reward: -0.010 [-1.000, 1.000], mean acti
326774/2500000: episode: 283, duration: 18.086s, episode steps: 2472, steps per second: 137, episode reward: -7.000, mean reward: -0.003 [-1.000, 1.000], mean actic
328285/2500000: episode: 284, duration: 11.110s, episode steps: 1511, steps per second: 136, episode reward: -16.000, mean reward: -0.011 [-1.000, 1.000], mean acti
330458/2500000: episode: 285, duration: 15.697s, episode steps: 2173, steps per second: 138, episode reward: -12.000, mean reward: -0.006 [-1.000, 1.000], mean acti
332174/2500000: episode: 286, duration: 12.675s, episode steps: 1716, steps per second: 135, episode reward: -14.000, mean reward: -0.008 [-1.000, 1.000], mean acti
334129/2500000: episode: 287, duration: 14.037s, episode steps: 1955, steps per second: 139, episode reward: -12.000, mean reward: -0.006 [-1.000, 1.000], mean acti
336081/2500000: episode: 288, duration: 14.644s, episode steps: 1952, steps per second: 133, episode reward: -14.000, mean reward: -0.007 [-1.000, 1.000], mean acti
338003/2500000: episode: 289, duration: 13.613s, episode steps: 1922, steps per second: 141, episode reward: -13.000, mean reward: -0.007 [-1.000, 1.000], mean acti
339742/2500000: episode: 290, duration: 12.783s, episode steps: 1739, steps per second: 136, episode reward: -16.000, mean reward: -0.009 [-1.000, 1.000], mean acti
341736/2500000: episode: 291, duration: 14.290s, episode steps: 1994, steps per second: 140, episode reward: -13.000, mean reward: -0.007 [-1.000, 1.000], mean acti
343838/2500000: episode: 292, duration: 15.542s, episode steps: 2102, steps per second: 135, episode reward: -12.000, mean reward: -0.006 [-1.000, 1.000], mean acti
346131/2500000: episode: 293, duration: 16.583s, episode steps: 2293, steps per second: 138, episode reward: -11.000, mean reward: -0.005 [-1.000, 1.000], mean acti
348007/2500000: episode: 294, duration: 13.625s, episode steps: 1876, steps per second: 138, episode reward: -14.000, mean reward: -0.007 [-1.000, 1.000], mean acti
349670/2500000: episode: 295, duration: 11.798s, episode steps: 1663, steps per second: 141, episode reward: -15.000, mean reward: -0.009 [-1.000, 1.000], mean acti
351732/2500000: episode: 296, duration: 15.072s, episode steps: 2062, steps per second: 137, episode reward: -10.000, mean reward: -0.005 [-1.000, 1.000], mean acti
353844/2500000: episode: 297, duration: 15.536s, episode steps: 2112, steps per second: 136, episode reward: -11.000, mean reward: -0.005 [-1.000, 1.000], mean acti
355922/2500000: episode: 298, duration: 15.146s, episode steps: 2078, steps per second: 137, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
358404/2500000: episode: 299, duration: 18.226s, episode steps: 2482, steps per second: 136, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
360671/2500000: episode: 300, duration: 16.481s, episode steps: 2267, steps per second: 138, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
362813/2500000: episode: 301, duration: 15.601s, episode steps: 2142, steps per second: 137, episode reward: -11.000, mean reward: -0.005 [-1.000, 1.000], mean acti
364342/2500000: episode: 302, duration: 11.381s, episode steps: 1529, steps per second: 134, episode reward: -17.000, mean reward: -0.011 [-1.000, 1.000], mean acti
366189/2500000: episode: 303, duration: 13.409s, episode steps: 1847, steps per second: 138, episode reward: -13.000, mean reward: -0.007 [-1.000, 1.000], mean acti
368281/2500000: episode: 304, duration: 15.442s, episode steps: 2092, steps per second: 135, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
370369/2500000: episode: 305, duration: 15.058s, episode steps: 2088, steps per second: 139, episode reward: -12.000, mean reward: -0.006 [-1.000, 1.000], mean acti
372462/2500000: episode: 306, duration: 15.199s, episode steps: 2093, steps per second: 138, episode reward: -10.000, mean reward: -0.005 [-1.000, 1.000], mean acti
374447/2500000: episode: 307, duration: 14.560s, episode steps: 1985, steps per second: 136, episode reward: -12.000, mean reward: -0.006 [-1.000, 1.000], mean acti
376898/2500000: episode: 308, duration: 17.717s, episode steps: 2451, steps per second: 138, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
378104/2500000: episode: 309, duration: 8.705s, episode steps: 1206, steps per second: 139, episode reward: -18.000, mean reward: -0.015 [-1.000, 1.000], mean actic
380246/2500000: episode: 310, duration: 15.851s, episode steps: 2142, steps per second: 135, episode reward: -12.000, mean reward: -0.006 [-1.000, 1.000], mean acti
381869/2500000: episode: 311, duration: 11.929s, episode steps: 1623, steps per second: 136, episode reward: -17.000, mean reward: -0.010 [-1.000, 1.000], mean acti
383864/2500000: episode: 312, duration: 14.499s, episode steps: 1995, steps per second: 138, episode reward: -17.000, mean reward: -0.009 [-1.000, 1.000], mean acti
386276/2500000: episode: 313, duration: 17.480s, episode steps: 2412, steps per second: 138, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean acti
388744/2500000: episode: 314, duration: 18.217s, episode steps: 2468, steps per second: 135, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
390181/2500000: episode: 315, duration: 10.342s, episode steps: 1437, steps per second: 139, episode reward: -13.000, mean reward: -0.009 [-1.000, 1.000], mean acti
392294/2500000: episode: 316, duration: 15.657s, episode steps: 2113, steps per second: 135, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
393714/2500000: episode: 317, duration: 10.576s, episode steps: 1420, steps per second: 134, episode reward: -18.000, mean reward: -0.013 [-1.000, 1.000], mean acti
395907/2500000: episode: 318, duration: 16.077s, episode steps: 2193, steps per second: 136, episode reward: -11.000, mean reward: -0.005 [-1.000, 1.000], mean acti
397333/2500000: episode: 319, duration: 10.458s, episode steps: 1426, steps per second: 136, episode reward: -17.000, mean reward: -0.012 [-1.000, 1.000], mean acti
399289/2500000: episode: 320, duration: 14.320s, episode steps: 1956, steps per second: 137, episode reward: -14.000, mean reward: -0.007 [-1.000, 1.000], mean acti
400776/2500000: episode: 321, duration: 11.198s, episode steps: 1487, steps per second: 133, episode reward: -16.000, mean reward: -0.011 [-1.000, 1.000], mean acti
402427/2500000: episode: 322, duration: 12.313s, episode steps: 1651, steps per second: 134, episode reward: -15.000, mean reward: -0.009 [-1.000, 1.000], mean acti
404598/2500000: episode: 323, duration: 15.895s, episode steps: 2171, steps per second: 137, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
406332/2500000: episode: 324, duration: 12.908s, episode steps: 1734, steps per second: 134, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean acti
408324/2500000: episode: 325, duration: 14.752s, episode steps: 1992, steps per second: 135, episode reward: -18.000, mean reward: -0.009 [-1.000, 1.000], mean acti
410522/2500000: episode: 326, duration: 16.075s, episode steps: 2198, steps per second: 137, episode reward: -14.000, mean reward: -0.006 [-1.000, 1.000], mean acti
```

```
412555/2500000: episode: 327, duration: 14.863s, episode steps: 2033, steps per second: 137, episode reward: -15.000, mean reward: -0.007 [-1.000, 1.000], mean acti
414970/2500000: episode: 328, duration: 17.878s, episode steps: 2415, steps per second: 135, episode reward: -11.000, mean reward: -0.005 [-1.000, 1.000], mean acti
417265/2500000: episode: 329, duration: 16.826s, episode steps: 2295, steps per second: 136, episode reward: -11.000, mean reward: -0.005 [-1.000, 1.000], mean acti
419355/2500000: episode: 330, duration: 15.192s, episode steps: 2090, steps per second: 138, episode reward: -10.000, mean reward: -0.005 [-1.000, 1.000], mean acti
421324/2500000: episode: 331, duration: 14.592s, episode steps: 1969, steps per second: 135, episode reward: -15.000, mean reward: -0.008 [-1.000, 1.000], mean acti
423180/2500000: episode: 332, duration: 13.555s, episode steps: 1856, steps per second: 137, episode reward: -12.000, mean reward: -0.006 [-1.000, 1.000], mean acti
425027/2500000: episode: 333, duration: 13.705s, episode steps: 1847, steps per second: 135, episode reward: -16.000, mean reward: -0.009 [-1.000, 1.000], mean acti
427211/2500000: episode: 334, duration: 16.017s, episode steps: 2184, steps per second: 136, episode reward: -14.000, mean reward: -0.006 [-1.000, 1.000], mean acti
429351/2500000: episode: 335, duration: 15.762s, episode steps: 2140, steps per second: 136, episode reward: -14.000, mean reward: -0.007 [-1.000, 1.000], mean acti
431992/2500000: episode: 336, duration: 19.639s, episode steps: 2641, steps per second: 134, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
433761/2500000: episode: 337, duration: 12.978s, episode steps: 1769, steps per second: 136, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean acti
436219/2500000: episode: 338, duration: 18.356s, episode steps: 2458, steps per second: 134, episode reward: -9.000, mean reward: -0.004 [-1.000, 1.000], mean actic
438003/2500000: episode: 339, duration: 13.164s, episode steps: 1784, steps per second: 136, episode reward: -16.000, mean reward: -0.009 [-1.000, 1.000], mean acti
440814/2500000: episode: 340, duration: 20.669s, episode steps: 2811, steps per second: 136, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
443565/2500000: episode: 341, duration: 20.586s, episode steps: 2751, steps per second: 134, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
445842/2500000: episode: 342, duration: 16.831s, episode steps: 2277, steps per second: 135, episode reward: -19.000, mean reward: -0.008 [-1.000, 1.000], mean acti
447124/2500000: episode: 343, duration: 9.439s, episode steps: 1282, steps per second: 136, episode reward: -19.000, mean reward: -0.015 [-1.000, 1.000], mean actic
449576/2500000: episode: 344, duration: 18.061s, episode steps: 2452, steps per second: 136, episode reward: -14.000, mean reward: -0.006 [-1.000, 1.000], mean acti
451578/2500000: episode: 345, duration: 14.746s, episode steps: 2002, steps per second: 136, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
454464/2500000: episode: 346, duration: 21.293s, episode steps: 2886, steps per second: 136, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
457023/2500000: episode: 347, duration: 19.100s, episode steps: 2559, steps per second: 134, episode reward: -9.000, mean reward: -0.004 [-1.000, 1.000], mean actic
459188/2500000: episode: 348, duration: 15.771s, episode steps: 2165, steps per second: 137, episode reward: -12.000, mean reward: -0.006 [-1.000, 1.000], mean acti
461066/2500000: episode: 349, duration: 13.760s, episode steps: 1878, steps per second: 136, episode reward: -18.000, mean reward: -0.010 [-1.000, 1.000], mean acti
462732/2500000: episode: 350, duration: 12.186s, episode steps: 1666, steps per second: 137, episode reward: -16.000, mean reward: -0.010 [-1.000, 1.000], mean acti
465062/2500000: episode: 351, duration: 17.203s, episode steps: 2330, steps per second: 135, episode reward: -15.000, mean reward: -0.006 [-1.000, 1.000], mean acti
467125/2500000: episode: 352, duration: 15.307s, episode steps: 2063, steps per second: 135, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
468815/2500000: episode: 353, duration: 12.492s, episode steps: 1690, steps per second: 135, episode reward: -17.000, mean reward: -0.010 [-1.000, 1.000], mean acti
471336/2500000: episode: 354, duration: 18.804s, episode steps: 2521, steps per second: 134, episode reward: -15.000, mean reward: -0.006 [-1.000, 1.000], mean acti
474360/2500000: episode: 355, duration: 22.463s, episode steps: 3024, steps per second: 135, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
476026/2500000: episode: 356, duration: 12.530s, episode steps: 1666, steps per second: 133, episode reward: -18.000, mean reward: -0.011 [-1.000, 1.000], mean acti
478252/2500000: episode: 357, duration: 16.632s, episode steps: 2226, steps per second: 134, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
480656/2500000: episode: 358, duration: 17.484s, episode steps: 2404, steps per second: 137, episode reward: -11.000, mean reward: -0.005 [-1.000, 1.000], mean acti
482566/2500000: episode: 359, duration: 13.853s, episode steps: 1910, steps per second: 138, episode reward: -15.000, mean reward: -0.008 [-1.000, 1.000], mean acti
485679/2500000: episode: 360, duration: 22.903s, episode steps: 3113, steps per second: 136, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
489224/2500000: episode: 361, duration: 25.998s, episode steps: 3545, steps per second: 136, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
491532/2500000: episode: 362, duration: 17.163s, episode steps: 2308, steps per second: 134, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
494261/2500000: episode: 363, duration: 20.386s, episode steps: 2729, steps per second: 134, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
497092/2500000: episode: 364, duration: 20.775s, episode steps: 2831, steps per second: 136, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
499403/2500000: episode: 365, duration: 17.340s, episode steps: 2311, steps per second: 133, episode reward: -11.000, mean reward: -0.005 [-1.000, 1.000], mean acti
502335/2500000: episode: 366, duration: 21.392s, episode steps: 2932, steps per second: 137, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
504150/2500000: episode: 367, duration: 13.366s, episode steps: 1815, steps per second: 136, episode reward: -17.000, mean reward: -0.009 [-1.000, 1.000], mean acti
506339/2500000: episode: 368, duration: 15.954s, episode steps: 2189, steps per second: 137, episode reward: -16.000, mean reward: -0.007 [-1.000, 1.000], mean acti
508601/2500000: episode: 369, duration: 16.502s, episode steps: 2262, steps per second: 137, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
511175/2500000: episode: 370, duration: 18.893s, episode steps: 2574, steps per second: 136, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
514513/2500000: episode: 371, duration: 24.387s, episode steps: 3338, steps per second: 137, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
517278/2500000: episode: 372, duration: 20.349s, episode steps: 2765, steps per second: 136, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
519086/2500000: episode: 373, duration: 13.776s, episode steps: 1808, steps per second: 131, episode reward: -16.000, mean reward: -0.009 [-1.000, 1.000], mean acti
522132/2500000: episode: 374, duration: 22.848s, episode steps: 3046, steps per second: 133, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
524866/2500000: episode: 375, duration: 20.067s, episode steps: 2734, steps per second: 136, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
527466/2500000: episode: 376, duration: 19.653s, episode steps: 2600, steps per second: 132, episode reward: -7.000, mean reward: -0.003 [-1.000, 1.000], mean actic
529899/2500000: episode: 377, duration: 18.033s, episode steps: 2433, steps per second: 135, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
532392/2500000: episode: 378, duration: 18.374s, episode steps: 2493, steps per second: 136, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
534504/2500000: episode: 379, duration: 15.842s, episode steps: 2112, steps per second: 133, episode reward: -12.000, mean reward: -0.006 [-1.000, 1.000], mean acti
536886/2500000: episode: 380, duration: 17.502s, episode steps: 2382, steps per second: 136, episode reward: -9.000, mean reward: -0.004 [-1.000, 1.000], mean actic
539932/2500000: episode: 381, duration: 22.672s, episode steps: 3046, steps per second: 134, episode reward: -5.000, mean reward: -0.002 [-1.000, 1.000], mean actic
541191/2500000: episode: 382, duration: 9.486s, episode steps: 1259, steps per second: 133, episode reward: -18.000, mean reward: -0.014 [-1.000, 1.000], mean actic
542940/2500000: episode: 383, duration: 12.934s, episode steps: 1749, steps per second: 135, episode reward: -16.000, mean reward: -0.009 [-1.000, 1.000], mean acti
545633/2500000: episode: 384, duration: 20.116s, episode steps: 2693, steps per second: 134, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
548173/2500000: episode: 385, duration: 19.032s, episode steps: 2540, steps per second: 133, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
550870/2500000: episode: 386, duration: 19.992s, episode steps: 2697, steps per second: 135, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
552978/2500000: episode: 387, duration: 15.636s, episode steps: 2108, steps per second: 135, episode reward: -11.000, mean reward: -0.005 [-1.000, 1.000], mean acti
555930/2500000: episode: 388, duration: 22.239s, episode steps: 2952, steps per second: 133, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
558463/2500000: episode: 389, duration: 19.101s, episode steps: 2533, steps per second: 133, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
560980/2500000: episode: 390, duration: 18.879s, episode steps: 2517, steps per second: 133, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
563782/2500000: episode: 391, duration: 20.959s, episode steps: 2802, steps per second: 134, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
566009/2500000: episode: 392, duration: 16.729s, episode steps: 2227, steps per second: 133, episode reward: -11.000, mean reward: -0.005 [-1.000, 1.000], mean acti
569030/2500000: episode: 393, duration: 23.046s, episode steps: 3021, steps per second: 131, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
572248/2500000: episode: 394, duration: 24.414s, episode steps: 3218, steps per second: 132, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
574856/2500000: episode: 395, duration: 19.702s, episode steps: 2608, steps per second: 132, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
577693/2500000: episode: 396, duration: 21.316s, episode steps: 2837, steps per second: 133, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
580835/2500000: episode: 397, duration: 23.668s, episode steps: 3142, steps per second: 133, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
583777/2500000: episode: 398, duration: 22.100s, episode steps: 2942, steps per second: 133, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
586659/2500000: episode: 399, duration: 21.490s, episode steps: 2882, steps per second: 134, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
589369/2500000: episode: 400, duration: 20.480s, episode steps: 2710, steps per second: 132, episode reward: -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
592283/2500000: episode: 401, duration: 21.998s, episode steps: 2914, steps per second: 132, episode reward: -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
595114/2500000: episode: 402, duration: 21.338s, episode steps: 2831, steps per second: 133, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
597886/2500000: episode: 403, duration: 20.302s, episode steps: 2772, steps per second: 137, episode reward: -7.000, mean reward: -0.003 [-1.000, 1.000], mean actic
600595/2500000: episode: 404, duration: 20.236s, episode steps: 2709, steps per second: 134, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
603448/2500000: episode: 405, duration: 21.265s, episode steps: 2853, steps per second: 134, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
606339/2500000: episode: 406, duration: 21.412s, episode steps: 2891, steps per second: 135, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
608402/2500000: episode: 407, duration: 15.524s, episode steps: 2063, steps per second: 133, episode reward: -14.000, mean reward: -0.007 [-1.000, 1.000], mean acti
611202/2500000: episode: 408, duration: 21.122s, episode steps: 2800, steps per second: 133, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
613549/2500000: episode: 409, duration: 17.444s, episode steps: 2347, steps per second: 135, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
615798/2500000: episode: 410, duration: 16.705s, episode steps: 2249, steps per second: 135, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
618380/2500000: episode: 411, duration: 19.235s, episode steps: 2582, steps per second: 134, episode reward: -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
621235/2500000: episode: 412, duration: 21.145s, episode steps: 2855, steps per second: 135, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
624605/2500000: episode: 413, duration: 24.988s, episode steps: 3370, steps per second: 135, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
627353/2500000: episode: 414, duration: 20.653s, episode steps: 2748, steps per second: 133, episode reward: -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
630208/2500000: episode: 415, duration: 21.433s, episode steps: 2855, steps per second: 133, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
633312/2500000: episode: 416, duration: 23.336s, episode steps: 3104, steps per second: 133, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
636001/2500000: episode: 417, duration: 20.170s, episode steps: 2689, steps per second: 133, episode reward: -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
639107/2500000: episode: 418, duration: 23.210s, episode steps: 3106, steps per second: 134, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
641954/2500000: episode: 419, duration: 21.517s, episode steps: 2847, steps per second: 132, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
644368/2500000: episode: 420, duration: 18.127s, episode steps: 2414, steps per second: 133, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
646481/2500000: episode: 421, duration: 15.694s, episode steps: 2113, steps per second: 135, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
648946/2500000: episode: 422, duration: 18.441s, episode steps: 2465, steps per second: 134, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
651847/2500000: episode: 423, duration: 22.039s, episode steps: 2901, steps per second: 132, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
654518/2500000: episode: 424, duration: 19.880s, episode steps: 2671, steps per second: 134, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
657966/2500000: episode: 425, duration: 25.511s, episode steps: 3448, steps per second: 135, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
660766/2500000: episode: 426, duration: 21.037s, episode steps: 2800, steps per second: 133, episode reward: -14.000, mean reward: -0.005 [-1.000, 1.000], mean acti
663589/2500000: episode: 427, duration: 20.950s, episode steps: 2823, steps per second: 135, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
667091/2500000: episode: 428, duration: 26.065s, episode steps: 3502, steps per second: 134, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
669268/2500000: episode: 429, duration: 16.338s, episode steps: 2177, steps per second: 133, episode reward: -14.000, mean reward: -0.006 [-1.000, 1.000], mean acti
671844/2500000: episode: 430, duration: 19.228s, episode steps: 2576, steps per second: 134, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean actic
674673/2500000: episode: 431, duration: 21.144s, episode steps: 2829, steps per second: 134, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
677170/2500000: episode: 432, duration: 18.635s, episode steps: 2497, steps per second: 134, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
679498/2500000: episode: 433, duration: 17.432s, episode steps: 2328, steps per second: 134, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
682955/2500000: episode: 434, duration: 25.678s, episode steps: 3457, steps per second: 135, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
685538/2500000: episode: 435, duration: 19.088s, episode steps: 2583, steps per second: 135, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
```

```
687484/2500000: episode: 436, duration: 13.982s, episode steps: 1946, steps per second: 139, episode reward: -16.000, mean reward: -0.008 [-1.000, 1.000], mean acti
690276/2500000: episode: 437, duration: 20.571s, episode steps: 2792, steps per second: 136, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
693452/2500000: episode: 438, duration: 23.945s, episode steps: 3176, steps per second: 133, episode reward: -11.000, mean reward: -0.003 [-1.000, 1.000], mean acti
696937/2500000: episode: 439, duration: 25.630s, episode steps: 3485, steps per second: 136, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean acti
700130/2500000: episode: 440, duration: 23.232s, episode steps: 3193, steps per second: 137, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean acti
702786/2500000: episode: 441, duration: 19.820s, episode steps: 2656, steps per second: 134, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean acti
705461/2500000: episode: 442, duration: 19.609s, episode steps: 2675, steps per second: 136, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean acti
708342/2500000: episode: 443, duration: 21.798s, episode steps: 2881, steps per second: 132, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
711378/2500000: episode: 444, duration: 22.241s, episode steps: 3036, steps per second: 137, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean acti
714335/2500000: episode: 445, duration: 21.393s, episode steps: 2957, steps per second: 138, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean acti
716630/2500000: episode: 446, duration: 17.000s, episode steps: 2295, steps per second: 135, episode reward: -11.000, mean reward: -0.005 [-1.000, 1.000], mean acti
719560/2500000: episode: 447, duration: 21.402s, episode steps: 2930, steps per second: 137, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean acti
723081/2500000: episode: 448, duration: 26.164s, episode steps: 3521, steps per second: 135, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean acti
725832/2500000: episode: 449, duration: 20.167s, episode steps: 2751, steps per second: 136, episode reward: -10.000, mean reward: -0.004 [-1.000, 1.000], mean acti
728163/2500000: episode: 450, duration: 16.959s, episode steps: 2331, steps per second: 137, episode reward: -14.000, mean reward: -0.006 [-1.000, 1.000], mean acti
732112/2500000: episode: 451, duration: 28.748s, episode steps: 3949, steps per second: 137, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean acti
735863/2500000: episode: 452, duration: 27.975s, episode steps: 3751, steps per second: 134, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean acti
738148/2500000: episode: 453, duration: 16.975s, episode steps: 2285, steps per second: 135, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
740779/2500000: episode: 454, duration: 19.466s, episode steps: 2631, steps per second: 135, episode reward: -15.000, mean reward: -0.006 [-1.000, 1.000], mean acti
743016/2500000: episode: 455, duration: 16.361s, episode steps: 2237, steps per second: 137, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
745551/2500000: episode: 456, duration: 18.330s, episode steps: 2535, steps per second: 138, episode reward: -13.000, mean reward: -0.005 [-1.000, 1.000], mean acti
747766/2500000: episode: 457, duration: 16.503s, episode steps: 2215, steps per second: 134, episode reward: -13.000, mean reward: -0.006 [-1.000, 1.000], mean acti
750564/2500000: episode: 458, duration: 20.548s, episode steps: 2798, steps per second: 136, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean acti
753989/2500000: episode: 459, duration: 24.889s, episode steps: 3425, steps per second: 138, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean acti
757028/2500000: episode: 460, duration: 22.449s, episode steps: 3039, steps per second: 135, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean acti
759947/2500000: episode: 461, duration: 21.463s, episode steps: 2919, steps per second: 136, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean acti
763650/2500000: episode: 462, duration: 27.221s, episode steps: 3703, steps per second: 136, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean acti
767128/2500000: episode: 463, duration: 25.528s, episode steps: 3478, steps per second: 136, episode reward: -5.000, mean reward: -0.001 [-1.000, 1.000], mean acti
769276/2500000: episode: 464, duration: 15.784s, episode steps: 2148, steps per second: 136, episode reward: -12.000, mean reward: -0.006 [-1.000, 1.000], mean acti
772691/2500000: episode: 465, duration: 25.042s, episode steps: 3415, steps per second: 136, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean acti
775067/2500000: episode: 466, duration: 17.750s, episode steps: 2376, steps per second: 134, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
778024/2500000: episode: 467, duration: 21.713s, episode steps: 2957, steps per second: 136, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean acti
780580/2500000: episode: 468, duration: 19.031s, episode steps: 2556, steps per second: 134, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
784053/2500000: episode: 469, duration: 25.702s, episode steps: 3473, steps per second: 135, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean acti
787606/2500000: episode: 470, duration: 26.216s, episode steps: 3553, steps per second: 136, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean acti
790984/2500000: episode: 471, duration: 25.389s, episode steps: 3378, steps per second: 133, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean acti
793899/2500000: episode: 472, duration: 21.642s, episode steps: 2915, steps per second: 135, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean acti
796566/2500000: episode: 473, duration: 20.001s, episode steps: 2667, steps per second: 133, episode reward: 7.000, mean reward: 0.003 [-1.000, 1.000], mean acti
800399/2500000: episode: 474, duration: 28.493s, episode steps: 3833, steps per second: 135, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean acti
803489/2500000: episode: 475, duration: 22.631s, episode steps: 3090, steps per second: 137, episode reward: -5.000, mean reward: -0.002 [-1.000, 1.000], mean acti
807225/2500000: episode: 476, duration: 28.183s, episode steps: 3736, steps per second: 133, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean acti
810474/2500000: episode: 477, duration: 24.433s, episode steps: 3249, steps per second: 133, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean acti
813942/2500000: episode: 478, duration: 25.926s, episode steps: 3468, steps per second: 134, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean acti
817442/2500000: episode: 479, duration: 26.815s, episode steps: 3500, steps per second: 131, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean acti
821047/2500000: episode: 480, duration: 27.057s, episode steps: 3605, steps per second: 133, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean acti
824652/2500000: episode: 481, duration: 27.295s, episode steps: 3605, steps per second: 132, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean acti
828475/2500000: episode: 482, duration: 28.947s, episode steps: 3823, steps per second: 132, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean acti
832286/2500000: episode: 483, duration: 28.560s, episode steps: 3811, steps per second: 133, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean acti
835968/2500000: episode: 484, duration: 27.732s, episode steps: 3682, steps per second: 133, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean acti
838356/2500000: episode: 485, duration: 18.240s, episode steps: 2388, steps per second: 131, episode reward: -12.000, mean reward: -0.005 [-1.000, 1.000], mean acti
841050/2500000: episode: 486, duration: 20.250s, episode steps: 2694, steps per second: 133, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean acti
844932/2500000: episode: 487, duration: 29.489s, episode steps: 3882, steps per second: 132, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean acti
848353/2500000: episode: 488, duration: 25.917s, episode steps: 3421, steps per second: 132, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean acti
850815/2500000: episode: 489, duration: 18.626s, episode steps: 2462, steps per second: 132, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean acti
855207/2500000: episode: 490, duration: 33.383s, episode steps: 4392, steps per second: 132, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean acti
858382/2500000: episode: 491, duration: 24.175s, episode steps: 3175, steps per second: 131, episode reward: -5.000, mean reward: -0.002 [-1.000, 1.000], mean acti
862355/2500000: episode: 492, duration: 30.164s, episode steps: 3973, steps per second: 132, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean acti
865947/2500000: episode: 493, duration: 27.071s, episode steps: 3592, steps per second: 133, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean acti
869390/2500000: episode: 494, duration: 26.045s, episode steps: 3443, steps per second: 132, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean acti
873143/2500000: episode: 495, duration: 28.512s, episode steps: 3753, steps per second: 132, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean acti
876938/2500000: episode: 496, duration: 28.765s, episode steps: 3795, steps per second: 132, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean acti
879541/2500000: episode: 497, duration: 19.845s, episode steps: 2603, steps per second: 131, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean acti
882799/2500000: episode: 498, duration: 24.640s, episode steps: 3258, steps per second: 132, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean acti
886415/2500000: episode: 499, duration: 27.478s, episode steps: 3616, steps per second: 132, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean acti
889744/2500000: episode: 500, duration: 25.169s, episode steps: 3329, steps per second: 132, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean acti
892815/2500000: episode: 501, duration: 23.108s, episode steps: 3071, steps per second: 133, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean acti
896832/2500000: episode: 502, duration: 30.396s, episode steps: 4017, steps per second: 132, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean acti
900284/2500000: episode: 503, duration: 25.871s, episode steps: 3452, steps per second: 133, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean acti
904264/2500000: episode: 504, duration: 29.574s, episode steps: 3980, steps per second: 135, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean acti
907662/2500000: episode: 505, duration: 25.242s, episode steps: 3398, steps per second: 135, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean acti
910834/2500000: episode: 506, duration: 23.598s, episode steps: 3172, steps per second: 134, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean acti
913773/2500000: episode: 507, duration: 22.230s, episode steps: 2939, steps per second: 132, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean acti
917017/2500000: episode: 508, duration: 24.299s, episode steps: 3244, steps per second: 134, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean acti
920959/2500000: episode: 509, duration: 29.118s, episode steps: 3942, steps per second: 135, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean acti
924230/2500000: episode: 510, duration: 24.377s, episode steps: 3271, steps per second: 134, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean acti
926650/2500000: episode: 511, duration: 18.129s, episode steps: 2420, steps per second: 133, episode reward: -16.000, mean reward: -0.007 [-1.000, 1.000], mean acti
929495/2500000: episode: 512, duration: 20.861s, episode steps: 2845, steps per second: 136, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean acti
933312/2500000: episode: 513, duration: 28.128s, episode steps: 3817, steps per second: 136, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean acti
936200/2500000: episode: 514, duration: 21.202s, episode steps: 2888, steps per second: 136, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
939382/2500000: episode: 515, duration: 23.424s, episode steps: 3182, steps per second: 136, episode reward: -8.000, mean reward: -0.003 [-1.000, 1.000], mean acti
942032/2500000: episode: 516, duration: 19.907s, episode steps: 2650, steps per second: 133, episode reward: -14.000, mean reward: -0.006 [-1.000, 1.000], mean acti
945886/2500000: episode: 517, duration: 27.687s, episode steps: 3854, steps per second: 139, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean acti
949179/2500000: episode: 518, duration: 23.574s, episode steps: 3293, steps per second: 140, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean acti
952970/2500000: episode: 519, duration: 27.377s, episode steps: 3791, steps per second: 138, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean acti
956769/2500000: episode: 520, duration: 26.927s, episode steps: 3799, steps per second: 141, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean acti
960086/2500000: episode: 521, duration: 23.628s, episode steps: 3317, steps per second: 140, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean acti
963208/2500000: episode: 522, duration: 22.187s, episode steps: 3122, steps per second: 141, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean acti
966361/2500000: episode: 523, duration: 22.583s, episode steps: 3153, steps per second: 140, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean acti
969581/2500000: episode: 524, duration: 23.239s, episode steps: 3220, steps per second: 139, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean acti
972368/2500000: episode: 525, duration: 20.141s, episode steps: 2787, steps per second: 138, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean acti
975654/2500000: episode: 526, duration: 23.369s, episode steps: 3286, steps per second: 141, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean acti
978478/2500000: episode: 527, duration: 20.282s, episode steps: 2824, steps per second: 139, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean acti
981812/2500000: episode: 528, duration: 24.067s, episode steps: 3334, steps per second: 139, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean acti
984986/2500000: episode: 529, duration: 23.161s, episode steps: 3174, steps per second: 137, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean acti
988094/2500000: episode: 530, duration: 21.799s, episode steps: 3108, steps per second: 143, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean acti
991612/2500000: episode: 531, duration: 24.926s, episode steps: 3518, steps per second: 141, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean acti
994832/2500000: episode: 532, duration: 22.759s, episode steps: 3220, steps per second: 141, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean acti
998157/2500000: episode: 533, duration: 23.736s, episode steps: 3325, steps per second: 140, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean acti
1001694/2500000: episode: 534, duration: 25.230s, episode steps: 3537, steps per second: 140, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean acti
1005225/2500000: episode: 535, duration: 25.059s, episode steps: 3531, steps per second: 141, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean acti
1008724/2500000: episode: 536, duration: 24.903s, episode steps: 3499, steps per second: 141, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean acti
1012091/2500000: episode: 537, duration: 24.564s, episode steps: 3367, steps per second: 137, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean acti
1015288/2500000: episode: 538, duration: 22.898s, episode steps: 3197, steps per second: 140, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean acti
1018121/2500000: episode: 539, duration: 20.188s, episode steps: 2833, steps per second: 140, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean acti
1021417/2500000: episode: 540, duration: 23.273s, episode steps: 3296, steps per second: 142, episode reward: -5.000, mean reward: -0.002 [-1.000, 1.000], mean acti
1023874/2500000: episode: 541, duration: 17.611s, episode steps: 2457, steps per second: 140, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean acti
1027658/2500000: episode: 542, duration: 27.384s, episode steps: 3784, steps per second: 138, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean acti
1030721/2500000: episode: 543, duration: 21.839s, episode steps: 3063, steps per second: 140, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean acti
1033722/2500000: episode: 544, duration: 21.523s, episode steps: 3001, steps per second: 139, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean acti
```

```
1036557/2500000: episode: 545, duration: 20.065s, episode steps: 2835, steps per second: 141, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1039874/2500000: episode: 546, duration: 23.605s, episode steps: 3317, steps per second: 141, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1042806/2500000: episode: 547, duration: 21.030s, episode steps: 2932, steps per second: 139, episode reward: -9.000, mean reward: -0.003 [-1.000, 1.000], mean actic
1046357/2500000: episode: 548, duration: 25.458s, episode steps: 3551, steps per second: 139, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1049086/2500000: episode: 549, duration: 19.328s, episode steps: 2729, steps per second: 141, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action:
1052112/2500000: episode: 550, duration: 21.473s, episode steps: 3026, steps per second: 141, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1055147/2500000: episode: 551, duration: 21.460s, episode steps: 3035, steps per second: 141, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1057876/2500000: episode: 552, duration: 19.247s, episode steps: 2729, steps per second: 142, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1061036/2500000: episode: 553, duration: 22.094s, episode steps: 3160, steps per second: 143, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1063555/2500000: episode: 554, duration: 17.452s, episode steps: 2519, steps per second: 144, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean actic
1066758/2500000: episode: 555, duration: 21.979s, episode steps: 3203, steps per second: 146, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1069967/2500000: episode: 556, duration: 22.661s, episode steps: 3209, steps per second: 142, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1073263/2500000: episode: 557, duration: 23.191s, episode steps: 3296, steps per second: 142, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1075838/2500000: episode: 558, duration: 18.043s, episode steps: 2575, steps per second: 143, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean actio
1079076/2500000: episode: 559, duration: 23.317s, episode steps: 3238, steps per second: 139, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1082833/2500000: episode: 560, duration: 26.985s, episode steps: 3757, steps per second: 139, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1086765/2500000: episode: 561, duration: 28.687s, episode steps: 3932, steps per second: 137, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1090557/2500000: episode: 562, duration: 27.812s, episode steps: 3792, steps per second: 136, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1094002/2500000: episode: 563, duration: 25.204s, episode steps: 3445, steps per second: 137, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1097271/2500000: episode: 564, duration: 24.191s, episode steps: 3269, steps per second: 135, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1099513/2500000: episode: 565, duration: 16.514s, episode steps: 2242, steps per second: 136, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean actio
1102438/2500000: episode: 566, duration: 21.763s, episode steps: 2925, steps per second: 134, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1105460/2500000: episode: 567, duration: 22.331s, episode steps: 3022, steps per second: 135, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1108709/2500000: episode: 568, duration: 24.007s, episode steps: 3249, steps per second: 135, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1111716/2500000: episode: 569, duration: 22.298s, episode steps: 3007, steps per second: 135, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1114907/2500000: episode: 570, duration: 24.237s, episode steps: 3191, steps per second: 132, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1117597/2500000: episode: 571, duration: 20.100s, episode steps: 2690, steps per second: 134, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean actio
1120484/2500000: episode: 572, duration: 21.622s, episode steps: 2887, steps per second: 134, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1123779/2500000: episode: 573, duration: 24.877s, episode steps: 3295, steps per second: 132, episode reward: -5.000, mean reward: -0.002 [-1.000, 1.000], mean actic
1126936/2500000: episode: 574, duration: 23.505s, episode steps: 3157, steps per second: 134, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1131323/2500000: episode: 575, duration: 32.828s, episode steps: 4387, steps per second: 134, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1134730/2500000: episode: 576, duration: 25.583s, episode steps: 3407, steps per second: 133, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1138653/2500000: episode: 577, duration: 29.553s, episode steps: 3923, steps per second: 133, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1142027/2500000: episode: 578, duration: 25.377s, episode steps: 3374, steps per second: 133, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1145211/2500000: episode: 579, duration: 23.608s, episode steps: 3184, steps per second: 135, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1148344/2500000: episode: 580, duration: 23.243s, episode steps: 3133, steps per second: 135, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
1151016/2500000: episode: 581, duration: 19.865s, episode steps: 2672, steps per second: 135, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean actio
1154101/2500000: episode: 582, duration: 23.041s, episode steps: 3085, steps per second: 134, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1157993/2500000: episode: 583, duration: 28.960s, episode steps: 3892, steps per second: 134, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1160757/2500000: episode: 584, duration: 20.189s, episode steps: 2764, steps per second: 137, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1163883/2500000: episode: 585, duration: 23.430s, episode steps: 3126, steps per second: 133, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1166862/2500000: episode: 586, duration: 22.127s, episode steps: 2979, steps per second: 135, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1170311/2500000: episode: 587, duration: 25.670s, episode steps: 3449, steps per second: 134, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1174085/2500000: episode: 588, duration: 28.157s, episode steps: 3774, steps per second: 134, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1177440/2500000: episode: 589, duration: 25.196s, episode steps: 3355, steps per second: 133, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
1180752/2500000: episode: 590, duration: 24.567s, episode steps: 3312, steps per second: 135, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1184559/2500000: episode: 591, duration: 28.261s, episode steps: 3807, steps per second: 135, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1187426/2500000: episode: 592, duration: 21.352s, episode steps: 2867, steps per second: 134, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1190139/2500000: episode: 593, duration: 20.268s, episode steps: 2713, steps per second: 134, episode reward: 14.000, mean reward: 0.005 [-1.000, 1.000], mean actio
1193014/2500000: episode: 594, duration: 21.510s, episode steps: 2875, steps per second: 134, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1196224/2500000: episode: 595, duration: 24.159s, episode steps: 3210, steps per second: 133, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1199138/2500000: episode: 596, duration: 21.727s, episode steps: 2914, steps per second: 134, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean actio
1202665/2500000: episode: 597, duration: 26.408s, episode steps: 3527, steps per second: 134, episode reward: -7.000, mean reward: -0.002 [-1.000, 1.000], mean actic
1206863/2500000: episode: 598, duration: 30.948s, episode steps: 4198, steps per second: 136, episode reward: -5.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1210306/2500000: episode: 599, duration: 25.706s, episode steps: 3443, steps per second: 134, episode reward: -8.000, mean reward: -0.002 [-1.000, 1.000], mean actic
1214203/2500000: episode: 600, duration: 29.241s, episode steps: 3897, steps per second: 133, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1217896/2500000: episode: 601, duration: 27.343s, episode steps: 3693, steps per second: 135, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1221111/2500000: episode: 602, duration: 23.889s, episode steps: 3215, steps per second: 135, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
1224584/2500000: episode: 603, duration: 25.535s, episode steps: 3473, steps per second: 136, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1227624/2500000: episode: 604, duration: 22.758s, episode steps: 3040, steps per second: 134, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1230230/2500000: episode: 605, duration: 19.606s, episode steps: 2606, steps per second: 133, episode reward: -6.000, mean reward: -0.002 [-1.000, 1.000], mean actic
1234387/2500000: episode: 606, duration: 30.833s, episode steps: 4157, steps per second: 135, episode reward: -5.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1237606/2500000: episode: 607, duration: 24.239s, episode steps: 3219, steps per second: 133, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1240165/2500000: episode: 608, duration: 19.215s, episode steps: 2559, steps per second: 133, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
1242960/2500000: episode: 609, duration: 20.984s, episode steps: 2795, steps per second: 133, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1246470/2500000: episode: 610, duration: 26.458s, episode steps: 3510, steps per second: 133, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1249692/2500000: episode: 611, duration: 24.207s, episode steps: 3222, steps per second: 133, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1252614/2500000: episode: 612, duration: 22.113s, episode steps: 2922, steps per second: 132, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1256111/2500000: episode: 613, duration: 26.424s, episode steps: 3497, steps per second: 132, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1260856/2500000: episode: 614, duration: 35.999s, episode steps: 4745, steps per second: 132, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1264122/2500000: episode: 615, duration: 24.886s, episode steps: 3266, steps per second: 131, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1267149/2500000: episode: 616, duration: 23.192s, episode steps: 3027, steps per second: 131, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean actio
1270214/2500000: episode: 617, duration: 23.112s, episode steps: 3065, steps per second: 133, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean actio
1274201/2500000: episode: 618, duration: 30.124s, episode steps: 3987, steps per second: 132, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1277713/2500000: episode: 619, duration: 26.754s, episode steps: 3512, steps per second: 131, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1280852/2500000: episode: 620, duration: 24.223s, episode steps: 3139, steps per second: 130, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean actio
1284061/2500000: episode: 621, duration: 24.346s, episode steps: 3209, steps per second: 132, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean actio
1286273/2500000: episode: 622, duration: 16.697s, episode steps: 2212, steps per second: 132, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean actio
1290263/2500000: episode: 623, duration: 30.331s, episode steps: 3990, steps per second: 132, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1294051/2500000: episode: 624, duration: 28.636s, episode steps: 3788, steps per second: 132, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1297241/2500000: episode: 625, duration: 24.300s, episode steps: 3190, steps per second: 131, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean actio
1300783/2500000: episode: 626, duration: 27.164s, episode steps: 3542, steps per second: 130, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1304447/2500000: episode: 627, duration: 27.638s, episode steps: 3664, steps per second: 133, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1307793/2500000: episode: 628, duration: 25.521s, episode steps: 3346, steps per second: 131, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1312311/2500000: episode: 629, duration: 34.207s, episode steps: 4518, steps per second: 132, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1315573/2500000: episode: 630, duration: 24.704s, episode steps: 3262, steps per second: 132, episode reward: -5.000, mean reward: -0.002 [-1.000, 1.000], mean actic
1319227/2500000: episode: 631, duration: 27.697s, episode steps: 3654, steps per second: 132, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1324053/2500000: episode: 632, duration: 36.794s, episode steps: 4826, steps per second: 131, episode reward: -2.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1327211/2500000: episode: 633, duration: 23.890s, episode steps: 3158, steps per second: 132, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1330353/2500000: episode: 634, duration: 23.814s, episode steps: 3142, steps per second: 132, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1333132/2500000: episode: 635, duration: 20.937s, episode steps: 2779, steps per second: 133, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1337344/2500000: episode: 636, duration: 31.641s, episode steps: 4212, steps per second: 133, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1340173/2500000: episode: 637, duration: 21.300s, episode steps: 2829, steps per second: 133, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean actio
1343223/2500000: episode: 638, duration: 23.066s, episode steps: 3050, steps per second: 132, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean actio
1346075/2500000: episode: 639, duration: 21.656s, episode steps: 2852, steps per second: 132, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean actio
1350687/2500000: episode: 640, duration: 34.749s, episode steps: 4612, steps per second: 133, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1354359/2500000: episode: 641, duration: 27.596s, episode steps: 3672, steps per second: 133, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1357386/2500000: episode: 642, duration: 23.087s, episode steps: 3027, steps per second: 131, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1361163/2500000: episode: 643, duration: 28.866s, episode steps: 3777, steps per second: 131, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1363368/2500000: episode: 644, duration: 16.618s, episode steps: 2205, steps per second: 133, episode reward: 13.000, mean reward: 0.006 [-1.000, 1.000], mean actio
1366576/2500000: episode: 645, duration: 24.386s, episode steps: 3208, steps per second: 132, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
1370113/2500000: episode: 646, duration: 26.979s, episode steps: 3537, steps per second: 131, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1372940/2500000: episode: 647, duration: 21.336s, episode steps: 2827, steps per second: 132, episode reward: -12.000, mean reward: -0.004 [-1.000, 1.000], mean acti
1375937/2500000: episode: 648, duration: 22.828s, episode steps: 2997, steps per second: 131, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1380532/2500000: episode: 649, duration: 34.612s, episode steps: 4595, steps per second: 133, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1383721/2500000: episode: 650, duration: 23.685s, episode steps: 3189, steps per second: 135, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1387184/2500000: episode: 651, duration: 26.161s, episode steps: 3463, steps per second: 132, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1390259/2500000: episode: 652, duration: 22.817s, episode steps: 3075, steps per second: 135, episode reward: -10.000, mean reward: -0.003 [-1.000, 1.000], mean acti
1393048/2500000: episode: 653, duration: 21.022s, episode steps: 2789, steps per second: 133, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
```

```
1395657/2500000: episode: 654, duration: 19.644s, episode steps: 2609, steps per second: 133, episode reward: 7.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1399975/2500000: episode: 655, duration: 32.602s, episode steps: 4318, steps per second: 132, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1403043/2500000: episode: 656, duration: 23.271s, episode steps: 3068, steps per second: 132, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1406589/2500000: episode: 657, duration: 26.723s, episode steps: 3546, steps per second: 133, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1410532/2500000: episode: 658, duration: 29.757s, episode steps: 3943, steps per second: 133, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1412666/2500000: episode: 659, duration: 15.965s, episode steps: 2134, steps per second: 134, episode reward: 18.000, mean reward: 0.008 [-1.000, 1.000], mean actior
1416134/2500000: episode: 660, duration: 25.957s, episode steps: 3468, steps per second: 134, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1418359/2500000: episode: 661, duration: 16.592s, episode steps: 2225, steps per second: 134, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean actior
1421772/2500000: episode: 662, duration: 25.621s, episode steps: 3413, steps per second: 133, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1425510/2500000: episode: 663, duration: 28.067s, episode steps: 3738, steps per second: 133, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1429529/2500000: episode: 664, duration: 29.854s, episode steps: 4019, steps per second: 135, episode reward: 2.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1431701/2500000: episode: 665, duration: 16.537s, episode steps: 2172, steps per second: 131, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean actior
1433937/2500000: episode: 666, duration: 16.818s, episode steps: 2236, steps per second: 133, episode reward: 18.000, mean reward: 0.008 [-1.000, 1.000], mean actior
1437397/2500000: episode: 667, duration: 25.917s, episode steps: 3460, steps per second: 134, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1440807/2500000: episode: 668, duration: 25.831s, episode steps: 3410, steps per second: 132, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1444238/2500000: episode: 669, duration: 26.103s, episode steps: 3431, steps per second: 131, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1447168/2500000: episode: 670, duration: 21.628s, episode steps: 2930, steps per second: 135, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean actior
1450194/2500000: episode: 671, duration: 22.776s, episode steps: 3026, steps per second: 133, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1454719/2500000: episode: 672, duration: 33.922s, episode steps: 4525, steps per second: 133, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1458582/2500000: episode: 673, duration: 28.525s, episode steps: 3863, steps per second: 135, episode reward: -5.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1461258/2500000: episode: 674, duration: 20.014s, episode steps: 2676, steps per second: 134, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1464809/2500000: episode: 675, duration: 26.110s, episode steps: 3551, steps per second: 136, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean actior
1468134/2500000: episode: 676, duration: 24.341s, episode steps: 3325, steps per second: 137, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1471245/2500000: episode: 677, duration: 22.810s, episode steps: 3111, steps per second: 136, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1474695/2500000: episode: 678, duration: 25.480s, episode steps: 3450, steps per second: 135, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1477928/2500000: episode: 679, duration: 23.647s, episode steps: 3233, steps per second: 137, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean actior
1480284/2500000: episode: 680, duration: 17.443s, episode steps: 2356, steps per second: 135, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean actior
1483286/2500000: episode: 681, duration: 22.040s, episode steps: 3002, steps per second: 136, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1485303/2500000: episode: 682, duration: 15.072s, episode steps: 2017, steps per second: 134, episode reward: 17.000, mean reward: 0.008 [-1.000, 1.000], mean actior
1488080/2500000: episode: 683, duration: 20.615s, episode steps: 2777, steps per second: 135, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1491480/2500000: episode: 684, duration: 25.565s, episode steps: 3400, steps per second: 133, episode reward: 11.000, mean reward: 0.003 [-1.000, 1.000], mean actior
1494648/2500000: episode: 685, duration: 23.762s, episode steps: 3168, steps per second: 133, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean actior
1497781/2500000: episode: 686, duration: 23.620s, episode steps: 3133, steps per second: 133, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1501524/2500000: episode: 687, duration: 28.564s, episode steps: 3743, steps per second: 131, episode reward: -5.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1504754/2500000: episode: 688, duration: 24.411s, episode steps: 3230, steps per second: 132, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1507246/2500000: episode: 689, duration: 18.588s, episode steps: 2492, steps per second: 134, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean actior
1509936/2500000: episode: 690, duration: 20.456s, episode steps: 2690, steps per second: 132, episode reward: 15.000, mean reward: 0.006 [-1.000, 1.000], mean actior
1513244/2500000: episode: 691, duration: 24.816s, episode steps: 3308, steps per second: 133, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1516248/2500000: episode: 692, duration: 22.611s, episode steps: 3004, steps per second: 133, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1519477/2500000: episode: 693, duration: 24.300s, episode steps: 3229, steps per second: 133, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1523251/2500000: episode: 694, duration: 27.998s, episode steps: 3774, steps per second: 135, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1525575/2500000: episode: 695, duration: 17.935s, episode steps: 2324, steps per second: 130, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean actior
1529228/2500000: episode: 696, duration: 27.439s, episode steps: 3653, steps per second: 133, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1533804/2500000: episode: 697, duration: 34.262s, episode steps: 4576, steps per second: 134, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1537556/2500000: episode: 698, duration: 28.361s, episode steps: 3752, steps per second: 132, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1540792/2500000: episode: 699, duration: 24.176s, episode steps: 3236, steps per second: 134, episode reward: 14.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1543790/2500000: episode: 700, duration: 22.782s, episode steps: 2998, steps per second: 132, episode reward: 13.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1546343/2500000: episode: 701, duration: 18.937s, episode steps: 2553, steps per second: 135, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean actior
1549231/2500000: episode: 702, duration: 21.571s, episode steps: 2888, steps per second: 134, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1551972/2500000: episode: 703, duration: 20.497s, episode steps: 2741, steps per second: 134, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1555489/2500000: episode: 704, duration: 26.272s, episode steps: 3517, steps per second: 134, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1558527/2500000: episode: 705, duration: 22.649s, episode steps: 3038, steps per second: 134, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1561745/2500000: episode: 706, duration: 24.212s, episode steps: 3218, steps per second: 133, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean actior
1564058/2500000: episode: 707, duration: 17.365s, episode steps: 2313, steps per second: 133, episode reward: 13.000, mean reward: 0.006 [-1.000, 1.000], mean actior
1566400/2500000: episode: 708, duration: 18.043s, episode steps: 2342, steps per second: 130, episode reward: 13.000, mean reward: 0.006 [-1.000, 1.000], mean actior
1569884/2500000: episode: 709, duration: 26.008s, episode steps: 3484, steps per second: 134, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1572500/2500000: episode: 710, duration: 19.640s, episode steps: 2616, steps per second: 133, episode reward: -11.000, mean reward: -0.004 [-1.000, 1.000], mean acti
1575366/2500000: episode: 711, duration: 21.572s, episode steps: 2866, steps per second: 133, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean actior
1577881/2500000: episode: 712, duration: 18.798s, episode steps: 2515, steps per second: 134, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean actior
1581732/2500000: episode: 713, duration: 28.642s, episode steps: 3851, steps per second: 134, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1585133/2500000: episode: 714, duration: 25.557s, episode steps: 3401, steps per second: 133, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1587253/2500000: episode: 715, duration: 15.841s, episode steps: 2120, steps per second: 134, episode reward: 18.000, mean reward: 0.008 [-1.000, 1.000], mean actior
1590222/2500000: episode: 716, duration: 22.162s, episode steps: 2969, steps per second: 134, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean actior
1593791/2500000: episode: 717, duration: 26.550s, episode steps: 3569, steps per second: 134, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1597146/2500000: episode: 718, duration: 24.760s, episode steps: 3355, steps per second: 135, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1599934/2500000: episode: 719, duration: 20.921s, episode steps: 2788, steps per second: 133, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1603601/2500000: episode: 720, duration: 27.351s, episode steps: 3667, steps per second: 134, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1607073/2500000: episode: 721, duration: 26.038s, episode steps: 3472, steps per second: 133, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1610752/2500000: episode: 722, duration: 27.725s, episode steps: 3679, steps per second: 133, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1613040/2500000: episode: 723, duration: 17.005s, episode steps: 2288, steps per second: 135, episode reward: 18.000, mean reward: 0.008 [-1.000, 1.000], mean actior
1616820/2500000: episode: 724, duration: 28.149s, episode steps: 3780, steps per second: 134, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1620119/2500000: episode: 725, duration: 24.583s, episode steps: 3299, steps per second: 134, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1622589/2500000: episode: 726, duration: 18.515s, episode steps: 2470, steps per second: 133, episode reward: 15.000, mean reward: 0.006 [-1.000, 1.000], mean actior
1625616/2500000: episode: 727, duration: 22.522s, episode steps: 3027, steps per second: 134, episode reward: 14.000, mean reward: 0.005 [-1.000, 1.000], mean actior
1628380/2500000: episode: 728, duration: 20.644s, episode steps: 2764, steps per second: 134, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1631218/2500000: episode: 729, duration: 20.976s, episode steps: 2838, steps per second: 135, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1634051/2500000: episode: 730, duration: 20.796s, episode steps: 2833, steps per second: 136, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean actior
1637402/2500000: episode: 731, duration: 24.349s, episode steps: 3351, steps per second: 138, episode reward: 11.000, mean reward: 0.003 [-1.000, 1.000], mean actior
1640732/2500000: episode: 732, duration: 25.017s, episode steps: 3330, steps per second: 133, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1644219/2500000: episode: 733, duration: 26.059s, episode steps: 3487, steps per second: 134, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1646888/2500000: episode: 734, duration: 19.579s, episode steps: 2669, steps per second: 136, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean actior
1650327/2500000: episode: 735, duration: 25.816s, episode steps: 3439, steps per second: 133, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1653446/2500000: episode: 736, duration: 22.676s, episode steps: 3119, steps per second: 138, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean actior
1657205/2500000: episode: 737, duration: 27.771s, episode steps: 3759, steps per second: 135, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1659428/2500000: episode: 738, duration: 16.250s, episode steps: 2223, steps per second: 137, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean actior
1664132/2500000: episode: 739, duration: 34.775s, episode steps: 4704, steps per second: 135, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1668354/2500000: episode: 740, duration: 30.814s, episode steps: 4222, steps per second: 137, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1672228/2500000: episode: 741, duration: 28.209s, episode steps: 3874, steps per second: 137, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actic
1674824/2500000: episode: 742, duration: 18.770s, episode steps: 2596, steps per second: 138, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean actior
1678253/2500000: episode: 743, duration: 24.535s, episode steps: 3429, steps per second: 140, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1680946/2500000: episode: 744, duration: 19.258s, episode steps: 2693, steps per second: 140, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1683564/2500000: episode: 745, duration: 18.885s, episode steps: 2618, steps per second: 139, episode reward: 14.000, mean reward: 0.005 [-1.000, 1.000], mean actior
1686486/2500000: episode: 746, duration: 20.830s, episode steps: 2922, steps per second: 140, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1689488/2500000: episode: 747, duration: 21.838s, episode steps: 3002, steps per second: 137, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1691863/2500000: episode: 748, duration: 17.266s, episode steps: 2375, steps per second: 138, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1693932/2500000: episode: 749, duration: 15.190s, episode steps: 2069, steps per second: 136, episode reward: 16.000, mean reward: 0.008 [-1.000, 1.000], mean actior
1696838/2500000: episode: 750, duration: 21.270s, episode steps: 2906, steps per second: 137, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1699992/2500000: episode: 751, duration: 23.359s, episode steps: 3154, steps per second: 135, episode reward: 2.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1702830/2500000: episode: 752, duration: 20.900s, episode steps: 2838, steps per second: 136, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1707086/2500000: episode: 753, duration: 31.840s, episode steps: 4256, steps per second: 134, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1709432/2500000: episode: 754, duration: 17.617s, episode steps: 2346, steps per second: 133, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean actior
1712063/2500000: episode: 755, duration: 19.580s, episode steps: 2631, steps per second: 134, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1715059/2500000: episode: 756, duration: 22.275s, episode steps: 2996, steps per second: 135, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean actior
1717826/2500000: episode: 757, duration: 20.553s, episode steps: 2767, steps per second: 135, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean actior
1720040/2500000: episode: 758, duration: 16.522s, episode steps: 2214, steps per second: 134, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean actior
1722809/2500000: episode: 759, duration: 20.523s, episode steps: 2769, steps per second: 135, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1724970/2500000: episode: 760, duration: 16.063s, episode steps: 2161, steps per second: 135, episode reward: 17.000, mean reward: 0.008 [-1.000, 1.000], mean actior
1728896/2500000: episode: 761, duration: 28.984s, episode steps: 3926, steps per second: 135, episode reward: -2.000, mean reward: -0.001 [-1.000, 1.000], mean actic
1731375/2500000: episode: 762, duration: 18.300s, episode steps: 2479, steps per second: 135, episode reward: 16.000, mean reward: 0.006 [-1.000, 1.000], mean actior
```
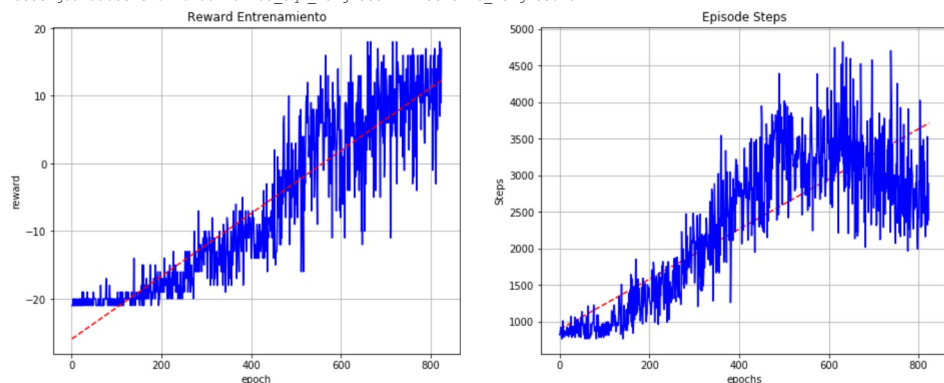
```
1734049/2500000: episode: 763, duration: 20.189s, episode steps: 2674, steps per second: 132, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
1737492/2500000: episode: 764, duration: 25.502s, episode steps: 3443, steps per second: 135, episode reward: 1.000, mean reward: 0.000 [-1.000, 1.000], mean action:
1740094/2500000: episode: 765, duration: 19.420s, episode steps: 2602, steps per second: 134, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1742533/2500000: episode: 766, duration: 18.094s, episode steps: 2439, steps per second: 135, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean action
1745136/2500000: episode: 767, duration: 19.389s, episode steps: 2603, steps per second: 134, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1748832/2500000: episode: 768, duration: 27.420s, episode steps: 3696, steps per second: 135, episode reward: -4.000, mean reward: -0.001 [-1.000, 1.000], mean actio
1751008/2500000: episode: 769, duration: 15.953s, episode steps: 2176, steps per second: 136, episode reward: 13.000, mean reward: 0.006 [-1.000, 1.000], mean action
1753529/2500000: episode: 770, duration: 18.671s, episode steps: 2521, steps per second: 135, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1755915/2500000: episode: 771, duration: 17.904s, episode steps: 2386, steps per second: 133, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean action
1758869/2500000: episode: 772, duration: 21.801s, episode steps: 2954, steps per second: 135, episode reward: 13.000, mean reward: 0.004 [-1.000, 1.000], mean action
1761225/2500000: episode: 773, duration: 17.494s, episode steps: 2356, steps per second: 135, episode reward: 13.000, mean reward: 0.006 [-1.000, 1.000], mean action
1763878/2500000: episode: 774, duration: 19.659s, episode steps: 2653, steps per second: 135, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean action
1766606/2500000: episode: 775, duration: 20.259s, episode steps: 2728, steps per second: 135, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1769859/2500000: episode: 776, duration: 24.506s, episode steps: 3253, steps per second: 133, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1771832/2500000: episode: 777, duration: 14.625s, episode steps: 1964, steps per second: 134, episode reward: 16.000, mean reward: 0.008 [-1.000, 1.000], mean action
1775729/2500000: episode: 778, duration: 28.937s, episode steps: 3906, steps per second: 135, episode reward: 5.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1778219/2500000: episode: 779, duration: 18.541s, episode steps: 2490, steps per second: 134, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
1780487/2500000: episode: 780, duration: 16.827s, episode steps: 2268, steps per second: 135, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean action
1783580/2500000: episode: 781, duration: 22.926s, episode steps: 3093, steps per second: 135, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1785731/2500000: episode: 782, duration: 16.013s, episode steps: 2151, steps per second: 134, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean action
1788264/2500000: episode: 783, duration: 19.226s, episode steps: 2533, steps per second: 132, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
1791097/2500000: episode: 784, duration: 20.909s, episode steps: 2833, steps per second: 135, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1794449/2500000: episode: 785, duration: 25.008s, episode steps: 3352, steps per second: 134, episode reward: 4.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1797542/2500000: episode: 786, duration: 23.038s, episode steps: 3093, steps per second: 134, episode reward: 3.000, mean reward: 0.001 [-1.000, 1.000], mean action:
1800557/2500000: episode: 787, duration: 22.543s, episode steps: 3015, steps per second: 134, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1802984/2500000: episode: 788, duration: 18.125s, episode steps: 2427, steps per second: 134, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
1805252/2500000: episode: 789, duration: 16.876s, episode steps: 2268, steps per second: 134, episode reward: 17.000, mean reward: 0.007 [-1.000, 1.000], mean action
1807785/2500000: episode: 790, duration: 18.720s, episode steps: 2533, steps per second: 135, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean action
1810605/2500000: episode: 791, duration: 21.103s, episode steps: 2820, steps per second: 134, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1813139/2500000: episode: 792, duration: 18.949s, episode steps: 2534, steps per second: 134, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
1815787/2500000: episode: 793, duration: 19.866s, episode steps: 2648, steps per second: 133, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean action
1818354/2500000: episode: 794, duration: 19.514s, episode steps: 2567, steps per second: 132, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1821252/2500000: episode: 795, duration: 21.610s, episode steps: 2898, steps per second: 134, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1823823/2500000: episode: 796, duration: 18.986s, episode steps: 2571, steps per second: 135, episode reward: 10.000, mean reward: 0.004 [-1.000, 1.000], mean action
1826962/2500000: episode: 797, duration: 23.491s, episode steps: 3139, steps per second: 134, episode reward: 8.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1829131/2500000: episode: 798, duration: 16.356s, episode steps: 2169, steps per second: 133, episode reward: 15.000, mean reward: 0.007 [-1.000, 1.000], mean action
1831123/2500000: episode: 799, duration: 14.855s, episode steps: 1992, steps per second: 134, episode reward: 16.000, mean reward: 0.008 [-1.000, 1.000], mean action
1833538/2500000: episode: 800, duration: 18.042s, episode steps: 2415, steps per second: 134, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean action
1835954/2500000: episode: 801, duration: 18.135s, episode steps: 2416, steps per second: 133, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
1838873/2500000: episode: 802, duration: 21.748s, episode steps: 2919, steps per second: 134, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1841785/2500000: episode: 803, duration: 21.863s, episode steps: 2912, steps per second: 133, episode reward: 11.000, mean reward: 0.004 [-1.000, 1.000], mean action
1845810/2500000: episode: 804, duration: 30.020s, episode steps: 4025, steps per second: 134, episode reward: -1.000, mean reward: -0.000 [-1.000, 1.000], mean actio
1849151/2500000: episode: 805, duration: 24.995s, episode steps: 3341, steps per second: 134, episode reward: 1.000, mean reward: 0 [-1.000, 1.000], mean action:
1851438/2500000: episode: 806, duration: 17.207s, episode steps: 2287, steps per second: 133, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean action
1854839/2500000: episode: 807, duration: 25.476s, episode steps: 3401, steps per second: 133, episode reward: 8.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1857153/2500000: episode: 808, duration: 17.364s, episode steps: 2314, steps per second: 133, episode reward: 17.000, mean reward: 0.007 [-1.000, 1.000], mean action
1860016/2500000: episode: 809, duration: 21.894s, episode steps: 2863, steps per second: 131, episode reward: 10.000, mean reward: 0.003 [-1.000, 1.000], mean action
1862659/2500000: episode: 810, duration: 19.696s, episode steps: 2643, steps per second: 134, episode reward: 6.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1866145/2500000: episode: 811, duration: 26.075s, episode steps: 3486, steps per second: 134, episode reward: -3.000, mean reward: -0.001 [-1.000, 1.000], mean actio
1868627/2500000: episode: 812, duration: 18.725s, episode steps: 2482, steps per second: 133, episode reward: 13.000, mean reward: 0.005 [-1.000, 1.000], mean action
1871128/2500000: episode: 813, duration: 18.752s, episode steps: 2501, steps per second: 133, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
1874319/2500000: episode: 814, duration: 23.803s, episode steps: 3191, steps per second: 134, episode reward: 5.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1876648/2500000: episode: 815, duration: 17.074s, episode steps: 2329, steps per second: 136, episode reward: 14.000, mean reward: 0.006 [-1.000, 1.000], mean action
1878808/2500000: episode: 816, duration: 16.100s, episode steps: 2160, steps per second: 134, episode reward: 16.000, mean reward: 0.007 [-1.000, 1.000], mean action
1881117/2500000: episode: 817, duration: 17.227s, episode steps: 2309, steps per second: 134, episode reward: 12.000, mean reward: 0.005 [-1.000, 1.000], mean action
1883826/2500000: episode: 818, duration: 19.992s, episode steps: 2709, steps per second: 136, episode reward: 12.000, mean reward: 0.004 [-1.000, 1.000], mean action
1886158/2500000: episode: 819, duration: 17.140s, episode steps: 2332, steps per second: 136, episode reward: 18.000, mean reward: 0.008 [-1.000, 1.000], mean action
1889687/2500000: episode: 820, duration: 25.968s, episode steps: 3529, steps per second: 136, episode reward: 7.000, mean reward: 0.002 [-1.000, 1.000], mean action:
1892011/2500000: episode: 821, duration: 17.210s, episode steps: 2324, steps per second: 135, episode reward: 15.000, mean reward: 0.006 [-1.000, 1.000], mean action
1894905/2500000: episode: 822, duration: 21.257s, episode steps: 2894, steps per second: 136, episode reward: 9.000, mean reward: 0.003 [-1.000, 1.000], mean action:
1897294/2500000: episode: 823, duration: 17.323s, episode steps: 2389, steps per second: 138, episode reward: 17.000, mean reward: 0.007 [-1.000, 1.000], mean action
done, took 14076.594 seconds
Pesos guardados en: 202001131203_dqn_PongDeterministic-v0_weights.h5f
```


Reward Entrenamiento / Episode Steps

```
Testing for 10 episodes ...
Episode 1: reward: 11.000, steps: 2666
Episode 2: reward: 10.000, steps: 2934
Episode 3: reward: 12.000, steps: 2405
Episode 4: reward: 14.000, steps: 2438
Episode 5: reward: 10.000, steps: 2849
Episode 6: reward: 12.000, steps: 2568
Episode 7: reward: 13.000, steps: 2623
Episode 8: reward: 11.000, steps: 2557
Episode 9: reward: 7.000, steps: 3097
Episode 10: reward: 17.000, steps: 2148
mean_reward: 11.7
mean_steps: 2628.5
(11.7, 2628.5)
```