

MP5_DQNLSTM_EC

May 1, 2024

1 Creating a DQN LSTM Agent

```
[ ]: %matplotlib inline

import sys
import gym
import torch
import pylab
import random
import numpy as np
from collections import deque
from datetime import datetime
from copy import deepcopy
import torch.nn as nn
import torch.optim as optim
import torch.nn.functional as F
from torch.autograd import Variable
from utils import find_max_lives, check_live, get_frame, get_init_state
from model import DQN, DQN_LSTM
from config import *

import matplotlib.pyplot as plt
%load_ext autoreload
%autoreload 2
```

The autoreload extension is already loaded. To reload it, use:

```
%reload_ext autoreload
```

```
[ ]: env = gym.make('BreakoutDeterministic-v4')
state = env.reset()
```

A.L.E: Arcade Learning Environment (version 0.8.1+53f58b7)

[Powered by Stella]

```
[ ]: number_lives = find_max_lives(env)
state_size = env.observation_space.shape
action_size = 3 #fire, left, and right
```

Now we will create a DQN agent that uses LSTM rather than past frames as history. We augment the experience replay to contain previous few (state, action, reward, next state) tuples rather than just one (state, action, reward, next state) tuple so it can work with LSTMs. Use the previous tuples to generate the current hidden and context vector for LSTM. Essentially, when you get a sample from replay buffer during training, start with the first tuple and generate hidden and context vector from this and pass it onto the next tuple. Do so consequitively till you reach the last tuple, where you will make Q value predictions. Training loop remains nearly the same.

```
[ ]: from agent_lstm import Agent_LSTM
agent = Agent_LSTM(action_size)
evaluation_reward = deque(maxlen=evaluation_reward_length)
frame = 0
memory_size = 0
```

```
[ ]: EPISODES = 20000
print(EPISODES)
```

20000

```
[ ]: # num_layer = 3
HISTORY_SIZE = 1
rewards, episodes = [], []
best_eval_reward = 0
halt_flag = 0

for e in range(EPISODES):
    done = False
    score = 0

    history = np.zeros([HISTORY_SIZE + 1, 84, 84], dtype=np.uint8)
    step = 0
    state, _ = env.reset()
    next_state = state
    life = number_lives
    hidden = None

    get_init_state(history, state, HISTORY_SIZE)

    while not done:
        step += 1
        frame += 1

        # Perform a fire action if ball is no longer on screen to continue onto
        ↪ next life
        if step > 1 and len(np.unique(next_state[:189] == state[:189])) < 2:
            action = torch.tensor([[0]]).cuda()
        else:
```

```

        action, hidden = agent.get_action(np.float32(history[:1, :, :]) /
↪255., hidden)
        state = next_state
        next_state, reward, terminated, truncated, info = env.step(action + 1)
        done = terminated or truncated

        frame_next_state = get_frame(next_state)
        history[1, :, :] = frame_next_state
        terminal_state = check_live(life, info['lives'])

        life = info['lives']
        r = reward

        # Store the transition in memory
        agent.memory.push(deepcopy(frame_next_state), action.cpu(), r,
↪terminal_state)
        # Start training after random sample generation
        if(frame >= train_frame):
            agent.train_policy_net(frame)
            score += reward
            history[:1, :, :] = history[1:, :, :]

        if done:
            evaluation_reward.append(score)
            rewards.append(np.mean(evaluation_reward))
            episodes.append(e)
            pylab.plot(episodes, rewards, 'b')
            pylab.xlabel('Episodes')
            pylab.ylabel('Rewards')
            pylab.title('Episodes vs Reward')
            pylab.savefig("./save_graph/breakout_dqn_lstm.png") # save graph
↪for training visualization

            # every episode, plot the play time
            print("episode:", e, "  score:", score, "  memory length:",
                  len(agent.memory), "  epsilon:", agent.epsilon, "  steps:",
↪step,
                  "  lr:", agent.optimizer.param_groups[0]['lr'], "  "
↪evaluation reward:", np.mean(evaluation_reward))

            # if the mean of scores of last 100 episode is bigger than 8 save
↪model

            ### Change this save condition to whatever you prefer ###
            if np.mean(evaluation_reward) > 8.0 and np.mean(evaluation_reward)
↪ best_eval_reward:

```

```

    torch.save(agent.policy_net, "./save_model/breakout_dqn_lstm.
pth")

    best_eval_reward = np.mean(evaluation_reward)
    if np.mean(evaluation_reward) > 8.2:
        halt_flag = 1
        break
if halt_flag == 1:
    break

```

```

episode: 0    score: 0.0    memory length: 123    epsilon: 1.0    steps: 123    lr:
0.0001    evaluation reward: 0.0
episode: 1    score: 2.0    memory length: 321    epsilon: 1.0    steps: 198    lr:
0.0001    evaluation reward: 1.0
episode: 2    score: 0.0    memory length: 444    epsilon: 1.0    steps: 123    lr:
0.0001    evaluation reward: 0.6666666666666666
episode: 3    score: 3.0    memory length: 709    epsilon: 1.0    steps: 265    lr:
0.0001    evaluation reward: 1.25
episode: 4    score: 4.0    memory length: 968    epsilon: 1.0    steps: 259    lr:
0.0001    evaluation reward: 1.8
episode: 5    score: 0.0    memory length: 1091    epsilon: 1.0    steps: 123
lr: 0.0001    evaluation reward: 1.5
episode: 6    score: 1.0    memory length: 1261    epsilon: 1.0    steps: 170
lr: 0.0001    evaluation reward: 1.4285714285714286
episode: 7    score: 3.0    memory length: 1507    epsilon: 1.0    steps: 246
lr: 0.0001    evaluation reward: 1.625
episode: 8    score: 1.0    memory length: 1678    epsilon: 1.0    steps: 171
lr: 0.0001    evaluation reward: 1.5555555555555556
episode: 9    score: 0.0    memory length: 1801    epsilon: 1.0    steps: 123
lr: 0.0001    evaluation reward: 1.4
episode: 10   score: 1.0    memory length: 1954    epsilon: 1.0    steps: 153
lr: 0.0001    evaluation reward: 1.3636363636363635
episode: 11   score: 0.0    memory length: 2077    epsilon: 1.0    steps: 123
lr: 0.0001    evaluation reward: 1.25
episode: 12   score: 0.0    memory length: 2199    epsilon: 1.0    steps: 122
lr: 0.0001    evaluation reward: 1.1538461538461537
episode: 13   score: 3.0    memory length: 2464    epsilon: 1.0    steps: 265
lr: 0.0001    evaluation reward: 1.2857142857142858
episode: 14   score: 0.0    memory length: 2587    epsilon: 1.0    steps: 123
lr: 0.0001    evaluation reward: 1.2
episode: 15   score: 1.0    memory length: 2738    epsilon: 1.0    steps: 151
lr: 0.0001    evaluation reward: 1.1875
episode: 16   score: 2.0    memory length: 2937    epsilon: 1.0    steps: 199
lr: 0.0001    evaluation reward: 1.2352941176470589
episode: 17   score: 1.0    memory length: 3108    epsilon: 1.0    steps: 171
lr: 0.0001    evaluation reward: 1.2222222222222223
episode: 18   score: 2.0    memory length: 3306    epsilon: 1.0    steps: 198
lr: 0.0001    evaluation reward: 1.263157894736842

```

| | | | | |
|-------------|---------------------------------------|---------------------|--------------|------------|
| episode: 19 | score: 0.0 | memory length: 3429 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.2 | | | |
| episode: 20 | score: 1.0 | memory length: 3598 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.1904761904761905 | | | |
| episode: 21 | score: 0.0 | memory length: 3721 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.1363636363636365 | | | |
| episode: 22 | score: 3.0 | memory length: 3946 | epsilon: 1.0 | steps: 225 |
| lr: 0.0001 | evaluation reward: 1.2173913043478262 | | | |
| episode: 23 | score: 2.0 | memory length: 4164 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.25 | | | |
| episode: 24 | score: 0.0 | memory length: 4287 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.2 | | | |
| episode: 25 | score: 2.0 | memory length: 4505 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.2307692307692308 | | | |
| episode: 26 | score: 2.0 | memory length: 4689 | epsilon: 1.0 | steps: 184 |
| lr: 0.0001 | evaluation reward: 1.2592592592592593 | | | |
| episode: 27 | score: 1.0 | memory length: 4858 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.25 | | | |
| episode: 28 | score: 4.0 | memory length: 5176 | epsilon: 1.0 | steps: 318 |
| lr: 0.0001 | evaluation reward: 1.3448275862068966 | | | |
| episode: 29 | score: 2.0 | memory length: 5393 | epsilon: 1.0 | steps: 217 |
| lr: 0.0001 | evaluation reward: 1.3666666666666667 | | | |
| episode: 30 | score: 1.0 | memory length: 5544 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.3548387096774193 | | | |
| episode: 31 | score: 0.0 | memory length: 5666 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.3125 | | | |
| episode: 32 | score: 4.0 | memory length: 5943 | epsilon: 1.0 | steps: 277 |
| lr: 0.0001 | evaluation reward: 1.393939393939394 | | | |
| episode: 33 | score: 0.0 | memory length: 6066 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3529411764705883 | | | |
| episode: 34 | score: 2.0 | memory length: 6263 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.3714285714285714 | | | |
| episode: 35 | score: 1.0 | memory length: 6432 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.3611111111111112 | | | |
| episode: 36 | score: 3.0 | memory length: 6659 | epsilon: 1.0 | steps: 227 |
| lr: 0.0001 | evaluation reward: 1.4054054054054055 | | | |
| episode: 37 | score: 0.0 | memory length: 6782 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.368421052631579 | | | |
| episode: 38 | score: 3.0 | memory length: 7033 | epsilon: 1.0 | steps: 251 |
| lr: 0.0001 | evaluation reward: 1.4102564102564104 | | | |
| episode: 39 | score: 1.0 | memory length: 7184 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 40 | score: 3.0 | memory length: 7451 | epsilon: 1.0 | steps: 267 |
| lr: 0.0001 | evaluation reward: 1.4390243902439024 | | | |
| episode: 41 | score: 1.0 | memory length: 7622 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.4285714285714286 | | | |
| episode: 42 | score: 1.0 | memory length: 7773 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.4186046511627908 | | | |

| | | | | |
|-------------|---------------------------------------|----------------------|--------------|------------|
| episode: 43 | score: 2.0 | memory length: 7991 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.4318181818181819 | | | |
| episode: 44 | score: 0.0 | memory length: 8114 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 45 | score: 0.0 | memory length: 8237 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3695652173913044 | | | |
| episode: 46 | score: 0.0 | memory length: 8360 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3404255319148937 | | | |
| episode: 47 | score: 4.0 | memory length: 8656 | epsilon: 1.0 | steps: 296 |
| lr: 0.0001 | evaluation reward: 1.3958333333333333 | | | |
| episode: 48 | score: 0.0 | memory length: 8778 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.3673469387755102 | | | |
| episode: 49 | score: 1.0 | memory length: 8929 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 50 | score: 3.0 | memory length: 9194 | epsilon: 1.0 | steps: 265 |
| lr: 0.0001 | evaluation reward: 1.392156862745098 | | | |
| episode: 51 | score: 1.0 | memory length: 9363 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.3846153846153846 | | | |
| episode: 52 | score: 0.0 | memory length: 9486 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3584905660377358 | | | |
| episode: 53 | score: 0.0 | memory length: 9609 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3333333333333333 | | | |
| episode: 54 | score: 3.0 | memory length: 9842 | epsilon: 1.0 | steps: 233 |
| lr: 0.0001 | evaluation reward: 1.3636363636363635 | | | |
| episode: 55 | score: 2.0 | memory length: 10040 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.375 | | | |
| episode: 56 | score: 2.0 | memory length: 10240 | epsilon: 1.0 | steps: 200 |
| lr: 0.0001 | evaluation reward: 1.3859649122807018 | | | |
| episode: 57 | score: 0.0 | memory length: 10363 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3620689655172413 | | | |
| episode: 58 | score: 1.0 | memory length: 10531 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.3559322033898304 | | | |
| episode: 59 | score: 0.0 | memory length: 10654 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3333333333333333 | | | |
| episode: 60 | score: 0.0 | memory length: 10777 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3114754098360655 | | | |
| episode: 61 | score: 2.0 | memory length: 10997 | epsilon: 1.0 | steps: 220 |
| lr: 0.0001 | evaluation reward: 1.3225806451612903 | | | |
| episode: 62 | score: 0.0 | memory length: 11120 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3015873015873016 | | | |
| episode: 63 | score: 1.0 | memory length: 11288 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.296875 | | | |
| episode: 64 | score: 4.0 | memory length: 11563 | epsilon: 1.0 | steps: 275 |
| lr: 0.0001 | evaluation reward: 1.3384615384615384 | | | |
| episode: 65 | score: 2.0 | memory length: 11761 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.3484848484848484 | | | |
| episode: 66 | score: 4.0 | memory length: 12075 | epsilon: 1.0 | steps: 314 |
| lr: 0.0001 | evaluation reward: 1.3880597014925373 | | | |

| | | | | |
|-------------|---------------------------------------|----------------------|--------------|------------|
| episode: 67 | score: 2.0 | memory length: 12274 | epsilon: 1.0 | steps: 199 |
| lr: 0.0001 | evaluation reward: 1.3970588235294117 | | | |
| episode: 68 | score: 4.0 | memory length: 12528 | epsilon: 1.0 | steps: 254 |
| lr: 0.0001 | evaluation reward: 1.434782608695652 | | | |
| episode: 69 | score: 2.0 | memory length: 12708 | epsilon: 1.0 | steps: 180 |
| lr: 0.0001 | evaluation reward: 1.4428571428571428 | | | |
| episode: 70 | score: 2.0 | memory length: 12907 | epsilon: 1.0 | steps: 199 |
| lr: 0.0001 | evaluation reward: 1.4507042253521127 | | | |
| episode: 71 | score: 1.0 | memory length: 13075 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.4444444444444444 | | | |
| episode: 72 | score: 4.0 | memory length: 13340 | epsilon: 1.0 | steps: 265 |
| lr: 0.0001 | evaluation reward: 1.4794520547945205 | | | |
| episode: 73 | score: 2.0 | memory length: 13538 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.4864864864864864 | | | |
| episode: 74 | score: 0.0 | memory length: 13661 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4666666666666666 | | | |
| episode: 75 | score: 2.0 | memory length: 13862 | epsilon: 1.0 | steps: 201 |
| lr: 0.0001 | evaluation reward: 1.4736842105263157 | | | |
| episode: 76 | score: 0.0 | memory length: 13985 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4545454545454546 | | | |
| episode: 77 | score: 0.0 | memory length: 14107 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.435897435897436 | | | |
| episode: 78 | score: 1.0 | memory length: 14276 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.4303797468354431 | | | |
| episode: 79 | score: 2.0 | memory length: 14495 | epsilon: 1.0 | steps: 219 |
| lr: 0.0001 | evaluation reward: 1.4375 | | | |
| episode: 80 | score: 1.0 | memory length: 14647 | epsilon: 1.0 | steps: 152 |
| lr: 0.0001 | evaluation reward: 1.4320987654320987 | | | |
| episode: 81 | score: 0.0 | memory length: 14770 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4146341463414633 | | | |
| episode: 82 | score: 3.0 | memory length: 14995 | epsilon: 1.0 | steps: 225 |
| lr: 0.0001 | evaluation reward: 1.4337349397590362 | | | |
| episode: 83 | score: 1.0 | memory length: 15165 | epsilon: 1.0 | steps: 170 |
| lr: 0.0001 | evaluation reward: 1.4285714285714286 | | | |
| episode: 84 | score: 0.0 | memory length: 15288 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.411764705882353 | | | |
| episode: 85 | score: 0.0 | memory length: 15410 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.3953488372093024 | | | |
| episode: 86 | score: 2.0 | memory length: 15607 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.4022988505747127 | | | |
| episode: 87 | score: 1.0 | memory length: 15775 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.3977272727272727 | | | |
| episode: 88 | score: 2.0 | memory length: 15995 | epsilon: 1.0 | steps: 220 |
| lr: 0.0001 | evaluation reward: 1.404494382022472 | | | |
| episode: 89 | score: 0.0 | memory length: 16117 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.3888888888888888 | | | |
| episode: 90 | score: 1.0 | memory length: 16286 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.3846153846153846 | | | |

| | | | | |
|--------------|---------------------------------------|----------------------|--------------|------------|
| episode: 91 | score: 2.0 | memory length: 16483 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.391304347826087 | | | |
| episode: 92 | score: 2.0 | memory length: 16684 | epsilon: 1.0 | steps: 201 |
| lr: 0.0001 | evaluation reward: 1.3978494623655915 | | | |
| episode: 93 | score: 4.0 | memory length: 16954 | epsilon: 1.0 | steps: 270 |
| lr: 0.0001 | evaluation reward: 1.425531914893617 | | | |
| episode: 94 | score: 0.0 | memory length: 17077 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4105263157894736 | | | |
| episode: 95 | score: 0.0 | memory length: 17200 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3958333333333333 | | | |
| episode: 96 | score: 1.0 | memory length: 17351 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.3917525773195876 | | | |
| episode: 97 | score: 3.0 | memory length: 17577 | epsilon: 1.0 | steps: 226 |
| lr: 0.0001 | evaluation reward: 1.4081632653061225 | | | |
| episode: 98 | score: 2.0 | memory length: 17775 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.4141414141414141 | | | |
| episode: 99 | score: 1.0 | memory length: 17945 | epsilon: 1.0 | steps: 170 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 100 | score: 0.0 | memory length: 18068 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 101 | score: 3.0 | memory length: 18317 | epsilon: 1.0 | steps: 249 |
| lr: 0.0001 | evaluation reward: 1.42 | | | |
| episode: 102 | score: 1.0 | memory length: 18489 | epsilon: 1.0 | steps: 172 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 103 | score: 0.0 | memory length: 18611 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 104 | score: 3.0 | memory length: 18878 | epsilon: 1.0 | steps: 267 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |
| episode: 105 | score: 2.0 | memory length: 19059 | epsilon: 1.0 | steps: 181 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 106 | score: 0.0 | memory length: 19182 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 107 | score: 0.0 | memory length: 19304 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.37 | | | |
| episode: 108 | score: 2.0 | memory length: 19502 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.38 | | | |
| episode: 109 | score: 1.0 | memory length: 19672 | epsilon: 1.0 | steps: 170 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |
| episode: 110 | score: 0.0 | memory length: 19795 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.38 | | | |
| episode: 111 | score: 3.0 | memory length: 20045 | epsilon: 1.0 | steps: 250 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 112 | score: 0.0 | memory length: 20168 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 113 | score: 0.0 | memory length: 20290 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.38 | | | |
| episode: 114 | score: 1.0 | memory length: 20461 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 115 | score: 2.0 | memory length: 20659 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 116 | score: 3.0 | memory length: 20928 | epsilon: 1.0 | steps: 269 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 117 | score: 3.0 | memory length: 21174 | epsilon: 1.0 | steps: 246 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 118 | score: 0.0 | memory length: 21297 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 119 | score: 0.0 | memory length: 21420 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 120 | score: 0.0 | memory length: 21543 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 121 | score: 1.0 | memory length: 21714 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 122 | score: 3.0 | memory length: 21947 | epsilon: 1.0 | steps: 233 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 123 | score: 0.0 | memory length: 22069 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |
| episode: 124 | score: 1.0 | memory length: 22237 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 125 | score: 2.0 | memory length: 22435 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 126 | score: 3.0 | memory length: 22661 | epsilon: 1.0 | steps: 226 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 127 | score: 2.0 | memory length: 22858 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.42 | | | |
| episode: 128 | score: 2.0 | memory length: 23074 | epsilon: 1.0 | steps: 216 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 129 | score: 2.0 | memory length: 23271 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 130 | score: 1.0 | memory length: 23423 | epsilon: 1.0 | steps: 152 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 131 | score: 4.0 | memory length: 23679 | epsilon: 1.0 | steps: 256 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 132 | score: 0.0 | memory length: 23802 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 133 | score: 2.0 | memory length: 23999 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.42 | | | |
| episode: 134 | score: 0.0 | memory length: 24122 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 135 | score: 4.0 | memory length: 24401 | epsilon: 1.0 | steps: 279 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 136 | score: 4.0 | memory length: 24714 | epsilon: 1.0 | steps: 313 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 137 | score: 0.0 | memory length: 24836 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 138 | score: 0.0 | memory length: 24959 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 139 | score: 3.0 | memory length: 25208 | epsilon: 1.0 | steps: 249 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 140 | score: 0.0 | memory length: 25331 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 141 | score: 0.0 | memory length: 25453 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |
| episode: 142 | score: 1.0 | memory length: 25622 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |
| episode: 143 | score: 1.0 | memory length: 25773 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.38 | | | |
| episode: 144 | score: 0.0 | memory length: 25895 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.38 | | | |
| episode: 145 | score: 2.0 | memory length: 26093 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 146 | score: 3.0 | memory length: 26338 | epsilon: 1.0 | steps: 245 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 147 | score: 0.0 | memory length: 26460 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |
| episode: 148 | score: 2.0 | memory length: 26660 | epsilon: 1.0 | steps: 200 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 149 | score: 4.0 | memory length: 26936 | epsilon: 1.0 | steps: 276 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 150 | score: 0.0 | memory length: 27059 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 151 | score: 4.0 | memory length: 27334 | epsilon: 1.0 | steps: 275 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 152 | score: 2.0 | memory length: 27552 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 153 | score: 1.0 | memory length: 27721 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 154 | score: 0.0 | memory length: 27844 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 155 | score: 2.0 | memory length: 28062 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 156 | score: 2.0 | memory length: 28259 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 157 | score: 3.0 | memory length: 28486 | epsilon: 1.0 | steps: 227 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 158 | score: 0.0 | memory length: 28609 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 159 | score: 0.0 | memory length: 28732 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 160 | score: 1.0 | memory length: 28901 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 161 | score: 7.0 | memory length: 29163 | epsilon: 1.0 | steps: 262 |
| lr: 0.0001 | evaluation reward: 1.52 | | | |
| episode: 162 | score: 1.0 | memory length: 29331 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.53 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 163 | score: 3.0 | memory length: 29576 | epsilon: 1.0 | steps: 245 |
| lr: 0.0001 | evaluation reward: 1.55 | | | |
| episode: 164 | score: 0.0 | memory length: 29699 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 165 | score: 0.0 | memory length: 29822 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 166 | score: 1.0 | memory length: 29972 | epsilon: 1.0 | steps: 150 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 167 | score: 2.0 | memory length: 30187 | epsilon: 1.0 | steps: 215 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 168 | score: 2.0 | memory length: 30404 | epsilon: 1.0 | steps: 217 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 169 | score: 1.0 | memory length: 30555 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 170 | score: 0.0 | memory length: 30678 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 171 | score: 0.0 | memory length: 30801 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 172 | score: 0.0 | memory length: 30924 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 173 | score: 2.0 | memory length: 31142 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 174 | score: 0.0 | memory length: 31265 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 175 | score: 0.0 | memory length: 31387 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.34 | | | |
| episode: 176 | score: 0.0 | memory length: 31510 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.34 | | | |
| episode: 177 | score: 1.0 | memory length: 31681 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 178 | score: 2.0 | memory length: 31878 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 179 | score: 0.0 | memory length: 32000 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.34 | | | |
| episode: 180 | score: 2.0 | memory length: 32200 | epsilon: 1.0 | steps: 200 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 181 | score: 1.0 | memory length: 32369 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 182 | score: 0.0 | memory length: 32491 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 183 | score: 1.0 | memory length: 32660 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 184 | score: 2.0 | memory length: 32877 | epsilon: 1.0 | steps: 217 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 185 | score: 2.0 | memory length: 33075 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.37 | | | |
| episode: 186 | score: 0.0 | memory length: 33198 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 187 | score: 0.0 | memory length: 33321 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.34 | | | |
| episode: 188 | score: 0.0 | memory length: 33444 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.32 | | | |
| episode: 189 | score: 3.0 | memory length: 33690 | epsilon: 1.0 | steps: 246 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 190 | score: 2.0 | memory length: 33907 | epsilon: 1.0 | steps: 217 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 191 | score: 2.0 | memory length: 34105 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 192 | score: 2.0 | memory length: 34307 | epsilon: 1.0 | steps: 202 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 193 | score: 1.0 | memory length: 34460 | epsilon: 1.0 | steps: 153 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 194 | score: 4.0 | memory length: 34755 | epsilon: 1.0 | steps: 295 |
| lr: 0.0001 | evaluation reward: 1.37 | | | |
| episode: 195 | score: 0.0 | memory length: 34878 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.37 | | | |
| episode: 196 | score: 0.0 | memory length: 35001 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 197 | score: 2.0 | memory length: 35219 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 198 | score: 2.0 | memory length: 35399 | epsilon: 1.0 | steps: 180 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 199 | score: 1.0 | memory length: 35570 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 200 | score: 0.0 | memory length: 35693 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 201 | score: 3.0 | memory length: 35936 | epsilon: 1.0 | steps: 243 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 202 | score: 4.0 | memory length: 36230 | epsilon: 1.0 | steps: 294 |
| lr: 0.0001 | evaluation reward: 1.38 | | | |
| episode: 203 | score: 5.0 | memory length: 36527 | epsilon: 1.0 | steps: 297 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 204 | score: 1.0 | memory length: 36677 | epsilon: 1.0 | steps: 150 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 205 | score: 0.0 | memory length: 36799 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |
| episode: 206 | score: 0.0 | memory length: 36922 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |
| episode: 207 | score: 3.0 | memory length: 37148 | epsilon: 1.0 | steps: 226 |
| lr: 0.0001 | evaluation reward: 1.42 | | | |
| episode: 208 | score: 0.0 | memory length: 37271 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 209 | score: 2.0 | memory length: 37469 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 210 | score: 2.0 | memory length: 37690 | epsilon: 1.0 | steps: 221 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 211 | score: 1.0 | memory length: 37859 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 212 | score: 4.0 | memory length: 38134 | epsilon: 1.0 | steps: 275 |
| lr: 0.0001 | evaluation reward: 1.45 | | | |
| episode: 213 | score: 4.0 | memory length: 38447 | epsilon: 1.0 | steps: 313 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 214 | score: 1.0 | memory length: 38598 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 215 | score: 1.0 | memory length: 38767 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.48 | | | |
| episode: 216 | score: 0.0 | memory length: 38889 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.45 | | | |
| episode: 217 | score: 1.0 | memory length: 39060 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 218 | score: 0.0 | memory length: 39182 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 219 | score: 2.0 | memory length: 39399 | epsilon: 1.0 | steps: 217 |
| lr: 0.0001 | evaluation reward: 1.45 | | | |
| episode: 220 | score: 3.0 | memory length: 39648 | epsilon: 1.0 | steps: 249 |
| lr: 0.0001 | evaluation reward: 1.48 | | | |
| episode: 221 | score: 1.0 | memory length: 39799 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.48 | | | |
| episode: 222 | score: 2.0 | memory length: 40015 | epsilon: 1.0 | steps: 216 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 223 | score: 1.0 | memory length: 40187 | epsilon: 1.0 | steps: 172 |
| lr: 0.0001 | evaluation reward: 1.48 | | | |
| episode: 224 | score: 3.0 | memory length: 40434 | epsilon: 1.0 | steps: 247 |
| lr: 0.0001 | evaluation reward: 1.5 | | | |
| episode: 225 | score: 1.0 | memory length: 40603 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 226 | score: 3.0 | memory length: 40848 | epsilon: 1.0 | steps: 245 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 227 | score: 2.0 | memory length: 41046 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 228 | score: 0.0 | memory length: 41169 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 229 | score: 2.0 | memory length: 41367 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 230 | score: 2.0 | memory length: 41566 | epsilon: 1.0 | steps: 199 |
| lr: 0.0001 | evaluation reward: 1.48 | | | |
| episode: 231 | score: 1.0 | memory length: 41737 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.45 | | | |
| episode: 232 | score: 1.0 | memory length: 41906 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 233 | score: 2.0 | memory length: 42125 | epsilon: 1.0 | steps: 219 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 234 | score: 3.0 | memory length: 42390 | epsilon: 1.0 | steps: 265 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 235 | score: 0.0 | memory length: 42513 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.45 | | | |
| episode: 236 | score: 1.0 | memory length: 42663 | epsilon: 1.0 | steps: 150 |
| lr: 0.0001 | evaluation reward: 1.42 | | | |
| episode: 237 | score: 2.0 | memory length: 42881 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 238 | score: 0.0 | memory length: 43004 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 239 | score: 1.0 | memory length: 43172 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.42 | | | |
| episode: 240 | score: 0.0 | memory length: 43295 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.42 | | | |
| episode: 241 | score: 0.0 | memory length: 43418 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.42 | | | |
| episode: 242 | score: 0.0 | memory length: 43541 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.41 | | | |
| episode: 243 | score: 3.0 | memory length: 43767 | epsilon: 1.0 | steps: 226 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 244 | score: 3.0 | memory length: 44011 | epsilon: 1.0 | steps: 244 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 245 | score: 0.0 | memory length: 44134 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |
| episode: 246 | score: 2.0 | memory length: 44353 | epsilon: 1.0 | steps: 219 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 247 | score: 0.0 | memory length: 44475 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 248 | score: 1.0 | memory length: 44626 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.42 | | | |
| episode: 249 | score: 1.0 | memory length: 44794 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |
| episode: 250 | score: 1.0 | memory length: 44965 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 251 | score: 1.0 | memory length: 45117 | epsilon: 1.0 | steps: 152 |
| lr: 0.0001 | evaluation reward: 1.37 | | | |
| episode: 252 | score: 1.0 | memory length: 45268 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 253 | score: 0.0 | memory length: 45391 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 254 | score: 2.0 | memory length: 45610 | epsilon: 1.0 | steps: 219 |
| lr: 0.0001 | evaluation reward: 1.37 | | | |
| episode: 255 | score: 1.0 | memory length: 45761 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 256 | score: 1.0 | memory length: 45912 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 257 | score: 1.0 | memory length: 46063 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 258 | score: 0.0 | memory length: 46185 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 259 | score: 0.0 | memory length: 46308 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 260 | score: 1.0 | memory length: 46478 | epsilon: 1.0 | steps: 170 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 261 | score: 1.0 | memory length: 46647 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.27 | | | |
| episode: 262 | score: 1.0 | memory length: 46797 | epsilon: 1.0 | steps: 150 |
| lr: 0.0001 | evaluation reward: 1.27 | | | |
| episode: 263 | score: 0.0 | memory length: 46920 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.24 | | | |
| episode: 264 | score: 1.0 | memory length: 47088 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.25 | | | |
| episode: 265 | score: 2.0 | memory length: 47286 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.27 | | | |
| episode: 266 | score: 3.0 | memory length: 47558 | epsilon: 1.0 | steps: 272 |
| lr: 0.0001 | evaluation reward: 1.29 | | | |
| episode: 267 | score: 0.0 | memory length: 47680 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.27 | | | |
| episode: 268 | score: 4.0 | memory length: 47977 | epsilon: 1.0 | steps: 297 |
| lr: 0.0001 | evaluation reward: 1.29 | | | |
| episode: 269 | score: 1.0 | memory length: 48128 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.29 | | | |
| episode: 270 | score: 0.0 | memory length: 48250 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.29 | | | |
| episode: 271 | score: 1.0 | memory length: 48418 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.3 | | | |
| episode: 272 | score: 0.0 | memory length: 48540 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.3 | | | |
| episode: 273 | score: 2.0 | memory length: 48722 | epsilon: 1.0 | steps: 182 |
| lr: 0.0001 | evaluation reward: 1.3 | | | |
| episode: 274 | score: 3.0 | memory length: 48953 | epsilon: 1.0 | steps: 231 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 275 | score: 1.0 | memory length: 49103 | epsilon: 1.0 | steps: 150 |
| lr: 0.0001 | evaluation reward: 1.34 | | | |
| episode: 276 | score: 3.0 | memory length: 49350 | epsilon: 1.0 | steps: 247 |
| lr: 0.0001 | evaluation reward: 1.37 | | | |
| episode: 277 | score: 0.0 | memory length: 49473 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 278 | score: 2.0 | memory length: 49671 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 279 | score: 0.0 | memory length: 49794 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 280 | score: 1.0 | memory length: 49963 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 281 | score: 4.0 | memory length: 50249 | epsilon: 1.0 | steps: 286 |
| lr: 0.0001 | evaluation reward: 1.38 | | | |
| episode: 282 | score: 0.0 | memory length: 50371 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.38 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 283 | score: 1.0 | memory length: 50521 | epsilon: 1.0 | steps: 150 |
| lr: 0.0001 | evaluation reward: 1.38 | | | |
| episode: 284 | score: 0.0 | memory length: 50644 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 285 | score: 2.0 | memory length: 50842 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 286 | score: 1.0 | memory length: 50993 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.37 | | | |
| episode: 287 | score: 3.0 | memory length: 51221 | epsilon: 1.0 | steps: 228 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 288 | score: 0.0 | memory length: 51343 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.4 | | | |
| episode: 289 | score: 2.0 | memory length: 51541 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |
| episode: 290 | score: 0.0 | memory length: 51664 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.37 | | | |
| episode: 291 | score: 0.0 | memory length: 51786 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 292 | score: 3.0 | memory length: 52029 | epsilon: 1.0 | steps: 243 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 293 | score: 1.0 | memory length: 52198 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 294 | score: 0.0 | memory length: 52321 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.32 | | | |
| episode: 295 | score: 2.0 | memory length: 52538 | epsilon: 1.0 | steps: 217 |
| lr: 0.0001 | evaluation reward: 1.34 | | | |
| episode: 296 | score: 0.0 | memory length: 52661 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.34 | | | |
| episode: 297 | score: 1.0 | memory length: 52830 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 298 | score: 3.0 | memory length: 53076 | epsilon: 1.0 | steps: 246 |
| lr: 0.0001 | evaluation reward: 1.34 | | | |
| episode: 299 | score: 0.0 | memory length: 53198 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 300 | score: 2.0 | memory length: 53414 | epsilon: 1.0 | steps: 216 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 301 | score: 1.0 | memory length: 53564 | epsilon: 1.0 | steps: 150 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 302 | score: 3.0 | memory length: 53811 | epsilon: 1.0 | steps: 247 |
| lr: 0.0001 | evaluation reward: 1.32 | | | |
| episode: 303 | score: 0.0 | memory length: 53934 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.27 | | | |
| episode: 304 | score: 2.0 | memory length: 54134 | epsilon: 1.0 | steps: 200 |
| lr: 0.0001 | evaluation reward: 1.28 | | | |
| episode: 305 | score: 3.0 | memory length: 54385 | epsilon: 1.0 | steps: 251 |
| lr: 0.0001 | evaluation reward: 1.31 | | | |
| episode: 306 | score: 2.0 | memory length: 54583 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 307 | score: 0.0 | memory length: 54706 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3 | | | |
| episode: 308 | score: 0.0 | memory length: 54829 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.3 | | | |
| episode: 309 | score: 1.0 | memory length: 54980 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.29 | | | |
| episode: 310 | score: 1.0 | memory length: 55131 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.28 | | | |
| episode: 311 | score: 1.0 | memory length: 55284 | epsilon: 1.0 | steps: 153 |
| lr: 0.0001 | evaluation reward: 1.28 | | | |
| episode: 312 | score: 1.0 | memory length: 55453 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.25 | | | |
| episode: 313 | score: 0.0 | memory length: 55576 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.21 | | | |
| episode: 314 | score: 1.0 | memory length: 55747 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.21 | | | |
| episode: 315 | score: 2.0 | memory length: 55945 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.22 | | | |
| episode: 316 | score: 0.0 | memory length: 56068 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.22 | | | |
| episode: 317 | score: 2.0 | memory length: 56283 | epsilon: 1.0 | steps: 215 |
| lr: 0.0001 | evaluation reward: 1.23 | | | |
| episode: 318 | score: 2.0 | memory length: 56481 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.25 | | | |
| episode: 319 | score: 2.0 | memory length: 56699 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.25 | | | |
| episode: 320 | score: 5.0 | memory length: 57026 | epsilon: 1.0 | steps: 327 |
| lr: 0.0001 | evaluation reward: 1.27 | | | |
| episode: 321 | score: 2.0 | memory length: 57224 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.28 | | | |
| episode: 322 | score: 3.0 | memory length: 57491 | epsilon: 1.0 | steps: 267 |
| lr: 0.0001 | evaluation reward: 1.29 | | | |
| episode: 323 | score: 5.0 | memory length: 57780 | epsilon: 1.0 | steps: 289 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 324 | score: 1.0 | memory length: 57930 | epsilon: 1.0 | steps: 150 |
| lr: 0.0001 | evaluation reward: 1.31 | | | |
| episode: 325 | score: 2.0 | memory length: 58128 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.32 | | | |
| episode: 326 | score: 0.0 | memory length: 58250 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.29 | | | |
| episode: 327 | score: 1.0 | memory length: 58403 | epsilon: 1.0 | steps: 153 |
| lr: 0.0001 | evaluation reward: 1.28 | | | |
| episode: 328 | score: 2.0 | memory length: 58600 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.3 | | | |
| episode: 329 | score: 0.0 | memory length: 58722 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.28 | | | |
| episode: 330 | score: 3.0 | memory length: 58988 | epsilon: 1.0 | steps: 266 |
| lr: 0.0001 | evaluation reward: 1.29 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 331 | score: 0.0 | memory length: 59111 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.28 | | | |
| episode: 332 | score: 1.0 | memory length: 59280 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.28 | | | |
| episode: 333 | score: 0.0 | memory length: 59403 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.26 | | | |
| episode: 334 | score: 1.0 | memory length: 59555 | epsilon: 1.0 | steps: 152 |
| lr: 0.0001 | evaluation reward: 1.24 | | | |
| episode: 335 | score: 0.0 | memory length: 59678 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.24 | | | |
| episode: 336 | score: 3.0 | memory length: 59925 | epsilon: 1.0 | steps: 247 |
| lr: 0.0001 | evaluation reward: 1.26 | | | |
| episode: 337 | score: 0.0 | memory length: 60048 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.24 | | | |
| episode: 338 | score: 1.0 | memory length: 60216 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.25 | | | |
| episode: 339 | score: 1.0 | memory length: 60388 | epsilon: 1.0 | steps: 172 |
| lr: 0.0001 | evaluation reward: 1.25 | | | |
| episode: 340 | score: 4.0 | memory length: 60682 | epsilon: 1.0 | steps: 294 |
| lr: 0.0001 | evaluation reward: 1.29 | | | |
| episode: 341 | score: 2.0 | memory length: 60882 | epsilon: 1.0 | steps: 200 |
| lr: 0.0001 | evaluation reward: 1.31 | | | |
| episode: 342 | score: 2.0 | memory length: 61068 | epsilon: 1.0 | steps: 186 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 343 | score: 0.0 | memory length: 61190 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.3 | | | |
| episode: 344 | score: 1.0 | memory length: 61362 | epsilon: 1.0 | steps: 172 |
| lr: 0.0001 | evaluation reward: 1.28 | | | |
| episode: 345 | score: 0.0 | memory length: 61484 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.28 | | | |
| episode: 346 | score: 1.0 | memory length: 61653 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.27 | | | |
| episode: 347 | score: 0.0 | memory length: 61776 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.27 | | | |
| episode: 348 | score: 0.0 | memory length: 61899 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.26 | | | |
| episode: 349 | score: 0.0 | memory length: 62021 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.25 | | | |
| episode: 350 | score: 2.0 | memory length: 62218 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.26 | | | |
| episode: 351 | score: 0.0 | memory length: 62341 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.25 | | | |
| episode: 352 | score: 0.0 | memory length: 62463 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.24 | | | |
| episode: 353 | score: 0.0 | memory length: 62585 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.24 | | | |
| episode: 354 | score: 2.0 | memory length: 62783 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.24 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 355 | score: 3.0 | memory length: 63017 | epsilon: 1.0 | steps: 234 |
| lr: 0.0001 | evaluation reward: 1.26 | | | |
| episode: 356 | score: 2.0 | memory length: 63215 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.27 | | | |
| episode: 357 | score: 3.0 | memory length: 63444 | epsilon: 1.0 | steps: 229 |
| lr: 0.0001 | evaluation reward: 1.29 | | | |
| episode: 358 | score: 0.0 | memory length: 63566 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.29 | | | |
| episode: 359 | score: 2.0 | memory length: 63764 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.31 | | | |
| episode: 360 | score: 4.0 | memory length: 64081 | epsilon: 1.0 | steps: 317 |
| lr: 0.0001 | evaluation reward: 1.34 | | | |
| episode: 361 | score: 2.0 | memory length: 64299 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 362 | score: 1.0 | memory length: 64469 | epsilon: 1.0 | steps: 170 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 363 | score: 1.0 | memory length: 64619 | epsilon: 1.0 | steps: 150 |
| lr: 0.0001 | evaluation reward: 1.36 | | | |
| episode: 364 | score: 0.0 | memory length: 64742 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.35 | | | |
| episode: 365 | score: 1.0 | memory length: 64910 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.34 | | | |
| episode: 366 | score: 0.0 | memory length: 65033 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.31 | | | |
| episode: 367 | score: 2.0 | memory length: 65253 | epsilon: 1.0 | steps: 220 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 368 | score: 2.0 | memory length: 65473 | epsilon: 1.0 | steps: 220 |
| lr: 0.0001 | evaluation reward: 1.31 | | | |
| episode: 369 | score: 1.0 | memory length: 65644 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.31 | | | |
| episode: 370 | score: 1.0 | memory length: 65796 | epsilon: 1.0 | steps: 152 |
| lr: 0.0001 | evaluation reward: 1.32 | | | |
| episode: 371 | score: 1.0 | memory length: 65968 | epsilon: 1.0 | steps: 172 |
| lr: 0.0001 | evaluation reward: 1.32 | | | |
| episode: 372 | score: 1.0 | memory length: 66137 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.33 | | | |
| episode: 373 | score: 1.0 | memory length: 66306 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.32 | | | |
| episode: 374 | score: 10.0 | memory length: 66706 | epsilon: 1.0 | steps: 400 |
| lr: 0.0001 | evaluation reward: 1.39 | | | |
| episode: 375 | score: 0.0 | memory length: 66828 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.38 | | | |
| episode: 376 | score: 3.0 | memory length: 67074 | epsilon: 1.0 | steps: 246 |
| lr: 0.0001 | evaluation reward: 1.38 | | | |
| episode: 377 | score: 4.0 | memory length: 67369 | epsilon: 1.0 | steps: 295 |
| lr: 0.0001 | evaluation reward: 1.42 | | | |
| episode: 378 | score: 2.0 | memory length: 67585 | epsilon: 1.0 | steps: 216 |
| lr: 0.0001 | evaluation reward: 1.42 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 379 | score: 7.0 | memory length: 67887 | epsilon: 1.0 | steps: 302 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 380 | score: 1.0 | memory length: 68038 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 381 | score: 2.0 | memory length: 68256 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 382 | score: 4.0 | memory length: 68534 | epsilon: 1.0 | steps: 278 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 383 | score: 1.0 | memory length: 68703 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 384 | score: 0.0 | memory length: 68825 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 385 | score: 5.0 | memory length: 69182 | epsilon: 1.0 | steps: 357 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 386 | score: 3.0 | memory length: 69427 | epsilon: 1.0 | steps: 245 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 387 | score: 2.0 | memory length: 69624 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.55 | | | |
| episode: 388 | score: 0.0 | memory length: 69747 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.55 | | | |
| episode: 389 | score: 3.0 | memory length: 69993 | epsilon: 1.0 | steps: 246 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 390 | score: 2.0 | memory length: 70193 | epsilon: 1.0 | steps: 200 |
| lr: 0.0001 | evaluation reward: 1.58 | | | |
| episode: 391 | score: 0.0 | memory length: 70316 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.58 | | | |
| episode: 392 | score: 1.0 | memory length: 70486 | epsilon: 1.0 | steps: 170 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 393 | score: 2.0 | memory length: 70704 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.57 | | | |
| episode: 394 | score: 2.0 | memory length: 70922 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.59 | | | |
| episode: 395 | score: 0.0 | memory length: 71045 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.57 | | | |
| episode: 396 | score: 0.0 | memory length: 71168 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.57 | | | |
| episode: 397 | score: 0.0 | memory length: 71291 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 398 | score: 1.0 | memory length: 71460 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 399 | score: 2.0 | memory length: 71676 | epsilon: 1.0 | steps: 216 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 400 | score: 2.0 | memory length: 71874 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 401 | score: 2.0 | memory length: 72071 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.57 | | | |
| episode: 402 | score: 0.0 | memory length: 72194 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 403 | score: 3.0 | memory length: 72457 | epsilon: 1.0 | steps: 263 |
| lr: 0.0001 | evaluation reward: 1.57 | | | |
| episode: 404 | score: 3.0 | memory length: 72702 | epsilon: 1.0 | steps: 245 |
| lr: 0.0001 | evaluation reward: 1.58 | | | |
| episode: 405 | score: 0.0 | memory length: 72825 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.55 | | | |
| episode: 406 | score: 2.0 | memory length: 73023 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.55 | | | |
| episode: 407 | score: 1.0 | memory length: 73193 | epsilon: 1.0 | steps: 170 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 408 | score: 4.0 | memory length: 73488 | epsilon: 1.0 | steps: 295 |
| lr: 0.0001 | evaluation reward: 1.6 | | | |
| episode: 409 | score: 3.0 | memory length: 73716 | epsilon: 1.0 | steps: 228 |
| lr: 0.0001 | evaluation reward: 1.62 | | | |
| episode: 410 | score: 1.0 | memory length: 73884 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.62 | | | |
| episode: 411 | score: 2.0 | memory length: 74101 | epsilon: 1.0 | steps: 217 |
| lr: 0.0001 | evaluation reward: 1.63 | | | |
| episode: 412 | score: 2.0 | memory length: 74317 | epsilon: 1.0 | steps: 216 |
| lr: 0.0001 | evaluation reward: 1.64 | | | |
| episode: 413 | score: 5.0 | memory length: 74664 | epsilon: 1.0 | steps: 347 |
| lr: 0.0001 | evaluation reward: 1.69 | | | |
| episode: 414 | score: 1.0 | memory length: 74833 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.69 | | | |
| episode: 415 | score: 2.0 | memory length: 75017 | epsilon: 1.0 | steps: 184 |
| lr: 0.0001 | evaluation reward: 1.69 | | | |
| episode: 416 | score: 0.0 | memory length: 75140 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.69 | | | |
| episode: 417 | score: 2.0 | memory length: 75321 | epsilon: 1.0 | steps: 181 |
| lr: 0.0001 | evaluation reward: 1.69 | | | |
| episode: 418 | score: 0.0 | memory length: 75444 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.67 | | | |
| episode: 419 | score: 0.0 | memory length: 75567 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.65 | | | |
| episode: 420 | score: 0.0 | memory length: 75690 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.6 | | | |
| episode: 421 | score: 1.0 | memory length: 75842 | epsilon: 1.0 | steps: 152 |
| lr: 0.0001 | evaluation reward: 1.59 | | | |
| episode: 422 | score: 0.0 | memory length: 75964 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 423 | score: 2.0 | memory length: 76162 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.53 | | | |
| episode: 424 | score: 2.0 | memory length: 76379 | epsilon: 1.0 | steps: 217 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 425 | score: 1.0 | memory length: 76548 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.53 | | | |
| episode: 426 | score: 1.0 | memory length: 76699 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 427 | score: 3.0 | memory length: 76946 | epsilon: 1.0 | steps: 247 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 428 | score: 3.0 | memory length: 77192 | epsilon: 1.0 | steps: 246 |
| lr: 0.0001 | evaluation reward: 1.57 | | | |
| episode: 429 | score: 0.0 | memory length: 77314 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.57 | | | |
| episode: 430 | score: 2.0 | memory length: 77512 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 431 | score: 0.0 | memory length: 77635 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 432 | score: 4.0 | memory length: 77892 | epsilon: 1.0 | steps: 257 |
| lr: 0.0001 | evaluation reward: 1.59 | | | |
| episode: 433 | score: 2.0 | memory length: 78090 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.61 | | | |
| episode: 434 | score: 0.0 | memory length: 78213 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.6 | | | |
| episode: 435 | score: 0.0 | memory length: 78336 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.6 | | | |
| episode: 436 | score: 0.0 | memory length: 78459 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.57 | | | |
| episode: 437 | score: 1.0 | memory length: 78611 | epsilon: 1.0 | steps: 152 |
| lr: 0.0001 | evaluation reward: 1.58 | | | |
| episode: 438 | score: 2.0 | memory length: 78832 | epsilon: 1.0 | steps: 221 |
| lr: 0.0001 | evaluation reward: 1.59 | | | |
| episode: 439 | score: 1.0 | memory length: 79001 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.59 | | | |
| episode: 440 | score: 0.0 | memory length: 79124 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.55 | | | |
| episode: 441 | score: 0.0 | memory length: 79246 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.53 | | | |
| episode: 442 | score: 0.0 | memory length: 79369 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 443 | score: 2.0 | memory length: 79589 | epsilon: 1.0 | steps: 220 |
| lr: 0.0001 | evaluation reward: 1.53 | | | |
| episode: 444 | score: 3.0 | memory length: 79832 | epsilon: 1.0 | steps: 243 |
| lr: 0.0001 | evaluation reward: 1.55 | | | |
| episode: 445 | score: 2.0 | memory length: 80030 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.57 | | | |
| episode: 446 | score: 1.0 | memory length: 80199 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.57 | | | |
| episode: 447 | score: 4.0 | memory length: 80497 | epsilon: 1.0 | steps: 298 |
| lr: 0.0001 | evaluation reward: 1.61 | | | |
| episode: 448 | score: 3.0 | memory length: 80725 | epsilon: 1.0 | steps: 228 |
| lr: 0.0001 | evaluation reward: 1.64 | | | |
| episode: 449 | score: 2.0 | memory length: 80922 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.66 | | | |
| episode: 450 | score: 1.0 | memory length: 81072 | epsilon: 1.0 | steps: 150 |
| lr: 0.0001 | evaluation reward: 1.65 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 451 | score: 0.0 | memory length: 81195 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.65 | | | |
| episode: 452 | score: 2.0 | memory length: 81393 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.67 | | | |
| episode: 453 | score: 0.0 | memory length: 81516 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.67 | | | |
| episode: 454 | score: 4.0 | memory length: 81832 | epsilon: 1.0 | steps: 316 |
| lr: 0.0001 | evaluation reward: 1.69 | | | |
| episode: 455 | score: 1.0 | memory length: 81983 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.67 | | | |
| episode: 456 | score: 0.0 | memory length: 82106 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.65 | | | |
| episode: 457 | score: 0.0 | memory length: 82229 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.62 | | | |
| episode: 458 | score: 0.0 | memory length: 82352 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.62 | | | |
| episode: 459 | score: 2.0 | memory length: 82550 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.62 | | | |
| episode: 460 | score: 0.0 | memory length: 82672 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.58 | | | |
| episode: 461 | score: 4.0 | memory length: 82983 | epsilon: 1.0 | steps: 311 |
| lr: 0.0001 | evaluation reward: 1.6 | | | |
| episode: 462 | score: 1.0 | memory length: 83155 | epsilon: 1.0 | steps: 172 |
| lr: 0.0001 | evaluation reward: 1.6 | | | |
| episode: 463 | score: 1.0 | memory length: 83324 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.6 | | | |
| episode: 464 | score: 0.0 | memory length: 83447 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.6 | | | |
| episode: 465 | score: 2.0 | memory length: 83645 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.61 | | | |
| episode: 466 | score: 1.0 | memory length: 83796 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.62 | | | |
| episode: 467 | score: 2.0 | memory length: 83994 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.62 | | | |
| episode: 468 | score: 1.0 | memory length: 84164 | epsilon: 1.0 | steps: 170 |
| lr: 0.0001 | evaluation reward: 1.61 | | | |
| episode: 469 | score: 1.0 | memory length: 84335 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.61 | | | |
| episode: 470 | score: 3.0 | memory length: 84581 | epsilon: 1.0 | steps: 246 |
| lr: 0.0001 | evaluation reward: 1.63 | | | |
| episode: 471 | score: 3.0 | memory length: 84825 | epsilon: 1.0 | steps: 244 |
| lr: 0.0001 | evaluation reward: 1.65 | | | |
| episode: 472 | score: 3.0 | memory length: 85053 | epsilon: 1.0 | steps: 228 |
| lr: 0.0001 | evaluation reward: 1.67 | | | |
| episode: 473 | score: 6.0 | memory length: 85395 | epsilon: 1.0 | steps: 342 |
| lr: 0.0001 | evaluation reward: 1.72 | | | |
| episode: 474 | score: 2.0 | memory length: 85613 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.64 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 475 | score: 4.0 | memory length: 85890 | epsilon: 1.0 | steps: 277 |
| lr: 0.0001 | evaluation reward: 1.68 | | | |
| episode: 476 | score: 0.0 | memory length: 86013 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.65 | | | |
| episode: 477 | score: 0.0 | memory length: 86135 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.61 | | | |
| episode: 478 | score: 0.0 | memory length: 86258 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.59 | | | |
| episode: 479 | score: 2.0 | memory length: 86456 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 480 | score: 1.0 | memory length: 86625 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 481 | score: 3.0 | memory length: 86853 | epsilon: 1.0 | steps: 228 |
| lr: 0.0001 | evaluation reward: 1.55 | | | |
| episode: 482 | score: 3.0 | memory length: 87079 | epsilon: 1.0 | steps: 226 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 483 | score: 0.0 | memory length: 87201 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.53 | | | |
| episode: 484 | score: 3.0 | memory length: 87426 | epsilon: 1.0 | steps: 225 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 485 | score: 0.0 | memory length: 87548 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 486 | score: 3.0 | memory length: 87775 | epsilon: 1.0 | steps: 227 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 487 | score: 1.0 | memory length: 87926 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.5 | | | |
| episode: 488 | score: 1.0 | memory length: 88077 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 489 | score: 3.0 | memory length: 88320 | epsilon: 1.0 | steps: 243 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 490 | score: 1.0 | memory length: 88489 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.5 | | | |
| episode: 491 | score: 0.0 | memory length: 88612 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.5 | | | |
| episode: 492 | score: 2.0 | memory length: 88812 | epsilon: 1.0 | steps: 200 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 493 | score: 1.0 | memory length: 88963 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.5 | | | |
| episode: 494 | score: 2.0 | memory length: 89161 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.5 | | | |
| episode: 495 | score: 1.0 | memory length: 89332 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 496 | score: 0.0 | memory length: 89455 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 497 | score: 0.0 | memory length: 89578 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 498 | score: 3.0 | memory length: 89845 | epsilon: 1.0 | steps: 267 |
| lr: 0.0001 | evaluation reward: 1.53 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 499 | score: 5.0 | memory length: 90164 | epsilon: 1.0 | steps: 319 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 500 | score: 0.0 | memory length: 90287 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 501 | score: 2.0 | memory length: 90503 | epsilon: 1.0 | steps: 216 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 502 | score: 4.0 | memory length: 90778 | epsilon: 1.0 | steps: 275 |
| lr: 0.0001 | evaluation reward: 1.58 | | | |
| episode: 503 | score: 1.0 | memory length: 90947 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 504 | score: 1.0 | memory length: 91117 | epsilon: 1.0 | steps: 170 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 505 | score: 2.0 | memory length: 91314 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |
| episode: 506 | score: 0.0 | memory length: 91437 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 507 | score: 1.0 | memory length: 91588 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 508 | score: 1.0 | memory length: 91757 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.51 | | | |
| episode: 509 | score: 0.0 | memory length: 91880 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.48 | | | |
| episode: 510 | score: 0.0 | memory length: 92003 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 511 | score: 3.0 | memory length: 92252 | epsilon: 1.0 | steps: 249 |
| lr: 0.0001 | evaluation reward: 1.48 | | | |
| episode: 512 | score: 4.0 | memory length: 92526 | epsilon: 1.0 | steps: 274 |
| lr: 0.0001 | evaluation reward: 1.5 | | | |
| episode: 513 | score: 1.0 | memory length: 92698 | epsilon: 1.0 | steps: 172 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 514 | score: 0.0 | memory length: 92821 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.45 | | | |
| episode: 515 | score: 0.0 | memory length: 92944 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 516 | score: 4.0 | memory length: 93222 | epsilon: 1.0 | steps: 278 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 517 | score: 4.0 | memory length: 93498 | epsilon: 1.0 | steps: 276 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 518 | score: 0.0 | memory length: 93621 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 519 | score: 1.0 | memory length: 93790 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.5 | | | |
| episode: 520 | score: 2.0 | memory length: 93970 | epsilon: 1.0 | steps: 180 |
| lr: 0.0001 | evaluation reward: 1.52 | | | |
| episode: 521 | score: 4.0 | memory length: 94266 | epsilon: 1.0 | steps: 296 |
| lr: 0.0001 | evaluation reward: 1.55 | | | |
| episode: 522 | score: 1.0 | memory length: 94436 | epsilon: 1.0 | steps: 170 |
| lr: 0.0001 | evaluation reward: 1.56 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 523 | score: 0.0 | memory length: 94559 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.54 | | | |
| episode: 524 | score: 1.0 | memory length: 94710 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.53 | | | |
| episode: 525 | score: 1.0 | memory length: 94879 | epsilon: 1.0 | steps: 169 |
| lr: 0.0001 | evaluation reward: 1.53 | | | |
| episode: 526 | score: 0.0 | memory length: 95002 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.52 | | | |
| episode: 527 | score: 0.0 | memory length: 95125 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 528 | score: 0.0 | memory length: 95247 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 529 | score: 0.0 | memory length: 95370 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 530 | score: 1.0 | memory length: 95541 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.45 | | | |
| episode: 531 | score: 3.0 | memory length: 95773 | epsilon: 1.0 | steps: 232 |
| lr: 0.0001 | evaluation reward: 1.48 | | | |
| episode: 532 | score: 1.0 | memory length: 95944 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.45 | | | |
| episode: 533 | score: 0.0 | memory length: 96067 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.43 | | | |
| episode: 534 | score: 2.0 | memory length: 96288 | epsilon: 1.0 | steps: 221 |
| lr: 0.0001 | evaluation reward: 1.45 | | | |
| episode: 535 | score: 1.0 | memory length: 96439 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 536 | score: 1.0 | memory length: 96590 | epsilon: 1.0 | steps: 151 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 537 | score: 0.0 | memory length: 96713 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 538 | score: 2.0 | memory length: 96931 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 539 | score: 2.0 | memory length: 97128 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 540 | score: 2.0 | memory length: 97325 | epsilon: 1.0 | steps: 197 |
| lr: 0.0001 | evaluation reward: 1.49 | | | |
| episode: 541 | score: 1.0 | memory length: 97493 | epsilon: 1.0 | steps: 168 |
| lr: 0.0001 | evaluation reward: 1.5 | | | |
| episode: 542 | score: 0.0 | memory length: 97615 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.5 | | | |
| episode: 543 | score: 2.0 | memory length: 97834 | epsilon: 1.0 | steps: 219 |
| lr: 0.0001 | evaluation reward: 1.5 | | | |
| episode: 544 | score: 0.0 | memory length: 97957 | epsilon: 1.0 | steps: 123 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 545 | score: 3.0 | memory length: 98206 | epsilon: 1.0 | steps: 249 |
| lr: 0.0001 | evaluation reward: 1.48 | | | |
| episode: 546 | score: 0.0 | memory length: 98328 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |

| | | | | |
|--------------|-------------------------|----------------------|--------------|------------|
| episode: 547 | score: 2.0 | memory length: 98546 | epsilon: 1.0 | steps: 218 |
| lr: 0.0001 | evaluation reward: 1.45 | | | |
| episode: 548 | score: 4.0 | memory length: 98839 | epsilon: 1.0 | steps: 293 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 549 | score: 2.0 | memory length: 99021 | epsilon: 1.0 | steps: 182 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 550 | score: 2.0 | memory length: 99238 | epsilon: 1.0 | steps: 217 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 551 | score: 0.0 | memory length: 99360 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.47 | | | |
| episode: 552 | score: 1.0 | memory length: 99531 | epsilon: 1.0 | steps: 171 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 553 | score: 0.0 | memory length: 99653 | epsilon: 1.0 | steps: 122 |
| lr: 0.0001 | evaluation reward: 1.46 | | | |
| episode: 554 | score: 2.0 | memory length: 99851 | epsilon: 1.0 | steps: 198 |
| lr: 0.0001 | evaluation reward: 1.44 | | | |

/u/qilong/QA_generator/CS444/mp5/memory.py:66: FutureWarning: The input object of type 'Tensor' is an array-like implementing one of the corresponding protocols (`__array__`, `__array_interface__` or `__array_struct__`); but not a sequence (or 0-D). In the future, this object will be coerced as if it was first converted using `np.array(obj)`. To retain the old behaviour, you have to either modify the type 'Tensor', or assign to an empty array created with `np.empty(correct_shape, dtype=object)`.

sample = np.array(sample, dtype=object)

/u/qilong/QA_generator/CS444/mp5/agent_lstm.py:53: FutureWarning: The input object of type 'Tensor' is an array-like implementing one of the corresponding protocols (`__array__`, `__array_interface__` or `__array_struct__`); but not a sequence (or 0-D). In the future, this object will be coerced as if it was first converted using `np.array(obj)`. To retain the old behaviour, you have to either modify the type 'Tensor', or assign to an empty array created with `np.empty(correct_shape, dtype=object)`.

mini_batch = np.array(mini_batch, dtype=object).transpose()

| | | | |
|--------------|------------|-------------------------|-----------------------------|
| episode: 555 | score: 3.0 | memory length: 100081 | epsilon: 0.9998376400000035 |
| steps: 230 | lr: 0.0001 | evaluation reward: 1.46 | |
| episode: 556 | score: 3.0 | memory length: 100327 | epsilon: 0.9993505600000141 |
| steps: 246 | lr: 0.0001 | evaluation reward: 1.49 | |
| episode: 557 | score: 3.0 | memory length: 100576 | epsilon: 0.9988575400000248 |
| steps: 249 | lr: 0.0001 | evaluation reward: 1.52 | |
| episode: 558 | score: 1.0 | memory length: 100744 | epsilon: 0.9985249000000032 |
| steps: 168 | lr: 0.0001 | evaluation reward: 1.53 | |
| episode: 559 | score: 2.0 | memory length: 100942 | epsilon: 0.9981328600000405 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.53 | |
| episode: 560 | score: 0.0 | memory length: 101065 | epsilon: 0.9978893200000458 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.53 | |
| episode: 561 | score: 4.0 | memory length: 101385 | epsilon: 0.9972557200000596 |
| steps: 320 | lr: 0.0001 | evaluation reward: 1.53 | |
| episode: 562 | score: 0.0 | memory length: 101507 | epsilon: 0.9970141600000648 |

| | | |
|--------------|------------|---|
| steps: 122 | lr: 0.0001 | evaluation reward: 1.52 |
| episode: 563 | score: 5.0 | memory length: 101831 epsilon: 0.9963726400000787 |
| steps: 324 | lr: 0.0001 | evaluation reward: 1.56 |
| episode: 564 | score: 2.0 | memory length: 102028 epsilon: 0.9959825800000872 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.58 |
| episode: 565 | score: 1.0 | memory length: 102199 epsilon: 0.9956440000000946 |
| steps: 171 | lr: 0.0001 | evaluation reward: 1.57 |
| episode: 566 | score: 0.0 | memory length: 102321 epsilon: 0.9954024400000998 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.56 |
| episode: 567 | score: 0.0 | memory length: 102444 epsilon: 0.9951589000001051 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.54 |
| episode: 568 | score: 1.0 | memory length: 102597 epsilon: 0.9948559600001117 |
| steps: 153 | lr: 0.0001 | evaluation reward: 1.54 |
| episode: 569 | score: 3.0 | memory length: 102825 epsilon: 0.9944045200001215 |
| steps: 228 | lr: 0.0001 | evaluation reward: 1.56 |
| episode: 570 | score: 3.0 | memory length: 103069 epsilon: 0.993921400000132 |
| steps: 244 | lr: 0.0001 | evaluation reward: 1.56 |
| episode: 571 | score: 3.0 | memory length: 103334 epsilon: 0.9933967000001434 |
| steps: 265 | lr: 0.0001 | evaluation reward: 1.56 |
| episode: 572 | score: 4.0 | memory length: 103633 epsilon: 0.9928046800001562 |
| steps: 299 | lr: 0.0001 | evaluation reward: 1.57 |
| episode: 573 | score: 1.0 | memory length: 103802 epsilon: 0.9924700600001635 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.52 |
| episode: 574 | score: 1.0 | memory length: 103953 epsilon: 0.99217108000017 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.51 |
| episode: 575 | score: 0.0 | memory length: 104075 epsilon: 0.9919295200001752 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.47 |
| episode: 576 | score: 2.0 | memory length: 104296 epsilon: 0.9914919400001847 |
| steps: 221 | lr: 0.0001 | evaluation reward: 1.49 |
| episode: 577 | score: 0.0 | memory length: 104418 epsilon: 0.99125038000019 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.49 |
| episode: 578 | score: 2.0 | memory length: 104615 epsilon: 0.9908603200001984 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.51 |
| episode: 579 | score: 3.0 | memory length: 104860 epsilon: 0.990375220000209 |
| steps: 245 | lr: 0.0001 | evaluation reward: 1.52 |
| episode: 580 | score: 1.0 | memory length: 105029 epsilon: 0.9900406000002162 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.52 |
| episode: 581 | score: 1.0 | memory length: 105182 epsilon: 0.9897376600002228 |
| steps: 153 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 582 | score: 1.0 | memory length: 105333 epsilon: 0.9894386800002293 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.48 |
| episode: 583 | score: 1.0 | memory length: 105502 epsilon: 0.9891040600002365 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.49 |
| episode: 584 | score: 2.0 | memory length: 105717 epsilon: 0.9886783600002458 |
| steps: 215 | lr: 0.0001 | evaluation reward: 1.48 |
| episode: 585 | score: 4.0 | memory length: 106000 epsilon: 0.988118020000258 |
| steps: 283 | lr: 0.0001 | evaluation reward: 1.52 |
| episode: 586 | score: 2.0 | memory length: 106198 epsilon: 0.9877259800002665 |

| | | |
|--------------|------------|---|
| steps: 198 | lr: 0.0001 | evaluation reward: 1.51 |
| episode: 587 | score: 1.0 | memory length: 106366 epsilon: 0.9873933400002737 |
| steps: 168 | lr: 0.0001 | evaluation reward: 1.51 |
| episode: 588 | score: 0.0 | memory length: 106489 epsilon: 0.987149800000279 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 589 | score: 1.0 | memory length: 106640 epsilon: 0.9868508200002855 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.48 |
| episode: 590 | score: 2.0 | memory length: 106840 epsilon: 0.986454820000294 |
| steps: 200 | lr: 0.0001 | evaluation reward: 1.49 |
| episode: 591 | score: 2.0 | memory length: 107037 epsilon: 0.9860647600003025 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.51 |
| episode: 592 | score: 0.0 | memory length: 107159 epsilon: 0.9858232000003078 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.49 |
| episode: 593 | score: 0.0 | memory length: 107282 epsilon: 0.985579660000313 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.48 |
| episode: 594 | score: 0.0 | memory length: 107405 epsilon: 0.9853361200003183 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.46 |
| episode: 595 | score: 2.0 | memory length: 107602 epsilon: 0.9849460600003268 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.47 |
| episode: 596 | score: 2.0 | memory length: 107782 epsilon: 0.9845896600003345 |
| steps: 180 | lr: 0.0001 | evaluation reward: 1.49 |
| episode: 597 | score: 3.0 | memory length: 108010 epsilon: 0.9841382200003443 |
| steps: 228 | lr: 0.0001 | evaluation reward: 1.52 |
| episode: 598 | score: 1.0 | memory length: 108161 epsilon: 0.9838392400003508 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 599 | score: 0.0 | memory length: 108284 epsilon: 0.9835957000003561 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.45 |
| episode: 600 | score: 2.0 | memory length: 108482 epsilon: 0.9832036600003646 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.47 |
| episode: 601 | score: 2.0 | memory length: 108680 epsilon: 0.9828116200003731 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.47 |
| episode: 602 | score: 1.0 | memory length: 108851 epsilon: 0.9824730400003805 |
| steps: 171 | lr: 0.0001 | evaluation reward: 1.44 |
| episode: 603 | score: 2.0 | memory length: 109071 epsilon: 0.98203744000039 |
| steps: 220 | lr: 0.0001 | evaluation reward: 1.45 |
| episode: 604 | score: 2.0 | memory length: 109268 epsilon: 0.9816473800003984 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.46 |
| episode: 605 | score: 2.0 | memory length: 109465 epsilon: 0.9812573200004069 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.46 |
| episode: 606 | score: 0.0 | memory length: 109588 epsilon: 0.9810137800004122 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.46 |
| episode: 607 | score: 2.0 | memory length: 109806 epsilon: 0.9805821400004215 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.47 |
| episode: 608 | score: 3.0 | memory length: 110031 epsilon: 0.9801366400004312 |
| steps: 225 | lr: 0.0001 | evaluation reward: 1.49 |
| episode: 609 | score: 2.0 | memory length: 110230 epsilon: 0.9797426200004398 |
| steps: 199 | lr: 0.0001 | evaluation reward: 1.51 |
| episode: 610 | score: 4.0 | memory length: 110489 epsilon: 0.9792298000004509 |

| | | |
|--------------|------------|--|
| steps: 259 | lr: 0.0001 | evaluation reward: 1.55 |
| episode: 611 | score: 2.0 | memory length: 110671 epsilon: 0.9788694400004587 |
| steps: 182 | lr: 0.0001 | evaluation reward: 1.54 |
| episode: 612 | score: 0.0 | memory length: 110794 epsilon: 0.978625900000464 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 613 | score: 3.0 | memory length: 111039 epsilon: 0.9781408000004745 |
| steps: 245 | lr: 0.0001 | evaluation reward: 1.52 |
| episode: 614 | score: 0.0 | memory length: 111162 epsilon: 0.9778972600004798 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.52 |
| episode: 615 | score: 0.0 | memory length: 111285 epsilon: 0.9776537200004851 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.52 |
| episode: 616 | score: 2.0 | memory length: 111467 epsilon: 0.9772933600004929 |
| steps: 182 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 617 | score: 1.0 | memory length: 111635 epsilon: 0.9769607200005002 |
| steps: 168 | lr: 0.0001 | evaluation reward: 1.47 |
| episode: 618 | score: 1.0 | memory length: 111805 epsilon: 0.9766241200005075 |
| steps: 170 | lr: 0.0001 | evaluation reward: 1.48 |
| episode: 619 | score: 2.0 | memory length: 112003 epsilon: 0.976232080000516 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.49 |
| episode: 620 | score: 3.0 | memory length: 112270 epsilon: 0.9757034200005275 |
| steps: 267 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 621 | score: 0.0 | memory length: 112392 epsilon: 0.9754618600005327 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.46 |
| episode: 622 | score: 0.0 | memory length: 112515 epsilon: 0.975218320000538 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.45 |
| episode: 623 | score: 0.0 | memory length: 112638 epsilon: 0.9749747800005433 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.45 |
| episode: 624 | score: 1.0 | memory length: 112788 epsilon: 0.9746777800005497 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.45 |
| episode: 625 | score: 2.0 | memory length: 112986 epsilon: 0.9742857400005582 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.46 |
| episode: 626 | score: 1.0 | memory length: 113154 epsilon: 0.9739531000005655 |
| steps: 168 | lr: 0.0001 | evaluation reward: 1.47 |
| episode: 627 | score: 0.0 | memory length: 113276 epsilon: 0.9737115400005707 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.47 |
| episode: 628 | score: 1.0 | memory length: 113446 epsilon: 0.973374940000578 |
| steps: 170 | lr: 0.0001 | evaluation reward: 1.48 |
| episode: 629 | score: 0.0 | memory length: 113569 epsilon: 0.9731314000005833 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.48 |
| episode: 630 | score: 1.0 | memory length: 113739 epsilon: 0.9727948000005906 |
| steps: 170 | lr: 0.0001 | evaluation reward: 1.48 |
| episode: 631 | score: 1.0 | memory length: 113909 epsilon: 0.9724582000005979 |
| steps: 170 | lr: 0.0001 | evaluation reward: 1.46 |
| episode: 632 | score: 3.0 | memory length: 114156 epsilon: 0.9719691400006085 |
| steps: 247 | lr: 0.0001 | evaluation reward: 1.48 |
| episode: 633 | score: 2.0 | memory length: 114353 epsilon: 0.971579080000617 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 634 | score: 0.0 | memory length: 114476 epsilon: 0.9713355400006223 |

| | | |
|--------------|------------|---|
| steps: 123 | lr: 0.0001 | evaluation reward: 1.48 |
| episode: 635 | score: 3.0 | memory length: 114728 epsilon: 0.9708365800006331 |
| steps: 252 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 636 | score: 1.0 | memory length: 114879 epsilon: 0.9705376000006396 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 637 | score: 3.0 | memory length: 115142 epsilon: 0.9700168600006509 |
| steps: 263 | lr: 0.0001 | evaluation reward: 1.53 |
| episode: 638 | score: 0.0 | memory length: 115264 epsilon: 0.9697753000006561 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.51 |
| episode: 639 | score: 2.0 | memory length: 115482 epsilon: 0.9693436600006655 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.51 |
| episode: 640 | score: 0.0 | memory length: 115605 epsilon: 0.9691001200006708 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.49 |
| episode: 641 | score: 2.0 | memory length: 115823 epsilon: 0.9686684800006802 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 642 | score: 3.0 | memory length: 116089 epsilon: 0.9681418000006916 |
| steps: 266 | lr: 0.0001 | evaluation reward: 1.53 |
| episode: 643 | score: 4.0 | memory length: 116385 epsilon: 0.9675557200007043 |
| steps: 296 | lr: 0.0001 | evaluation reward: 1.55 |
| episode: 644 | score: 0.0 | memory length: 116507 epsilon: 0.9673141600007096 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.55 |
| episode: 645 | score: 3.0 | memory length: 116752 epsilon: 0.9668290600007201 |
| steps: 245 | lr: 0.0001 | evaluation reward: 1.55 |
| episode: 646 | score: 1.0 | memory length: 116924 epsilon: 0.9664885000007275 |
| steps: 172 | lr: 0.0001 | evaluation reward: 1.56 |
| episode: 647 | score: 2.0 | memory length: 117144 epsilon: 0.966052900000737 |
| steps: 220 | lr: 0.0001 | evaluation reward: 1.56 |
| episode: 648 | score: 1.0 | memory length: 117316 epsilon: 0.9657123400007444 |
| steps: 172 | lr: 0.0001 | evaluation reward: 1.53 |
| episode: 649 | score: 0.0 | memory length: 117438 epsilon: 0.9654707800007496 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.51 |
| episode: 650 | score: 0.0 | memory length: 117561 epsilon: 0.9652272400007549 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.49 |
| episode: 651 | score: 1.0 | memory length: 117711 epsilon: 0.9649302400007613 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 652 | score: 1.0 | memory length: 117862 epsilon: 0.9646312600007678 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 653 | score: 3.0 | memory length: 118131 epsilon: 0.9640986400007794 |
| steps: 269 | lr: 0.0001 | evaluation reward: 1.53 |
| episode: 654 | score: 2.0 | memory length: 118349 epsilon: 0.9636670000007888 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.53 |
| episode: 655 | score: 1.0 | memory length: 118500 epsilon: 0.9633680200007952 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.51 |
| episode: 656 | score: 1.0 | memory length: 118650 epsilon: 0.9630710200008017 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.49 |
| episode: 657 | score: 0.0 | memory length: 118773 epsilon: 0.962827480000807 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.46 |
| episode: 658 | score: 5.0 | memory length: 119096 epsilon: 0.9621879400008209 |

| | | | |
|--------------|------------|-------------------------|-----------------------------|
| steps: 323 | lr: 0.0001 | evaluation reward: 1.5 | |
| episode: 659 | score: 2.0 | memory length: 119314 | epsilon: 0.9617563000008302 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.5 | |
| episode: 660 | score: 3.0 | memory length: 119562 | epsilon: 0.9612652600008409 |
| steps: 248 | lr: 0.0001 | evaluation reward: 1.53 | |
| episode: 661 | score: 6.0 | memory length: 119880 | epsilon: 0.9606356200008546 |
| steps: 318 | lr: 0.0001 | evaluation reward: 1.55 | |
| episode: 662 | score: 1.0 | memory length: 120052 | epsilon: 0.960295060000862 |
| steps: 172 | lr: 0.0001 | evaluation reward: 1.56 | |
| episode: 663 | score: 1.0 | memory length: 120221 | epsilon: 0.9599604400008692 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.52 | |
| episode: 664 | score: 0.0 | memory length: 120344 | epsilon: 0.9597169000008745 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.5 | |
| episode: 665 | score: 0.0 | memory length: 120466 | epsilon: 0.9594753400008798 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.49 | |
| episode: 666 | score: 0.0 | memory length: 120589 | epsilon: 0.959231800000885 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.49 | |
| episode: 667 | score: 4.0 | memory length: 120886 | epsilon: 0.9586437400008978 |
| steps: 297 | lr: 0.0001 | evaluation reward: 1.53 | |
| episode: 668 | score: 0.0 | memory length: 121009 | epsilon: 0.9584002000009031 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.52 | |
| episode: 669 | score: 2.0 | memory length: 121208 | epsilon: 0.9580061800009116 |
| steps: 199 | lr: 0.0001 | evaluation reward: 1.51 | |
| episode: 670 | score: 1.0 | memory length: 121358 | epsilon: 0.9577091800009181 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.49 | |
| episode: 671 | score: 2.0 | memory length: 121556 | epsilon: 0.9573171400009266 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.48 | |
| episode: 672 | score: 3.0 | memory length: 121787 | epsilon: 0.9568597600009365 |
| steps: 231 | lr: 0.0001 | evaluation reward: 1.47 | |
| episode: 673 | score: 0.0 | memory length: 121910 | epsilon: 0.9566162200009418 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.46 | |
| episode: 674 | score: 2.0 | memory length: 122133 | epsilon: 0.9561746800009514 |
| steps: 223 | lr: 0.0001 | evaluation reward: 1.47 | |
| episode: 675 | score: 2.0 | memory length: 122331 | epsilon: 0.9557826400009599 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.49 | |
| episode: 676 | score: 0.0 | memory length: 122453 | epsilon: 0.9555410800009652 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.47 | |
| episode: 677 | score: 6.0 | memory length: 122780 | epsilon: 0.9548936200009792 |
| steps: 327 | lr: 0.0001 | evaluation reward: 1.53 | |
| episode: 678 | score: 1.0 | memory length: 122932 | epsilon: 0.9545926600009857 |
| steps: 152 | lr: 0.0001 | evaluation reward: 1.52 | |
| episode: 679 | score: 0.0 | memory length: 123054 | epsilon: 0.954351100000991 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.49 | |
| episode: 680 | score: 2.0 | memory length: 123271 | epsilon: 0.9539214400010003 |
| steps: 217 | lr: 0.0001 | evaluation reward: 1.5 | |
| episode: 681 | score: 2.0 | memory length: 123474 | epsilon: 0.953519500001009 |
| steps: 203 | lr: 0.0001 | evaluation reward: 1.51 | |
| episode: 682 | score: 4.0 | memory length: 123768 | epsilon: 0.9529373800010217 |

| | | |
|--------------|------------|--|
| steps: 294 | lr: 0.0001 | evaluation reward: 1.54 |
| episode: 683 | score: 0.0 | memory length: 123891 epsilon: 0.952693840001027 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.53 |
| episode: 684 | score: 2.0 | memory length: 124089 epsilon: 0.9523018000010355 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.53 |
| episode: 685 | score: 3.0 | memory length: 124317 epsilon: 0.9518503600010453 |
| steps: 228 | lr: 0.0001 | evaluation reward: 1.52 |
| episode: 686 | score: 0.0 | memory length: 124440 epsilon: 0.9516068200010506 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 687 | score: 1.0 | memory length: 124590 epsilon: 0.951309820001057 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.5 |
| episode: 688 | score: 3.0 | memory length: 124859 epsilon: 0.9507772000010686 |
| steps: 269 | lr: 0.0001 | evaluation reward: 1.53 |
| episode: 689 | score: 0.0 | memory length: 124982 epsilon: 0.9505336600010739 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.52 |
| episode: 690 | score: 4.0 | memory length: 125258 epsilon: 0.9499871800010857 |
| steps: 276 | lr: 0.0001 | evaluation reward: 1.54 |
| episode: 691 | score: 2.0 | memory length: 125455 epsilon: 0.9495971200010942 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.54 |
| episode: 692 | score: 4.0 | memory length: 125751 epsilon: 0.9490110400011069 |
| steps: 296 | lr: 0.0001 | evaluation reward: 1.58 |
| episode: 693 | score: 4.0 | memory length: 126066 epsilon: 0.9483873400011205 |
| steps: 315 | lr: 0.0001 | evaluation reward: 1.62 |
| episode: 694 | score: 0.0 | memory length: 126188 epsilon: 0.9481457800011257 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.62 |
| episode: 695 | score: 1.0 | memory length: 126357 epsilon: 0.947811160001133 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 696 | score: 3.0 | memory length: 126606 epsilon: 0.9473181400011437 |
| steps: 249 | lr: 0.0001 | evaluation reward: 1.62 |
| episode: 697 | score: 3.0 | memory length: 126831 epsilon: 0.9468726400011533 |
| steps: 225 | lr: 0.0001 | evaluation reward: 1.62 |
| episode: 698 | score: 1.0 | memory length: 126982 epsilon: 0.9465736600011598 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.62 |
| episode: 699 | score: 3.0 | memory length: 127229 epsilon: 0.9460846000011705 |
| steps: 247 | lr: 0.0001 | evaluation reward: 1.65 |
| episode: 700 | score: 1.0 | memory length: 127397 epsilon: 0.9457519600011777 |
| steps: 168 | lr: 0.0001 | evaluation reward: 1.64 |
| episode: 701 | score: 0.0 | memory length: 127519 epsilon: 0.9455104000011829 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.62 |
| episode: 702 | score: 0.0 | memory length: 127642 epsilon: 0.9452668600011882 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 703 | score: 2.0 | memory length: 127858 epsilon: 0.9448391800011975 |
| steps: 216 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 704 | score: 1.0 | memory length: 128009 epsilon: 0.944540200001204 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.6 |
| episode: 705 | score: 0.0 | memory length: 128132 epsilon: 0.9442966600012093 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.58 |
| episode: 706 | score: 1.0 | memory length: 128282 epsilon: 0.9439996600012157 |

| | | |
|--------------|------------|---|
| steps: 150 | lr: 0.0001 | evaluation reward: 1.59 |
| episode: 707 | score: 2.0 | memory length: 128479 epsilon: 0.9436096000012242 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.59 |
| episode: 708 | score: 1.0 | memory length: 128647 epsilon: 0.94327696000012314 |
| steps: 168 | lr: 0.0001 | evaluation reward: 1.57 |
| episode: 709 | score: 2.0 | memory length: 128844 epsilon: 0.94288690000012399 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.57 |
| episode: 710 | score: 4.0 | memory length: 129112 epsilon: 0.94235626000012514 |
| steps: 268 | lr: 0.0001 | evaluation reward: 1.57 |
| episode: 711 | score: 0.0 | memory length: 129235 epsilon: 0.94211272000012567 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.55 |
| episode: 712 | score: 1.0 | memory length: 129386 epsilon: 0.94181374000012632 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.56 |
| episode: 713 | score: 0.0 | memory length: 129508 epsilon: 0.94157218000012684 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.53 |
| episode: 714 | score: 2.0 | memory length: 129710 epsilon: 0.94117222000012771 |
| steps: 202 | lr: 0.0001 | evaluation reward: 1.55 |
| episode: 715 | score: 4.0 | memory length: 130025 epsilon: 0.94054852000012906 |
| steps: 315 | lr: 0.0001 | evaluation reward: 1.59 |
| episode: 716 | score: 2.0 | memory length: 130243 epsilon: 0.94011688000013 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.59 |
| episode: 717 | score: 1.0 | memory length: 130412 epsilon: 0.93978226000013073 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.59 |
| episode: 718 | score: 3.0 | memory length: 130681 epsilon: 0.93924964000013188 |
| steps: 269 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 719 | score: 2.0 | memory length: 130878 epsilon: 0.93885958000013273 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 720 | score: 3.0 | memory length: 131106 epsilon: 0.93840814000013371 |
| steps: 228 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 721 | score: 2.0 | memory length: 131325 epsilon: 0.93797452000013465 |
| steps: 219 | lr: 0.0001 | evaluation reward: 1.63 |
| episode: 722 | score: 7.0 | memory length: 131746 epsilon: 0.93714094000013646 |
| steps: 421 | lr: 0.0001 | evaluation reward: 1.7 |
| episode: 723 | score: 2.0 | memory length: 131946 epsilon: 0.93674494000013732 |
| steps: 200 | lr: 0.0001 | evaluation reward: 1.72 |
| episode: 724 | score: 2.0 | memory length: 132164 epsilon: 0.93631330000013826 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.73 |
| episode: 725 | score: 0.0 | memory length: 132287 epsilon: 0.93606976000013879 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 726 | score: 2.0 | memory length: 132484 epsilon: 0.93567970000013963 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.72 |
| episode: 727 | score: 5.0 | memory length: 132803 epsilon: 0.935048080000141 |
| steps: 319 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 728 | score: 0.0 | memory length: 132926 epsilon: 0.93480454000014153 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 729 | score: 2.0 | memory length: 133123 epsilon: 0.93441448000014238 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.78 |
| episode: 730 | score: 1.0 | memory length: 133293 epsilon: 0.93407788000014311 |

| | | |
|--------------|------------|--|
| steps: 170 | lr: 0.0001 | evaluation reward: 1.78 |
| episode: 731 | score: 0.0 | memory length: 133416 epsilon: 0.9338343400014364 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 732 | score: 1.0 | memory length: 133567 epsilon: 0.9335353600014429 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 733 | score: 4.0 | memory length: 133843 epsilon: 0.9329888800014547 |
| steps: 276 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 734 | score: 0.0 | memory length: 133966 epsilon: 0.93274534000146 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 735 | score: 3.0 | memory length: 134216 epsilon: 0.9322503400014708 |
| steps: 250 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 736 | score: 0.0 | memory length: 134338 epsilon: 0.932008780001476 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 737 | score: 0.0 | memory length: 134461 epsilon: 0.9317652400014813 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.73 |
| episode: 738 | score: 2.0 | memory length: 134679 epsilon: 0.9313336000014907 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 739 | score: 2.0 | memory length: 134877 epsilon: 0.9309415600014992 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 740 | score: 2.0 | memory length: 135095 epsilon: 0.9305099200015086 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 741 | score: 4.0 | memory length: 135391 epsilon: 0.9299238400015213 |
| steps: 296 | lr: 0.0001 | evaluation reward: 1.79 |
| episode: 742 | score: 1.0 | memory length: 135559 epsilon: 0.9295912000015285 |
| steps: 168 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 743 | score: 4.0 | memory length: 135833 epsilon: 0.9290486800015403 |
| steps: 274 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 744 | score: 9.0 | memory length: 136222 epsilon: 0.928278460001557 |
| steps: 389 | lr: 0.0001 | evaluation reward: 1.86 |
| episode: 745 | score: 2.0 | memory length: 136438 epsilon: 0.9278507800015663 |
| steps: 216 | lr: 0.0001 | evaluation reward: 1.85 |
| episode: 746 | score: 3.0 | memory length: 136666 epsilon: 0.9273993400015761 |
| steps: 228 | lr: 0.0001 | evaluation reward: 1.87 |
| episode: 747 | score: 2.0 | memory length: 136845 epsilon: 0.9270449200015838 |
| steps: 179 | lr: 0.0001 | evaluation reward: 1.87 |
| episode: 748 | score: 0.0 | memory length: 136967 epsilon: 0.926803360001589 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.86 |
| episode: 749 | score: 1.0 | memory length: 137135 epsilon: 0.9264707200015962 |
| steps: 168 | lr: 0.0001 | evaluation reward: 1.87 |
| episode: 750 | score: 0.0 | memory length: 137258 epsilon: 0.9262271800016015 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.87 |
| episode: 751 | score: 0.0 | memory length: 137381 epsilon: 0.9259836400016068 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.86 |
| episode: 752 | score: 2.0 | memory length: 137599 epsilon: 0.9255520000016162 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.87 |
| episode: 753 | score: 2.0 | memory length: 137817 epsilon: 0.9251203600016256 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.86 |
| episode: 754 | score: 3.0 | memory length: 138062 epsilon: 0.9246352600016361 |

| | | |
|--------------|------------|--|
| steps: 245 | lr: 0.0001 | evaluation reward: 1.87 |
| episode: 755 | score: 2.0 | memory length: 138277 epsilon: 0.9242095600016453 |
| steps: 215 | lr: 0.0001 | evaluation reward: 1.88 |
| episode: 756 | score: 1.0 | memory length: 138445 epsilon: 0.9238769200016526 |
| steps: 168 | lr: 0.0001 | evaluation reward: 1.88 |
| episode: 757 | score: 0.0 | memory length: 138567 epsilon: 0.9236353600016578 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.88 |
| episode: 758 | score: 0.0 | memory length: 138689 epsilon: 0.923393800001663 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.83 |
| episode: 759 | score: 2.0 | memory length: 138908 epsilon: 0.9229601800016725 |
| steps: 219 | lr: 0.0001 | evaluation reward: 1.83 |
| episode: 760 | score: 1.0 | memory length: 139077 epsilon: 0.9226255600016797 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.81 |
| episode: 761 | score: 4.0 | memory length: 139391 epsilon: 0.9220038400016932 |
| steps: 314 | lr: 0.0001 | evaluation reward: 1.79 |
| episode: 762 | score: 0.0 | memory length: 139513 epsilon: 0.9217622800016985 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.78 |
| episode: 763 | score: 1.0 | memory length: 139664 epsilon: 0.921463300001705 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.78 |
| episode: 764 | score: 3.0 | memory length: 139913 epsilon: 0.9209702800017157 |
| steps: 249 | lr: 0.0001 | evaluation reward: 1.81 |
| episode: 765 | score: 5.0 | memory length: 140223 epsilon: 0.920356480001729 |
| steps: 310 | lr: 0.0001 | evaluation reward: 1.86 |
| episode: 766 | score: 2.0 | memory length: 140443 epsilon: 0.9199208800017384 |
| steps: 220 | lr: 0.0001 | evaluation reward: 1.88 |
| episode: 767 | score: 3.0 | memory length: 140711 epsilon: 0.91939024000175 |
| steps: 268 | lr: 0.0001 | evaluation reward: 1.87 |
| episode: 768 | score: 3.0 | memory length: 140958 epsilon: 0.9189011800017606 |
| steps: 247 | lr: 0.0001 | evaluation reward: 1.9 |
| episode: 769 | score: 2.0 | memory length: 141175 epsilon: 0.9184715200017699 |
| steps: 217 | lr: 0.0001 | evaluation reward: 1.9 |
| episode: 770 | score: 2.0 | memory length: 141373 epsilon: 0.9180794800017784 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.91 |
| episode: 771 | score: 0.0 | memory length: 141496 epsilon: 0.9178359400017837 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.89 |
| episode: 772 | score: 0.0 | memory length: 141619 epsilon: 0.917592400001789 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.86 |
| episode: 773 | score: 4.0 | memory length: 141918 epsilon: 0.9170003800018018 |
| steps: 299 | lr: 0.0001 | evaluation reward: 1.9 |
| episode: 774 | score: 1.0 | memory length: 142088 epsilon: 0.9166637800018091 |
| steps: 170 | lr: 0.0001 | evaluation reward: 1.89 |
| episode: 775 | score: 4.0 | memory length: 142399 epsilon: 0.9160480000018225 |
| steps: 311 | lr: 0.0001 | evaluation reward: 1.91 |
| episode: 776 | score: 2.0 | memory length: 142578 epsilon: 0.9156935800018302 |
| steps: 179 | lr: 0.0001 | evaluation reward: 1.93 |
| episode: 777 | score: 0.0 | memory length: 142701 epsilon: 0.9154500400018355 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.87 |
| episode: 778 | score: 2.0 | memory length: 142899 epsilon: 0.915058000001844 |

| | | | |
|--------------|------------|-------------------------|-----------------------------|
| steps: 198 | lr: 0.0001 | evaluation reward: 1.88 | |
| episode: 779 | score: 2.0 | memory length: 143115 | epsilon: 0.9146303200018533 |
| steps: 216 | lr: 0.0001 | evaluation reward: 1.9 | |
| episode: 780 | score: 0.0 | memory length: 143238 | epsilon: 0.9143867800018586 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.88 | |
| episode: 781 | score: 2.0 | memory length: 143436 | epsilon: 0.9139947400018671 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.88 | |
| episode: 782 | score: 1.0 | memory length: 143607 | epsilon: 0.9136561600018744 |
| steps: 171 | lr: 0.0001 | evaluation reward: 1.85 | |
| episode: 783 | score: 2.0 | memory length: 143825 | epsilon: 0.9132245200018838 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.87 | |
| episode: 784 | score: 0.0 | memory length: 143947 | epsilon: 0.912982960001889 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.85 | |
| episode: 785 | score: 2.0 | memory length: 144144 | epsilon: 0.9125929000018975 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.84 | |
| episode: 786 | score: 0.0 | memory length: 144267 | epsilon: 0.9123493600019028 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.84 | |
| episode: 787 | score: 0.0 | memory length: 144390 | epsilon: 0.9121058200019081 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.83 | |
| episode: 788 | score: 1.0 | memory length: 144559 | epsilon: 0.9117712000019154 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.81 | |
| episode: 789 | score: 2.0 | memory length: 144757 | epsilon: 0.9113791600019239 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.83 | |
| episode: 790 | score: 0.0 | memory length: 144880 | epsilon: 0.9111356200019292 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.79 | |
| episode: 791 | score: 2.0 | memory length: 145078 | epsilon: 0.9107435800019377 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.79 | |
| episode: 792 | score: 0.0 | memory length: 145200 | epsilon: 0.9105020200019429 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.75 | |
| episode: 793 | score: 0.0 | memory length: 145322 | epsilon: 0.9102604600019482 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.71 | |
| episode: 794 | score: 0.0 | memory length: 145445 | epsilon: 0.9100169200019534 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.71 | |
| episode: 795 | score: 2.0 | memory length: 145663 | epsilon: 0.9095852800019628 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.72 | |
| episode: 796 | score: 2.0 | memory length: 145861 | epsilon: 0.9091932400019713 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.71 | |
| episode: 797 | score: 1.0 | memory length: 146032 | epsilon: 0.9088546600019787 |
| steps: 171 | lr: 0.0001 | evaluation reward: 1.69 | |
| episode: 798 | score: 2.0 | memory length: 146247 | epsilon: 0.9084289600019879 |
| steps: 215 | lr: 0.0001 | evaluation reward: 1.7 | |
| episode: 799 | score: 2.0 | memory length: 146445 | epsilon: 0.9080369200019964 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.69 | |
| episode: 800 | score: 0.0 | memory length: 146568 | epsilon: 0.9077933800020017 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.68 | |
| episode: 801 | score: 2.0 | memory length: 146766 | epsilon: 0.9074013400020102 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.7 | |
| episode: 802 | score: 4.0 | memory length: 147044 | epsilon: 0.9068509000020222 |

| | | |
|--------------|------------|--|
| steps: 278 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 803 | score: 2.0 | memory length: 147260 epsilon: 0.9064232200020315 |
| steps: 216 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 804 | score: 2.0 | memory length: 147461 epsilon: 0.9060252400020401 |
| steps: 201 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 805 | score: 0.0 | memory length: 147583 epsilon: 0.9057836800020453 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 806 | score: 2.0 | memory length: 147798 epsilon: 0.9053579800020546 |
| steps: 215 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 807 | score: 2.0 | memory length: 147996 epsilon: 0.9049659400020631 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 808 | score: 0.0 | memory length: 148119 epsilon: 0.9047224000020684 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 809 | score: 2.0 | memory length: 148316 epsilon: 0.9043323400020769 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 810 | score: 0.0 | memory length: 148438 epsilon: 0.9040907800020821 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 811 | score: 0.0 | memory length: 148560 epsilon: 0.9038492200020873 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 812 | score: 2.0 | memory length: 148777 epsilon: 0.9034195600020967 |
| steps: 217 | lr: 0.0001 | evaluation reward: 1.72 |
| episode: 813 | score: 3.0 | memory length: 149005 epsilon: 0.9029681200021065 |
| steps: 228 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 814 | score: 1.0 | memory length: 149177 epsilon: 0.9026275600021139 |
| steps: 172 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 815 | score: 3.0 | memory length: 149408 epsilon: 0.9021701800021238 |
| steps: 231 | lr: 0.0001 | evaluation reward: 1.73 |
| episode: 816 | score: 2.0 | memory length: 149606 epsilon: 0.9017781400021323 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.73 |
| episode: 817 | score: 3.0 | memory length: 149831 epsilon: 0.901332640002142 |
| steps: 225 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 818 | score: 1.0 | memory length: 150003 epsilon: 0.9009920800021494 |
| steps: 172 | lr: 0.0001 | evaluation reward: 1.73 |
| episode: 819 | score: 2.0 | memory length: 150183 epsilon: 0.9006356800021571 |
| steps: 180 | lr: 0.0001 | evaluation reward: 1.73 |
| episode: 820 | score: 8.0 | memory length: 150530 epsilon: 0.899948620002172 |
| steps: 347 | lr: 0.0001 | evaluation reward: 1.78 |
| episode: 821 | score: 3.0 | memory length: 150773 epsilon: 0.8994674800021825 |
| steps: 243 | lr: 0.0001 | evaluation reward: 1.79 |
| episode: 822 | score: 2.0 | memory length: 150989 epsilon: 0.8990398000021917 |
| steps: 216 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 823 | score: 0.0 | memory length: 151112 epsilon: 0.898796260002197 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.72 |
| episode: 824 | score: 1.0 | memory length: 151263 epsilon: 0.8984972800022035 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 825 | score: 2.0 | memory length: 151461 epsilon: 0.898105240002212 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.73 |
| episode: 826 | score: 3.0 | memory length: 151708 epsilon: 0.8976161800022227 |

| | | |
|--------------|------------|--|
| steps: 247 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 827 | score: 2.0 | memory length: 151909 epsilon: 0.8972182000022313 |
| steps: 201 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 828 | score: 4.0 | memory length: 152203 epsilon: 0.8966360800022439 |
| steps: 294 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 829 | score: 3.0 | memory length: 152431 epsilon: 0.8961846400022537 |
| steps: 228 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 830 | score: 2.0 | memory length: 152648 epsilon: 0.8957549800022631 |
| steps: 217 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 831 | score: 0.0 | memory length: 152771 epsilon: 0.8955114400022683 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 832 | score: 0.0 | memory length: 152894 epsilon: 0.8952679000022736 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 833 | score: 3.0 | memory length: 153140 epsilon: 0.8947808200022842 |
| steps: 246 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 834 | score: 1.0 | memory length: 153309 epsilon: 0.8944462000022915 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 835 | score: 2.0 | memory length: 153506 epsilon: 0.8940561400022999 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 836 | score: 2.0 | memory length: 153685 epsilon: 0.8937017200023076 |
| steps: 179 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 837 | score: 1.0 | memory length: 153853 epsilon: 0.8933690800023149 |
| steps: 168 | lr: 0.0001 | evaluation reward: 1.78 |
| episode: 838 | score: 3.0 | memory length: 154101 epsilon: 0.8928780400023255 |
| steps: 248 | lr: 0.0001 | evaluation reward: 1.79 |
| episode: 839 | score: 0.0 | memory length: 154224 epsilon: 0.8926345000023308 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 840 | score: 3.0 | memory length: 154468 epsilon: 0.8921513800023413 |
| steps: 244 | lr: 0.0001 | evaluation reward: 1.78 |
| episode: 841 | score: 1.0 | memory length: 154618 epsilon: 0.8918543800023477 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 842 | score: 1.0 | memory length: 154788 epsilon: 0.891517780002355 |
| steps: 170 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 843 | score: 2.0 | memory length: 154985 epsilon: 0.8911277200023635 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.73 |
| episode: 844 | score: 2.0 | memory length: 155183 epsilon: 0.890735680002372 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.66 |
| episode: 845 | score: 3.0 | memory length: 155409 epsilon: 0.8902882000023817 |
| steps: 226 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 846 | score: 0.0 | memory length: 155532 epsilon: 0.890044660002387 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.64 |
| episode: 847 | score: 2.0 | memory length: 155730 epsilon: 0.8896526200023955 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.64 |
| episode: 848 | score: 3.0 | memory length: 155956 epsilon: 0.8892051400024052 |
| steps: 226 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 849 | score: 4.0 | memory length: 156249 epsilon: 0.8886250000024178 |
| steps: 293 | lr: 0.0001 | evaluation reward: 1.7 |
| episode: 850 | score: 0.0 | memory length: 156372 epsilon: 0.8883814600024231 |

| | | |
|--------------|------------|--|
| steps: 123 | lr: 0.0001 | evaluation reward: 1.7 |
| episode: 851 | score: 7.0 | memory length: 156784 epsilon: 0.8875657000024408 |
| steps: 412 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 852 | score: 2.0 | memory length: 156982 epsilon: 0.8871736600024493 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 853 | score: 0.0 | memory length: 157105 epsilon: 0.8869301200024546 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 854 | score: 2.0 | memory length: 157303 epsilon: 0.8865380800024631 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 855 | score: 2.0 | memory length: 157502 epsilon: 0.8861440600024717 |
| steps: 199 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 856 | score: 2.0 | memory length: 157699 epsilon: 0.8857540000024802 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 857 | score: 1.0 | memory length: 157868 epsilon: 0.8854193800024874 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 858 | score: 1.0 | memory length: 158039 epsilon: 0.8850808000024948 |
| steps: 171 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 859 | score: 1.0 | memory length: 158208 epsilon: 0.884746180002502 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 860 | score: 3.0 | memory length: 158456 epsilon: 0.8842551400025127 |
| steps: 248 | lr: 0.0001 | evaluation reward: 1.78 |
| episode: 861 | score: 2.0 | memory length: 158654 epsilon: 0.8838631000025212 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 862 | score: 3.0 | memory length: 158898 epsilon: 0.8833799800025317 |
| steps: 244 | lr: 0.0001 | evaluation reward: 1.79 |
| episode: 863 | score: 2.0 | memory length: 159095 epsilon: 0.8829899200025402 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.8 |
| episode: 864 | score: 0.0 | memory length: 159218 epsilon: 0.8827463800025455 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 865 | score: 2.0 | memory length: 159416 epsilon: 0.882354340002554 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 866 | score: 5.0 | memory length: 159718 epsilon: 0.881756380002567 |
| steps: 302 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 867 | score: 3.0 | memory length: 159989 epsilon: 0.8812198000025786 |
| steps: 271 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 868 | score: 2.0 | memory length: 160207 epsilon: 0.880788160002588 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 869 | score: 2.0 | memory length: 160386 epsilon: 0.8804337400025957 |
| steps: 179 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 870 | score: 0.0 | memory length: 160508 epsilon: 0.8801921800026009 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 871 | score: 2.0 | memory length: 160706 epsilon: 0.8798001400026094 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 872 | score: 2.0 | memory length: 160906 epsilon: 0.879404140002618 |
| steps: 200 | lr: 0.0001 | evaluation reward: 1.78 |
| episode: 873 | score: 0.0 | memory length: 161029 epsilon: 0.8791606000026233 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 874 | score: 2.0 | memory length: 161226 epsilon: 0.8787705400026318 |

| | | |
|--------------|------------|---|
| steps: 197 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 875 | score: 0.0 | memory length: 161348 epsilon: 0.878528980002637 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 876 | score: 0.0 | memory length: 161470 epsilon: 0.8782874200026423 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.69 |
| episode: 877 | score: 2.0 | memory length: 161668 epsilon: 0.8778953800026508 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 878 | score: 2.0 | memory length: 161868 epsilon: 0.8774993800026594 |
| steps: 200 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 879 | score: 0.0 | memory length: 161990 epsilon: 0.8772578200026646 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.69 |
| episode: 880 | score: 0.0 | memory length: 162113 epsilon: 0.8770142800026699 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.69 |
| episode: 881 | score: 3.0 | memory length: 162362 epsilon: 0.8765212600026806 |
| steps: 249 | lr: 0.0001 | evaluation reward: 1.7 |
| episode: 882 | score: 2.0 | memory length: 162580 epsilon: 0.87608962000269 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 883 | score: 1.0 | memory length: 162731 epsilon: 0.8757906400026965 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.7 |
| episode: 884 | score: 2.0 | memory length: 162931 epsilon: 0.8753946400027051 |
| steps: 200 | lr: 0.0001 | evaluation reward: 1.72 |
| episode: 885 | score: 1.0 | memory length: 163103 epsilon: 0.8750540800027125 |
| steps: 172 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 886 | score: 2.0 | memory length: 163283 epsilon: 0.8746976800027202 |
| steps: 180 | lr: 0.0001 | evaluation reward: 1.73 |
| episode: 887 | score: 3.0 | memory length: 163530 epsilon: 0.8742086200027308 |
| steps: 247 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 888 | score: 0.0 | memory length: 163653 epsilon: 0.8739650800027361 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 889 | score: 2.0 | memory length: 163855 epsilon: 0.8735651200027448 |
| steps: 202 | lr: 0.0001 | evaluation reward: 1.75 |
| episode: 890 | score: 1.0 | memory length: 164005 epsilon: 0.8732681200027512 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.76 |
| episode: 891 | score: 0.0 | memory length: 164128 epsilon: 0.8730245800027565 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 892 | score: 0.0 | memory length: 164251 epsilon: 0.8727810400027618 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 893 | score: 4.0 | memory length: 164508 epsilon: 0.8722721800027728 |
| steps: 257 | lr: 0.0001 | evaluation reward: 1.78 |
| episode: 894 | score: 1.0 | memory length: 164678 epsilon: 0.8719355800027802 |
| steps: 170 | lr: 0.0001 | evaluation reward: 1.79 |
| episode: 895 | score: 2.0 | memory length: 164875 epsilon: 0.8715455200027886 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.79 |
| episode: 896 | score: 2.0 | memory length: 165073 epsilon: 0.8711534800027971 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.79 |
| episode: 897 | score: 2.0 | memory length: 165270 epsilon: 0.8707634200028056 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.8 |
| episode: 898 | score: 3.0 | memory length: 165539 epsilon: 0.8702308000028172 |

| | | |
|--------------|------------|--|
| steps: 269 | lr: 0.0001 | evaluation reward: 1.81 |
| episode: 899 | score: 2.0 | memory length: 165737 epsilon: 0.8698387600028257 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.81 |
| episode: 900 | score: 2.0 | memory length: 165934 epsilon: 0.8694487000028341 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.83 |
| episode: 901 | score: 0.0 | memory length: 166057 epsilon: 0.8692051600028394 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.81 |
| episode: 902 | score: 3.0 | memory length: 166286 epsilon: 0.8687517400028493 |
| steps: 229 | lr: 0.0001 | evaluation reward: 1.8 |
| episode: 903 | score: 2.0 | memory length: 166485 epsilon: 0.8683577200028578 |
| steps: 199 | lr: 0.0001 | evaluation reward: 1.8 |
| episode: 904 | score: 0.0 | memory length: 166608 epsilon: 0.8681141800028631 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.78 |
| episode: 905 | score: 1.0 | memory length: 166777 epsilon: 0.8677795600028704 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.79 |
| episode: 906 | score: 0.0 | memory length: 166900 epsilon: 0.8675360200028757 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 907 | score: 2.0 | memory length: 167097 epsilon: 0.8671459600028841 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.77 |
| episode: 908 | score: 2.0 | memory length: 167294 epsilon: 0.8667559000028926 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.79 |
| episode: 909 | score: 3.0 | memory length: 167539 epsilon: 0.8662708000029031 |
| steps: 245 | lr: 0.0001 | evaluation reward: 1.8 |
| episode: 910 | score: 3.0 | memory length: 167767 epsilon: 0.8658193600029129 |
| steps: 228 | lr: 0.0001 | evaluation reward: 1.83 |
| episode: 911 | score: 3.0 | memory length: 167993 epsilon: 0.8653718800029226 |
| steps: 226 | lr: 0.0001 | evaluation reward: 1.86 |
| episode: 912 | score: 4.0 | memory length: 168290 epsilon: 0.8647838200029354 |
| steps: 297 | lr: 0.0001 | evaluation reward: 1.88 |
| episode: 913 | score: 1.0 | memory length: 168441 epsilon: 0.8644848400029419 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.86 |
| episode: 914 | score: 1.0 | memory length: 168610 epsilon: 0.8641502200029492 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.86 |
| episode: 915 | score: 2.0 | memory length: 168808 epsilon: 0.8637581800029577 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.85 |
| episode: 916 | score: 0.0 | memory length: 168931 epsilon: 0.863514640002963 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.83 |
| episode: 917 | score: 2.0 | memory length: 169130 epsilon: 0.8631206200029715 |
| steps: 199 | lr: 0.0001 | evaluation reward: 1.82 |
| episode: 918 | score: 0.0 | memory length: 169253 epsilon: 0.8628770800029768 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.81 |
| episode: 919 | score: 0.0 | memory length: 169375 epsilon: 0.862635520002982 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.79 |
| episode: 920 | score: 0.0 | memory length: 169497 epsilon: 0.8623939600029873 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 921 | score: 1.0 | memory length: 169648 epsilon: 0.8620949800029938 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.69 |
| episode: 922 | score: 0.0 | memory length: 169771 epsilon: 0.8618514400029991 |

| | | |
|--------------|------------|--|
| steps: 123 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 923 | score: 0.0 | memory length: 169894 epsilon: 0.8616079000030044 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 924 | score: 1.0 | memory length: 170066 epsilon: 0.8612673400030117 |
| steps: 172 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 925 | score: 6.0 | memory length: 170463 epsilon: 0.8604812800030288 |
| steps: 397 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 926 | score: 0.0 | memory length: 170585 epsilon: 0.8602397200030341 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.68 |
| episode: 927 | score: 0.0 | memory length: 170708 epsilon: 0.8599961800030393 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.66 |
| episode: 928 | score: 3.0 | memory length: 170955 epsilon: 0.85950712000305 |
| steps: 247 | lr: 0.0001 | evaluation reward: 1.65 |
| episode: 929 | score: 0.0 | memory length: 171077 epsilon: 0.8592655600030552 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.62 |
| episode: 930 | score: 1.0 | memory length: 171227 epsilon: 0.8589685600030617 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 931 | score: 3.0 | memory length: 171453 epsilon: 0.8585210800030714 |
| steps: 226 | lr: 0.0001 | evaluation reward: 1.64 |
| episode: 932 | score: 3.0 | memory length: 171698 epsilon: 0.8580359800030819 |
| steps: 245 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 933 | score: 2.0 | memory length: 171920 epsilon: 0.8575964200030914 |
| steps: 222 | lr: 0.0001 | evaluation reward: 1.66 |
| episode: 934 | score: 3.0 | memory length: 172164 epsilon: 0.8571133000031019 |
| steps: 244 | lr: 0.0001 | evaluation reward: 1.68 |
| episode: 935 | score: 0.0 | memory length: 172286 epsilon: 0.8568717400031072 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.66 |
| episode: 936 | score: 2.0 | memory length: 172483 epsilon: 0.8564816800031156 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.66 |
| episode: 937 | score: 0.0 | memory length: 172606 epsilon: 0.8562381400031209 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.65 |
| episode: 938 | score: 3.0 | memory length: 172831 epsilon: 0.8557926400031306 |
| steps: 225 | lr: 0.0001 | evaluation reward: 1.65 |
| episode: 939 | score: 3.0 | memory length: 173077 epsilon: 0.8553055600031412 |
| steps: 246 | lr: 0.0001 | evaluation reward: 1.68 |
| episode: 940 | score: 1.0 | memory length: 173248 epsilon: 0.8549669800031485 |
| steps: 171 | lr: 0.0001 | evaluation reward: 1.66 |
| episode: 941 | score: 3.0 | memory length: 173494 epsilon: 0.8544799000031591 |
| steps: 246 | lr: 0.0001 | evaluation reward: 1.68 |
| episode: 942 | score: 2.0 | memory length: 173692 epsilon: 0.8540878600031676 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.69 |
| episode: 943 | score: 1.0 | memory length: 173864 epsilon: 0.853747300003175 |
| steps: 172 | lr: 0.0001 | evaluation reward: 1.68 |
| episode: 944 | score: 3.0 | memory length: 174112 epsilon: 0.8532562600031857 |
| steps: 248 | lr: 0.0001 | evaluation reward: 1.69 |
| episode: 945 | score: 1.0 | memory length: 174283 epsilon: 0.852917680003193 |
| steps: 171 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 946 | score: 0.0 | memory length: 174405 epsilon: 0.8526761200031983 |

| | | |
|--------------|------------|--|
| steps: 122 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 947 | score: 2.0 | memory length: 174602 epsilon: 0.8522860600032067 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 948 | score: 2.0 | memory length: 174800 epsilon: 0.8518940200032152 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.66 |
| episode: 949 | score: 0.0 | memory length: 174922 epsilon: 0.8516524600032205 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.62 |
| episode: 950 | score: 3.0 | memory length: 175186 epsilon: 0.8511297400032318 |
| steps: 264 | lr: 0.0001 | evaluation reward: 1.65 |
| episode: 951 | score: 0.0 | memory length: 175308 epsilon: 0.8508881800032371 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.58 |
| episode: 952 | score: 3.0 | memory length: 175535 epsilon: 0.8504387200032468 |
| steps: 227 | lr: 0.0001 | evaluation reward: 1.59 |
| episode: 953 | score: 0.0 | memory length: 175658 epsilon: 0.8501951800032521 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.59 |
| episode: 954 | score: 3.0 | memory length: 175885 epsilon: 0.8497457200032619 |
| steps: 227 | lr: 0.0001 | evaluation reward: 1.6 |
| episode: 955 | score: 1.0 | memory length: 176054 epsilon: 0.8494111000032691 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.59 |
| episode: 956 | score: 2.0 | memory length: 176272 epsilon: 0.8489794600032785 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.59 |
| episode: 957 | score: 3.0 | memory length: 176539 epsilon: 0.84845080000329 |
| steps: 267 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 958 | score: 3.0 | memory length: 176764 epsilon: 0.8480053000032997 |
| steps: 225 | lr: 0.0001 | evaluation reward: 1.63 |
| episode: 959 | score: 1.0 | memory length: 176933 epsilon: 0.8476706800033069 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.63 |
| episode: 960 | score: 1.0 | memory length: 177083 epsilon: 0.8473736800033134 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 961 | score: 2.0 | memory length: 177265 epsilon: 0.8470133200033212 |
| steps: 182 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 962 | score: 4.0 | memory length: 177559 epsilon: 0.8464312000033338 |
| steps: 294 | lr: 0.0001 | evaluation reward: 1.62 |
| episode: 963 | score: 0.0 | memory length: 177681 epsilon: 0.8461896400033391 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.6 |
| episode: 964 | score: 1.0 | memory length: 177852 epsilon: 0.8458510600033464 |
| steps: 171 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 965 | score: 2.0 | memory length: 178049 epsilon: 0.8454610000033549 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 966 | score: 1.0 | memory length: 178218 epsilon: 0.8451263800033622 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.57 |
| episode: 967 | score: 3.0 | memory length: 178464 epsilon: 0.8446393000033727 |
| steps: 246 | lr: 0.0001 | evaluation reward: 1.57 |
| episode: 968 | score: 3.0 | memory length: 178710 epsilon: 0.8441522200033833 |
| steps: 246 | lr: 0.0001 | evaluation reward: 1.58 |
| episode: 969 | score: 0.0 | memory length: 178832 epsilon: 0.8439106600033885 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.56 |
| episode: 970 | score: 2.0 | memory length: 179029 epsilon: 0.843520600003397 |

| | | |
|--------------|------------|--|
| steps: 197 | lr: 0.0001 | evaluation reward: 1.58 |
| episode: 971 | score: 2.0 | memory length: 179227 epsilon: 0.8431285600034055 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.58 |
| episode: 972 | score: 0.0 | memory length: 179349 epsilon: 0.8428870000034108 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.56 |
| episode: 973 | score: 2.0 | memory length: 179570 epsilon: 0.8424494200034203 |
| steps: 221 | lr: 0.0001 | evaluation reward: 1.58 |
| episode: 974 | score: 2.0 | memory length: 179767 epsilon: 0.8420593600034287 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.58 |
| episode: 975 | score: 1.0 | memory length: 179935 epsilon: 0.841726720003436 |
| steps: 168 | lr: 0.0001 | evaluation reward: 1.59 |
| episode: 976 | score: 4.0 | memory length: 180231 epsilon: 0.8411406400034487 |
| steps: 296 | lr: 0.0001 | evaluation reward: 1.63 |
| episode: 977 | score: 0.0 | memory length: 180354 epsilon: 0.840897100003454 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 978 | score: 1.0 | memory length: 180523 epsilon: 0.8405624800034612 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.6 |
| episode: 979 | score: 1.0 | memory length: 180673 epsilon: 0.8402654800034677 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 980 | score: 2.0 | memory length: 180871 epsilon: 0.8398734400034762 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.63 |
| episode: 981 | score: 1.0 | memory length: 181022 epsilon: 0.8395744600034827 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.61 |
| episode: 982 | score: 5.0 | memory length: 181348 epsilon: 0.8389289800034967 |
| steps: 326 | lr: 0.0001 | evaluation reward: 1.64 |
| episode: 983 | score: 4.0 | memory length: 181644 epsilon: 0.8383429000035094 |
| steps: 296 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 984 | score: 2.0 | memory length: 181841 epsilon: 0.8379528400035179 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 985 | score: 1.0 | memory length: 181992 epsilon: 0.8376538600035244 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 986 | score: 2.0 | memory length: 182208 epsilon: 0.8372261800035337 |
| steps: 216 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 987 | score: 3.0 | memory length: 182434 epsilon: 0.8367787000035434 |
| steps: 226 | lr: 0.0001 | evaluation reward: 1.67 |
| episode: 988 | score: 2.0 | memory length: 182632 epsilon: 0.8363866600035519 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.69 |
| episode: 989 | score: 3.0 | memory length: 182858 epsilon: 0.8359391800035616 |
| steps: 226 | lr: 0.0001 | evaluation reward: 1.7 |
| episode: 990 | score: 2.0 | memory length: 183057 epsilon: 0.8355451600035702 |
| steps: 199 | lr: 0.0001 | evaluation reward: 1.71 |
| episode: 991 | score: 1.0 | memory length: 183226 epsilon: 0.8352105400035774 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.72 |
| episode: 992 | score: 2.0 | memory length: 183447 epsilon: 0.8347729600035869 |
| steps: 221 | lr: 0.0001 | evaluation reward: 1.74 |
| episode: 993 | score: 3.0 | memory length: 183676 epsilon: 0.8343195400035968 |
| steps: 229 | lr: 0.0001 | evaluation reward: 1.73 |
| episode: 994 | score: 0.0 | memory length: 183799 epsilon: 0.834076000003602 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 123 | lr: 0.0001 | evaluation reward: 1.72 | |
| episode: 995 | score: 3.0 | memory length: 184070 | epsilon: 0.8335394200036137 |
| steps: 271 | lr: 0.0001 | evaluation reward: 1.73 | |
| episode: 996 | score: 4.0 | memory length: 184346 | epsilon: 0.8329929400036256 |
| steps: 276 | lr: 0.0001 | evaluation reward: 1.75 | |
| episode: 997 | score: 1.0 | memory length: 184517 | epsilon: 0.8326543600036329 |
| steps: 171 | lr: 0.0001 | evaluation reward: 1.74 | |
| episode: 998 | score: 4.0 | memory length: 184809 | epsilon: 0.8320762000036455 |
| steps: 292 | lr: 0.0001 | evaluation reward: 1.75 | |
| episode: 999 | score: 1.0 | memory length: 184981 | epsilon: 0.8317356400036529 |
| steps: 172 | lr: 0.0001 | evaluation reward: 1.74 | |
| episode: 1000 | score: 1.0 | memory length: 185150 | epsilon: 0.8314010200036601 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.73 | |
| episode: 1001 | score: 0.0 | memory length: 185272 | epsilon: 0.8311594600036654 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.73 | |
| episode: 1002 | score: 0.0 | memory length: 185395 | epsilon: 0.8309159200036706 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.7 | |
| episode: 1003 | score: 2.0 | memory length: 185592 | epsilon: 0.8305258600036791 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.7 | |
| episode: 1004 | score: 3.0 | memory length: 185856 | epsilon: 0.8300031400036905 |
| steps: 264 | lr: 0.0001 | evaluation reward: 1.73 | |
| episode: 1005 | score: 2.0 | memory length: 186056 | epsilon: 0.8296071400036991 |
| steps: 200 | lr: 0.0001 | evaluation reward: 1.74 | |
| episode: 1006 | score: 3.0 | memory length: 186301 | epsilon: 0.8291220400037096 |
| steps: 245 | lr: 0.0001 | evaluation reward: 1.77 | |
| episode: 1007 | score: 3.0 | memory length: 186526 | epsilon: 0.8286765400037193 |
| steps: 225 | lr: 0.0001 | evaluation reward: 1.78 | |
| episode: 1008 | score: 2.0 | memory length: 186744 | epsilon: 0.8282449000037286 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.78 | |
| episode: 1009 | score: 2.0 | memory length: 186942 | epsilon: 0.8278528600037371 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.77 | |
| episode: 1010 | score: 3.0 | memory length: 187189 | epsilon: 0.8273638000037478 |
| steps: 247 | lr: 0.0001 | evaluation reward: 1.77 | |
| episode: 1011 | score: 2.0 | memory length: 187386 | epsilon: 0.8269737400037562 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.76 | |
| episode: 1012 | score: 1.0 | memory length: 187536 | epsilon: 0.8266767400037627 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.73 | |
| episode: 1013 | score: 0.0 | memory length: 187659 | epsilon: 0.826433200003768 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.72 | |
| episode: 1014 | score: 1.0 | memory length: 187809 | epsilon: 0.8261362000037744 |
| steps: 150 | lr: 0.0001 | evaluation reward: 1.72 | |
| episode: 1015 | score: 3.0 | memory length: 188034 | epsilon: 0.8256907000037841 |
| steps: 225 | lr: 0.0001 | evaluation reward: 1.73 | |
| episode: 1016 | score: 1.0 | memory length: 188185 | epsilon: 0.8253917200037906 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.74 | |
| episode: 1017 | score: 3.0 | memory length: 188434 | epsilon: 0.8248987000038013 |
| steps: 249 | lr: 0.0001 | evaluation reward: 1.75 | |
| episode: 1018 | score: 2.0 | memory length: 188634 | epsilon: 0.8245027000038099 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 200 | lr: 0.0001 | evaluation reward: 1.77 | |
| episode: 1019 | score: 2.0 | memory length: 188831 | epsilon: 0.8241126400038183 |
| steps: 197 | lr: 0.0001 | evaluation reward: 1.79 | |
| episode: 1020 | score: 3.0 | memory length: 189062 | epsilon: 0.8236552600038283 |
| steps: 231 | lr: 0.0001 | evaluation reward: 1.82 | |
| episode: 1021 | score: 1.0 | memory length: 189213 | epsilon: 0.8233562800038348 |
| steps: 151 | lr: 0.0001 | evaluation reward: 1.82 | |
| episode: 1022 | score: 4.0 | memory length: 189490 | epsilon: 0.8228078200038467 |
| steps: 277 | lr: 0.0001 | evaluation reward: 1.86 | |
| episode: 1023 | score: 0.0 | memory length: 189613 | epsilon: 0.822564280003852 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.86 | |
| episode: 1024 | score: 4.0 | memory length: 189905 | epsilon: 0.8219861200038645 |
| steps: 292 | lr: 0.0001 | evaluation reward: 1.89 | |
| episode: 1025 | score: 4.0 | memory length: 190179 | epsilon: 0.8214436000038763 |
| steps: 274 | lr: 0.0001 | evaluation reward: 1.87 | |
| episode: 1026 | score: 2.0 | memory length: 190377 | epsilon: 0.8210515600038848 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.89 | |
| episode: 1027 | score: 3.0 | memory length: 190646 | epsilon: 0.8205189400038964 |
| steps: 269 | lr: 0.0001 | evaluation reward: 1.92 | |
| episode: 1028 | score: 0.0 | memory length: 190769 | epsilon: 0.8202754000039016 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.89 | |
| episode: 1029 | score: 2.0 | memory length: 190987 | epsilon: 0.819843760003911 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.91 | |
| episode: 1030 | score: 0.0 | memory length: 191109 | epsilon: 0.8196022000039163 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.9 | |
| episode: 1031 | score: 4.0 | memory length: 191404 | epsilon: 0.8190181000039289 |
| steps: 295 | lr: 0.0001 | evaluation reward: 1.91 | |
| episode: 1032 | score: 3.0 | memory length: 191649 | epsilon: 0.8185330000039395 |
| steps: 245 | lr: 0.0001 | evaluation reward: 1.91 | |
| episode: 1033 | score: 2.0 | memory length: 191866 | epsilon: 0.8181033400039488 |
| steps: 217 | lr: 0.0001 | evaluation reward: 1.91 | |
| episode: 1034 | score: 3.0 | memory length: 192135 | epsilon: 0.8175707200039604 |
| steps: 269 | lr: 0.0001 | evaluation reward: 1.91 | |
| episode: 1035 | score: 0.0 | memory length: 192257 | epsilon: 0.8173291600039656 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.91 | |
| episode: 1036 | score: 4.0 | memory length: 192516 | epsilon: 0.8168163400039767 |
| steps: 259 | lr: 0.0001 | evaluation reward: 1.93 | |
| episode: 1037 | score: 2.0 | memory length: 192734 | epsilon: 0.8163847000039861 |
| steps: 218 | lr: 0.0001 | evaluation reward: 1.95 | |
| episode: 1038 | score: 6.0 | memory length: 193114 | epsilon: 0.8156323000040024 |
| steps: 380 | lr: 0.0001 | evaluation reward: 1.98 | |
| episode: 1039 | score: 0.0 | memory length: 193236 | epsilon: 0.8153907400040077 |
| steps: 122 | lr: 0.0001 | evaluation reward: 1.95 | |
| episode: 1040 | score: 3.0 | memory length: 193484 | epsilon: 0.8148997000040183 |
| steps: 248 | lr: 0.0001 | evaluation reward: 1.97 | |
| episode: 1041 | score: 2.0 | memory length: 193682 | epsilon: 0.8145076600040269 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.96 | |
| episode: 1042 | score: 4.0 | memory length: 193998 | epsilon: 0.8138819800040404 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 316 | lr: 0.0001 | evaluation reward: 1.98 | |
| episode: 1043 | score: 0.0 | memory length: 194121 | epsilon: 0.8136384400040457 |
| steps: 123 | lr: 0.0001 | evaluation reward: 1.97 | |
| episode: 1044 | score: 1.0 | memory length: 194290 | epsilon: 0.813303820004053 |
| steps: 169 | lr: 0.0001 | evaluation reward: 1.95 | |
| episode: 1045 | score: 3.0 | memory length: 194516 | epsilon: 0.8128563400040627 |
| steps: 226 | lr: 0.0001 | evaluation reward: 1.97 | |
| episode: 1046 | score: 2.0 | memory length: 194714 | epsilon: 0.8124643000040712 |
| steps: 198 | lr: 0.0001 | evaluation reward: 1.99 | |
| episode: 1047 | score: 4.0 | memory length: 194989 | epsilon: 0.811919800004083 |
| steps: 275 | lr: 0.0001 | evaluation reward: 2.01 | |
| episode: 1048 | score: 1.0 | memory length: 195140 | epsilon: 0.8116208200040895 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.0 | |
| episode: 1049 | score: 1.0 | memory length: 195291 | epsilon: 0.811321840004096 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.01 | |
| episode: 1050 | score: 3.0 | memory length: 195537 | epsilon: 0.8108347600041066 |
| steps: 246 | lr: 0.0001 | evaluation reward: 2.01 | |
| episode: 1051 | score: 2.0 | memory length: 195734 | epsilon: 0.8104447000041151 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.03 | |
| episode: 1052 | score: 4.0 | memory length: 196029 | epsilon: 0.8098606000041277 |
| steps: 295 | lr: 0.0001 | evaluation reward: 2.04 | |
| episode: 1053 | score: 2.0 | memory length: 196227 | epsilon: 0.8094685600041363 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.06 | |
| episode: 1054 | score: 2.0 | memory length: 196448 | epsilon: 0.8090309800041457 |
| steps: 221 | lr: 0.0001 | evaluation reward: 2.05 | |
| episode: 1055 | score: 1.0 | memory length: 196616 | epsilon: 0.808698340004153 |
| steps: 168 | lr: 0.0001 | evaluation reward: 2.05 | |
| episode: 1056 | score: 2.0 | memory length: 196816 | epsilon: 0.8083023400041616 |
| steps: 200 | lr: 0.0001 | evaluation reward: 2.05 | |
| episode: 1057 | score: 4.0 | memory length: 197082 | epsilon: 0.807775660004173 |
| steps: 266 | lr: 0.0001 | evaluation reward: 2.06 | |
| episode: 1058 | score: 2.0 | memory length: 197279 | epsilon: 0.8073856000041815 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.05 | |
| episode: 1059 | score: 2.0 | memory length: 197476 | epsilon: 0.8069955400041899 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.06 | |
| episode: 1060 | score: 2.0 | memory length: 197674 | epsilon: 0.8066035000041984 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.07 | |
| episode: 1061 | score: 0.0 | memory length: 197797 | epsilon: 0.8063599600042037 |
| steps: 123 | lr: 0.0001 | evaluation reward: 2.05 | |
| episode: 1062 | score: 2.0 | memory length: 197995 | epsilon: 0.8059679200042122 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.03 | |
| episode: 1063 | score: 4.0 | memory length: 198252 | epsilon: 0.8054590600042233 |
| steps: 257 | lr: 0.0001 | evaluation reward: 2.07 | |
| episode: 1064 | score: 3.0 | memory length: 198478 | epsilon: 0.805011580004233 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.09 | |
| episode: 1065 | score: 2.0 | memory length: 198679 | epsilon: 0.8046136000042416 |
| steps: 201 | lr: 0.0001 | evaluation reward: 2.09 | |
| episode: 1066 | score: 0.0 | memory length: 198802 | epsilon: 0.8043700600042469 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 123 | lr: 0.0001 | evaluation reward: 2.08 | |
| episode: 1067 | score: 4.0 | memory length: 199117 | epsilon: 0.8037463600042605 |
| steps: 315 | lr: 0.0001 | evaluation reward: 2.09 | |
| episode: 1068 | score: 2.0 | memory length: 199315 | epsilon: 0.803354320004269 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.08 | |
| episode: 1069 | score: 3.0 | memory length: 199562 | epsilon: 0.8028652600042796 |
| steps: 247 | lr: 0.0001 | evaluation reward: 2.11 | |
| episode: 1070 | score: 5.0 | memory length: 199863 | epsilon: 0.8022692800042925 |
| steps: 301 | lr: 0.0001 | evaluation reward: 2.14 | |
| episode: 1071 | score: 3.0 | memory length: 200088 | epsilon: 0.8018237800043022 |
| steps: 225 | lr: 0.0001 | evaluation reward: 2.15 | |
| episode: 1072 | score: 2.0 | memory length: 200285 | epsilon: 0.8014337200043107 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.17 | |
| episode: 1073 | score: 1.0 | memory length: 200456 | epsilon: 0.801095140004318 |
| steps: 171 | lr: 0.0001 | evaluation reward: 2.16 | |
| episode: 1074 | score: 4.0 | memory length: 200733 | epsilon: 0.8005466800043299 |
| steps: 277 | lr: 0.0001 | evaluation reward: 2.18 | |
| episode: 1075 | score: 2.0 | memory length: 200931 | epsilon: 0.8001546400043384 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.19 | |
| episode: 1076 | score: 2.0 | memory length: 201148 | epsilon: 0.7997249800043478 |
| steps: 217 | lr: 0.0001 | evaluation reward: 2.17 | |
| episode: 1077 | score: 0.0 | memory length: 201271 | epsilon: 0.7994814400043531 |
| steps: 123 | lr: 0.0001 | evaluation reward: 2.17 | |
| episode: 1078 | score: 3.0 | memory length: 201517 | epsilon: 0.7989943600043636 |
| steps: 246 | lr: 0.0001 | evaluation reward: 2.19 | |
| episode: 1079 | score: 2.0 | memory length: 201714 | epsilon: 0.7986043000043721 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.2 | |
| episode: 1080 | score: 1.0 | memory length: 201885 | epsilon: 0.7982657200043795 |
| steps: 171 | lr: 0.0001 | evaluation reward: 2.19 | |
| episode: 1081 | score: 1.0 | memory length: 202057 | epsilon: 0.7979251600043868 |
| steps: 172 | lr: 0.0001 | evaluation reward: 2.19 | |
| episode: 1082 | score: 3.0 | memory length: 202283 | epsilon: 0.7974776800043966 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.17 | |
| episode: 1083 | score: 3.0 | memory length: 202508 | epsilon: 0.7970321800044062 |
| steps: 225 | lr: 0.0001 | evaluation reward: 2.16 | |
| episode: 1084 | score: 0.0 | memory length: 202630 | epsilon: 0.7967906200044115 |
| steps: 122 | lr: 0.0001 | evaluation reward: 2.14 | |
| episode: 1085 | score: 5.0 | memory length: 202974 | epsilon: 0.7961095000044263 |
| steps: 344 | lr: 0.0001 | evaluation reward: 2.18 | |
| episode: 1086 | score: 2.0 | memory length: 203191 | epsilon: 0.7956798400044356 |
| steps: 217 | lr: 0.0001 | evaluation reward: 2.18 | |
| episode: 1087 | score: 2.0 | memory length: 203408 | epsilon: 0.7952501800044449 |
| steps: 217 | lr: 0.0001 | evaluation reward: 2.17 | |
| episode: 1088 | score: 2.0 | memory length: 203605 | epsilon: 0.7948601200044534 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.17 | |
| episode: 1089 | score: 1.0 | memory length: 203756 | epsilon: 0.7945611400044599 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.15 | |
| episode: 1090 | score: 4.0 | memory length: 204071 | epsilon: 0.7939374400044734 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 315 | lr: 0.0001 | evaluation reward: 2.17 | |
| episode: 1091 | score: 2.0 | memory length: 204269 | epsilon: 0.7935454000044819 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.18 | |
| episode: 1092 | score: 4.0 | memory length: 204566 | epsilon: 0.7929573400044947 |
| steps: 297 | lr: 0.0001 | evaluation reward: 2.2 | |
| episode: 1093 | score: 3.0 | memory length: 204811 | epsilon: 0.7924722400045052 |
| steps: 245 | lr: 0.0001 | evaluation reward: 2.2 | |
| episode: 1094 | score: 5.0 | memory length: 205137 | epsilon: 0.7918267600045192 |
| steps: 326 | lr: 0.0001 | evaluation reward: 2.25 | |
| episode: 1095 | score: 3.0 | memory length: 205363 | epsilon: 0.791379280004529 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.25 | |
| episode: 1096 | score: 3.0 | memory length: 205590 | epsilon: 0.7909298200045387 |
| steps: 227 | lr: 0.0001 | evaluation reward: 2.24 | |
| episode: 1097 | score: 4.0 | memory length: 205864 | epsilon: 0.7903873000045505 |
| steps: 274 | lr: 0.0001 | evaluation reward: 2.27 | |
| episode: 1098 | score: 2.0 | memory length: 206082 | epsilon: 0.7899556600045599 |
| steps: 218 | lr: 0.0001 | evaluation reward: 2.25 | |
| episode: 1099 | score: 3.0 | memory length: 206308 | epsilon: 0.7895081800045696 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.27 | |
| episode: 1100 | score: 4.0 | memory length: 206587 | epsilon: 0.7889557600045816 |
| steps: 279 | lr: 0.0001 | evaluation reward: 2.3 | |
| episode: 1101 | score: 4.0 | memory length: 206866 | epsilon: 0.7884033400045936 |
| steps: 279 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1102 | score: 1.0 | memory length: 207036 | epsilon: 0.7880667400046009 |
| steps: 170 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1103 | score: 0.0 | memory length: 207159 | epsilon: 0.7878232000046062 |
| steps: 123 | lr: 0.0001 | evaluation reward: 2.33 | |
| episode: 1104 | score: 4.0 | memory length: 207473 | epsilon: 0.7872014800046196 |
| steps: 314 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1105 | score: 3.0 | memory length: 207743 | epsilon: 0.7866668800046313 |
| steps: 270 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1106 | score: 4.0 | memory length: 208021 | epsilon: 0.7861164400046432 |
| steps: 278 | lr: 0.0001 | evaluation reward: 2.36 | |
| episode: 1107 | score: 3.0 | memory length: 208246 | epsilon: 0.7856709400046529 |
| steps: 225 | lr: 0.0001 | evaluation reward: 2.36 | |
| episode: 1108 | score: 4.0 | memory length: 208533 | epsilon: 0.7851026800046652 |
| steps: 287 | lr: 0.0001 | evaluation reward: 2.38 | |
| episode: 1109 | score: 2.0 | memory length: 208730 | epsilon: 0.7847126200046737 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.38 | |
| episode: 1110 | score: 1.0 | memory length: 208880 | epsilon: 0.7844156200046801 |
| steps: 150 | lr: 0.0001 | evaluation reward: 2.36 | |
| episode: 1111 | score: 4.0 | memory length: 209145 | epsilon: 0.7838909200046915 |
| steps: 265 | lr: 0.0001 | evaluation reward: 2.38 | |
| episode: 1112 | score: 4.0 | memory length: 209440 | epsilon: 0.7833068200047042 |
| steps: 295 | lr: 0.0001 | evaluation reward: 2.41 | |
| episode: 1113 | score: 2.0 | memory length: 209619 | epsilon: 0.7829524000047119 |
| steps: 179 | lr: 0.0001 | evaluation reward: 2.43 | |
| episode: 1114 | score: 3.0 | memory length: 209866 | epsilon: 0.7824633400047225 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 247 | lr: 0.0001 | evaluation reward: 2.45 | |
| episode: 1115 | score: 2.0 | memory length: 210045 | epsilon: 0.7821089200047302 |
| steps: 179 | lr: 0.0001 | evaluation reward: 2.44 | |
| episode: 1116 | score: 1.0 | memory length: 210195 | epsilon: 0.7818119200047366 |
| steps: 150 | lr: 0.0001 | evaluation reward: 2.44 | |
| episode: 1117 | score: 0.0 | memory length: 210317 | epsilon: 0.7815703600047419 |
| steps: 122 | lr: 0.0001 | evaluation reward: 2.41 | |
| episode: 1118 | score: 2.0 | memory length: 210517 | epsilon: 0.7811743600047505 |
| steps: 200 | lr: 0.0001 | evaluation reward: 2.41 | |
| episode: 1119 | score: 3.0 | memory length: 210743 | epsilon: 0.7807268800047602 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.42 | |
| episode: 1120 | score: 2.0 | memory length: 210940 | epsilon: 0.7803368200047687 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.41 | |
| episode: 1121 | score: 4.0 | memory length: 211236 | epsilon: 0.7797507400047814 |
| steps: 296 | lr: 0.0001 | evaluation reward: 2.44 | |
| episode: 1122 | score: 4.0 | memory length: 211531 | epsilon: 0.7791666400047941 |
| steps: 295 | lr: 0.0001 | evaluation reward: 2.44 | |
| episode: 1123 | score: 0.0 | memory length: 211653 | epsilon: 0.7789250800047993 |
| steps: 122 | lr: 0.0001 | evaluation reward: 2.44 | |
| episode: 1124 | score: 4.0 | memory length: 211928 | epsilon: 0.7783805800048111 |
| steps: 275 | lr: 0.0001 | evaluation reward: 2.44 | |
| episode: 1125 | score: 2.0 | memory length: 212107 | epsilon: 0.7780261600048188 |
| steps: 179 | lr: 0.0001 | evaluation reward: 2.42 | |
| episode: 1126 | score: 4.0 | memory length: 212382 | epsilon: 0.7774816600048307 |
| steps: 275 | lr: 0.0001 | evaluation reward: 2.44 | |
| episode: 1127 | score: 3.0 | memory length: 212610 | epsilon: 0.7770302200048405 |
| steps: 228 | lr: 0.0001 | evaluation reward: 2.44 | |
| episode: 1128 | score: 4.0 | memory length: 212905 | epsilon: 0.7764461200048531 |
| steps: 295 | lr: 0.0001 | evaluation reward: 2.48 | |
| episode: 1129 | score: 2.0 | memory length: 213102 | epsilon: 0.7760560600048616 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.48 | |
| episode: 1130 | score: 5.0 | memory length: 213428 | epsilon: 0.7754105800048756 |
| steps: 326 | lr: 0.0001 | evaluation reward: 2.53 | |
| episode: 1131 | score: 0.0 | memory length: 213550 | epsilon: 0.7751690200048809 |
| steps: 122 | lr: 0.0001 | evaluation reward: 2.49 | |
| episode: 1132 | score: 0.0 | memory length: 213673 | epsilon: 0.7749254800048861 |
| steps: 123 | lr: 0.0001 | evaluation reward: 2.46 | |
| episode: 1133 | score: 0.0 | memory length: 213796 | epsilon: 0.7746819400048914 |
| steps: 123 | lr: 0.0001 | evaluation reward: 2.44 | |
| episode: 1134 | score: 1.0 | memory length: 213947 | epsilon: 0.7743829600048979 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.42 | |
| episode: 1135 | score: 2.0 | memory length: 214149 | epsilon: 0.7739830000049066 |
| steps: 202 | lr: 0.0001 | evaluation reward: 2.44 | |
| episode: 1136 | score: 3.0 | memory length: 214362 | epsilon: 0.7735612600049158 |
| steps: 213 | lr: 0.0001 | evaluation reward: 2.43 | |
| episode: 1137 | score: 2.0 | memory length: 214559 | epsilon: 0.7731712000049242 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.43 | |
| episode: 1138 | score: 1.0 | memory length: 214728 | epsilon: 0.7728365800049315 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 169 | lr: 0.0001 | evaluation reward: 2.38 | |
| episode: 1139 | score: 0.0 | memory length: 214850 | epsilon: 0.7725950200049367 |
| steps: 122 | lr: 0.0001 | evaluation reward: 2.38 | |
| episode: 1140 | score: 2.0 | memory length: 215047 | epsilon: 0.7722049600049452 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.37 | |
| episode: 1141 | score: 2.0 | memory length: 215244 | epsilon: 0.7718149000049537 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.37 | |
| episode: 1142 | score: 0.0 | memory length: 215367 | epsilon: 0.771571360004959 |
| steps: 123 | lr: 0.0001 | evaluation reward: 2.33 | |
| episode: 1143 | score: 2.0 | memory length: 215584 | epsilon: 0.7711417000049683 |
| steps: 217 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1144 | score: 2.0 | memory length: 215781 | epsilon: 0.7707516400049768 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.36 | |
| episode: 1145 | score: 2.0 | memory length: 215978 | epsilon: 0.7703615800049852 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1146 | score: 0.0 | memory length: 216100 | epsilon: 0.7701200200049905 |
| steps: 122 | lr: 0.0001 | evaluation reward: 2.33 | |
| episode: 1147 | score: 3.0 | memory length: 216343 | epsilon: 0.7696388800050009 |
| steps: 243 | lr: 0.0001 | evaluation reward: 2.32 | |
| episode: 1148 | score: 2.0 | memory length: 216541 | epsilon: 0.7692468400050094 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.33 | |
| episode: 1149 | score: 2.0 | memory length: 216739 | epsilon: 0.7688548000050179 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1150 | score: 3.0 | memory length: 216967 | epsilon: 0.7684033600050277 |
| steps: 228 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1151 | score: 2.0 | memory length: 217165 | epsilon: 0.7680113200050362 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1152 | score: 5.0 | memory length: 217467 | epsilon: 0.7674133600050492 |
| steps: 302 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1153 | score: 1.0 | memory length: 217617 | epsilon: 0.7671163600050557 |
| steps: 150 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1154 | score: 1.0 | memory length: 217787 | epsilon: 0.766779760005063 |
| steps: 170 | lr: 0.0001 | evaluation reward: 2.33 | |
| episode: 1155 | score: 4.0 | memory length: 218099 | epsilon: 0.7661620000050764 |
| steps: 312 | lr: 0.0001 | evaluation reward: 2.36 | |
| episode: 1156 | score: 0.0 | memory length: 218221 | epsilon: 0.7659204400050816 |
| steps: 122 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1157 | score: 2.0 | memory length: 218421 | epsilon: 0.7655244400050902 |
| steps: 200 | lr: 0.0001 | evaluation reward: 2.32 | |
| episode: 1158 | score: 3.0 | memory length: 218667 | epsilon: 0.7650373600051008 |
| steps: 246 | lr: 0.0001 | evaluation reward: 2.33 | |
| episode: 1159 | score: 3.0 | memory length: 218893 | epsilon: 0.7645898800051105 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1160 | score: 3.0 | memory length: 219105 | epsilon: 0.7641701200051196 |
| steps: 212 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1161 | score: 3.0 | memory length: 219331 | epsilon: 0.7637226400051293 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.38 | |
| episode: 1162 | score: 2.0 | memory length: 219552 | epsilon: 0.7632850600051388 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 221 | lr: 0.0001 | evaluation reward: 2.38 | |
| episode: 1163 | score: 2.0 | memory length: 219750 | epsilon: 0.7628930200051474 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.36 | |
| episode: 1164 | score: 4.0 | memory length: 220048 | epsilon: 0.7623029800051602 |
| steps: 298 | lr: 0.0001 | evaluation reward: 2.37 | |
| episode: 1165 | score: 5.0 | memory length: 220375 | epsilon: 0.7616555200051742 |
| steps: 327 | lr: 0.0001 | evaluation reward: 2.4 | |
| episode: 1166 | score: 1.0 | memory length: 220525 | epsilon: 0.7613585200051807 |
| steps: 150 | lr: 0.0001 | evaluation reward: 2.41 | |
| episode: 1167 | score: 1.0 | memory length: 220693 | epsilon: 0.7610258800051879 |
| steps: 168 | lr: 0.0001 | evaluation reward: 2.38 | |
| episode: 1168 | score: 8.0 | memory length: 221137 | epsilon: 0.760146760005207 |
| steps: 444 | lr: 0.0001 | evaluation reward: 2.44 | |
| episode: 1169 | score: 2.0 | memory length: 221335 | epsilon: 0.7597547200052155 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.43 | |
| episode: 1170 | score: 0.0 | memory length: 221458 | epsilon: 0.7595111800052208 |
| steps: 123 | lr: 0.0001 | evaluation reward: 2.38 | |
| episode: 1171 | score: 2.0 | memory length: 221674 | epsilon: 0.7590835000052301 |
| steps: 216 | lr: 0.0001 | evaluation reward: 2.37 | |
| episode: 1172 | score: 4.0 | memory length: 221946 | epsilon: 0.7585449400052418 |
| steps: 272 | lr: 0.0001 | evaluation reward: 2.39 | |
| episode: 1173 | score: 3.0 | memory length: 222174 | epsilon: 0.7580935000052516 |
| steps: 228 | lr: 0.0001 | evaluation reward: 2.41 | |
| episode: 1174 | score: 3.0 | memory length: 222418 | epsilon: 0.757610380005262 |
| steps: 244 | lr: 0.0001 | evaluation reward: 2.4 | |
| episode: 1175 | score: 3.0 | memory length: 222662 | epsilon: 0.7571272600052725 |
| steps: 244 | lr: 0.0001 | evaluation reward: 2.41 | |
| episode: 1176 | score: 2.0 | memory length: 222860 | epsilon: 0.756735220005281 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.41 | |
| episode: 1177 | score: 2.0 | memory length: 223057 | epsilon: 0.7563451600052895 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.43 | |
| episode: 1178 | score: 2.0 | memory length: 223275 | epsilon: 0.7559135200052989 |
| steps: 218 | lr: 0.0001 | evaluation reward: 2.42 | |
| episode: 1179 | score: 2.0 | memory length: 223473 | epsilon: 0.7555214800053074 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.42 | |
| episode: 1180 | score: 2.0 | memory length: 223693 | epsilon: 0.7550858800053168 |
| steps: 220 | lr: 0.0001 | evaluation reward: 2.43 | |
| episode: 1181 | score: 4.0 | memory length: 223969 | epsilon: 0.7545394000053287 |
| steps: 276 | lr: 0.0001 | evaluation reward: 2.46 | |
| episode: 1182 | score: 2.0 | memory length: 224166 | epsilon: 0.7541493400053372 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.45 | |
| episode: 1183 | score: 5.0 | memory length: 224509 | epsilon: 0.7534702000053519 |
| steps: 343 | lr: 0.0001 | evaluation reward: 2.47 | |
| episode: 1184 | score: 3.0 | memory length: 224753 | epsilon: 0.7529870800053624 |
| steps: 244 | lr: 0.0001 | evaluation reward: 2.5 | |
| episode: 1185 | score: 2.0 | memory length: 224971 | epsilon: 0.7525554400053718 |
| steps: 218 | lr: 0.0001 | evaluation reward: 2.47 | |
| episode: 1186 | score: 3.0 | memory length: 225198 | epsilon: 0.7521059800053815 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 227 | lr: 0.0001 | evaluation reward: 2.48 | |
| episode: 1187 | score: 3.0 | memory length: 225423 | epsilon: 0.7516604800053912 |
| steps: 225 | lr: 0.0001 | evaluation reward: 2.49 | |
| episode: 1188 | score: 5.0 | memory length: 225746 | epsilon: 0.7510209400054051 |
| steps: 323 | lr: 0.0001 | evaluation reward: 2.52 | |
| episode: 1189 | score: 3.0 | memory length: 225995 | epsilon: 0.7505279200054158 |
| steps: 249 | lr: 0.0001 | evaluation reward: 2.54 | |
| episode: 1190 | score: 1.0 | memory length: 226166 | epsilon: 0.7501893400054231 |
| steps: 171 | lr: 0.0001 | evaluation reward: 2.51 | |
| episode: 1191 | score: 0.0 | memory length: 226289 | epsilon: 0.7499458000054284 |
| steps: 123 | lr: 0.0001 | evaluation reward: 2.49 | |
| episode: 1192 | score: 2.0 | memory length: 226486 | epsilon: 0.7495557400054369 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.47 | |
| episode: 1193 | score: 2.0 | memory length: 226684 | epsilon: 0.7491637000054454 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.46 | |
| episode: 1194 | score: 2.0 | memory length: 226902 | epsilon: 0.7487320600054548 |
| steps: 218 | lr: 0.0001 | evaluation reward: 2.43 | |
| episode: 1195 | score: 3.0 | memory length: 227128 | epsilon: 0.7482845800054645 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.43 | |
| episode: 1196 | score: 3.0 | memory length: 227397 | epsilon: 0.7477519600054761 |
| steps: 269 | lr: 0.0001 | evaluation reward: 2.43 | |
| episode: 1197 | score: 2.0 | memory length: 227594 | epsilon: 0.7473619000054845 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.41 | |
| episode: 1198 | score: 2.0 | memory length: 227791 | epsilon: 0.746971840005493 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.41 | |
| episode: 1199 | score: 3.0 | memory length: 228035 | epsilon: 0.7464887200055035 |
| steps: 244 | lr: 0.0001 | evaluation reward: 2.41 | |
| episode: 1200 | score: 2.0 | memory length: 228232 | epsilon: 0.746098660005512 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.39 | |
| episode: 1201 | score: 2.0 | memory length: 228449 | epsilon: 0.7456690000055213 |
| steps: 217 | lr: 0.0001 | evaluation reward: 2.37 | |
| episode: 1202 | score: 1.0 | memory length: 228618 | epsilon: 0.7453343800055285 |
| steps: 169 | lr: 0.0001 | evaluation reward: 2.37 | |
| episode: 1203 | score: 2.0 | memory length: 228816 | epsilon: 0.744942340005537 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.39 | |
| episode: 1204 | score: 4.0 | memory length: 229113 | epsilon: 0.7443542800055498 |
| steps: 297 | lr: 0.0001 | evaluation reward: 2.39 | |
| episode: 1205 | score: 1.0 | memory length: 229282 | epsilon: 0.7440196600055571 |
| steps: 169 | lr: 0.0001 | evaluation reward: 2.37 | |
| episode: 1206 | score: 2.0 | memory length: 229480 | epsilon: 0.7436276200055656 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1207 | score: 3.0 | memory length: 229727 | epsilon: 0.7431385600055762 |
| steps: 247 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1208 | score: 3.0 | memory length: 229971 | epsilon: 0.7426554400055867 |
| steps: 244 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1209 | score: 2.0 | memory length: 230168 | epsilon: 0.7422653800055952 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1210 | score: 4.0 | memory length: 230446 | epsilon: 0.7417149400056071 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 278 | lr: 0.0001 | evaluation reward: 2.37 | |
| episode: 1211 | score: 2.0 | memory length: 230644 | epsilon: 0.7413229000056156 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1212 | score: 2.0 | memory length: 230842 | epsilon: 0.7409308600056241 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.33 | |
| episode: 1213 | score: 1.0 | memory length: 231013 | epsilon: 0.7405922800056315 |
| steps: 171 | lr: 0.0001 | evaluation reward: 2.32 | |
| episode: 1214 | score: 2.0 | memory length: 231211 | epsilon: 0.74020024000564 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.31 | |
| episode: 1215 | score: 2.0 | memory length: 231408 | epsilon: 0.7398101800056485 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.31 | |
| episode: 1216 | score: 0.0 | memory length: 231530 | epsilon: 0.7395686200056537 |
| steps: 122 | lr: 0.0001 | evaluation reward: 2.3 | |
| episode: 1217 | score: 3.0 | memory length: 231800 | epsilon: 0.7390340200056653 |
| steps: 270 | lr: 0.0001 | evaluation reward: 2.33 | |
| episode: 1218 | score: 2.0 | memory length: 232000 | epsilon: 0.7386380200056739 |
| steps: 200 | lr: 0.0001 | evaluation reward: 2.33 | |
| episode: 1219 | score: 4.0 | memory length: 232313 | epsilon: 0.7380182800056874 |
| steps: 313 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1220 | score: 6.0 | memory length: 232667 | epsilon: 0.7373173600057026 |
| steps: 354 | lr: 0.0001 | evaluation reward: 2.38 | |
| episode: 1221 | score: 1.0 | memory length: 232818 | epsilon: 0.7370183800057091 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1222 | score: 4.0 | memory length: 233134 | epsilon: 0.7363927000057227 |
| steps: 316 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1223 | score: 2.0 | memory length: 233331 | epsilon: 0.7360026400057311 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.37 | |
| episode: 1224 | score: 3.0 | memory length: 233578 | epsilon: 0.7355135800057417 |
| steps: 247 | lr: 0.0001 | evaluation reward: 2.36 | |
| episode: 1225 | score: 2.0 | memory length: 233778 | epsilon: 0.7351175800057503 |
| steps: 200 | lr: 0.0001 | evaluation reward: 2.36 | |
| episode: 1226 | score: 2.0 | memory length: 233975 | epsilon: 0.7347275200057588 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.34 | |
| episode: 1227 | score: 4.0 | memory length: 234251 | epsilon: 0.7341810400057707 |
| steps: 276 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1228 | score: 1.0 | memory length: 234402 | epsilon: 0.7338820600057772 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.32 | |
| episode: 1229 | score: 3.0 | memory length: 234632 | epsilon: 0.733426660005787 |
| steps: 230 | lr: 0.0001 | evaluation reward: 2.33 | |
| episode: 1230 | score: 7.0 | memory length: 235037 | epsilon: 0.7326247600058045 |
| steps: 405 | lr: 0.0001 | evaluation reward: 2.35 | |
| episode: 1231 | score: 5.0 | memory length: 235363 | epsilon: 0.7319792800058185 |
| steps: 326 | lr: 0.0001 | evaluation reward: 2.4 | |
| episode: 1232 | score: 3.0 | memory length: 235609 | epsilon: 0.731492200005829 |
| steps: 246 | lr: 0.0001 | evaluation reward: 2.43 | |
| episode: 1233 | score: 2.0 | memory length: 235807 | epsilon: 0.7311001600058376 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.45 | |
| episode: 1234 | score: 1.0 | memory length: 235978 | epsilon: 0.7307615800058449 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 171 | lr: 0.0001 | evaluation reward: 2.45 | |
| episode: 1235 | score: 4.0 | memory length: 236254 | epsilon: 0.7302151000058568 |
| steps: 276 | lr: 0.0001 | evaluation reward: 2.47 | |
| episode: 1236 | score: 2.0 | memory length: 236451 | epsilon: 0.7298250400058652 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.46 | |
| episode: 1237 | score: 4.0 | memory length: 236717 | epsilon: 0.7292983600058767 |
| steps: 266 | lr: 0.0001 | evaluation reward: 2.48 | |
| episode: 1238 | score: 2.0 | memory length: 236915 | epsilon: 0.7289063200058852 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.49 | |
| episode: 1239 | score: 2.0 | memory length: 237112 | epsilon: 0.7285162600058936 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.51 | |
| episode: 1240 | score: 3.0 | memory length: 237358 | epsilon: 0.7280291800059042 |
| steps: 246 | lr: 0.0001 | evaluation reward: 2.52 | |
| episode: 1241 | score: 4.0 | memory length: 237641 | epsilon: 0.7274688400059164 |
| steps: 283 | lr: 0.0001 | evaluation reward: 2.54 | |
| episode: 1242 | score: 3.0 | memory length: 237889 | epsilon: 0.726977800005927 |
| steps: 248 | lr: 0.0001 | evaluation reward: 2.57 | |
| episode: 1243 | score: 2.0 | memory length: 238087 | epsilon: 0.7265857600059356 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.57 | |
| episode: 1244 | score: 3.0 | memory length: 238313 | epsilon: 0.7261382800059453 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.58 | |
| episode: 1245 | score: 5.0 | memory length: 238674 | epsilon: 0.7254235000059608 |
| steps: 361 | lr: 0.0001 | evaluation reward: 2.61 | |
| episode: 1246 | score: 2.0 | memory length: 238856 | epsilon: 0.7250631400059686 |
| steps: 182 | lr: 0.0001 | evaluation reward: 2.63 | |
| episode: 1247 | score: 2.0 | memory length: 239037 | epsilon: 0.7247047600059764 |
| steps: 181 | lr: 0.0001 | evaluation reward: 2.62 | |
| episode: 1248 | score: 4.0 | memory length: 239317 | epsilon: 0.7241503600059884 |
| steps: 280 | lr: 0.0001 | evaluation reward: 2.64 | |
| episode: 1249 | score: 3.0 | memory length: 239583 | epsilon: 0.7236236800059999 |
| steps: 266 | lr: 0.0001 | evaluation reward: 2.65 | |
| episode: 1250 | score: 2.0 | memory length: 239781 | epsilon: 0.7232316400060084 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.64 | |
| episode: 1251 | score: 1.0 | memory length: 239932 | epsilon: 0.7229326600060149 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.63 | |
| episode: 1252 | score: 2.0 | memory length: 240113 | epsilon: 0.7225742800060226 |
| steps: 181 | lr: 0.0001 | evaluation reward: 2.6 | |
| episode: 1253 | score: 2.0 | memory length: 240292 | epsilon: 0.7222198600060303 |
| steps: 179 | lr: 0.0001 | evaluation reward: 2.61 | |
| episode: 1254 | score: 3.0 | memory length: 240519 | epsilon: 0.7217704000060401 |
| steps: 227 | lr: 0.0001 | evaluation reward: 2.63 | |
| episode: 1255 | score: 3.0 | memory length: 240750 | epsilon: 0.72131302000605 |
| steps: 231 | lr: 0.0001 | evaluation reward: 2.62 | |
| episode: 1256 | score: 2.0 | memory length: 240947 | epsilon: 0.7209229600060585 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.64 | |
| episode: 1257 | score: 7.0 | memory length: 241266 | epsilon: 0.7202913400060722 |
| steps: 319 | lr: 0.0001 | evaluation reward: 2.69 | |
| episode: 1258 | score: 3.0 | memory length: 241495 | epsilon: 0.719837920006082 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 229 | lr: 0.0001 | evaluation reward: 2.69 | |
| episode: 1259 | score: 3.0 | memory length: 241742 | epsilon: 0.7193488600060927 |
| steps: 247 | lr: 0.0001 | evaluation reward: 2.69 | |
| episode: 1260 | score: 4.0 | memory length: 242037 | epsilon: 0.7187647600061053 |
| steps: 295 | lr: 0.0001 | evaluation reward: 2.7 | |
| episode: 1261 | score: 3.0 | memory length: 242264 | epsilon: 0.7183153000061151 |
| steps: 227 | lr: 0.0001 | evaluation reward: 2.7 | |
| episode: 1262 | score: 1.0 | memory length: 242415 | epsilon: 0.7180163200061216 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.69 | |
| episode: 1263 | score: 4.0 | memory length: 242672 | epsilon: 0.7175074600061326 |
| steps: 257 | lr: 0.0001 | evaluation reward: 2.71 | |
| episode: 1264 | score: 2.0 | memory length: 242890 | epsilon: 0.717075820006142 |
| steps: 218 | lr: 0.0001 | evaluation reward: 2.69 | |
| episode: 1265 | score: 4.0 | memory length: 243164 | epsilon: 0.7165333000061538 |
| steps: 274 | lr: 0.0001 | evaluation reward: 2.68 | |
| episode: 1266 | score: 2.0 | memory length: 243362 | epsilon: 0.7161412600061623 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.69 | |
| episode: 1267 | score: 2.0 | memory length: 243559 | epsilon: 0.7157512000061708 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.7 | |
| episode: 1268 | score: 4.0 | memory length: 243856 | epsilon: 0.7151631400061835 |
| steps: 297 | lr: 0.0001 | evaluation reward: 2.66 | |
| episode: 1269 | score: 4.0 | memory length: 244113 | epsilon: 0.7146542800061946 |
| steps: 257 | lr: 0.0001 | evaluation reward: 2.68 | |
| episode: 1270 | score: 3.0 | memory length: 244339 | epsilon: 0.7142068000062043 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.71 | |
| episode: 1271 | score: 2.0 | memory length: 244537 | epsilon: 0.7138147600062128 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.71 | |
| episode: 1272 | score: 2.0 | memory length: 244719 | epsilon: 0.7134544000062206 |
| steps: 182 | lr: 0.0001 | evaluation reward: 2.69 | |
| episode: 1273 | score: 4.0 | memory length: 244994 | epsilon: 0.7129099000062324 |
| steps: 275 | lr: 0.0001 | evaluation reward: 2.7 | |
| episode: 1274 | score: 3.0 | memory length: 245237 | epsilon: 0.7124287600062429 |
| steps: 243 | lr: 0.0001 | evaluation reward: 2.7 | |
| episode: 1275 | score: 4.0 | memory length: 245558 | epsilon: 0.7117931800062567 |
| steps: 321 | lr: 0.0001 | evaluation reward: 2.71 | |
| episode: 1276 | score: 4.0 | memory length: 245837 | epsilon: 0.7112407600062687 |
| steps: 279 | lr: 0.0001 | evaluation reward: 2.73 | |
| episode: 1277 | score: 2.0 | memory length: 246034 | epsilon: 0.7108507000062771 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.73 | |
| episode: 1278 | score: 1.0 | memory length: 246185 | epsilon: 0.7105517200062836 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.72 | |
| episode: 1279 | score: 3.0 | memory length: 246429 | epsilon: 0.7100686000062941 |
| steps: 244 | lr: 0.0001 | evaluation reward: 2.73 | |
| episode: 1280 | score: 2.0 | memory length: 246646 | epsilon: 0.7096389400063035 |
| steps: 217 | lr: 0.0001 | evaluation reward: 2.73 | |
| episode: 1281 | score: 5.0 | memory length: 246952 | epsilon: 0.7090330600063166 |
| steps: 306 | lr: 0.0001 | evaluation reward: 2.74 | |
| episode: 1282 | score: 2.0 | memory length: 247154 | epsilon: 0.7086331000063253 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 202 | lr: 0.0001 | evaluation reward: 2.74 | |
| episode: 1283 | score: 0.0 | memory length: 247276 | epsilon: 0.7083915400063305 |
| steps: 122 | lr: 0.0001 | evaluation reward: 2.69 | |
| episode: 1284 | score: 3.0 | memory length: 247486 | epsilon: 0.7079757400063396 |
| steps: 210 | lr: 0.0001 | evaluation reward: 2.69 | |
| episode: 1285 | score: 4.0 | memory length: 247780 | epsilon: 0.7073936200063522 |
| steps: 294 | lr: 0.0001 | evaluation reward: 2.71 | |
| episode: 1286 | score: 4.0 | memory length: 248073 | epsilon: 0.7068134800063648 |
| steps: 293 | lr: 0.0001 | evaluation reward: 2.72 | |
| episode: 1287 | score: 2.0 | memory length: 248255 | epsilon: 0.7064531200063726 |
| steps: 182 | lr: 0.0001 | evaluation reward: 2.71 | |
| episode: 1288 | score: 2.0 | memory length: 248452 | epsilon: 0.7060630600063811 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.68 | |
| episode: 1289 | score: 3.0 | memory length: 248679 | epsilon: 0.7056136000063908 |
| steps: 227 | lr: 0.0001 | evaluation reward: 2.68 | |
| episode: 1290 | score: 1.0 | memory length: 248831 | epsilon: 0.7053126400063974 |
| steps: 152 | lr: 0.0001 | evaluation reward: 2.68 | |
| episode: 1291 | score: 3.0 | memory length: 249057 | epsilon: 0.7048651600064071 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.71 | |
| episode: 1292 | score: 3.0 | memory length: 249303 | epsilon: 0.7043780800064177 |
| steps: 246 | lr: 0.0001 | evaluation reward: 2.72 | |
| episode: 1293 | score: 1.0 | memory length: 249471 | epsilon: 0.7040454400064249 |
| steps: 168 | lr: 0.0001 | evaluation reward: 2.71 | |
| episode: 1294 | score: 3.0 | memory length: 249717 | epsilon: 0.7035583600064355 |
| steps: 246 | lr: 0.0001 | evaluation reward: 2.72 | |
| episode: 1295 | score: 2.0 | memory length: 249935 | epsilon: 0.7031267200064448 |
| steps: 218 | lr: 0.0001 | evaluation reward: 2.71 | |
| episode: 1296 | score: 3.0 | memory length: 250163 | epsilon: 0.7026752800064546 |
| steps: 228 | lr: 0.0001 | evaluation reward: 2.71 | |
| episode: 1297 | score: 5.0 | memory length: 250487 | epsilon: 0.7020337600064686 |
| steps: 324 | lr: 0.0001 | evaluation reward: 2.74 | |
| episode: 1298 | score: 3.0 | memory length: 250734 | epsilon: 0.7015447000064792 |
| steps: 247 | lr: 0.0001 | evaluation reward: 2.75 | |
| episode: 1299 | score: 3.0 | memory length: 250959 | epsilon: 0.7010992000064888 |
| steps: 225 | lr: 0.0001 | evaluation reward: 2.75 | |
| episode: 1300 | score: 2.0 | memory length: 251156 | epsilon: 0.7007091400064973 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.75 | |
| episode: 1301 | score: 7.0 | memory length: 251551 | epsilon: 0.6999270400065143 |
| steps: 395 | lr: 0.0001 | evaluation reward: 2.8 | |
| episode: 1302 | score: 3.0 | memory length: 251795 | epsilon: 0.6994439200065248 |
| steps: 244 | lr: 0.0001 | evaluation reward: 2.82 | |
| episode: 1303 | score: 5.0 | memory length: 252118 | epsilon: 0.6988043800065387 |
| steps: 323 | lr: 0.0001 | evaluation reward: 2.85 | |
| episode: 1304 | score: 2.0 | memory length: 252316 | epsilon: 0.6984123400065472 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.83 | |
| episode: 1305 | score: 2.0 | memory length: 252513 | epsilon: 0.6980222800065556 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.84 | |
| episode: 1306 | score: 2.0 | memory length: 252731 | epsilon: 0.697590640006565 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 218 | lr: 0.0001 | evaluation reward: 2.84 | |
| episode: 1307 | score: 3.0 | memory length: 252978 | epsilon: 0.6971015800065756 |
| steps: 247 | lr: 0.0001 | evaluation reward: 2.84 | |
| episode: 1308 | score: 5.0 | memory length: 253328 | epsilon: 0.6964085800065907 |
| steps: 350 | lr: 0.0001 | evaluation reward: 2.86 | |
| episode: 1309 | score: 2.0 | memory length: 253510 | epsilon: 0.6960482200065985 |
| steps: 182 | lr: 0.0001 | evaluation reward: 2.86 | |
| episode: 1310 | score: 0.0 | memory length: 253632 | epsilon: 0.6958066600066037 |
| steps: 122 | lr: 0.0001 | evaluation reward: 2.82 | |
| episode: 1311 | score: 4.0 | memory length: 253899 | epsilon: 0.6952780000066152 |
| steps: 267 | lr: 0.0001 | evaluation reward: 2.84 | |
| episode: 1312 | score: 3.0 | memory length: 254124 | epsilon: 0.6948325000066249 |
| steps: 225 | lr: 0.0001 | evaluation reward: 2.85 | |
| episode: 1313 | score: 4.0 | memory length: 254388 | epsilon: 0.6943097800066362 |
| steps: 264 | lr: 0.0001 | evaluation reward: 2.88 | |
| episode: 1314 | score: 5.0 | memory length: 254708 | epsilon: 0.69367618000665 |
| steps: 320 | lr: 0.0001 | evaluation reward: 2.91 | |
| episode: 1315 | score: 2.0 | memory length: 254905 | epsilon: 0.6932861200066585 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.91 | |
| episode: 1316 | score: 2.0 | memory length: 255103 | epsilon: 0.692894080006667 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.93 | |
| episode: 1317 | score: 2.0 | memory length: 255285 | epsilon: 0.6925337200066748 |
| steps: 182 | lr: 0.0001 | evaluation reward: 2.92 | |
| episode: 1318 | score: 4.0 | memory length: 255581 | epsilon: 0.6919476400066875 |
| steps: 296 | lr: 0.0001 | evaluation reward: 2.94 | |
| episode: 1319 | score: 2.0 | memory length: 255779 | epsilon: 0.691555600006696 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.92 | |
| episode: 1320 | score: 2.0 | memory length: 255976 | epsilon: 0.6911655400067045 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.88 | |
| episode: 1321 | score: 1.0 | memory length: 256144 | epsilon: 0.6908329000067117 |
| steps: 168 | lr: 0.0001 | evaluation reward: 2.88 | |
| episode: 1322 | score: 3.0 | memory length: 256390 | epsilon: 0.6903458200067223 |
| steps: 246 | lr: 0.0001 | evaluation reward: 2.87 | |
| episode: 1323 | score: 2.0 | memory length: 256587 | epsilon: 0.6899557600067308 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.87 | |
| episode: 1324 | score: 3.0 | memory length: 256833 | epsilon: 0.6894686800067413 |
| steps: 246 | lr: 0.0001 | evaluation reward: 2.87 | |
| episode: 1325 | score: 5.0 | memory length: 257154 | epsilon: 0.6888331000067551 |
| steps: 321 | lr: 0.0001 | evaluation reward: 2.9 | |
| episode: 1326 | score: 1.0 | memory length: 257304 | epsilon: 0.6885361000067616 |
| steps: 150 | lr: 0.0001 | evaluation reward: 2.89 | |
| episode: 1327 | score: 3.0 | memory length: 257529 | epsilon: 0.6880906000067712 |
| steps: 225 | lr: 0.0001 | evaluation reward: 2.88 | |
| episode: 1328 | score: 4.0 | memory length: 257804 | epsilon: 0.6875461000067831 |
| steps: 275 | lr: 0.0001 | evaluation reward: 2.91 | |
| episode: 1329 | score: 2.0 | memory length: 258023 | epsilon: 0.6871124800067925 |
| steps: 219 | lr: 0.0001 | evaluation reward: 2.9 | |
| episode: 1330 | score: 4.0 | memory length: 258318 | epsilon: 0.6865283800068052 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 295 | lr: 0.0001 | evaluation reward: 2.87 | |
| episode: 1331 | score: 6.0 | memory length: 258671 | epsilon: 0.6858294400068203 |
| steps: 353 | lr: 0.0001 | evaluation reward: 2.88 | |
| episode: 1332 | score: 5.0 | memory length: 259016 | epsilon: 0.6851463400068352 |
| steps: 345 | lr: 0.0001 | evaluation reward: 2.9 | |
| episode: 1333 | score: 3.0 | memory length: 259246 | epsilon: 0.684690940006845 |
| steps: 230 | lr: 0.0001 | evaluation reward: 2.91 | |
| episode: 1334 | score: 5.0 | memory length: 259569 | epsilon: 0.6840514000068589 |
| steps: 323 | lr: 0.0001 | evaluation reward: 2.95 | |
| episode: 1335 | score: 4.0 | memory length: 259848 | epsilon: 0.6834989800068709 |
| steps: 279 | lr: 0.0001 | evaluation reward: 2.95 | |
| episode: 1336 | score: 4.0 | memory length: 260144 | epsilon: 0.6829129000068836 |
| steps: 296 | lr: 0.0001 | evaluation reward: 2.97 | |
| episode: 1337 | score: 4.0 | memory length: 260458 | epsilon: 0.6822911800068971 |
| steps: 314 | lr: 0.0001 | evaluation reward: 2.97 | |
| episode: 1338 | score: 2.0 | memory length: 260638 | epsilon: 0.6819347800069049 |
| steps: 180 | lr: 0.0001 | evaluation reward: 2.97 | |
| episode: 1339 | score: 4.0 | memory length: 260934 | epsilon: 0.6813487000069176 |
| steps: 296 | lr: 0.0001 | evaluation reward: 2.99 | |
| episode: 1340 | score: 2.0 | memory length: 261131 | epsilon: 0.6809586400069261 |
| steps: 197 | lr: 0.0001 | evaluation reward: 2.98 | |
| episode: 1341 | score: 3.0 | memory length: 261361 | epsilon: 0.680503240006936 |
| steps: 230 | lr: 0.0001 | evaluation reward: 2.97 | |
| episode: 1342 | score: 4.0 | memory length: 261620 | epsilon: 0.6799904200069471 |
| steps: 259 | lr: 0.0001 | evaluation reward: 2.98 | |
| episode: 1343 | score: 3.0 | memory length: 261846 | epsilon: 0.6795429400069568 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.99 | |
| episode: 1344 | score: 2.0 | memory length: 262044 | epsilon: 0.6791509000069653 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.98 | |
| episode: 1345 | score: 4.0 | memory length: 262319 | epsilon: 0.6786064000069771 |
| steps: 275 | lr: 0.0001 | evaluation reward: 2.97 | |
| episode: 1346 | score: 1.0 | memory length: 262469 | epsilon: 0.6783094000069836 |
| steps: 150 | lr: 0.0001 | evaluation reward: 2.96 | |
| episode: 1347 | score: 2.0 | memory length: 262667 | epsilon: 0.6779173600069921 |
| steps: 198 | lr: 0.0001 | evaluation reward: 2.96 | |
| episode: 1348 | score: 1.0 | memory length: 262818 | epsilon: 0.6776183800069986 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.93 | |
| episode: 1349 | score: 4.0 | memory length: 263060 | epsilon: 0.677139220007009 |
| steps: 242 | lr: 0.0001 | evaluation reward: 2.94 | |
| episode: 1350 | score: 4.0 | memory length: 263317 | epsilon: 0.67663036000702 |
| steps: 257 | lr: 0.0001 | evaluation reward: 2.96 | |
| episode: 1351 | score: 3.0 | memory length: 263562 | epsilon: 0.6761452600070306 |
| steps: 245 | lr: 0.0001 | evaluation reward: 2.98 | |
| episode: 1352 | score: 1.0 | memory length: 263713 | epsilon: 0.6758462800070371 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.97 | |
| episode: 1353 | score: 4.0 | memory length: 264010 | epsilon: 0.6752582200070498 |
| steps: 297 | lr: 0.0001 | evaluation reward: 2.99 | |
| episode: 1354 | score: 0.0 | memory length: 264133 | epsilon: 0.6750146800070551 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 123 | lr: 0.0001 | evaluation reward: 2.96 | |
| episode: 1355 | score: 3.0 | memory length: 264399 | epsilon: 0.6744880000070665 |
| steps: 266 | lr: 0.0001 | evaluation reward: 2.96 | |
| episode: 1356 | score: 6.0 | memory length: 264772 | epsilon: 0.6737494600070826 |
| steps: 373 | lr: 0.0001 | evaluation reward: 3.0 | |
| episode: 1357 | score: 3.0 | memory length: 264998 | epsilon: 0.6733019800070923 |
| steps: 226 | lr: 0.0001 | evaluation reward: 2.96 | |
| episode: 1358 | score: 3.0 | memory length: 265223 | epsilon: 0.672856480007102 |
| steps: 225 | lr: 0.0001 | evaluation reward: 2.96 | |
| episode: 1359 | score: 4.0 | memory length: 265514 | epsilon: 0.6722803000071145 |
| steps: 291 | lr: 0.0001 | evaluation reward: 2.97 | |
| episode: 1360 | score: 1.0 | memory length: 265665 | epsilon: 0.671981320007121 |
| steps: 151 | lr: 0.0001 | evaluation reward: 2.94 | |
| episode: 1361 | score: 3.0 | memory length: 265908 | epsilon: 0.6715001800071314 |
| steps: 243 | lr: 0.0001 | evaluation reward: 2.94 | |
| episode: 1362 | score: 3.0 | memory length: 266175 | epsilon: 0.6709715200071429 |
| steps: 267 | lr: 0.0001 | evaluation reward: 2.96 | |
| episode: 1363 | score: 3.0 | memory length: 266388 | epsilon: 0.670549780007152 |
| steps: 213 | lr: 0.0001 | evaluation reward: 2.95 | |
| episode: 1364 | score: 3.0 | memory length: 266615 | epsilon: 0.6701003200071618 |
| steps: 227 | lr: 0.0001 | evaluation reward: 2.96 | |
| episode: 1365 | score: 3.0 | memory length: 266843 | epsilon: 0.6696488800071716 |
| steps: 228 | lr: 0.0001 | evaluation reward: 2.95 | |
| episode: 1366 | score: 4.0 | memory length: 267104 | epsilon: 0.6691321000071828 |
| steps: 261 | lr: 0.0001 | evaluation reward: 2.97 | |
| episode: 1367 | score: 3.0 | memory length: 267329 | epsilon: 0.6686866000071925 |
| steps: 225 | lr: 0.0001 | evaluation reward: 2.98 | |
| episode: 1368 | score: 1.0 | memory length: 267481 | epsilon: 0.668385640007199 |
| steps: 152 | lr: 0.0001 | evaluation reward: 2.95 | |
| episode: 1369 | score: 3.0 | memory length: 267706 | epsilon: 0.6679401400072087 |
| steps: 225 | lr: 0.0001 | evaluation reward: 2.94 | |
| episode: 1370 | score: 0.0 | memory length: 267829 | epsilon: 0.667696600007214 |
| steps: 123 | lr: 0.0001 | evaluation reward: 2.91 | |
| episode: 1371 | score: 4.0 | memory length: 268108 | epsilon: 0.667144180007226 |
| steps: 279 | lr: 0.0001 | evaluation reward: 2.93 | |
| episode: 1372 | score: 2.0 | memory length: 268325 | epsilon: 0.6667145200072353 |
| steps: 217 | lr: 0.0001 | evaluation reward: 2.93 | |
| episode: 1373 | score: 3.0 | memory length: 268572 | epsilon: 0.6662254600072459 |
| steps: 247 | lr: 0.0001 | evaluation reward: 2.92 | |
| episode: 1374 | score: 4.0 | memory length: 268870 | epsilon: 0.6656354200072587 |
| steps: 298 | lr: 0.0001 | evaluation reward: 2.93 | |
| episode: 1375 | score: 5.0 | memory length: 269191 | epsilon: 0.6649998400072725 |
| steps: 321 | lr: 0.0001 | evaluation reward: 2.94 | |
| episode: 1376 | score: 6.0 | memory length: 269582 | epsilon: 0.6642256600072893 |
| steps: 391 | lr: 0.0001 | evaluation reward: 2.96 | |
| episode: 1377 | score: 4.0 | memory length: 269856 | epsilon: 0.6636831400073011 |
| steps: 274 | lr: 0.0001 | evaluation reward: 2.98 | |
| episode: 1378 | score: 5.0 | memory length: 270181 | epsilon: 0.6630396400073151 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 325 | lr: 0.0001 | evaluation reward: 3.02 | |
| episode: 1379 | score: 2.0 | memory length: 270362 | epsilon: 0.6626812600073229 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.01 | |
| episode: 1380 | score: 4.0 | memory length: 270619 | epsilon: 0.6621724000073339 |
| steps: 257 | lr: 0.0001 | evaluation reward: 3.03 | |
| episode: 1381 | score: 2.0 | memory length: 270800 | epsilon: 0.6618140200073417 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.0 | |
| episode: 1382 | score: 2.0 | memory length: 270997 | epsilon: 0.6614239600073502 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.0 | |
| episode: 1383 | score: 4.0 | memory length: 271313 | epsilon: 0.6607982800073637 |
| steps: 316 | lr: 0.0001 | evaluation reward: 3.04 | |
| episode: 1384 | score: 3.0 | memory length: 271541 | epsilon: 0.6603468400073735 |
| steps: 228 | lr: 0.0001 | evaluation reward: 3.04 | |
| episode: 1385 | score: 4.0 | memory length: 271817 | epsilon: 0.6598003600073854 |
| steps: 276 | lr: 0.0001 | evaluation reward: 3.04 | |
| episode: 1386 | score: 2.0 | memory length: 272015 | epsilon: 0.6594083200073939 |
| steps: 198 | lr: 0.0001 | evaluation reward: 3.02 | |
| episode: 1387 | score: 3.0 | memory length: 272263 | epsilon: 0.6589172800074046 |
| steps: 248 | lr: 0.0001 | evaluation reward: 3.03 | |
| episode: 1388 | score: 2.0 | memory length: 272460 | epsilon: 0.658527220007413 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.03 | |
| episode: 1389 | score: 3.0 | memory length: 272686 | epsilon: 0.6580797400074228 |
| steps: 226 | lr: 0.0001 | evaluation reward: 3.03 | |
| episode: 1390 | score: 4.0 | memory length: 272960 | epsilon: 0.6575372200074345 |
| steps: 274 | lr: 0.0001 | evaluation reward: 3.06 | |
| episode: 1391 | score: 3.0 | memory length: 273187 | epsilon: 0.6570877600074443 |
| steps: 227 | lr: 0.0001 | evaluation reward: 3.06 | |
| episode: 1392 | score: 3.0 | memory length: 273433 | epsilon: 0.6566006800074549 |
| steps: 246 | lr: 0.0001 | evaluation reward: 3.06 | |
| episode: 1393 | score: 6.0 | memory length: 273808 | epsilon: 0.655858180007471 |
| steps: 375 | lr: 0.0001 | evaluation reward: 3.11 | |
| episode: 1394 | score: 6.0 | memory length: 274170 | epsilon: 0.6551414200074865 |
| steps: 362 | lr: 0.0001 | evaluation reward: 3.14 | |
| episode: 1395 | score: 4.0 | memory length: 274426 | epsilon: 0.6546345400074975 |
| steps: 256 | lr: 0.0001 | evaluation reward: 3.16 | |
| episode: 1396 | score: 2.0 | memory length: 274624 | epsilon: 0.6542425000075061 |
| steps: 198 | lr: 0.0001 | evaluation reward: 3.15 | |
| episode: 1397 | score: 5.0 | memory length: 274939 | epsilon: 0.6536188000075196 |
| steps: 315 | lr: 0.0001 | evaluation reward: 3.15 | |
| episode: 1398 | score: 5.0 | memory length: 275260 | epsilon: 0.6529832200075334 |
| steps: 321 | lr: 0.0001 | evaluation reward: 3.17 | |
| episode: 1399 | score: 3.0 | memory length: 275490 | epsilon: 0.6525278200075433 |
| steps: 230 | lr: 0.0001 | evaluation reward: 3.17 | |
| episode: 1400 | score: 4.0 | memory length: 275752 | epsilon: 0.6520090600075545 |
| steps: 262 | lr: 0.0001 | evaluation reward: 3.19 | |
| episode: 1401 | score: 4.0 | memory length: 276036 | epsilon: 0.6514467400075667 |
| steps: 284 | lr: 0.0001 | evaluation reward: 3.16 | |
| episode: 1402 | score: 3.0 | memory length: 276247 | epsilon: 0.6510289600075758 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 211 | lr: 0.0001 | evaluation reward: 3.16 | |
| episode: 1403 | score: 5.0 | memory length: 276548 | epsilon: 0.6504329800075888 |
| steps: 301 | lr: 0.0001 | evaluation reward: 3.16 | |
| episode: 1404 | score: 6.0 | memory length: 276884 | epsilon: 0.6497677000076032 |
| steps: 336 | lr: 0.0001 | evaluation reward: 3.2 | |
| episode: 1405 | score: 4.0 | memory length: 277142 | epsilon: 0.6492568600076143 |
| steps: 258 | lr: 0.0001 | evaluation reward: 3.22 | |
| episode: 1406 | score: 4.0 | memory length: 277438 | epsilon: 0.648670780007627 |
| steps: 296 | lr: 0.0001 | evaluation reward: 3.24 | |
| episode: 1407 | score: 5.0 | memory length: 277733 | epsilon: 0.6480866800076397 |
| steps: 295 | lr: 0.0001 | evaluation reward: 3.26 | |
| episode: 1408 | score: 1.0 | memory length: 277905 | epsilon: 0.6477461200076471 |
| steps: 172 | lr: 0.0001 | evaluation reward: 3.22 | |
| episode: 1409 | score: 2.0 | memory length: 278084 | epsilon: 0.6473917000076548 |
| steps: 179 | lr: 0.0001 | evaluation reward: 3.22 | |
| episode: 1410 | score: 4.0 | memory length: 278376 | epsilon: 0.6468135400076673 |
| steps: 292 | lr: 0.0001 | evaluation reward: 3.26 | |
| episode: 1411 | score: 0.0 | memory length: 278498 | epsilon: 0.6465719800076726 |
| steps: 122 | lr: 0.0001 | evaluation reward: 3.22 | |
| episode: 1412 | score: 4.0 | memory length: 278798 | epsilon: 0.6459779800076855 |
| steps: 300 | lr: 0.0001 | evaluation reward: 3.23 | |
| episode: 1413 | score: 2.0 | memory length: 278995 | epsilon: 0.6455879200076939 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.21 | |
| episode: 1414 | score: 3.0 | memory length: 279220 | epsilon: 0.6451424200077036 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.19 | |
| episode: 1415 | score: 1.0 | memory length: 279370 | epsilon: 0.6448454200077101 |
| steps: 150 | lr: 0.0001 | evaluation reward: 3.18 | |
| episode: 1416 | score: 3.0 | memory length: 279580 | epsilon: 0.6444296200077191 |
| steps: 210 | lr: 0.0001 | evaluation reward: 3.19 | |
| episode: 1417 | score: 4.0 | memory length: 279839 | epsilon: 0.6439168000077302 |
| steps: 259 | lr: 0.0001 | evaluation reward: 3.21 | |
| episode: 1418 | score: 3.0 | memory length: 280085 | epsilon: 0.6434297200077408 |
| steps: 246 | lr: 0.0001 | evaluation reward: 3.2 | |
| episode: 1419 | score: 4.0 | memory length: 280359 | epsilon: 0.6428872000077526 |
| steps: 274 | lr: 0.0001 | evaluation reward: 3.22 | |
| episode: 1420 | score: 2.0 | memory length: 280556 | epsilon: 0.642497140007761 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.22 | |
| episode: 1421 | score: 4.0 | memory length: 280831 | epsilon: 0.6419526400077729 |
| steps: 275 | lr: 0.0001 | evaluation reward: 3.25 | |
| episode: 1422 | score: 4.0 | memory length: 281106 | epsilon: 0.6414081400077847 |
| steps: 275 | lr: 0.0001 | evaluation reward: 3.26 | |
| episode: 1423 | score: 3.0 | memory length: 281316 | epsilon: 0.6409923400077937 |
| steps: 210 | lr: 0.0001 | evaluation reward: 3.27 | |
| episode: 1424 | score: 4.0 | memory length: 281597 | epsilon: 0.6404359600078058 |
| steps: 281 | lr: 0.0001 | evaluation reward: 3.28 | |
| episode: 1425 | score: 2.0 | memory length: 281796 | epsilon: 0.6400419400078143 |
| steps: 199 | lr: 0.0001 | evaluation reward: 3.25 | |
| episode: 1426 | score: 4.0 | memory length: 282093 | epsilon: 0.6394538800078271 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 297 | lr: 0.0001 | evaluation reward: 3.28 | |
| episode: 1427 | score: 3.0 | memory length: 282318 | epsilon: 0.6390083800078368 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.28 | |
| episode: 1428 | score: 2.0 | memory length: 282499 | epsilon: 0.6386500000078446 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.26 | |
| episode: 1429 | score: 3.0 | memory length: 282709 | epsilon: 0.6382342000078536 |
| steps: 210 | lr: 0.0001 | evaluation reward: 3.27 | |
| episode: 1430 | score: 1.0 | memory length: 282860 | epsilon: 0.6379352200078601 |
| steps: 151 | lr: 0.0001 | evaluation reward: 3.24 | |
| episode: 1431 | score: 4.0 | memory length: 283114 | epsilon: 0.637432300007871 |
| steps: 254 | lr: 0.0001 | evaluation reward: 3.22 | |
| episode: 1432 | score: 4.0 | memory length: 283390 | epsilon: 0.6368858200078829 |
| steps: 276 | lr: 0.0001 | evaluation reward: 3.21 | |
| episode: 1433 | score: 2.0 | memory length: 283587 | epsilon: 0.6364957600078913 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.2 | |
| episode: 1434 | score: 2.0 | memory length: 283768 | epsilon: 0.6361373800078991 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.17 | |
| episode: 1435 | score: 3.0 | memory length: 283993 | epsilon: 0.6356918800079088 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.16 | |
| episode: 1436 | score: 5.0 | memory length: 284301 | epsilon: 0.635082040007922 |
| steps: 308 | lr: 0.0001 | evaluation reward: 3.17 | |
| episode: 1437 | score: 2.0 | memory length: 284500 | epsilon: 0.6346880200079306 |
| steps: 199 | lr: 0.0001 | evaluation reward: 3.15 | |
| episode: 1438 | score: 3.0 | memory length: 284727 | epsilon: 0.6342385600079403 |
| steps: 227 | lr: 0.0001 | evaluation reward: 3.16 | |
| episode: 1439 | score: 3.0 | memory length: 284952 | epsilon: 0.63379306000795 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.15 | |
| episode: 1440 | score: 5.0 | memory length: 285296 | epsilon: 0.6331119400079648 |
| steps: 344 | lr: 0.0001 | evaluation reward: 3.18 | |
| episode: 1441 | score: 6.0 | memory length: 285668 | epsilon: 0.6323753800079808 |
| steps: 372 | lr: 0.0001 | evaluation reward: 3.21 | |
| episode: 1442 | score: 3.0 | memory length: 285893 | epsilon: 0.6319298800079904 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.2 | |
| episode: 1443 | score: 5.0 | memory length: 286209 | epsilon: 0.631304200008004 |
| steps: 316 | lr: 0.0001 | evaluation reward: 3.22 | |
| episode: 1444 | score: 2.0 | memory length: 286390 | epsilon: 0.6309458200080118 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.22 | |
| episode: 1445 | score: 3.0 | memory length: 286615 | epsilon: 0.6305003200080215 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.21 | |
| episode: 1446 | score: 1.0 | memory length: 286766 | epsilon: 0.630201340008028 |
| steps: 151 | lr: 0.0001 | evaluation reward: 3.21 | |
| episode: 1447 | score: 3.0 | memory length: 287015 | epsilon: 0.6297083200080387 |
| steps: 249 | lr: 0.0001 | evaluation reward: 3.22 | |
| episode: 1448 | score: 5.0 | memory length: 287300 | epsilon: 0.6291440200080509 |
| steps: 285 | lr: 0.0001 | evaluation reward: 3.26 | |
| episode: 1449 | score: 3.0 | memory length: 287513 | epsilon: 0.6287222800080601 |
| steps: 213 | lr: 0.0001 | evaluation reward: 3.25 | |
| episode: 1450 | score: 5.0 | memory length: 287816 | epsilon: 0.6281223400080731 |

| | | | |
|--------------------|-------------|-------------------------|-----------------------------|
| steps: 303 | lr: 0.0001 | evaluation reward: 3.26 | |
| episode: 1451 | score: 4.0 | memory length: 288112 | epsilon: 0.6275362600080858 |
| steps: 296 | lr: 0.0001 | evaluation reward: 3.27 | |
| episode: 1452 | score: 3.0 | memory length: 288359 | epsilon: 0.6270472000080964 |
| steps: 247 | lr: 0.0001 | evaluation reward: 3.29 | |
| episode: 1453 | score: 3.0 | memory length: 288610 | epsilon: 0.6265502200081072 |
| steps: 251 | lr: 0.0001 | evaluation reward: 3.28 | |
| episode: 1454 | score: 3.0 | memory length: 288856 | epsilon: 0.6260631400081178 |
| steps: 246 | lr: 0.0001 | evaluation reward: 3.31 | |
| episode: 1455 | score: 2.0 | memory length: 289074 | epsilon: 0.6256315000081272 |
| steps: 218 | lr: 0.0001 | evaluation reward: 3.3 | |
| episode: 1456 | score: 3.0 | memory length: 289299 | epsilon: 0.6251860000081368 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.27 | |
| episode: 1457 | score: 4.0 | memory length: 289578 | epsilon: 0.6246335800081488 |
| steps: 279 | lr: 0.0001 | evaluation reward: 3.28 | |
| episode: 1458 | score: 3.0 | memory length: 289804 | epsilon: 0.6241861000081586 |
| steps: 226 | lr: 0.0001 | evaluation reward: 3.28 | |
| episode: 1459 | score: 3.0 | memory length: 290030 | epsilon: 0.6237386200081683 |
| steps: 226 | lr: 0.0001 | evaluation reward: 3.27 | |
| episode: 1460 | score: 3.0 | memory length: 290300 | epsilon: 0.6232040200081799 |
| steps: 270 | lr: 0.0001 | evaluation reward: 3.29 | |
| episode: 1461 | score: 3.0 | memory length: 290546 | epsilon: 0.6227169400081904 |
| steps: 246 | lr: 0.0001 | evaluation reward: 3.29 | |
| episode: 1462 | score: 2.0 | memory length: 290727 | epsilon: 0.6223585600081982 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.28 | |
| episode: 1463 | score: 4.0 | memory length: 291014 | epsilon: 0.6217903000082106 |
| steps: 287 | lr: 0.0001 | evaluation reward: 3.29 | |
| episode: 1464 | score: 3.0 | memory length: 291224 | epsilon: 0.6213745000082196 |
| steps: 210 | lr: 0.0001 | evaluation reward: 3.29 | |
| episode: 1465 | score: 5.0 | memory length: 291527 | epsilon: 0.6207745600082326 |
| steps: 303 | lr: 0.0001 | evaluation reward: 3.31 | |
| episode: 1466 | score: 4.0 | memory length: 291823 | epsilon: 0.6201884800082453 |
| steps: 296 | lr: 0.0001 | evaluation reward: 3.31 | |
| episode: 1467 | score: 5.0 | memory length: 292166 | epsilon: 0.6195093400082601 |
| steps: 343 | lr: 0.0001 | evaluation reward: 3.33 | |
| episode: 1468 | score: 1.0 | memory length: 292335 | epsilon: 0.6191747200082673 |
| steps: 169 | lr: 0.0001 | evaluation reward: 3.33 | |
| episode: 1469 | score: 5.0 | memory length: 292658 | epsilon: 0.6185351800082812 |
| steps: 323 | lr: 0.0001 | evaluation reward: 3.35 | |
| episode: 1470 | score: 3.0 | memory length: 292883 | epsilon: 0.6180896800082909 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.38 | |
| episode: 1471 | score: 10.0 | memory length: 293324 | epsilon: |
| 0.6172165000083099 | steps: 441 | lr: 0.0001 | evaluation reward: 3.44 |
| episode: 1472 | score: 6.0 | memory length: 293663 | epsilon: 0.6165452800083244 |
| steps: 339 | lr: 0.0001 | evaluation reward: 3.48 | |
| episode: 1473 | score: 4.0 | memory length: 293937 | epsilon: 0.6160027600083362 |
| steps: 274 | lr: 0.0001 | evaluation reward: 3.49 | |
| episode: 1474 | score: 2.0 | memory length: 294134 | epsilon: 0.6156127000083447 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 197 | lr: 0.0001 | evaluation reward: 3.47 | |
| episode: 1475 | score: 3.0 | memory length: 294362 | epsilon: 0.6151612600083545 |
| steps: 228 | lr: 0.0001 | evaluation reward: 3.45 | |
| episode: 1476 | score: 2.0 | memory length: 294560 | epsilon: 0.614769220008363 |
| steps: 198 | lr: 0.0001 | evaluation reward: 3.41 | |
| episode: 1477 | score: 3.0 | memory length: 294770 | epsilon: 0.614353420008372 |
| steps: 210 | lr: 0.0001 | evaluation reward: 3.4 | |
| episode: 1478 | score: 4.0 | memory length: 295037 | epsilon: 0.6138247600083835 |
| steps: 267 | lr: 0.0001 | evaluation reward: 3.39 | |
| episode: 1479 | score: 3.0 | memory length: 295247 | epsilon: 0.6134089600083925 |
| steps: 210 | lr: 0.0001 | evaluation reward: 3.4 | |
| episode: 1480 | score: 7.0 | memory length: 295611 | epsilon: 0.6126882400084082 |
| steps: 364 | lr: 0.0001 | evaluation reward: 3.43 | |
| episode: 1481 | score: 5.0 | memory length: 295904 | epsilon: 0.6121081000084208 |
| steps: 293 | lr: 0.0001 | evaluation reward: 3.46 | |
| episode: 1482 | score: 5.0 | memory length: 296216 | epsilon: 0.6114903400084342 |
| steps: 312 | lr: 0.0001 | evaluation reward: 3.49 | |
| episode: 1483 | score: 4.0 | memory length: 296496 | epsilon: 0.6109359400084462 |
| steps: 280 | lr: 0.0001 | evaluation reward: 3.49 | |
| episode: 1484 | score: 1.0 | memory length: 296667 | epsilon: 0.6105973600084535 |
| steps: 171 | lr: 0.0001 | evaluation reward: 3.47 | |
| episode: 1485 | score: 2.0 | memory length: 296884 | epsilon: 0.6101677000084629 |
| steps: 217 | lr: 0.0001 | evaluation reward: 3.45 | |
| episode: 1486 | score: 4.0 | memory length: 297179 | epsilon: 0.6095836000084756 |
| steps: 295 | lr: 0.0001 | evaluation reward: 3.47 | |
| episode: 1487 | score: 4.0 | memory length: 297453 | epsilon: 0.6090410800084873 |
| steps: 274 | lr: 0.0001 | evaluation reward: 3.48 | |
| episode: 1488 | score: 5.0 | memory length: 297762 | epsilon: 0.6084292600085006 |
| steps: 309 | lr: 0.0001 | evaluation reward: 3.51 | |
| episode: 1489 | score: 2.0 | memory length: 297960 | epsilon: 0.6080372200085091 |
| steps: 198 | lr: 0.0001 | evaluation reward: 3.5 | |
| episode: 1490 | score: 4.0 | memory length: 298258 | epsilon: 0.6074471800085219 |
| steps: 298 | lr: 0.0001 | evaluation reward: 3.5 | |
| episode: 1491 | score: 5.0 | memory length: 298564 | epsilon: 0.6068413000085351 |
| steps: 306 | lr: 0.0001 | evaluation reward: 3.52 | |
| episode: 1492 | score: 4.0 | memory length: 298838 | epsilon: 0.6062987800085469 |
| steps: 274 | lr: 0.0001 | evaluation reward: 3.53 | |
| episode: 1493 | score: 3.0 | memory length: 299065 | epsilon: 0.6058493200085566 |
| steps: 227 | lr: 0.0001 | evaluation reward: 3.5 | |
| episode: 1494 | score: 3.0 | memory length: 299296 | epsilon: 0.6053919400085666 |
| steps: 231 | lr: 0.0001 | evaluation reward: 3.47 | |
| episode: 1495 | score: 6.0 | memory length: 299660 | epsilon: 0.6046712200085822 |
| steps: 364 | lr: 0.0001 | evaluation reward: 3.49 | |
| episode: 1511 | score: 4.0 | memory length: 303803 | epsilon: 0.5964680800087603 |
| steps: 275 | lr: 0.0001 | evaluation reward: 3.49 | |
| episode: 1512 | score: 5.0 | memory length: 304098 | epsilon: 0.595883980008773 |
| steps: 295 | lr: 0.0001 | evaluation reward: 3.5 | |
| episode: 1513 | score: 5.0 | memory length: 304427 | epsilon: 0.5952325600087871 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 329 | lr: 0.0001 | evaluation reward: 3.53 | |
| episode: 1514 | score: 4.0 | memory length: 304685 | epsilon: 0.5947217200087982 |
| steps: 258 | lr: 0.0001 | evaluation reward: 3.54 | |
| episode: 1515 | score: 4.0 | memory length: 304960 | epsilon: 0.59417722000881 |
| steps: 275 | lr: 0.0001 | evaluation reward: 3.57 | |
| episode: 1516 | score: 3.0 | memory length: 305186 | epsilon: 0.5937297400088197 |
| steps: 226 | lr: 0.0001 | evaluation reward: 3.57 | |
| episode: 1517 | score: 2.0 | memory length: 305365 | epsilon: 0.5933753200088274 |
| steps: 179 | lr: 0.0001 | evaluation reward: 3.55 | |
| episode: 1518 | score: 3.0 | memory length: 305612 | epsilon: 0.592886260008838 |
| steps: 247 | lr: 0.0001 | evaluation reward: 3.55 | |
| episode: 1519 | score: 4.0 | memory length: 305887 | epsilon: 0.5923417600088499 |
| steps: 275 | lr: 0.0001 | evaluation reward: 3.55 | |
| episode: 1520 | score: 4.0 | memory length: 306165 | epsilon: 0.5917913200088618 |
| steps: 278 | lr: 0.0001 | evaluation reward: 3.57 | |
| episode: 1521 | score: 2.0 | memory length: 306363 | epsilon: 0.5913992800088703 |
| steps: 198 | lr: 0.0001 | evaluation reward: 3.55 | |
| episode: 1522 | score: 5.0 | memory length: 306686 | epsilon: 0.5907597400088842 |
| steps: 323 | lr: 0.0001 | evaluation reward: 3.56 | |
| episode: 1523 | score: 5.0 | memory length: 306991 | epsilon: 0.5901558400088973 |
| steps: 305 | lr: 0.0001 | evaluation reward: 3.58 | |
| episode: 1524 | score: 2.0 | memory length: 307173 | epsilon: 0.5897954800089051 |
| steps: 182 | lr: 0.0001 | evaluation reward: 3.56 | |
| episode: 1525 | score: 4.0 | memory length: 307453 | epsilon: 0.5892410800089172 |
| steps: 280 | lr: 0.0001 | evaluation reward: 3.58 | |
| episode: 1526 | score: 3.0 | memory length: 307662 | epsilon: 0.5888272600089262 |
| steps: 209 | lr: 0.0001 | evaluation reward: 3.57 | |
| episode: 1527 | score: 4.0 | memory length: 307937 | epsilon: 0.588282760008938 |
| steps: 275 | lr: 0.0001 | evaluation reward: 3.58 | |
| episode: 1528 | score: 4.0 | memory length: 308196 | epsilon: 0.5877699400089491 |
| steps: 259 | lr: 0.0001 | evaluation reward: 3.6 | |
| episode: 1529 | score: 4.0 | memory length: 308461 | epsilon: 0.5872452400089605 |
| steps: 265 | lr: 0.0001 | evaluation reward: 3.61 | |
| episode: 1530 | score: 5.0 | memory length: 308774 | epsilon: 0.586625500008974 |
| steps: 313 | lr: 0.0001 | evaluation reward: 3.65 | |
| episode: 1531 | score: 4.0 | memory length: 309036 | epsilon: 0.5861067400089852 |
| steps: 262 | lr: 0.0001 | evaluation reward: 3.65 | |
| episode: 1532 | score: 5.0 | memory length: 309384 | epsilon: 0.5854177000090002 |
| steps: 348 | lr: 0.0001 | evaluation reward: 3.66 | |
| episode: 1533 | score: 3.0 | memory length: 309609 | epsilon: 0.5849722000090098 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.67 | |
| episode: 1534 | score: 1.0 | memory length: 309760 | epsilon: 0.5846732200090163 |
| steps: 151 | lr: 0.0001 | evaluation reward: 3.66 | |
| episode: 1535 | score: 2.0 | memory length: 309957 | epsilon: 0.5842831600090248 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.65 | |
| episode: 1536 | score: 4.0 | memory length: 310213 | epsilon: 0.5837762800090358 |
| steps: 256 | lr: 0.0001 | evaluation reward: 3.64 | |
| episode: 1537 | score: 3.0 | memory length: 310441 | epsilon: 0.5833248400090456 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 228 | lr: 0.0001 | evaluation reward: 3.65 | |
| episode: 1538 | score: 4.0 | memory length: 310707 | epsilon: 0.582798160009057 |
| steps: 266 | lr: 0.0001 | evaluation reward: 3.66 | |
| episode: 1539 | score: 3.0 | memory length: 310935 | epsilon: 0.5823467200090668 |
| steps: 228 | lr: 0.0001 | evaluation reward: 3.66 | |
| episode: 1540 | score: 7.0 | memory length: 311310 | epsilon: 0.581604220009083 |
| steps: 375 | lr: 0.0001 | evaluation reward: 3.68 | |
| episode: 1541 | score: 3.0 | memory length: 311536 | epsilon: 0.5811567400090927 |
| steps: 226 | lr: 0.0001 | evaluation reward: 3.65 | |
| episode: 1542 | score: 2.0 | memory length: 311734 | epsilon: 0.5807647000091012 |
| steps: 198 | lr: 0.0001 | evaluation reward: 3.64 | |
| episode: 1543 | score: 4.0 | memory length: 311981 | epsilon: 0.5802756400091118 |
| steps: 247 | lr: 0.0001 | evaluation reward: 3.63 | |
| episode: 1544 | score: 3.0 | memory length: 312196 | epsilon: 0.579849940009121 |
| steps: 215 | lr: 0.0001 | evaluation reward: 3.64 | |
| episode: 1545 | score: 2.0 | memory length: 312393 | epsilon: 0.5794598800091295 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.63 | |
| episode: 1546 | score: 1.0 | memory length: 312563 | epsilon: 0.5791232800091368 |
| steps: 170 | lr: 0.0001 | evaluation reward: 3.63 | |
| episode: 1547 | score: 1.0 | memory length: 312713 | epsilon: 0.5788262800091433 |
| steps: 150 | lr: 0.0001 | evaluation reward: 3.61 | |
| episode: 1548 | score: 2.0 | memory length: 312929 | epsilon: 0.5783986000091526 |
| steps: 216 | lr: 0.0001 | evaluation reward: 3.58 | |
| episode: 1549 | score: 3.0 | memory length: 313142 | epsilon: 0.5779768600091617 |
| steps: 213 | lr: 0.0001 | evaluation reward: 3.58 | |
| episode: 1550 | score: 2.0 | memory length: 313358 | epsilon: 0.577549180009171 |
| steps: 216 | lr: 0.0001 | evaluation reward: 3.55 | |
| episode: 1551 | score: 6.0 | memory length: 313758 | epsilon: 0.5767571800091882 |
| steps: 400 | lr: 0.0001 | evaluation reward: 3.57 | |
| episode: 1552 | score: 4.0 | memory length: 314019 | epsilon: 0.5762404000091994 |
| steps: 261 | lr: 0.0001 | evaluation reward: 3.58 | |
| episode: 1553 | score: 3.0 | memory length: 314268 | epsilon: 0.5757473800092101 |
| steps: 249 | lr: 0.0001 | evaluation reward: 3.58 | |
| episode: 1554 | score: 3.0 | memory length: 314493 | epsilon: 0.5753018800092198 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.58 | |
| episode: 1555 | score: 5.0 | memory length: 314817 | epsilon: 0.5746603600092337 |
| steps: 324 | lr: 0.0001 | evaluation reward: 3.61 | |
| episode: 1556 | score: 5.0 | memory length: 315142 | epsilon: 0.5740168600092477 |
| steps: 325 | lr: 0.0001 | evaluation reward: 3.63 | |
| episode: 1557 | score: 3.0 | memory length: 315368 | epsilon: 0.5735693800092574 |
| steps: 226 | lr: 0.0001 | evaluation reward: 3.62 | |
| episode: 1558 | score: 4.0 | memory length: 315663 | epsilon: 0.5729852800092701 |
| steps: 295 | lr: 0.0001 | evaluation reward: 3.63 | |
| episode: 1559 | score: 2.0 | memory length: 315845 | epsilon: 0.5726249200092779 |
| steps: 182 | lr: 0.0001 | evaluation reward: 3.62 | |
| episode: 1560 | score: 2.0 | memory length: 316027 | epsilon: 0.5722645600092857 |
| steps: 182 | lr: 0.0001 | evaluation reward: 3.61 | |
| episode: 1561 | score: 4.0 | memory length: 316270 | epsilon: 0.5717834200092962 |

| | | | |
|---------------|-------------|-------------------------|-----------------------------|
| steps: 243 | lr: 0.0001 | evaluation reward: 3.62 | |
| episode: 1562 | score: 4.0 | memory length: 316563 | epsilon: 0.5712032800093088 |
| steps: 293 | lr: 0.0001 | evaluation reward: 3.64 | |
| episode: 1563 | score: 2.0 | memory length: 316760 | epsilon: 0.5708132200093172 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.62 | |
| episode: 1564 | score: 3.0 | memory length: 316989 | epsilon: 0.5703598000093271 |
| steps: 229 | lr: 0.0001 | evaluation reward: 3.62 | |
| episode: 1565 | score: 6.0 | memory length: 317365 | epsilon: 0.5696153200093432 |
| steps: 376 | lr: 0.0001 | evaluation reward: 3.63 | |
| episode: 1566 | score: 4.0 | memory length: 317659 | epsilon: 0.5690332000093559 |
| steps: 294 | lr: 0.0001 | evaluation reward: 3.63 | |
| episode: 1567 | score: 2.0 | memory length: 317877 | epsilon: 0.5686015600093652 |
| steps: 218 | lr: 0.0001 | evaluation reward: 3.6 | |
| episode: 1568 | score: 2.0 | memory length: 318074 | epsilon: 0.5682115000093737 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.61 | |
| episode: 1569 | score: 4.0 | memory length: 318351 | epsilon: 0.5676630400093856 |
| steps: 277 | lr: 0.0001 | evaluation reward: 3.6 | |
| episode: 1570 | score: 2.0 | memory length: 318549 | epsilon: 0.5672710000093941 |
| steps: 198 | lr: 0.0001 | evaluation reward: 3.59 | |
| episode: 1571 | score: 5.0 | memory length: 318871 | epsilon: 0.566633440009408 |
| steps: 322 | lr: 0.0001 | evaluation reward: 3.54 | |
| episode: 1572 | score: 4.0 | memory length: 319158 | epsilon: 0.5660651800094203 |
| steps: 287 | lr: 0.0001 | evaluation reward: 3.52 | |
| episode: 1573 | score: 2.0 | memory length: 319355 | epsilon: 0.5656751200094288 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.5 | |
| episode: 1574 | score: 10.0 | memory length: 319713 | epsilon: 0.5649662800094442 |
| steps: 358 | lr: 0.0001 | evaluation reward: 3.58 | |
| episode: 1575 | score: 7.0 | memory length: 320139 | epsilon: 0.5641228000094625 |
| steps: 426 | lr: 0.0001 | evaluation reward: 3.62 | |
| episode: 1576 | score: 3.0 | memory length: 320351 | epsilon: 0.5637030400094716 |
| steps: 212 | lr: 0.0001 | evaluation reward: 3.63 | |
| episode: 1577 | score: 7.0 | memory length: 320743 | epsilon: 0.5629268800094884 |
| steps: 392 | lr: 0.0001 | evaluation reward: 3.67 | |
| episode: 1578 | score: 6.0 | memory length: 321070 | epsilon: 0.5622794200095025 |
| steps: 327 | lr: 0.0001 | evaluation reward: 3.69 | |
| episode: 1579 | score: 4.0 | memory length: 321364 | epsilon: 0.5616973000095151 |
| steps: 294 | lr: 0.0001 | evaluation reward: 3.7 | |
| episode: 1580 | score: 3.0 | memory length: 321610 | epsilon: 0.5612102200095257 |
| steps: 246 | lr: 0.0001 | evaluation reward: 3.66 | |
| episode: 1581 | score: 6.0 | memory length: 321948 | epsilon: 0.5605409800095402 |
| steps: 338 | lr: 0.0001 | evaluation reward: 3.67 | |
| episode: 1582 | score: 4.0 | memory length: 322228 | epsilon: 0.5599865800095523 |
| steps: 280 | lr: 0.0001 | evaluation reward: 3.66 | |
| episode: 1583 | score: 5.0 | memory length: 322539 | epsilon: 0.5593708000095656 |
| steps: 311 | lr: 0.0001 | evaluation reward: 3.67 | |
| episode: 1584 | score: 5.0 | memory length: 322857 | epsilon: 0.5587411600095793 |
| steps: 318 | lr: 0.0001 | evaluation reward: 3.71 | |
| episode: 1585 | score: 3.0 | memory length: 323101 | epsilon: 0.5582580400095898 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 244 | lr: 0.0001 | evaluation reward: 3.72 | |
| episode: 1586 | score: 5.0 | memory length: 323391 | epsilon: 0.5576838400096022 |
| steps: 290 | lr: 0.0001 | evaluation reward: 3.73 | |
| episode: 1587 | score: 3.0 | memory length: 323620 | epsilon: 0.5572304200096121 |
| steps: 229 | lr: 0.0001 | evaluation reward: 3.72 | |
| episode: 1588 | score: 3.0 | memory length: 323847 | epsilon: 0.5567809600096219 |
| steps: 227 | lr: 0.0001 | evaluation reward: 3.7 | |
| episode: 1589 | score: 3.0 | memory length: 324114 | epsilon: 0.5562523000096333 |
| steps: 267 | lr: 0.0001 | evaluation reward: 3.71 | |
| episode: 1590 | score: 3.0 | memory length: 324325 | epsilon: 0.5558345200096424 |
| steps: 211 | lr: 0.0001 | evaluation reward: 3.7 | |
| episode: 1591 | score: 6.0 | memory length: 324685 | epsilon: 0.5551217200096579 |
| steps: 360 | lr: 0.0001 | evaluation reward: 3.71 | |
| episode: 1592 | score: 5.0 | memory length: 325029 | epsilon: 0.5544406000096727 |
| steps: 344 | lr: 0.0001 | evaluation reward: 3.72 | |
| episode: 1593 | score: 1.0 | memory length: 325179 | epsilon: 0.5541436000096791 |
| steps: 150 | lr: 0.0001 | evaluation reward: 3.7 | |
| episode: 1594 | score: 3.0 | memory length: 325426 | epsilon: 0.5536545400096897 |
| steps: 247 | lr: 0.0001 | evaluation reward: 3.7 | |
| episode: 1595 | score: 4.0 | memory length: 325691 | epsilon: 0.5531298400097011 |
| steps: 265 | lr: 0.0001 | evaluation reward: 3.68 | |
| episode: 1596 | score: 5.0 | memory length: 326019 | epsilon: 0.5524804000097152 |
| steps: 328 | lr: 0.0001 | evaluation reward: 3.7 | |
| episode: 1597 | score: 4.0 | memory length: 326276 | epsilon: 0.5519715400097263 |
| steps: 257 | lr: 0.0001 | evaluation reward: 3.72 | |
| episode: 1598 | score: 6.0 | memory length: 326615 | epsilon: 0.5513003200097408 |
| steps: 339 | lr: 0.0001 | evaluation reward: 3.75 | |
| episode: 1599 | score: 7.0 | memory length: 327009 | epsilon: 0.5505202000097578 |
| steps: 394 | lr: 0.0001 | evaluation reward: 3.79 | |
| episode: 1600 | score: 1.0 | memory length: 327177 | epsilon: 0.550187560009765 |
| steps: 168 | lr: 0.0001 | evaluation reward: 3.75 | |
| episode: 1601 | score: 8.0 | memory length: 327597 | epsilon: 0.549355960009783 |
| steps: 420 | lr: 0.0001 | evaluation reward: 3.8 | |
| episode: 1602 | score: 2.0 | memory length: 327796 | epsilon: 0.5489619400097916 |
| steps: 199 | lr: 0.0001 | evaluation reward: 3.76 | |
| episode: 1603 | score: 7.0 | memory length: 328045 | epsilon: 0.5484689200098023 |
| steps: 249 | lr: 0.0001 | evaluation reward: 3.79 | |
| episode: 1604 | score: 3.0 | memory length: 328273 | epsilon: 0.5480174800098121 |
| steps: 228 | lr: 0.0001 | evaluation reward: 3.79 | |
| episode: 1605 | score: 4.0 | memory length: 328555 | epsilon: 0.5474591200098242 |
| steps: 282 | lr: 0.0001 | evaluation reward: 3.8 | |
| episode: 1606 | score: 4.0 | memory length: 328851 | epsilon: 0.5468730400098369 |
| steps: 296 | lr: 0.0001 | evaluation reward: 3.81 | |
| episode: 1607 | score: 8.0 | memory length: 329245 | epsilon: 0.5460929200098539 |
| steps: 394 | lr: 0.0001 | evaluation reward: 3.84 | |
| episode: 1608 | score: 0.0 | memory length: 329367 | epsilon: 0.5458513600098591 |
| steps: 122 | lr: 0.0001 | evaluation reward: 3.81 | |
| episode: 1609 | score: 4.0 | memory length: 329642 | epsilon: 0.5453068600098709 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 275 | lr: 0.0001 | evaluation reward: 3.81 | |
| episode: 1610 | score: 2.0 | memory length: 329840 | epsilon: 0.5449148200098795 |
| steps: 198 | lr: 0.0001 | evaluation reward: 3.8 | |
| episode: 1611 | score: 6.0 | memory length: 330233 | epsilon: 0.5441366800098963 |
| steps: 393 | lr: 0.0001 | evaluation reward: 3.82 | |
| episode: 1612 | score: 4.0 | memory length: 330533 | epsilon: 0.5435426800099092 |
| steps: 300 | lr: 0.0001 | evaluation reward: 3.81 | |
| episode: 1613 | score: 4.0 | memory length: 330792 | epsilon: 0.5430298600099204 |
| steps: 259 | lr: 0.0001 | evaluation reward: 3.8 | |
| episode: 1614 | score: 2.0 | memory length: 330974 | epsilon: 0.5426695000099282 |
| steps: 182 | lr: 0.0001 | evaluation reward: 3.78 | |
| episode: 1615 | score: 2.0 | memory length: 331155 | epsilon: 0.542311120009936 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.76 | |
| episode: 1616 | score: 2.0 | memory length: 331335 | epsilon: 0.5419547200099437 |
| steps: 180 | lr: 0.0001 | evaluation reward: 3.75 | |
| episode: 1617 | score: 2.0 | memory length: 331517 | epsilon: 0.5415943600099515 |
| steps: 182 | lr: 0.0001 | evaluation reward: 3.75 | |
| episode: 1618 | score: 7.0 | memory length: 331934 | epsilon: 0.5407687000099695 |
| steps: 417 | lr: 0.0001 | evaluation reward: 3.79 | |
| episode: 1619 | score: 5.0 | memory length: 332254 | epsilon: 0.5401351000099832 |
| steps: 320 | lr: 0.0001 | evaluation reward: 3.8 | |
| episode: 1620 | score: 4.0 | memory length: 332510 | epsilon: 0.5396282200099942 |
| steps: 256 | lr: 0.0001 | evaluation reward: 3.8 | |
| episode: 1621 | score: 4.0 | memory length: 332805 | epsilon: 0.5390441200100069 |
| steps: 295 | lr: 0.0001 | evaluation reward: 3.82 | |
| episode: 1622 | score: 3.0 | memory length: 333052 | epsilon: 0.5385550600100175 |
| steps: 247 | lr: 0.0001 | evaluation reward: 3.8 | |
| episode: 1623 | score: 4.0 | memory length: 333328 | epsilon: 0.5380085800100294 |
| steps: 276 | lr: 0.0001 | evaluation reward: 3.79 | |
| episode: 1624 | score: 3.0 | memory length: 333574 | epsilon: 0.53752150001004 |
| steps: 246 | lr: 0.0001 | evaluation reward: 3.8 | |
| episode: 1625 | score: 6.0 | memory length: 333916 | epsilon: 0.5368443400100547 |
| steps: 342 | lr: 0.0001 | evaluation reward: 3.82 | |
| episode: 1626 | score: 5.0 | memory length: 334261 | epsilon: 0.5361612400100695 |
| steps: 345 | lr: 0.0001 | evaluation reward: 3.84 | |
| episode: 1627 | score: 3.0 | memory length: 334490 | epsilon: 0.5357078200100793 |
| steps: 229 | lr: 0.0001 | evaluation reward: 3.83 | |
| episode: 1628 | score: 3.0 | memory length: 334735 | epsilon: 0.5352227200100899 |
| steps: 245 | lr: 0.0001 | evaluation reward: 3.82 | |
| episode: 1629 | score: 4.0 | memory length: 335014 | epsilon: 0.5346703000101019 |
| steps: 279 | lr: 0.0001 | evaluation reward: 3.82 | |
| episode: 1630 | score: 3.0 | memory length: 335246 | epsilon: 0.5342109400101118 |
| steps: 232 | lr: 0.0001 | evaluation reward: 3.8 | |
| episode: 1631 | score: 3.0 | memory length: 335472 | epsilon: 0.5337634600101215 |
| steps: 226 | lr: 0.0001 | evaluation reward: 3.79 | |
| episode: 1632 | score: 4.0 | memory length: 335769 | epsilon: 0.5331754000101343 |
| steps: 297 | lr: 0.0001 | evaluation reward: 3.78 | |
| episode: 1633 | score: 0.0 | memory length: 335891 | epsilon: 0.5329338400101395 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 122 | lr: 0.0001 | evaluation reward: 3.75 | |
| episode: 1634 | score: 6.0 | memory length: 336234 | epsilon: 0.5322547000101543 |
| steps: 343 | lr: 0.0001 | evaluation reward: 3.8 | |
| episode: 1635 | score: 3.0 | memory length: 336462 | epsilon: 0.5318032600101641 |
| steps: 228 | lr: 0.0001 | evaluation reward: 3.81 | |
| episode: 1636 | score: 4.0 | memory length: 336719 | epsilon: 0.5312944000101751 |
| steps: 257 | lr: 0.0001 | evaluation reward: 3.81 | |
| episode: 1637 | score: 2.0 | memory length: 336916 | epsilon: 0.5309043400101836 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.8 | |
| episode: 1638 | score: 6.0 | memory length: 337260 | epsilon: 0.5302232200101984 |
| steps: 344 | lr: 0.0001 | evaluation reward: 3.82 | |
| episode: 1639 | score: 3.0 | memory length: 337508 | epsilon: 0.529732180010209 |
| steps: 248 | lr: 0.0001 | evaluation reward: 3.82 | |
| episode: 1640 | score: 2.0 | memory length: 337705 | epsilon: 0.5293421200102175 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.77 | |
| episode: 1641 | score: 3.0 | memory length: 337931 | epsilon: 0.5288946400102272 |
| steps: 226 | lr: 0.0001 | evaluation reward: 3.77 | |
| episode: 1642 | score: 3.0 | memory length: 338156 | epsilon: 0.5284491400102369 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.78 | |
| episode: 1643 | score: 2.0 | memory length: 338353 | epsilon: 0.5280590800102454 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.76 | |
| episode: 1644 | score: 1.0 | memory length: 338504 | epsilon: 0.5277601000102519 |
| steps: 151 | lr: 0.0001 | evaluation reward: 3.74 | |
| episode: 1645 | score: 5.0 | memory length: 338807 | epsilon: 0.5271601600102649 |
| steps: 303 | lr: 0.0001 | evaluation reward: 3.77 | |
| episode: 1646 | score: 3.0 | memory length: 339050 | epsilon: 0.5266790200102753 |
| steps: 243 | lr: 0.0001 | evaluation reward: 3.79 | |
| episode: 1647 | score: 0.0 | memory length: 339172 | epsilon: 0.5264374600102806 |
| steps: 122 | lr: 0.0001 | evaluation reward: 3.78 | |
| episode: 1648 | score: 3.0 | memory length: 339419 | epsilon: 0.5259484000102912 |
| steps: 247 | lr: 0.0001 | evaluation reward: 3.79 | |
| episode: 1649 | score: 3.0 | memory length: 339645 | epsilon: 0.5255009200103009 |
| steps: 226 | lr: 0.0001 | evaluation reward: 3.79 | |
| episode: 1650 | score: 2.0 | memory length: 339827 | epsilon: 0.5251405600103087 |
| steps: 182 | lr: 0.0001 | evaluation reward: 3.79 | |
| episode: 1651 | score: 4.0 | memory length: 340103 | epsilon: 0.5245940800103206 |
| steps: 276 | lr: 0.0001 | evaluation reward: 3.77 | |
| episode: 1652 | score: 3.0 | memory length: 340328 | epsilon: 0.5241485800103303 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.76 | |
| episode: 1653 | score: 2.0 | memory length: 340509 | epsilon: 0.523790200010338 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.75 | |
| episode: 1654 | score: 3.0 | memory length: 340755 | epsilon: 0.5233031200103486 |
| steps: 246 | lr: 0.0001 | evaluation reward: 3.75 | |
| episode: 1655 | score: 2.0 | memory length: 340955 | epsilon: 0.5229071200103572 |
| steps: 200 | lr: 0.0001 | evaluation reward: 3.72 | |
| episode: 1656 | score: 4.0 | memory length: 341271 | epsilon: 0.5222814400103708 |
| steps: 316 | lr: 0.0001 | evaluation reward: 3.71 | |
| episode: 1657 | score: 8.0 | memory length: 341565 | epsilon: 0.5216993200103834 |

| | | | |
|---------------|------------|-------------------------|-----------------------------|
| steps: 294 | lr: 0.0001 | evaluation reward: 3.76 | |
| episode: 1658 | score: 3.0 | memory length: 341811 | epsilon: 0.521212240010394 |
| steps: 246 | lr: 0.0001 | evaluation reward: 3.75 | |
| episode: 1659 | score: 4.0 | memory length: 342085 | epsilon: 0.5206697200104058 |
| steps: 274 | lr: 0.0001 | evaluation reward: 3.77 | |
| episode: 1660 | score: 3.0 | memory length: 342311 | epsilon: 0.5202222400104155 |
| steps: 226 | lr: 0.0001 | evaluation reward: 3.78 | |
| episode: 1661 | score: 3.0 | memory length: 342536 | epsilon: 0.5197767400104252 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.77 | |
| episode: 1662 | score: 2.0 | memory length: 342716 | epsilon: 0.5194203400104329 |
| steps: 180 | lr: 0.0001 | evaluation reward: 3.75 | |
| episode: 1663 | score: 2.0 | memory length: 342897 | epsilon: 0.5190619600104407 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.75 | |
| episode: 1664 | score: 3.0 | memory length: 343122 | epsilon: 0.5186164600104504 |
| steps: 225 | lr: 0.0001 | evaluation reward: 3.75 | |
| episode: 1665 | score: 3.0 | memory length: 343348 | epsilon: 0.5181689800104601 |
| steps: 226 | lr: 0.0001 | evaluation reward: 3.72 | |
| episode: 1666 | score: 5.0 | memory length: 343654 | epsilon: 0.5175631000104732 |
| steps: 306 | lr: 0.0001 | evaluation reward: 3.73 | |
| episode: 1667 | score: 2.0 | memory length: 343834 | epsilon: 0.517206700010481 |
| steps: 180 | lr: 0.0001 | evaluation reward: 3.73 | |
| episode: 1668 | score: 2.0 | memory length: 344033 | epsilon: 0.5168126800104895 |
| steps: 199 | lr: 0.0001 | evaluation reward: 3.73 | |
| episode: 1669 | score: 2.0 | memory length: 344215 | epsilon: 0.5164523200104973 |
| steps: 182 | lr: 0.0001 | evaluation reward: 3.71 | |
| episode: 1670 | score: 2.0 | memory length: 344432 | epsilon: 0.5160226600105067 |
| steps: 217 | lr: 0.0001 | evaluation reward: 3.71 | |
| episode: 1671 | score: 3.0 | memory length: 344645 | epsilon: 0.5156009200105158 |
| steps: 213 | lr: 0.0001 | evaluation reward: 3.69 | |
| episode: 1672 | score: 4.0 | memory length: 344923 | epsilon: 0.5150504800105278 |
| steps: 278 | lr: 0.0001 | evaluation reward: 3.69 | |
| episode: 1673 | score: 6.0 | memory length: 345286 | epsilon: 0.5143317400105434 |
| steps: 363 | lr: 0.0001 | evaluation reward: 3.73 | |
| episode: 1674 | score: 6.0 | memory length: 345603 | epsilon: 0.513704080010557 |
| steps: 317 | lr: 0.0001 | evaluation reward: 3.69 | |
| episode: 1675 | score: 2.0 | memory length: 345801 | epsilon: 0.5133120400105655 |
| steps: 198 | lr: 0.0001 | evaluation reward: 3.64 | |
| episode: 1676 | score: 2.0 | memory length: 345982 | epsilon: 0.5129536600105733 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.63 | |
| episode: 1677 | score: 3.0 | memory length: 346229 | epsilon: 0.5124646000105839 |
| steps: 247 | lr: 0.0001 | evaluation reward: 3.59 | |
| episode: 1678 | score: 5.0 | memory length: 346521 | epsilon: 0.5118864400105965 |
| steps: 292 | lr: 0.0001 | evaluation reward: 3.58 | |
| episode: 1679 | score: 6.0 | memory length: 346867 | epsilon: 0.5112013600106113 |
| steps: 346 | lr: 0.0001 | evaluation reward: 3.6 | |
| episode: 1680 | score: 2.0 | memory length: 347064 | epsilon: 0.5108113000106198 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.59 | |
| episode: 1681 | score: 2.0 | memory length: 347264 | epsilon: 0.5104153000106284 |

| | | | |
|---------------------|------------|-------------------------|-----------------------------|
| steps: 200 | lr: 0.0001 | evaluation reward: 3.55 | |
| episode: 1682 | score: 3.0 | memory length: 347512 | epsilon: 0.5099242600106391 |
| steps: 248 | lr: 0.0001 | evaluation reward: 3.54 | |
| episode: 1683 | score: 2.0 | memory length: 347712 | epsilon: 0.5095282600106477 |
| steps: 200 | lr: 0.0001 | evaluation reward: 3.51 | |
| episode: 1684 | score: 3.0 | memory length: 347945 | epsilon: 0.5090669200106577 |
| steps: 233 | lr: 0.0001 | evaluation reward: 3.49 | |
| episode: 1685 | score: 6.0 | memory length: 348303 | epsilon: 0.5083580800106731 |
| steps: 358 | lr: 0.0001 | evaluation reward: 3.52 | |
| episode: 1686 | score: 2.0 | memory length: 348500 | epsilon: 0.5079680200106815 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.49 | |
| episode: 1687 | score: 4.0 | memory length: 348815 | epsilon: 0.5073443200106951 |
| steps: 315 | lr: 0.0001 | evaluation reward: 3.5 | |
| episode: 1688 | score: 2.0 | memory length: 349013 | epsilon: 0.5069522800107036 |
| steps: 198 | lr: 0.0001 | evaluation reward: 3.49 | |
| episode: 1689 | score: 3.0 | memory length: 349226 | epsilon: 0.5065305400107127 |
| steps: 213 | lr: 0.0001 | evaluation reward: 3.49 | |
| episode: 1690 | score: 5.0 | memory length: 349542 | epsilon: 0.5059048600107263 |
| steps: 316 | lr: 0.0001 | evaluation reward: 3.51 | |
| episode: 1691 | score: 2.0 | memory length: 349723 | epsilon: 0.5055464800107341 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.47 | |
| episode: 1692 | score: 2.0 | memory length: 349923 | epsilon: 0.5051504800107427 |
| steps: 200 | lr: 0.0001 | evaluation reward: 3.44 | |
| episode: 1693 | score: 2.0 | memory length: 350104 | epsilon: 0.5047921000107505 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.45 | |
| episode: 1694 | score: 4.0 | memory length: 350384 | epsilon: 0.5042377000107625 |
| steps: 280 | lr: 0.0001 | evaluation reward: 3.46 | |
| episode: 1695 | score: 4.0 | memory length: 350680 | epsilon: 0.5036516200107752 |
| steps: 296 | lr: 0.0001 | evaluation reward: 3.46 | |
| episode: 1696 | score: 2.0 | memory length: 350861 | epsilon: 0.503293240010783 |
| steps: 181 | lr: 0.0001 | evaluation reward: 3.43 | |
| episode: 1697 | score: 4.0 | memory length: 351121 | epsilon: 0.5027784400107942 |
| steps: 260 | lr: 0.0001 | evaluation reward: 3.43 | |
| episode: 1698 | score: 3.0 | memory length: 351369 | epsilon: 0.5022874000108049 |
| steps: 248 | lr: 0.0001 | evaluation reward: 3.4 | |
| episode: 1699 | score: 3.0 | memory length: 351597 | epsilon: 0.5018359600108147 |
| steps: 228 | lr: 0.0001 | evaluation reward: 3.36 | |
| episode: 1700 | score: 5.0 | memory length: 351921 | epsilon: 0.5011944400108286 |
| steps: 324 | lr: 0.0001 | evaluation reward: 3.4 | |
| episode: 1701 | score: 4.0 | memory length: 352237 | epsilon: 0.5005687600108422 |
| steps: 316 | lr: 0.0001 | evaluation reward: 3.36 | |
| episode: 1702 | score: 2.0 | memory length: 352436 | epsilon: 0.5001747400108507 |
| steps: 199 | lr: 0.0001 | evaluation reward: 3.36 | |
| episode: 1703 | score: 5.0 | memory length: 352718 | epsilon: |
| 0.49961638001085207 | steps: 282 | lr: 0.0001 | evaluation reward: 3.34 |
| episode: 1704 | score: 2.0 | memory length: 352915 | epsilon: 0.4992263200108496 |
| steps: 197 | lr: 0.0001 | evaluation reward: 3.33 | |
| episode: 1705 | score: 3.0 | memory length: 353161 | epsilon: 0.4987392400108465 |

```

steps: 246    lr: 0.0001    evaluation reward: 3.32
episode: 1706  score: 7.0    memory length: 353584    epsilon: 0.4979017000108412
steps: 423    lr: 0.0001    evaluation reward: 3.35
episode: 1707  score: 2.0    memory length: 353765    epsilon:
0.49754332001083895    steps: 181    lr: 0.0001    evaluation reward: 3.29
episode: 1708  score: 4.0    memory length: 354024    epsilon: 0.4970305000108357
steps: 259    lr: 0.0001    evaluation reward: 3.33
episode: 1709  score: 2.0    memory length: 354205    epsilon:
0.49667212001083344    steps: 181    lr: 0.0001    evaluation reward: 3.31
episode: 1710  score: 5.0    memory length: 354485    epsilon:
0.49611772001082993    steps: 280    lr: 0.0001    evaluation reward: 3.34
episode: 1711  score: 1.0    memory length: 354656    epsilon: 0.4957791400108278
steps: 171    lr: 0.0001    evaluation reward: 3.29
episode: 1712  score: 3.0    memory length: 354884    epsilon:
0.49532770001082493    steps: 228    lr: 0.0001    evaluation reward: 3.28
episode: 1713  score: 5.0    memory length: 355177    epsilon:
0.49474756001082126    steps: 293    lr: 0.0001    evaluation reward: 3.29
episode: 1714  score: 1.0    memory length: 355327    epsilon: 0.4944505600108194
steps: 150    lr: 0.0001    evaluation reward: 3.28
episode: 1715  score: 3.0    memory length: 355555    epsilon:
0.49399912001081653    steps: 228    lr: 0.0001    evaluation reward: 3.29
episode: 1716  score: 3.0    memory length: 355780    epsilon: 0.4935536200108137
steps: 225    lr: 0.0001    evaluation reward: 3.3
episode: 1717  score: 3.0    memory length: 355992    epsilon:
0.49313386001081105    steps: 212    lr: 0.0001    evaluation reward: 3.31
episode: 1718  score: 4.0    memory length: 356266    epsilon: 0.4925913400108076
steps: 274    lr: 0.0001    evaluation reward: 3.28
episode: 1719  score: 4.0    memory length: 356521    epsilon: 0.4920864400108044
steps: 255    lr: 0.0001    evaluation reward: 3.27
episode: 1720  score: 3.0    memory length: 356730    epsilon: 0.4916726200108018
steps: 209    lr: 0.0001    evaluation reward: 3.26
episode: 1721  score: 2.0    memory length: 356928    epsilon:
0.49128058001079933    steps: 198    lr: 0.0001    evaluation reward: 3.24
episode: 1722  score: 3.0    memory length: 357174    epsilon:
0.49079350001079625    steps: 246    lr: 0.0001    evaluation reward: 3.24
episode: 1723  score: 7.0    memory length: 357534    epsilon:
0.49008070001079174    steps: 360    lr: 0.0001    evaluation reward: 3.27
episode: 1724  score: 1.0    memory length: 357705    epsilon: 0.4897421200107896
steps: 171    lr: 0.0001    evaluation reward: 3.25
episode: 1725  score: 4.0    memory length: 357987    epsilon:
0.48918376001078606    steps: 282    lr: 0.0001    evaluation reward: 3.23
episode: 1726  score: 2.0    memory length: 358184    epsilon: 0.4887937000107836
steps: 197    lr: 0.0001    evaluation reward: 3.2
episode: 1727  score: 4.0    memory length: 358477    epsilon: 0.4882135600107799
steps: 293    lr: 0.0001    evaluation reward: 3.21
episode: 1728  score: 5.0    memory length: 358783    epsilon: 0.4876076800107761
steps: 306    lr: 0.0001    evaluation reward: 3.23
episode: 1729  score: 4.0    memory length: 359026    epsilon:

```

0.48712654001077305 steps: 243 lr: 0.0001 evaluation reward: 3.23
 episode: 1730 score: 3.0 memory length: 359238 epsilon: 0.4867067800107704
 steps: 212 lr: 0.0001 evaluation reward: 3.23
 episode: 1731 score: 10.0 memory length: 359677 epsilon:
 0.4858375600107649 steps: 439 lr: 0.0001 evaluation reward: 3.3
 episode: 1732 score: 4.0 memory length: 359952 epsilon:
 0.48529306001076145 steps: 275 lr: 0.0001 evaluation reward: 3.3
 episode: 1733 score: 5.0 memory length: 360298 epsilon: 0.4846079800107571
 steps: 346 lr: 0.0001 evaluation reward: 3.35
 episode: 1734 score: 2.0 memory length: 360479 epsilon:
 0.48424960001075484 steps: 181 lr: 0.0001 evaluation reward: 3.31
 episode: 1735 score: 4.0 memory length: 360776 epsilon: 0.4836615400107511
 steps: 297 lr: 0.0001 evaluation reward: 3.32
 episode: 1736 score: 3.0 memory length: 361002 epsilon: 0.4832140600107483
 steps: 226 lr: 0.0001 evaluation reward: 3.31
 episode: 1737 score: 7.0 memory length: 361396 epsilon:
 0.48243394001074336 steps: 394 lr: 0.0001 evaluation reward: 3.36
 episode: 1738 score: 5.0 memory length: 361700 epsilon:
 0.48183202001073955 steps: 304 lr: 0.0001 evaluation reward: 3.35
 episode: 1739 score: 6.0 memory length: 362059 epsilon:
 0.48112120001073505 steps: 359 lr: 0.0001 evaluation reward: 3.38
 episode: 1740 score: 6.0 memory length: 362412 epsilon:
 0.48042226001073063 steps: 353 lr: 0.0001 evaluation reward: 3.42
 episode: 1741 score: 3.0 memory length: 362641 epsilon:
 0.47996884001072776 steps: 229 lr: 0.0001 evaluation reward: 3.42
 episode: 1742 score: 5.0 memory length: 362936 epsilon:
 0.47938474001072406 steps: 295 lr: 0.0001 evaluation reward: 3.44
 episode: 1743 score: 2.0 memory length: 363135 epsilon:
 0.47899072001072157 steps: 199 lr: 0.0001 evaluation reward: 3.44
 episode: 1744 score: 5.0 memory length: 363448 epsilon:
 0.47837098001071765 steps: 313 lr: 0.0001 evaluation reward: 3.48
 episode: 1745 score: 4.0 memory length: 363708 epsilon: 0.4778561800107144
 steps: 260 lr: 0.0001 evaluation reward: 3.47
 episode: 1746 score: 6.0 memory length: 364066 epsilon: 0.4771473400107099
 steps: 358 lr: 0.0001 evaluation reward: 3.5
 episode: 1747 score: 5.0 memory length: 364388 epsilon: 0.4765097800107059
 steps: 322 lr: 0.0001 evaluation reward: 3.55
 episode: 1748 score: 3.0 memory length: 364637 epsilon:
 0.47601676001070276 steps: 249 lr: 0.0001 evaluation reward: 3.55
 episode: 1749 score: 5.0 memory length: 364947 epsilon:
 0.47540296001069887 steps: 310 lr: 0.0001 evaluation reward: 3.57
 episode: 1750 score: 0.0 memory length: 365069 epsilon:
 0.47516140001069734 steps: 122 lr: 0.0001 evaluation reward: 3.55
 episode: 1751 score: 4.0 memory length: 365343 epsilon: 0.4746188800106939
 steps: 274 lr: 0.0001 evaluation reward: 3.55
 episode: 1752 score: 2.0 memory length: 365545 epsilon: 0.4742189200106914
 steps: 202 lr: 0.0001 evaluation reward: 3.54
 episode: 1753 score: 6.0 memory length: 365888 epsilon: 0.4735397800106871

```

steps: 343    lr: 0.0001    evaluation reward: 3.58
episode: 1754  score: 4.0    memory length: 366182    epsilon: 0.4729576600106834
steps: 294    lr: 0.0001    evaluation reward: 3.59
episode: 1755  score: 6.0    memory length: 366547    epsilon:
0.47223496001067883    steps: 365    lr: 0.0001    evaluation reward: 3.63
episode: 1756  score: 4.0    memory length: 366842    epsilon:
0.47165086001067513    steps: 295    lr: 0.0001    evaluation reward: 3.63
episode: 1757  score: 6.0    memory length: 367163    epsilon: 0.4710152800106711
steps: 321    lr: 0.0001    evaluation reward: 3.61
episode: 1758  score: 4.0    memory length: 367435    epsilon: 0.4704767200106677
steps: 272    lr: 0.0001    evaluation reward: 3.62
episode: 1759  score: 5.0    memory length: 367749    epsilon:
0.46985500001066377    steps: 314    lr: 0.0001    evaluation reward: 3.63
episode: 1760  score: 3.0    memory length: 367995    epsilon: 0.4693679200106607
steps: 246    lr: 0.0001    evaluation reward: 3.63
episode: 1761  score: 3.0    memory length: 368264    epsilon: 0.4688353000106573
steps: 269    lr: 0.0001    evaluation reward: 3.63
episode: 1762  score: 2.0    memory length: 368461    epsilon:
0.46844524001065485    steps: 197    lr: 0.0001    evaluation reward: 3.63
episode: 1763  score: 4.0    memory length: 368738    epsilon: 0.4678967800106514
steps: 277    lr: 0.0001    evaluation reward: 3.65
episode: 1764  score: 4.0    memory length: 368997    epsilon:
0.46738396001064814    steps: 259    lr: 0.0001    evaluation reward: 3.66
episode: 1765  score: 4.0    memory length: 369256    epsilon: 0.4668711400106449
steps: 259    lr: 0.0001    evaluation reward: 3.67
episode: 1766  score: 8.0    memory length: 369699    epsilon:
0.46599400001063934    steps: 443    lr: 0.0001    evaluation reward: 3.7
episode: 1767  score: 7.0    memory length: 370083    epsilon:
0.46523368001063453    steps: 384    lr: 0.0001    evaluation reward: 3.75
episode: 1768  score: 2.0    memory length: 370280    epsilon:
0.46484362001063206    steps: 197    lr: 0.0001    evaluation reward: 3.75
episode: 1769  score: 4.0    memory length: 370572    epsilon: 0.4642654600106284
steps: 292    lr: 0.0001    evaluation reward: 3.77
episode: 1770  score: 4.0    memory length: 370869    epsilon: 0.4636774000106247
steps: 297    lr: 0.0001    evaluation reward: 3.79
episode: 1771  score: 3.0    memory length: 371094    epsilon:
0.46323190001062187    steps: 225    lr: 0.0001    evaluation reward: 3.79
episode: 1772  score: 5.0    memory length: 371419    epsilon: 0.4625884000106178
steps: 325    lr: 0.0001    evaluation reward: 3.8
episode: 1773  score: 6.0    memory length: 371755    epsilon: 0.4619231200106136
steps: 336    lr: 0.0001    evaluation reward: 3.8
episode: 1774  score: 4.0    memory length: 372022    epsilon:
0.46139446001061024    steps: 267    lr: 0.0001    evaluation reward: 3.78
episode: 1775  score: 2.0    memory length: 372223    epsilon: 0.4609964800106077
steps: 201    lr: 0.0001    evaluation reward: 3.78
episode: 1776  score: 2.0    memory length: 372404    epsilon:
0.46063810001060546    steps: 181    lr: 0.0001    evaluation reward: 3.78
episode: 1777  score: 3.0    memory length: 372650    epsilon: 0.4601510200106024

```

```

steps: 246    lr: 0.0001    evaluation reward: 3.78
episode: 1778    score: 6.0    memory length: 372971    epsilon:
0.45951544001059835    steps: 321    lr: 0.0001    evaluation reward: 3.79
episode: 1779    score: 3.0    memory length: 373217    epsilon:
0.45902836001059527    steps: 246    lr: 0.0001    evaluation reward: 3.76
episode: 1780    score: 5.0    memory length: 373544    epsilon: 0.4583809000105912
steps: 327    lr: 0.0001    evaluation reward: 3.79
episode: 1781    score: 2.0    memory length: 373741    epsilon: 0.4579908400105887
steps: 197    lr: 0.0001    evaluation reward: 3.79
episode: 1782    score: 4.0    memory length: 374020    epsilon: 0.4574384200105852
steps: 279    lr: 0.0001    evaluation reward: 3.8
episode: 1783    score: 3.0    memory length: 374248    epsilon:
0.45698698001058236    steps: 228    lr: 0.0001    evaluation reward: 3.81
episode: 1784    score: 2.0    memory length: 374446    epsilon: 0.4565949400105799
steps: 198    lr: 0.0001    evaluation reward: 3.8
episode: 1785    score: 4.0    memory length: 374722    epsilon: 0.4560484600105764
steps: 276    lr: 0.0001    evaluation reward: 3.78
episode: 1786    score: 4.0    memory length: 374987    epsilon: 0.4555237600105731
steps: 265    lr: 0.0001    evaluation reward: 3.8
episode: 1787    score: 5.0    memory length: 375295    epsilon:
0.45491392001056924    steps: 308    lr: 0.0001    evaluation reward: 3.81
episode: 1788    score: 2.0    memory length: 375476    epsilon:
0.45455554001056697    steps: 181    lr: 0.0001    evaluation reward: 3.81
episode: 1789    score: 2.0    memory length: 375673    epsilon: 0.4541654800105645
steps: 197    lr: 0.0001    evaluation reward: 3.8
episode: 1790    score: 6.0    memory length: 376007    epsilon: 0.4535041600105603
steps: 334    lr: 0.0001    evaluation reward: 3.81
episode: 1791    score: 7.0    memory length: 376330    epsilon: 0.4528646200105563
steps: 323    lr: 0.0001    evaluation reward: 3.86
episode: 1792    score: 6.0    memory length: 376687    epsilon: 0.4521577600105518
steps: 357    lr: 0.0001    evaluation reward: 3.9
episode: 1793    score: 4.0    memory length: 376980    epsilon:
0.45157762001054813    steps: 293    lr: 0.0001    evaluation reward: 3.92
episode: 1794    score: 7.0    memory length: 377392    epsilon:
0.45076186001054297    steps: 412    lr: 0.0001    evaluation reward: 3.95
episode: 1795    score: 5.0    memory length: 377700    epsilon: 0.4501520200105391
steps: 308    lr: 0.0001    evaluation reward: 3.96
episode: 1796    score: 4.0    memory length: 377975    epsilon:
0.44960752001053567    steps: 275    lr: 0.0001    evaluation reward: 3.98
episode: 1797    score: 2.0    memory length: 378157    epsilon: 0.4492471600105334
steps: 182    lr: 0.0001    evaluation reward: 3.96
episode: 1798    score: 3.0    memory length: 378370    epsilon: 0.4488254200105307
steps: 213    lr: 0.0001    evaluation reward: 3.96
episode: 1799    score: 6.0    memory length: 378726    epsilon:
0.44812054001052626    steps: 356    lr: 0.0001    evaluation reward: 3.99
episode: 1800    score: 4.0    memory length: 379022    epsilon:
0.44753446001052255    steps: 296    lr: 0.0001    evaluation reward: 3.98
episode: 1801    score: 3.0    memory length: 379292    epsilon:

```

0.44699986001051917 steps: 270 lr: 0.0001 evaluation reward: 3.97
 episode: 1802 score: 5.0 memory length: 379598 epsilon:
 0.44639398001051533 steps: 306 lr: 0.0001 evaluation reward: 4.0
 episode: 1803 score: 2.0 memory length: 379795 epsilon:
 0.44600392001051287 steps: 197 lr: 0.0001 evaluation reward: 3.97
 episode: 1804 score: 4.0 memory length: 380061 epsilon:
 0.44547724001050953 steps: 266 lr: 0.0001 evaluation reward: 3.99
 episode: 1805 score: 3.0 memory length: 380289 epsilon: 0.4450258000105067
 steps: 228 lr: 0.0001 evaluation reward: 3.99
 episode: 1806 score: 2.0 memory length: 380489 epsilon: 0.4446298000105042
 steps: 200 lr: 0.0001 evaluation reward: 3.94
 episode: 1807 score: 4.0 memory length: 380749 epsilon: 0.4441150000105009
 steps: 260 lr: 0.0001 evaluation reward: 3.96
 episode: 1808 score: 6.0 memory length: 381097 epsilon:
 0.44342596001049656 steps: 348 lr: 0.0001 evaluation reward: 3.98
 episode: 1809 score: 11.0 memory length: 381478 epsilon:
 0.4426715800104918 steps: 381 lr: 0.0001 evaluation reward: 4.07
 episode: 1810 score: 2.0 memory length: 381658 epsilon:
 0.44231518001048953 steps: 180 lr: 0.0001 evaluation reward: 4.04
 episode: 1811 score: 6.0 memory length: 381984 epsilon:
 0.44166970001048544 steps: 326 lr: 0.0001 evaluation reward: 4.09
 episode: 1812 score: 5.0 memory length: 382279 epsilon:
 0.44108560001048175 steps: 295 lr: 0.0001 evaluation reward: 4.11
 episode: 1813 score: 3.0 memory length: 382508 epsilon: 0.4406321800104789
 steps: 229 lr: 0.0001 evaluation reward: 4.09
 episode: 1814 score: 2.0 memory length: 382708 epsilon: 0.4402361800104764
 steps: 200 lr: 0.0001 evaluation reward: 4.1
 episode: 1815 score: 3.0 memory length: 382938 epsilon: 0.4397807800104735
 steps: 230 lr: 0.0001 evaluation reward: 4.1
 episode: 1816 score: 7.0 memory length: 383330 epsilon: 0.4390046200104686
 steps: 392 lr: 0.0001 evaluation reward: 4.14
 episode: 1817 score: 6.0 memory length: 383707 epsilon:
 0.43825816001046386 steps: 377 lr: 0.0001 evaluation reward: 4.17
 episode: 1818 score: 4.0 memory length: 383970 epsilon:
 0.43773742001046057 steps: 263 lr: 0.0001 evaluation reward: 4.17
 episode: 1819 score: 5.0 memory length: 384294 epsilon: 0.4370959000104565
 steps: 324 lr: 0.0001 evaluation reward: 4.18
 episode: 1820 score: 4.0 memory length: 384570 epsilon:
 0.43654942001045305 steps: 276 lr: 0.0001 evaluation reward: 4.19
 episode: 1821 score: 4.0 memory length: 384845 epsilon: 0.4360049200104496
 steps: 275 lr: 0.0001 evaluation reward: 4.21
 episode: 1822 score: 2.0 memory length: 385042 epsilon:
 0.43561486001044714 steps: 197 lr: 0.0001 evaluation reward: 4.2
 episode: 1823 score: 5.0 memory length: 385357 epsilon: 0.4349911600104432
 steps: 315 lr: 0.0001 evaluation reward: 4.18
 episode: 1824 score: 3.0 memory length: 385582 epsilon:
 0.43454566001044037 steps: 225 lr: 0.0001 evaluation reward: 4.2
 episode: 1825 score: 3.0 memory length: 385809 epsilon: 0.4340962000104375

```

steps: 227    lr: 0.0001    evaluation reward: 4.19
episode: 1826  score: 3.0    memory length: 386077    epsilon:
0.43356556001043417    steps: 268    lr: 0.0001    evaluation reward: 4.2
episode: 1827  score: 4.0    memory length: 386353    epsilon: 0.4330190800104307
steps: 276    lr: 0.0001    evaluation reward: 4.2
episode: 1828  score: 7.0    memory length: 386744    epsilon: 0.4322449000104258
steps: 391    lr: 0.0001    evaluation reward: 4.22
episode: 1829  score: 2.0    memory length: 386925    epsilon:
0.43188652001042355    steps: 181    lr: 0.0001    evaluation reward: 4.2
episode: 1830  score: 2.0    memory length: 387104    epsilon: 0.4315321000104213
steps: 179    lr: 0.0001    evaluation reward: 4.19
episode: 1831  score: 3.0    memory length: 387349    epsilon:
0.43104700001041824    steps: 245    lr: 0.0001    evaluation reward: 4.12
episode: 1832  score: 1.0    memory length: 387500    epsilon:
0.43074802001041634    steps: 151    lr: 0.0001    evaluation reward: 4.09
episode: 1833  score: 6.0    memory length: 387839    epsilon: 0.4300768000104121
steps: 339    lr: 0.0001    evaluation reward: 4.1
episode: 1834  score: 8.0    memory length: 388312    epsilon:
0.42914026001040617    steps: 473    lr: 0.0001    evaluation reward: 4.16
episode: 1835  score: 3.0    memory length: 388524    epsilon: 0.4287205000104035
steps: 212    lr: 0.0001    evaluation reward: 4.15
episode: 1836  score: 5.0    memory length: 388813    epsilon: 0.4281482800103999
steps: 289    lr: 0.0001    evaluation reward: 4.17
episode: 1837  score: 6.0    memory length: 389156    epsilon: 0.4274691400103956
steps: 343    lr: 0.0001    evaluation reward: 4.16
episode: 1838  score: 5.0    memory length: 389460    epsilon: 0.4268672200103918
steps: 304    lr: 0.0001    evaluation reward: 4.16
episode: 1839  score: 3.0    memory length: 389672    epsilon:
0.42644746001038913    steps: 212    lr: 0.0001    evaluation reward: 4.13
episode: 1840  score: 5.0    memory length: 389995    epsilon: 0.4258079200103851
steps: 323    lr: 0.0001    evaluation reward: 4.12
episode: 1841  score: 5.0    memory length: 390299    epsilon: 0.4252060000103813
steps: 304    lr: 0.0001    evaluation reward: 4.14
episode: 1842  score: 5.0    memory length: 390602    epsilon: 0.4246060600103775
steps: 303    lr: 0.0001    evaluation reward: 4.14
episode: 1843  score: 4.0    memory length: 390860    epsilon:
0.42409522001037425    steps: 258    lr: 0.0001    evaluation reward: 4.16
episode: 1844  score: 6.0    memory length: 391197    epsilon:
0.42342796001037003    steps: 337    lr: 0.0001    evaluation reward: 4.17
episode: 1845  score: 4.0    memory length: 391459    epsilon:
0.42290920001036675    steps: 262    lr: 0.0001    evaluation reward: 4.17
episode: 1846  score: 7.0    memory length: 391851    epsilon:
0.42213304001036184    steps: 392    lr: 0.0001    evaluation reward: 4.18
episode: 1847  score: 9.0    memory length: 392343    epsilon: 0.4211588800103557
steps: 492    lr: 0.0001    evaluation reward: 4.22
episode: 1848  score: 4.0    memory length: 392602    epsilon:
0.42064606001035243    steps: 259    lr: 0.0001    evaluation reward: 4.23
episode: 1849  score: 6.0    memory length: 392966    epsilon:

```


0.41992534001034787 steps: 364 lr: 0.0001 evaluation reward: 4.24
 episode: 1850 score: 4.0 memory length: 393264 epsilon:
 0.41933530001034414 steps: 298 lr: 0.0001 evaluation reward: 4.28
 episode: 1851 score: 4.0 memory length: 393529 epsilon: 0.4188106000103408
 steps: 265 lr: 0.0001 evaluation reward: 4.28
 episode: 1852 score: 3.0 memory length: 393758 epsilon:
 0.41835718001033795 steps: 229 lr: 0.0001 evaluation reward: 4.29
 episode: 1853 score: 5.0 memory length: 394081 epsilon: 0.4177176400103339
 steps: 323 lr: 0.0001 evaluation reward: 4.28
 episode: 1854 score: 5.0 memory length: 394420 epsilon:
 0.41704642001032965 steps: 339 lr: 0.0001 evaluation reward: 4.29
 episode: 1855 score: 3.0 memory length: 394667 epsilon:
 0.41655736001032656 steps: 247 lr: 0.0001 evaluation reward: 4.26
 episode: 1856 score: 3.0 memory length: 394914 epsilon:
 0.41606830001032347 steps: 247 lr: 0.0001 evaluation reward: 4.25
 episode: 1857 score: 5.0 memory length: 395237 epsilon: 0.4154287600103194
 steps: 323 lr: 0.0001 evaluation reward: 4.24
 episode: 1858 score: 5.0 memory length: 395542 epsilon: 0.4148248600103156
 steps: 305 lr: 0.0001 evaluation reward: 4.25
 episode: 1859 score: 3.0 memory length: 395767 epsilon: 0.4143793600103128
 steps: 225 lr: 0.0001 evaluation reward: 4.23
 episode: 1860 score: 6.0 memory length: 396109 epsilon: 0.4137022000103085
 steps: 342 lr: 0.0001 evaluation reward: 4.26
 episode: 1861 score: 6.0 memory length: 396483 epsilon: 0.4129616800103038
 steps: 374 lr: 0.0001 evaluation reward: 4.29
 episode: 1862 score: 5.0 memory length: 396776 epsilon:
 0.41238154001030014 steps: 293 lr: 0.0001 evaluation reward: 4.32
 episode: 1863 score: 3.0 memory length: 397001 epsilon: 0.4119360400102973
 steps: 225 lr: 0.0001 evaluation reward: 4.31
 episode: 1864 score: 2.0 memory length: 397198 epsilon:
 0.41154598001029485 steps: 197 lr: 0.0001 evaluation reward: 4.29
 episode: 1865 score: 4.0 memory length: 397456 epsilon: 0.4110351400102916
 steps: 258 lr: 0.0001 evaluation reward: 4.29
 episode: 1866 score: 2.0 memory length: 397653 epsilon:
 0.41064508001028915 steps: 197 lr: 0.0001 evaluation reward: 4.23
 episode: 1867 score: 4.0 memory length: 397931 epsilon:
 0.41009464001028567 steps: 278 lr: 0.0001 evaluation reward: 4.2
 episode: 1868 score: 4.0 memory length: 398205 epsilon:
 0.40955212001028224 steps: 274 lr: 0.0001 evaluation reward: 4.22
 episode: 1869 score: 2.0 memory length: 398402 epsilon:
 0.40916206001027977 steps: 197 lr: 0.0001 evaluation reward: 4.2
 episode: 1870 score: 4.0 memory length: 398682 epsilon:
 0.40860766001027626 steps: 280 lr: 0.0001 evaluation reward: 4.2
 episode: 1871 score: 5.0 memory length: 398990 epsilon: 0.4079978200102724
 steps: 308 lr: 0.0001 evaluation reward: 4.22
 episode: 1872 score: 3.0 memory length: 399200 epsilon: 0.4075820200102698
 steps: 210 lr: 0.0001 evaluation reward: 4.2
 episode: 1873 score: 8.0 memory length: 399617 epsilon:

0.40675636001026455 steps: 417 lr: 0.0001 evaluation reward: 4.22
 episode: 1874 score: 6.0 memory length: 399981 epsilon: 0.40603564001026
 steps: 364 lr: 0.0001 evaluation reward: 4.24
 episode: 1875 score: 2.0 memory length: 400180 epsilon: 0.4056416200102575
 steps: 199 lr: 4e-05 evaluation reward: 4.24
 episode: 1876 score: 3.0 memory length: 400392 epsilon:
 0.40522186001025484 steps: 212 lr: 4e-05 evaluation reward: 4.25
 episode: 1877 score: 3.0 memory length: 400602 epsilon: 0.4048060600102522
 steps: 210 lr: 4e-05 evaluation reward: 4.25
 episode: 1878 score: 4.0 memory length: 400842 epsilon: 0.4043308600102492
 steps: 240 lr: 4e-05 evaluation reward: 4.23
 episode: 1879 score: 7.0 memory length: 401250 epsilon: 0.4035230200102441
 steps: 408 lr: 4e-05 evaluation reward: 4.27
 episode: 1880 score: 4.0 memory length: 401524 epsilon:
 0.40298050001024066 steps: 274 lr: 4e-05 evaluation reward: 4.26
 episode: 1881 score: 6.0 memory length: 401887 epsilon: 0.4022617600102361
 steps: 363 lr: 4e-05 evaluation reward: 4.3
 episode: 1882 score: 4.0 memory length: 402161 epsilon: 0.4017192400102327
 steps: 274 lr: 4e-05 evaluation reward: 4.3
 episode: 1883 score: 12.0 memory length: 402463 epsilon:
 0.4011212800102289 steps: 302 lr: 4e-05 evaluation reward: 4.39
 episode: 1884 score: 2.0 memory length: 402644 epsilon:
 0.40076290001022663 steps: 181 lr: 4e-05 evaluation reward: 4.39
 episode: 1885 score: 3.0 memory length: 402875 epsilon:
 0.40030552001022374 steps: 231 lr: 4e-05 evaluation reward: 4.38
 episode: 1886 score: 4.0 memory length: 403134 epsilon: 0.3997927000102205
 steps: 259 lr: 4e-05 evaluation reward: 4.38
 episode: 1887 score: 4.0 memory length: 403390 epsilon: 0.3992858200102173
 steps: 256 lr: 4e-05 evaluation reward: 4.37
 episode: 1888 score: 2.0 memory length: 403588 epsilon: 0.3988937800102148
 steps: 198 lr: 4e-05 evaluation reward: 4.37
 episode: 1889 score: 6.0 memory length: 403933 epsilon: 0.3982106800102105
 steps: 345 lr: 4e-05 evaluation reward: 4.41
 episode: 1890 score: 3.0 memory length: 404146 epsilon: 0.3977889400102078
 steps: 213 lr: 4e-05 evaluation reward: 4.38
 episode: 1891 score: 5.0 memory length: 404475 epsilon: 0.3971375200102037
 steps: 329 lr: 4e-05 evaluation reward: 4.36
 episode: 1892 score: 4.0 memory length: 404769 epsilon: 0.3965554000102
 steps: 294 lr: 4e-05 evaluation reward: 4.34
 episode: 1893 score: 7.0 memory length: 405155 epsilon: 0.3957911200101952
 steps: 386 lr: 4e-05 evaluation reward: 4.37
 episode: 1894 score: 3.0 memory length: 405384 epsilon: 0.3953377000101923
 steps: 229 lr: 4e-05 evaluation reward: 4.33
 episode: 1895 score: 4.0 memory length: 405677 epsilon:
 0.39475756001018864 steps: 293 lr: 4e-05 evaluation reward: 4.32
 episode: 1896 score: 2.0 memory length: 405856 epsilon: 0.3944031400101864
 steps: 179 lr: 4e-05 evaluation reward: 4.3
 episode: 1897 score: 2.0 memory length: 406037 epsilon: 0.3940447600101841

```

steps: 181    lr: 4e-05    evaluation reward: 4.3
episode: 1898  score: 4.0    memory length: 406291    epsilon:
0.39354184001018094    steps: 254    lr: 4e-05    evaluation reward: 4.31
episode: 1899  score: 3.0    memory length: 406504    epsilon: 0.3931201000101783
steps: 213    lr: 4e-05    evaluation reward: 4.28
episode: 1900  score: 8.0    memory length: 406923    epsilon: 0.392290480010173
steps: 419    lr: 4e-05    evaluation reward: 4.32
episode: 1901  score: 6.0    memory length: 407334    epsilon: 0.3914767000101679
steps: 411    lr: 4e-05    evaluation reward: 4.35
episode: 1902  score: 5.0    memory length: 407639    epsilon:
0.39087280001016406    steps: 305    lr: 4e-05    evaluation reward: 4.35
episode: 1903  score: 5.0    memory length: 407946    epsilon: 0.3902649400101602
steps: 307    lr: 4e-05    evaluation reward: 4.38
episode: 1904  score: 3.0    memory length: 408175    epsilon:
0.38981152001015734    steps: 229    lr: 4e-05    evaluation reward: 4.37
episode: 1905  score: 4.0    memory length: 408431    epsilon:
0.38930464001015413    steps: 256    lr: 4e-05    evaluation reward: 4.38
episode: 1906  score: 7.0    memory length: 408816    epsilon: 0.3885423400101493
steps: 385    lr: 4e-05    evaluation reward: 4.43
episode: 1907  score: 5.0    memory length: 409162    epsilon: 0.387857260010145
steps: 346    lr: 4e-05    evaluation reward: 4.44
episode: 1908  score: 5.0    memory length: 409504    epsilon: 0.3871801000101407
steps: 342    lr: 4e-05    evaluation reward: 4.43
episode: 1909  score: 2.0    memory length: 409703    epsilon: 0.3867860800101382
steps: 199    lr: 4e-05    evaluation reward: 4.34
episode: 1910  score: 2.0    memory length: 409884    epsilon:
0.38642770001013593    steps: 181    lr: 4e-05    evaluation reward: 4.34
episode: 1911  score: 2.0    memory length: 410081    epsilon:
0.38603764001013346    steps: 197    lr: 4e-05    evaluation reward: 4.3
episode: 1912  score: 5.0    memory length: 410407    epsilon: 0.3853921600101294
steps: 326    lr: 4e-05    evaluation reward: 4.3
episode: 1913  score: 5.0    memory length: 410730    epsilon:
0.38475262001012533    steps: 323    lr: 4e-05    evaluation reward: 4.32
episode: 1914  score: 5.0    memory length: 411056    epsilon:
0.38410714001012125    steps: 326    lr: 4e-05    evaluation reward: 4.35
episode: 1915  score: 3.0    memory length: 411285    epsilon: 0.3836537200101184
steps: 229    lr: 4e-05    evaluation reward: 4.35
episode: 1916  score: 4.0    memory length: 411565    epsilon: 0.3830993200101149
steps: 280    lr: 4e-05    evaluation reward: 4.32
episode: 1917  score: 4.0    memory length: 411823    epsilon:
0.38258848001011164    steps: 258    lr: 4e-05    evaluation reward: 4.3
episode: 1918  score: 4.0    memory length: 412064    epsilon: 0.3821113000101086
steps: 241    lr: 4e-05    evaluation reward: 4.3
episode: 1919  score: 2.0    memory length: 412245    epsilon:
0.38175292001010636    steps: 181    lr: 4e-05    evaluation reward: 4.27
episode: 1920  score: 3.0    memory length: 412470    epsilon:
0.38130742001010354    steps: 225    lr: 4e-05    evaluation reward: 4.26
episode: 1921  score: 7.0    memory length: 412877    epsilon:

```

0.38050156001009844 steps: 407 lr: 4e-05 evaluation reward: 4.29
 episode: 1922 score: 6.0 memory length: 413200 epsilon: 0.3798620200100944
 steps: 323 lr: 4e-05 evaluation reward: 4.33
 episode: 1923 score: 2.0 memory length: 413381 epsilon: 0.3795036400100921
 steps: 181 lr: 4e-05 evaluation reward: 4.3
 episode: 1924 score: 4.0 memory length: 413641 epsilon:
 0.37898884001008887 steps: 260 lr: 4e-05 evaluation reward: 4.31
 episode: 1925 score: 3.0 memory length: 413886 epsilon: 0.3785037400100858
 steps: 245 lr: 4e-05 evaluation reward: 4.31
 episode: 1926 score: 5.0 memory length: 414211 epsilon: 0.3778602400100817
 steps: 325 lr: 4e-05 evaluation reward: 4.33
 episode: 1927 score: 4.0 memory length: 414505 epsilon:
 0.37727812001007804 steps: 294 lr: 4e-05 evaluation reward: 4.33
 episode: 1928 score: 4.0 memory length: 414762 epsilon: 0.3767692600100748
 steps: 257 lr: 4e-05 evaluation reward: 4.3
 episode: 1929 score: 7.0 memory length: 415009 epsilon:
 0.37628020001007173 steps: 247 lr: 4e-05 evaluation reward: 4.35
 episode: 1930 score: 4.0 memory length: 415271 epsilon:
 0.37576144001006845 steps: 262 lr: 4e-05 evaluation reward: 4.37
 episode: 1931 score: 3.0 memory length: 415520 epsilon:
 0.37526842001006533 steps: 249 lr: 4e-05 evaluation reward: 4.37
 episode: 1932 score: 3.0 memory length: 415768 epsilon: 0.3747773800100622
 steps: 248 lr: 4e-05 evaluation reward: 4.39
 episode: 1933 score: 4.0 memory length: 416011 epsilon: 0.3742962400100592
 steps: 243 lr: 4e-05 evaluation reward: 4.37
 episode: 1934 score: 4.0 memory length: 416267 epsilon:
 0.37378936001005597 steps: 256 lr: 4e-05 evaluation reward: 4.33
 episode: 1935 score: 3.0 memory length: 416476 epsilon:
 0.37337554001005335 steps: 209 lr: 4e-05 evaluation reward: 4.33
 episode: 1936 score: 7.0 memory length: 416849 epsilon: 0.3726370000100487
 steps: 373 lr: 4e-05 evaluation reward: 4.35
 episode: 1937 score: 3.0 memory length: 417079 epsilon: 0.3721816000100458
 steps: 230 lr: 4e-05 evaluation reward: 4.32
 episode: 1938 score: 2.0 memory length: 417261 epsilon: 0.3718212400100435
 steps: 182 lr: 4e-05 evaluation reward: 4.29
 episode: 1939 score: 3.0 memory length: 417487 epsilon: 0.3713737600100407
 steps: 226 lr: 4e-05 evaluation reward: 4.29
 episode: 1940 score: 2.0 memory length: 417668 epsilon: 0.3710153800100384
 steps: 181 lr: 4e-05 evaluation reward: 4.26
 episode: 1941 score: 4.0 memory length: 417906 epsilon:
 0.37054414001003544 steps: 238 lr: 4e-05 evaluation reward: 4.25
 episode: 1942 score: 7.0 memory length: 418265 epsilon:
 0.36983332001003094 steps: 359 lr: 4e-05 evaluation reward: 4.27
 episode: 1943 score: 4.0 memory length: 418523 epsilon: 0.3693224800100277
 steps: 258 lr: 4e-05 evaluation reward: 4.27
 episode: 1944 score: 2.0 memory length: 418720 epsilon:
 0.36893242001002524 steps: 197 lr: 4e-05 evaluation reward: 4.23
 episode: 1945 score: 3.0 memory length: 418965 epsilon: 0.3684473200100222

```

steps: 245    lr: 4e-05    evaluation reward: 4.22
episode: 1946  score: 2.0    memory length: 419164  epsilon: 0.3680533000100197
steps: 199    lr: 4e-05    evaluation reward: 4.17
episode: 1947  score: 2.0    memory length: 419345  epsilon: 0.3676949200100174
steps: 181    lr: 4e-05    evaluation reward: 4.1
episode: 1948  score: 4.0    memory length: 419640  epsilon: 0.3671108200100137
steps: 295    lr: 4e-05    evaluation reward: 4.1
episode: 1949  score: 4.0    memory length: 419912  epsilon: 0.3665722600100103
steps: 272    lr: 4e-05    evaluation reward: 4.08
episode: 1950  score: 2.0    memory length: 420093  epsilon:
0.36621388001000804  steps: 181    lr: 4e-05    evaluation reward: 4.06
episode: 1951  score: 2.0    memory length: 420274  epsilon: 0.3658555000100058
steps: 181    lr: 4e-05    evaluation reward: 4.04
episode: 1952  score: 7.0    memory length: 420646  epsilon: 0.3651189400100011
steps: 372    lr: 4e-05    evaluation reward: 4.08
episode: 1953  score: 3.0    memory length: 420893  epsilon: 0.3646298800099998
steps: 247    lr: 4e-05    evaluation reward: 4.06
episode: 1954  score: 3.0    memory length: 421120  epsilon: 0.3641804200099952
steps: 227    lr: 4e-05    evaluation reward: 4.04
episode: 1955  score: 3.0    memory length: 421368  epsilon:
0.36368938000999207  steps: 248    lr: 4e-05    evaluation reward: 4.04
episode: 1956  score: 2.0    memory length: 421550  epsilon: 0.3633290200099898
steps: 182    lr: 4e-05    evaluation reward: 4.03
episode: 1957  score: 2.0    memory length: 421731  epsilon: 0.3629706400099875
steps: 181    lr: 4e-05    evaluation reward: 4.0
episode: 1958  score: 2.0    memory length: 421913  epsilon:
0.36261028000998524  steps: 182    lr: 4e-05    evaluation reward: 3.97
episode: 1959  score: 3.0    memory length: 422143  epsilon:
0.36215488000998236  steps: 230    lr: 4e-05    evaluation reward: 3.97
episode: 1960  score: 6.0    memory length: 422485  epsilon: 0.3614777200099781
steps: 342    lr: 4e-05    evaluation reward: 3.97
episode: 1961  score: 6.0    memory length: 422848  epsilon:
0.36075898000997353  steps: 363    lr: 4e-05    evaluation reward: 3.97
episode: 1962  score: 3.0    memory length: 423073  epsilon: 0.3603134800099707
steps: 225    lr: 4e-05    evaluation reward: 3.95
episode: 1963  score: 3.0    memory length: 423283  epsilon: 0.3598976800099681
steps: 210    lr: 4e-05    evaluation reward: 3.95
episode: 1964  score: 5.0    memory length: 423609  epsilon: 0.359252200009964
steps: 326    lr: 4e-05    evaluation reward: 3.98
episode: 1965  score: 3.0    memory length: 423819  epsilon:
0.35883640000996136  steps: 210    lr: 4e-05    evaluation reward: 3.97
episode: 1966  score: 2.0    memory length: 424000  epsilon: 0.3584780200099591
steps: 181    lr: 4e-05    evaluation reward: 3.97
episode: 1967  score: 6.0    memory length: 424357  epsilon: 0.3577711600099546
steps: 357    lr: 4e-05    evaluation reward: 3.99
episode: 1968  score: 4.0    memory length: 424622  epsilon: 0.3572464600099513
steps: 265    lr: 4e-05    evaluation reward: 3.99
episode: 1969  score: 4.0    memory length: 424901  epsilon: 0.3566940400099478

```

```

steps: 279    lr: 4e-05    evaluation reward: 4.01
episode: 1970  score: 5.0    memory length: 425245    epsilon: 0.3560129200099435
steps: 344    lr: 4e-05    evaluation reward: 4.02
episode: 1971  score: 5.0    memory length: 425553    epsilon:
0.35540308000993964    steps: 308    lr: 4e-05    evaluation reward: 4.02
episode: 1972  score: 4.0    memory length: 425812    epsilon: 0.3548902600099364
steps: 259    lr: 4e-05    evaluation reward: 4.03
episode: 1973  score: 4.0    memory length: 426089    epsilon:
0.35434180000993293    steps: 277    lr: 4e-05    evaluation reward: 3.99
episode: 1974  score: 6.0    memory length: 426407    epsilon:
0.35371216000992894    steps: 318    lr: 4e-05    evaluation reward: 3.99
episode: 1975  score: 5.0    memory length: 426674    epsilon: 0.3531835000099256
steps: 267    lr: 4e-05    evaluation reward: 4.02
episode: 1976  score: 5.0    memory length: 426954    epsilon: 0.3526291000099221
steps: 280    lr: 4e-05    evaluation reward: 4.04
episode: 1977  score: 6.0    memory length: 427307    epsilon:
0.35193016000991767    steps: 353    lr: 4e-05    evaluation reward: 4.07
episode: 1978  score: 2.0    memory length: 427508    epsilon:
0.35153218000991515    steps: 201    lr: 4e-05    evaluation reward: 4.05
episode: 1979  score: 4.0    memory length: 427786    epsilon:
0.35098174000991167    steps: 278    lr: 4e-05    evaluation reward: 4.02
episode: 1980  score: 8.0    memory length: 428297    epsilon:
0.34996996000990527    steps: 511    lr: 4e-05    evaluation reward: 4.06
episode: 1981  score: 4.0    memory length: 428575    epsilon: 0.3494195200099018
steps: 278    lr: 4e-05    evaluation reward: 4.04
episode: 1982  score: 3.0    memory length: 428810    epsilon:
0.34895422000989884    steps: 235    lr: 4e-05    evaluation reward: 4.03
episode: 1983  score: 6.0    memory length: 429172    epsilon: 0.3482374600098943
steps: 362    lr: 4e-05    evaluation reward: 3.97
episode: 1984  score: 2.0    memory length: 429369    epsilon:
0.34784740000989184    steps: 197    lr: 4e-05    evaluation reward: 3.97
episode: 1985  score: 3.0    memory length: 429581    epsilon: 0.3474276400098892
steps: 212    lr: 4e-05    evaluation reward: 3.97
episode: 1986  score: 4.0    memory length: 429842    epsilon: 0.3469108600098859
steps: 261    lr: 4e-05    evaluation reward: 3.97
episode: 1987  score: 2.0    memory length: 430021    epsilon:
0.34655644000988367    steps: 179    lr: 4e-05    evaluation reward: 3.95
episode: 1988  score: 4.0    memory length: 430295    epsilon:
0.34601392000988024    steps: 274    lr: 4e-05    evaluation reward: 3.97
episode: 1989  score: 6.0    memory length: 430658    epsilon: 0.3452951800098757
steps: 363    lr: 4e-05    evaluation reward: 3.97
episode: 1990  score: 6.0    memory length: 431000    epsilon: 0.3446180200098714
steps: 342    lr: 4e-05    evaluation reward: 4.0
episode: 1991  score: 2.0    memory length: 431181    epsilon:
0.34425964000986914    steps: 181    lr: 4e-05    evaluation reward: 3.97
episode: 1992  score: 4.0    memory length: 431442    epsilon:
0.34374286000986587    steps: 261    lr: 4e-05    evaluation reward: 3.97
episode: 1993  score: 3.0    memory length: 431691    epsilon:

```

0.34324984000986275 steps: 249 lr: 4e-05 evaluation reward: 3.93
 episode: 1994 score: 4.0 memory length: 431948 epsilon:
 0.34274098000985953 steps: 257 lr: 4e-05 evaluation reward: 3.94
 episode: 1995 score: 4.0 memory length: 432208 epsilon: 0.3422261800098563
 steps: 260 lr: 4e-05 evaluation reward: 3.94
 episode: 1996 score: 3.0 memory length: 432434 epsilon:
 0.34177870000985344 steps: 226 lr: 4e-05 evaluation reward: 3.95
 episode: 1997 score: 2.0 memory length: 432616 epsilon:
 0.34141834000985116 steps: 182 lr: 4e-05 evaluation reward: 3.95
 episode: 1998 score: 6.0 memory length: 432979 epsilon: 0.3406996000098466
 steps: 363 lr: 4e-05 evaluation reward: 3.97
 episode: 1999 score: 6.0 memory length: 433314 epsilon: 0.3400363000098424
 steps: 335 lr: 4e-05 evaluation reward: 4.0
 episode: 2000 score: 3.0 memory length: 433539 epsilon: 0.3395908000098396
 steps: 225 lr: 4e-05 evaluation reward: 3.95
 episode: 2001 score: 3.0 memory length: 433768 epsilon:
 0.33913738000983673 steps: 229 lr: 4e-05 evaluation reward: 3.92
 episode: 2002 score: 6.0 memory length: 434104 epsilon: 0.3384721000098325
 steps: 336 lr: 4e-05 evaluation reward: 3.93
 episode: 2003 score: 2.0 memory length: 434286 epsilon:
 0.33811174000983024 steps: 182 lr: 4e-05 evaluation reward: 3.9
 episode: 2004 score: 6.0 memory length: 434649 epsilon: 0.3373930000098257
 steps: 363 lr: 4e-05 evaluation reward: 3.93
 episode: 2005 score: 2.0 memory length: 434830 epsilon: 0.3370346200098234
 steps: 181 lr: 4e-05 evaluation reward: 3.91
 episode: 2006 score: 2.0 memory length: 435012 epsilon:
 0.33667426000982115 steps: 182 lr: 4e-05 evaluation reward: 3.86
 episode: 2007 score: 6.0 memory length: 435375 epsilon: 0.3359555200098166
 steps: 363 lr: 4e-05 evaluation reward: 3.87
 episode: 2008 score: 3.0 memory length: 435603 epsilon:
 0.33550408000981374 steps: 228 lr: 4e-05 evaluation reward: 3.85
 episode: 2009 score: 4.0 memory length: 435848 epsilon: 0.3350189800098107
 steps: 245 lr: 4e-05 evaluation reward: 3.87
 episode: 2010 score: 4.0 memory length: 436123 epsilon:
 0.33447448000980723 steps: 275 lr: 4e-05 evaluation reward: 3.89
 episode: 2011 score: 2.0 memory length: 436304 epsilon:
 0.33411610000980496 steps: 181 lr: 4e-05 evaluation reward: 3.89
 episode: 2012 score: 5.0 memory length: 436599 epsilon:
 0.33353200000980127 steps: 295 lr: 4e-05 evaluation reward: 3.89
 episode: 2013 score: 5.0 memory length: 436914 epsilon: 0.3329083000097973
 steps: 315 lr: 4e-05 evaluation reward: 3.89
 episode: 2014 score: 4.0 memory length: 437171 epsilon: 0.3323994400097941
 steps: 257 lr: 4e-05 evaluation reward: 3.88
 episode: 2015 score: 3.0 memory length: 437401 epsilon: 0.3319440400097912
 steps: 230 lr: 4e-05 evaluation reward: 3.88
 episode: 2016 score: 3.0 memory length: 437629 epsilon:
 0.33149260000978836 steps: 228 lr: 4e-05 evaluation reward: 3.87
 episode: 2017 score: 2.0 memory length: 437811 epsilon: 0.3311322400097861

```

steps: 182    lr: 4e-05    evaluation reward: 3.85
episode: 2018  score: 5.0    memory length: 438103    epsilon: 0.3305540800097824
steps: 292    lr: 4e-05    evaluation reward: 3.86
episode: 2019  score: 9.0    memory length: 438572    epsilon:
0.32962546000977655    steps: 469    lr: 4e-05    evaluation reward: 3.93
episode: 2020  score: 4.0    memory length: 438815    epsilon: 0.3291443200097735
steps: 243    lr: 4e-05    evaluation reward: 3.94
episode: 2021  score: 2.0    memory length: 438997    epsilon: 0.3287839600097712
steps: 182    lr: 4e-05    evaluation reward: 3.89
episode: 2022  score: 5.0    memory length: 439323    epsilon:
0.32813848000976714    steps: 326    lr: 4e-05    evaluation reward: 3.88
episode: 2023  score: 5.0    memory length: 439634    epsilon:
0.32752270000976325    steps: 311    lr: 4e-05    evaluation reward: 3.91
episode: 2024  score: 3.0    memory length: 439881    epsilon:
0.32703364000976015    steps: 247    lr: 4e-05    evaluation reward: 3.9
episode: 2025  score: 3.0    memory length: 440093    epsilon: 0.3266138800097575
steps: 212    lr: 4e-05    evaluation reward: 3.9
episode: 2026  score: 3.0    memory length: 440319    epsilon:
0.32616640000975466    steps: 226    lr: 4e-05    evaluation reward: 3.88
episode: 2027  score: 3.0    memory length: 440546    epsilon: 0.3257169400097518
steps: 227    lr: 4e-05    evaluation reward: 3.87
episode: 2028  score: 2.0    memory length: 440725    epsilon: 0.3253625200097496
steps: 179    lr: 4e-05    evaluation reward: 3.85
episode: 2029  score: 2.0    memory length: 440904    epsilon:
0.32500810000974734    steps: 179    lr: 4e-05    evaluation reward: 3.8
episode: 2030  score: 5.0    memory length: 441209    epsilon: 0.3244042000097435
steps: 305    lr: 4e-05    evaluation reward: 3.81
episode: 2031  score: 3.0    memory length: 441417    epsilon: 0.3239923600097409
steps: 208    lr: 4e-05    evaluation reward: 3.81
episode: 2032  score: 4.0    memory length: 441678    epsilon:
0.32347558000973764    steps: 261    lr: 4e-05    evaluation reward: 3.82
episode: 2033  score: 5.0    memory length: 441993    epsilon: 0.3228518800097337
steps: 315    lr: 4e-05    evaluation reward: 3.83
episode: 2034  score: 2.0    memory length: 442173    epsilon:
0.32249548000973144    steps: 180    lr: 4e-05    evaluation reward: 3.81
episode: 2035  score: 4.0    memory length: 442449    epsilon: 0.321949000009728
steps: 276    lr: 4e-05    evaluation reward: 3.82
episode: 2036  score: 3.0    memory length: 442694    epsilon: 0.3214639000097249
steps: 245    lr: 4e-05    evaluation reward: 3.78
episode: 2037  score: 3.0    memory length: 442919    epsilon: 0.3210184000097221
steps: 225    lr: 4e-05    evaluation reward: 3.78
episode: 2038  score: 3.0    memory length: 443148    epsilon: 0.3205649800097192
steps: 229    lr: 4e-05    evaluation reward: 3.79
episode: 2039  score: 7.0    memory length: 443553    epsilon:
0.31976308000971415    steps: 405    lr: 4e-05    evaluation reward: 3.83
episode: 2040  score: 4.0    memory length: 443807    epsilon:
0.31926016000971097    steps: 254    lr: 4e-05    evaluation reward: 3.85
episode: 2041  score: 2.0    memory length: 443988    epsilon: 0.3189017800097087

```



```

steps: 181    lr: 4e-05    evaluation reward: 3.83
episode: 2042  score: 7.0  memory length: 444409  epsilon:
0.31806820000970343  steps: 421  lr: 4e-05    evaluation reward: 3.83
episode: 2043  score: 2.0  memory length: 444590  epsilon:
0.317709820000970116  steps: 181  lr: 4e-05    evaluation reward: 3.81
episode: 2044  score: 5.0  memory length: 444935  epsilon:
0.31702672000969684  steps: 345  lr: 4e-05    evaluation reward: 3.84
episode: 2045  score: 4.0  memory length: 445212  epsilon:
0.31647826000969337  steps: 277  lr: 4e-05    evaluation reward: 3.85
episode: 2046  score: 3.0  memory length: 445461  epsilon:
0.31598524000969025  steps: 249  lr: 4e-05    evaluation reward: 3.86
episode: 2047  score: 2.0  memory length: 445640  epsilon: 0.315630820009688
steps: 179    lr: 4e-05    evaluation reward: 3.86
episode: 2048  score: 5.0  memory length: 445946  epsilon: 0.3150249400096842
steps: 306    lr: 4e-05    evaluation reward: 3.87
episode: 2049  score: 7.0  memory length: 446358  epsilon: 0.314209180009679
steps: 412    lr: 4e-05    evaluation reward: 3.9
episode: 2050  score: 3.0  memory length: 446587  epsilon:
0.31375576000967614  steps: 229  lr: 4e-05    evaluation reward: 3.91
episode: 2051  score: 8.0  memory length: 447031  epsilon: 0.3128766400096706
steps: 444    lr: 4e-05    evaluation reward: 3.97
episode: 2052  score: 3.0  memory length: 447241  epsilon:
0.31246084000966795  steps: 210  lr: 4e-05    evaluation reward: 3.93
episode: 2053  score: 3.0  memory length: 447468  epsilon: 0.3120113800096651
steps: 227    lr: 4e-05    evaluation reward: 3.93
episode: 2054  score: 5.0  memory length: 447774  epsilon: 0.3114055000096613
steps: 306    lr: 4e-05    evaluation reward: 3.95
episode: 2055  score: 3.0  memory length: 448002  epsilon: 0.3109540600096584
steps: 228    lr: 4e-05    evaluation reward: 3.95
episode: 2056  score: 3.0  memory length: 448230  epsilon:
0.31050262000965556  steps: 228  lr: 4e-05    evaluation reward: 3.96
episode: 2057  score: 4.0  memory length: 448471  epsilon:
0.31002544000965254  steps: 241  lr: 4e-05    evaluation reward: 3.98
episode: 2058  score: 2.0  memory length: 448652  epsilon: 0.3096670600096503
steps: 181    lr: 4e-05    evaluation reward: 3.98
episode: 2059  score: 3.0  memory length: 448862  epsilon:
0.30925126000964764  steps: 210  lr: 4e-05    evaluation reward: 3.98
episode: 2060  score: 3.0  memory length: 449090  epsilon: 0.3087998200096448
steps: 228    lr: 4e-05    evaluation reward: 3.95
episode: 2061  score: 3.0  memory length: 449318  epsilon:
0.30834838000964193  steps: 228  lr: 4e-05    evaluation reward: 3.92
episode: 2062  score: 3.0  memory length: 449546  epsilon: 0.3078969400096391
steps: 228    lr: 4e-05    evaluation reward: 3.92
episode: 2063  score: 2.0  memory length: 449727  epsilon: 0.3075385600096368
steps: 181    lr: 4e-05    evaluation reward: 3.91
episode: 2064  score: 2.0  memory length: 449909  epsilon:
0.30717820000963453  steps: 182  lr: 4e-05    evaluation reward: 3.88
episode: 2065  score: 5.0  memory length: 450207  epsilon: 0.3065881600096308

```

```

steps: 298    lr: 4e-05    evaluation reward: 3.9
episode: 2066  score: 2.0    memory length: 450388    epsilon: 0.3062297800096285
steps: 181    lr: 4e-05    evaluation reward: 3.9
episode: 2067  score: 2.0    memory length: 450569    epsilon:
0.30587140000962626    steps: 181    lr: 4e-05    evaluation reward: 3.86
episode: 2068  score: 2.0    memory length: 450750    epsilon: 0.305513020009624
steps: 181    lr: 4e-05    evaluation reward: 3.84
episode: 2069  score: 4.0    memory length: 451004    epsilon: 0.3050101000096208
steps: 254    lr: 4e-05    evaluation reward: 3.84
episode: 2070  score: 5.0    memory length: 451292    epsilon: 0.3044398600096172
steps: 288    lr: 4e-05    evaluation reward: 3.84
episode: 2071  score: 3.0    memory length: 451503    epsilon:
0.30402208000961456    steps: 211    lr: 4e-05    evaluation reward: 3.82
episode: 2072  score: 6.0    memory length: 451846    epsilon:
0.30334294000961026    steps: 343    lr: 4e-05    evaluation reward: 3.84
episode: 2073  score: 4.0    memory length: 452087    epsilon:
0.30286576000960724    steps: 241    lr: 4e-05    evaluation reward: 3.84
episode: 2074  score: 4.0    memory length: 452364    epsilon: 0.3023173000096038
steps: 277    lr: 4e-05    evaluation reward: 3.82
episode: 2075  score: 2.0    memory length: 452561    epsilon: 0.3019272400096013
steps: 197    lr: 4e-05    evaluation reward: 3.79
episode: 2076  score: 5.0    memory length: 452867    epsilon:
0.30132136000959747    steps: 306    lr: 4e-05    evaluation reward: 3.79
episode: 2077  score: 7.0    memory length: 453246    epsilon: 0.3005709400095927
steps: 379    lr: 4e-05    evaluation reward: 3.8
episode: 2078  score: 4.0    memory length: 453524    epsilon:
0.30002050000958924    steps: 278    lr: 4e-05    evaluation reward: 3.82
episode: 2079  score: 6.0    memory length: 453877    epsilon: 0.2993215600095848
steps: 353    lr: 4e-05    evaluation reward: 3.84
episode: 2080  score: 5.0    memory length: 454184    epsilon: 0.298713700009581
steps: 307    lr: 4e-05    evaluation reward: 3.81
episode: 2081  score: 6.0    memory length: 454538    epsilon:
0.29801278000957654    steps: 354    lr: 4e-05    evaluation reward: 3.83
episode: 2082  score: 5.0    memory length: 454854    epsilon: 0.2973871000095726
steps: 316    lr: 4e-05    evaluation reward: 3.85
episode: 2083  score: 6.0    memory length: 455208    epsilon:
0.29668618000956815    steps: 354    lr: 4e-05    evaluation reward: 3.85
episode: 2084  score: 4.0    memory length: 455482    epsilon: 0.2961436600095647
steps: 274    lr: 4e-05    evaluation reward: 3.87
episode: 2085  score: 2.0    memory length: 455664    epsilon:
0.29578330000956243    steps: 182    lr: 4e-05    evaluation reward: 3.86
episode: 2086  score: 6.0    memory length: 455990    epsilon:
0.29513782000955835    steps: 326    lr: 4e-05    evaluation reward: 3.88
episode: 2087  score: 8.0    memory length: 456433    epsilon: 0.2942606800095528
steps: 443    lr: 4e-05    evaluation reward: 3.94
episode: 2088  score: 2.0    memory length: 456614    epsilon:
0.29390230000955053    steps: 181    lr: 4e-05    evaluation reward: 3.92
episode: 2089  score: 4.0    memory length: 456870    epsilon: 0.2933954200095473

```

```

steps: 256    lr: 4e-05    evaluation reward: 3.9
episode: 2090  score: 4.0    memory length: 457145    epsilon: 0.2928509200095439
steps: 275    lr: 4e-05    evaluation reward: 3.88
episode: 2091  score: 3.0    memory length: 457392    epsilon: 0.2923618600095408
steps: 247    lr: 4e-05    evaluation reward: 3.89
episode: 2092  score: 8.0    memory length: 457816    epsilon: 0.2915223400095355
steps: 424    lr: 4e-05    evaluation reward: 3.93
episode: 2093  score: 4.0    memory length: 458073    epsilon:
0.29101348000953225    steps: 257    lr: 4e-05    evaluation reward: 3.94
episode: 2094  score: 6.0    memory length: 458448    epsilon:
0.29027098000952756    steps: 375    lr: 4e-05    evaluation reward: 3.96
episode: 2095  score: 4.0    memory length: 458704    epsilon:
0.28976410000952435    steps: 256    lr: 4e-05    evaluation reward: 3.96
episode: 2096  score: 2.0    memory length: 458883    epsilon: 0.2894096800095221
steps: 179    lr: 4e-05    evaluation reward: 3.95
episode: 2097  score: 4.0    memory length: 459139    epsilon: 0.2889028000095189
steps: 256    lr: 4e-05    evaluation reward: 3.97
episode: 2098  score: 2.0    memory length: 459321    epsilon: 0.2885424400095166
steps: 182    lr: 4e-05    evaluation reward: 3.93
episode: 2099  score: 4.0    memory length: 459605    epsilon:
0.28798012000951306    steps: 284    lr: 4e-05    evaluation reward: 3.91
episode: 2100  score: 2.0    memory length: 459787    epsilon: 0.2876197600095108
steps: 182    lr: 4e-05    evaluation reward: 3.9
episode: 2101  score: 3.0    memory length: 459996    epsilon:
0.28720594000950816    steps: 209    lr: 4e-05    evaluation reward: 3.9
episode: 2102  score: 5.0    memory length: 460321    epsilon: 0.2865624400095041
steps: 325    lr: 4e-05    evaluation reward: 3.89
episode: 2103  score: 3.0    memory length: 460531    epsilon:
0.28614664000950146    steps: 210    lr: 4e-05    evaluation reward: 3.9
episode: 2104  score: 4.0    memory length: 460790    epsilon: 0.2856338200094982
steps: 259    lr: 4e-05    evaluation reward: 3.88
episode: 2105  score: 5.0    memory length: 461136    epsilon: 0.2849487400094939
steps: 346    lr: 4e-05    evaluation reward: 3.91
episode: 2106  score: 6.0    memory length: 461499    epsilon:
0.28423000000948934    steps: 363    lr: 4e-05    evaluation reward: 3.95
episode: 2107  score: 3.0    memory length: 461710    epsilon: 0.2838122200094867
steps: 211    lr: 4e-05    evaluation reward: 3.92
episode: 2108  score: 4.0    memory length: 461964    epsilon: 0.2833093000094835
steps: 254    lr: 4e-05    evaluation reward: 3.93
episode: 2109  score: 8.0    memory length: 462378    epsilon: 0.2824895800094783
steps: 414    lr: 4e-05    evaluation reward: 3.97
episode: 2110  score: 4.0    memory length: 462673    epsilon:
0.28190548000947463    steps: 295    lr: 4e-05    evaluation reward: 3.97
episode: 2111  score: 10.0   memory length: 463147    epsilon:
0.2809669600094687    steps: 474    lr: 4e-05    evaluation reward: 4.05
episode: 2112  score: 7.0    memory length: 463541    epsilon:
0.28018684000946376    steps: 394    lr: 4e-05    evaluation reward: 4.07
episode: 2113  score: 3.0    memory length: 463772    epsilon:

```

0.27972946000946086 steps: 231 lr: 4e-05 evaluation reward: 4.05
 episode: 2114 score: 6.0 memory length: 464114 epsilon: 0.2790523000094566
 steps: 342 lr: 4e-05 evaluation reward: 4.07
 episode: 2115 score: 5.0 memory length: 464405 epsilon:
 0.27847612000945293 steps: 291 lr: 4e-05 evaluation reward: 4.09
 episode: 2116 score: 2.0 memory length: 464585 epsilon: 0.2781197200094507
 steps: 180 lr: 4e-05 evaluation reward: 4.08
 episode: 2117 score: 5.0 memory length: 464878 epsilon: 0.277539580009447
 steps: 293 lr: 4e-05 evaluation reward: 4.11
 episode: 2118 score: 3.0 memory length: 465089 epsilon:
 0.27712180000944436 steps: 211 lr: 4e-05 evaluation reward: 4.09
 episode: 2119 score: 3.0 memory length: 465320 epsilon:
 0.27666442000944147 steps: 231 lr: 4e-05 evaluation reward: 4.03
 episode: 2120 score: 3.0 memory length: 465530 epsilon:
 0.27624862000943884 steps: 210 lr: 4e-05 evaluation reward: 4.02
 episode: 2121 score: 6.0 memory length: 465874 epsilon:
 0.27556750000943453 steps: 344 lr: 4e-05 evaluation reward: 4.06
 episode: 2122 score: 4.0 memory length: 466134 epsilon: 0.2750527000094313
 steps: 260 lr: 4e-05 evaluation reward: 4.05
 episode: 2123 score: 5.0 memory length: 466441 epsilon: 0.2744448400094274
 steps: 307 lr: 4e-05 evaluation reward: 4.05
 episode: 2124 score: 6.0 memory length: 466798 epsilon:
 0.27373798000942295 steps: 357 lr: 4e-05 evaluation reward: 4.08
 episode: 2125 score: 3.0 memory length: 467028 epsilon: 0.2732825800094201
 steps: 230 lr: 4e-05 evaluation reward: 4.08
 episode: 2126 score: 2.0 memory length: 467209 epsilon: 0.2729242000094178
 steps: 181 lr: 4e-05 evaluation reward: 4.07
 episode: 2127 score: 5.0 memory length: 467505 epsilon: 0.2723381200094141
 steps: 296 lr: 4e-05 evaluation reward: 4.09
 episode: 2128 score: 3.0 memory length: 467733 epsilon:
 0.27188668000941124 steps: 228 lr: 4e-05 evaluation reward: 4.1
 episode: 2129 score: 3.0 memory length: 467961 epsilon: 0.2714352400094084
 steps: 228 lr: 4e-05 evaluation reward: 4.11
 episode: 2130 score: 5.0 memory length: 468241 epsilon: 0.2708808400094049
 steps: 280 lr: 4e-05 evaluation reward: 4.11
 episode: 2131 score: 4.0 memory length: 468483 epsilon:
 0.27040168000940185 steps: 242 lr: 4e-05 evaluation reward: 4.12
 episode: 2132 score: 3.0 memory length: 468708 epsilon: 0.269956180009399
 steps: 225 lr: 4e-05 evaluation reward: 4.11
 episode: 2133 score: 3.0 memory length: 468953 epsilon:
 0.26947108000939596 steps: 245 lr: 4e-05 evaluation reward: 4.09
 episode: 2134 score: 5.0 memory length: 469280 epsilon:
 0.26882362000939186 steps: 327 lr: 4e-05 evaluation reward: 4.12
 episode: 2135 score: 2.0 memory length: 469461 epsilon: 0.2684652400093896
 steps: 181 lr: 4e-05 evaluation reward: 4.1
 episode: 2136 score: 8.0 memory length: 469901 epsilon: 0.2675940400093841
 steps: 440 lr: 4e-05 evaluation reward: 4.15
 episode: 2137 score: 4.0 memory length: 470162 epsilon: 0.2670772600093808

```

steps: 261    lr: 4e-05    evaluation reward: 4.16
episode: 2138  score: 4.0  memory length: 470440  epsilon:
0.26652682000937733  steps: 278    lr: 4e-05    evaluation reward: 4.17
episode: 2139  score: 11.0  memory length: 470826  epsilon:
0.2657625400093725  steps: 386    lr: 4e-05    evaluation reward: 4.21
episode: 2140  score: 6.0  memory length: 471166  epsilon:
0.26508934000936824  steps: 340    lr: 4e-05    evaluation reward: 4.23
episode: 2141  score: 3.0  memory length: 471396  epsilon:
0.26463394000936535  steps: 230    lr: 4e-05    evaluation reward: 4.24
episode: 2142  score: 7.0  memory length: 471763  epsilon:
0.26390728000936076  steps: 367    lr: 4e-05    evaluation reward: 4.24
episode: 2143  score: 7.0  memory length: 472157  epsilon: 0.2631271600093558
steps: 394    lr: 4e-05    evaluation reward: 4.29
episode: 2144  score: 3.0  memory length: 472385  epsilon:
0.26267572000935296  steps: 228    lr: 4e-05    evaluation reward: 4.27
episode: 2145  score: 7.0  memory length: 472752  epsilon:
0.26194906000934837  steps: 367    lr: 4e-05    evaluation reward: 4.3
episode: 2146  score: 5.0  memory length: 473059  epsilon: 0.2613412000093445
steps: 307    lr: 4e-05    evaluation reward: 4.32
episode: 2147  score: 4.0  memory length: 473333  epsilon: 0.2607986800093411
steps: 274    lr: 4e-05    evaluation reward: 4.34
episode: 2148  score: 5.0  memory length: 473613  epsilon: 0.2602442800093376
steps: 280    lr: 4e-05    evaluation reward: 4.34
episode: 2149  score: 3.0  memory length: 473861  epsilon: 0.2597532400093345
steps: 248    lr: 4e-05    evaluation reward: 4.3
episode: 2150  score: 5.0  memory length: 474189  epsilon:
0.25910380000933037  steps: 328    lr: 4e-05    evaluation reward: 4.32
episode: 2151  score: 2.0  memory length: 474368  epsilon: 0.2587493800093281
steps: 179    lr: 4e-05    evaluation reward: 4.26
episode: 2152  score: 7.0  memory length: 474758  epsilon:
0.25797718000932324  steps: 390    lr: 4e-05    evaluation reward: 4.3
episode: 2153  score: 3.0  memory length: 475005  epsilon:
0.25748812000932014  steps: 247    lr: 4e-05    evaluation reward: 4.3
episode: 2154  score: 3.0  memory length: 475254  epsilon: 0.256995100009317
steps: 249    lr: 4e-05    evaluation reward: 4.28
episode: 2155  score: 5.0  memory length: 475561  epsilon: 0.2563872400093132
steps: 307    lr: 4e-05    evaluation reward: 4.3
episode: 2156  score: 3.0  memory length: 475809  epsilon:
0.25589620000931007  steps: 248    lr: 4e-05    evaluation reward: 4.3
episode: 2157  score: 5.0  memory length: 476115  epsilon:
0.25529032000930624  steps: 306    lr: 4e-05    evaluation reward: 4.31
episode: 2158  score: 4.0  memory length: 476390  epsilon: 0.2547458200093028
steps: 275    lr: 4e-05    evaluation reward: 4.33
episode: 2159  score: 8.0  memory length: 476835  epsilon: 0.2538647200092972
steps: 445    lr: 4e-05    evaluation reward: 4.38
episode: 2160  score: 4.0  memory length: 477100  epsilon: 0.2533400200092939
steps: 265    lr: 4e-05    evaluation reward: 4.39
episode: 2161  score: 7.0  memory length: 477491  epsilon: 0.252565840009289

```

```

steps: 391    lr: 4e-05    evaluation reward: 4.43
episode: 2162  score: 3.0  memory length: 477703  epsilon:
0.25214608000928634  steps: 212  lr: 4e-05    evaluation reward: 4.43
episode: 2163  score: 5.0  memory length: 478010  epsilon: 0.2515382200092825
steps: 307    lr: 4e-05    evaluation reward: 4.46
episode: 2164  score: 2.0  memory length: 478191  epsilon:
0.25117984000928023  steps: 181  lr: 4e-05    evaluation reward: 4.46
episode: 2165  score: 8.0  memory length: 478635  epsilon:
0.25030072000927467  steps: 444  lr: 4e-05    evaluation reward: 4.49
episode: 2166  score: 6.0  memory length: 478977  epsilon:
0.24962356000927038  steps: 342  lr: 4e-05    evaluation reward: 4.53
episode: 2167  score: 8.0  memory length: 479418  epsilon:
0.24875038000926486  steps: 441  lr: 4e-05    evaluation reward: 4.59
episode: 2168  score: 5.0  memory length: 479714  epsilon:
0.24816430000926115  steps: 296  lr: 4e-05    evaluation reward: 4.62
episode: 2169  score: 2.0  memory length: 479895  epsilon:
0.24780592000925888  steps: 181  lr: 4e-05    evaluation reward: 4.6
episode: 2170  score: 5.0  memory length: 480218  epsilon:
0.24716638000925484  steps: 323  lr: 4e-05    evaluation reward: 4.6
episode: 2171  score: 5.0  memory length: 480524  epsilon: 0.246560500009251
steps: 306    lr: 4e-05    evaluation reward: 4.62
episode: 2172  score: 2.0  memory length: 480705  epsilon:
0.24620212000924874  steps: 181  lr: 4e-05    evaluation reward: 4.58
episode: 2173  score: 4.0  memory length: 480962  epsilon:
0.24569326000924552  steps: 257  lr: 4e-05    evaluation reward: 4.58
episode: 2174  score: 6.0  memory length: 481302  epsilon:
0.24502006000924126  steps: 340  lr: 4e-05    evaluation reward: 4.6
episode: 2175  score: 3.0  memory length: 481549  epsilon:
0.24453100000923816  steps: 247  lr: 4e-05    evaluation reward: 4.61
episode: 2176  score: 2.0  memory length: 481730  epsilon: 0.2441726200092359
steps: 181    lr: 4e-05    evaluation reward: 4.58
episode: 2177  score: 3.0  memory length: 481974  epsilon:
0.24368950000923284  steps: 244  lr: 4e-05    evaluation reward: 4.54
episode: 2178  score: 5.0  memory length: 482282  epsilon:
0.24307966000922898  steps: 308  lr: 4e-05    evaluation reward: 4.55
episode: 2179  score: 5.0  memory length: 482605  epsilon:
0.24244012000922494  steps: 323  lr: 4e-05    evaluation reward: 4.54
episode: 2180  score: 4.0  memory length: 482864  epsilon: 0.2419273000092217
steps: 259    lr: 4e-05    evaluation reward: 4.53
episode: 2181  score: 4.0  memory length: 483125  epsilon:
0.24141052000921842  steps: 261  lr: 4e-05    evaluation reward: 4.51
episode: 2182  score: 3.0  memory length: 483337  epsilon:
0.24099076000921577  steps: 212  lr: 4e-05    evaluation reward: 4.49
episode: 2183  score: 2.0  memory length: 483534  epsilon: 0.2406007000092133
steps: 197    lr: 4e-05    evaluation reward: 4.45
episode: 2184  score: 4.0  memory length: 483795  epsilon:
0.24008392000921003  steps: 261  lr: 4e-05    evaluation reward: 4.45
episode: 2185  score: 4.0  memory length: 484072  epsilon:

```

0.23953546000920656 steps: 277 lr: 4e-05 evaluation reward: 4.47
 episode: 2186 score: 4.0 memory length: 484346 epsilon:
 0.23899294000920313 steps: 274 lr: 4e-05 evaluation reward: 4.45
 episode: 2187 score: 3.0 memory length: 484571 epsilon: 0.2385474400092003
 steps: 225 lr: 4e-05 evaluation reward: 4.4
 episode: 2188 score: 4.0 memory length: 484830 epsilon:
 0.23803462000919706 steps: 259 lr: 4e-05 evaluation reward: 4.42
 episode: 2189 score: 4.0 memory length: 485091 epsilon: 0.2375178400091938
 steps: 261 lr: 4e-05 evaluation reward: 4.42
 episode: 2190 score: 3.0 memory length: 485318 epsilon:
 0.23706838000919095 steps: 227 lr: 4e-05 evaluation reward: 4.41
 episode: 2191 score: 4.0 memory length: 485577 epsilon: 0.2365555600091877
 steps: 259 lr: 4e-05 evaluation reward: 4.42
 episode: 2192 score: 2.0 memory length: 485758 epsilon:
 0.23619718000918544 steps: 181 lr: 4e-05 evaluation reward: 4.36
 episode: 2193 score: 5.0 memory length: 486103 epsilon:
 0.23551408000918111 steps: 345 lr: 4e-05 evaluation reward: 4.37
 episode: 2194 score: 3.0 memory length: 486328 epsilon: 0.2350685800091783
 steps: 225 lr: 4e-05 evaluation reward: 4.34
 episode: 2195 score: 4.0 memory length: 486605 epsilon:
 0.23452012000917483 steps: 277 lr: 4e-05 evaluation reward: 4.34
 episode: 2196 score: 3.0 memory length: 486851 epsilon:
 0.23403304000917174 steps: 246 lr: 4e-05 evaluation reward: 4.35
 episode: 2197 score: 4.0 memory length: 487111 epsilon: 0.2335182400091685
 steps: 260 lr: 4e-05 evaluation reward: 4.35
 episode: 2198 score: 8.0 memory length: 487424 epsilon:
 0.23289850000916457 steps: 313 lr: 4e-05 evaluation reward: 4.41
 episode: 2199 score: 4.0 memory length: 487717 epsilon: 0.2323183600091609
 steps: 293 lr: 4e-05 evaluation reward: 4.41
 episode: 2200 score: 3.0 memory length: 487929 epsilon:
 0.23189860000915824 steps: 212 lr: 4e-05 evaluation reward: 4.42
 episode: 2201 score: 4.0 memory length: 488207 epsilon:
 0.23134816000915476 steps: 278 lr: 4e-05 evaluation reward: 4.43
 episode: 2202 score: 2.0 memory length: 488388 epsilon: 0.2309897800091525
 steps: 181 lr: 4e-05 evaluation reward: 4.4
 episode: 2203 score: 2.0 memory length: 488569 epsilon:
 0.23063140000915022 steps: 181 lr: 4e-05 evaluation reward: 4.39
 episode: 2204 score: 3.0 memory length: 488799 epsilon:
 0.23017600000914734 steps: 230 lr: 4e-05 evaluation reward: 4.38
 episode: 2205 score: 4.0 memory length: 489058 epsilon: 0.2296631800091441
 steps: 259 lr: 4e-05 evaluation reward: 4.37
 episode: 2206 score: 5.0 memory length: 489356 epsilon:
 0.22907314000914036 steps: 298 lr: 4e-05 evaluation reward: 4.36
 episode: 2207 score: 5.0 memory length: 489652 epsilon:
 0.22848706000913666 steps: 296 lr: 4e-05 evaluation reward: 4.38
 episode: 2208 score: 6.0 memory length: 489948 epsilon:
 0.22790098000913295 steps: 296 lr: 4e-05 evaluation reward: 4.4
 episode: 2209 score: 3.0 memory length: 490177 epsilon:

0.22744756000913008 steps: 229 lr: 4e-05 evaluation reward: 4.35
 episode: 2210 score: 3.0 memory length: 490387 epsilon:
 0.22703176000912745 steps: 210 lr: 4e-05 evaluation reward: 4.34
 episode: 2211 score: 7.0 memory length: 490744 epsilon:
 0.22632490000912298 steps: 357 lr: 4e-05 evaluation reward: 4.31
 episode: 2212 score: 2.0 memory length: 490926 epsilon: 0.2259645400091207
 steps: 182 lr: 4e-05 evaluation reward: 4.26
 episode: 2213 score: 2.0 memory length: 491125 epsilon: 0.2255705200091182
 steps: 199 lr: 4e-05 evaluation reward: 4.25
 episode: 2214 score: 4.0 memory length: 491381 epsilon: 0.225063640009115
 steps: 256 lr: 4e-05 evaluation reward: 4.23
 episode: 2215 score: 4.0 memory length: 491657 epsilon:
 0.22451716000911154 steps: 276 lr: 4e-05 evaluation reward: 4.22
 episode: 2216 score: 6.0 memory length: 492020 epsilon: 0.223798420009107
 steps: 363 lr: 4e-05 evaluation reward: 4.26
 episode: 2217 score: 6.0 memory length: 492347 epsilon: 0.2231509600091029
 steps: 327 lr: 4e-05 evaluation reward: 4.27
 episode: 2218 score: 2.0 memory length: 492528 epsilon:
 0.22279258000910063 steps: 181 lr: 4e-05 evaluation reward: 4.26
 episode: 2219 score: 4.0 memory length: 492806 epsilon:
 0.22224214000909714 steps: 278 lr: 4e-05 evaluation reward: 4.27
 episode: 2220 score: 3.0 memory length: 493016 epsilon: 0.2218263400090945
 steps: 210 lr: 4e-05 evaluation reward: 4.27
 episode: 2221 score: 2.0 memory length: 493197 epsilon:
 0.22146796000909225 steps: 181 lr: 4e-05 evaluation reward: 4.23
 episode: 2222 score: 4.0 memory length: 493473 epsilon: 0.2209214800090888
 steps: 276 lr: 4e-05 evaluation reward: 4.23
 episode: 2223 score: 8.0 memory length: 493943 epsilon: 0.2199908800090829
 steps: 470 lr: 4e-05 evaluation reward: 4.26
 episode: 2224 score: 3.0 memory length: 494154 epsilon:
 0.21957310000908026 steps: 211 lr: 4e-05 evaluation reward: 4.23
 episode: 2225 score: 3.0 memory length: 494363 epsilon:
 0.21915928000907764 steps: 209 lr: 4e-05 evaluation reward: 4.23
 episode: 2226 score: 6.0 memory length: 494734 epsilon: 0.218424700009073
 steps: 371 lr: 4e-05 evaluation reward: 4.27
 episode: 2227 score: 4.0 memory length: 495014 epsilon:
 0.21787030000906948 steps: 280 lr: 4e-05 evaluation reward: 4.26
 episode: 2228 score: 4.0 memory length: 495275 epsilon:
 0.21735352000906621 steps: 261 lr: 4e-05 evaluation reward: 4.27
 episode: 2229 score: 5.0 memory length: 495563 epsilon: 0.2167832800090626
 steps: 288 lr: 4e-05 evaluation reward: 4.29
 episode: 2230 score: 2.0 memory length: 495745 epsilon:
 0.21642292000906033 steps: 182 lr: 4e-05 evaluation reward: 4.26
 episode: 2231 score: 4.0 memory length: 496004 epsilon:
 0.21591010000905708 steps: 259 lr: 4e-05 evaluation reward: 4.26
 episode: 2232 score: 4.0 memory length: 496263 epsilon:
 0.21539728000905384 steps: 259 lr: 4e-05 evaluation reward: 4.27
 episode: 2233 score: 4.0 memory length: 496505 epsilon: 0.2149181200090508


```

steps: 242    lr: 4e-05    evaluation reward: 4.28
episode: 2234  score: 6.0    memory length: 496850    epsilon:
0.21423502000904648    steps: 345    lr: 4e-05    evaluation reward: 4.29
episode: 2235  score: 2.0    memory length: 497031    epsilon:
0.21387664000904422    steps: 181    lr: 4e-05    evaluation reward: 4.29
episode: 2236  score: 3.0    memory length: 497239    epsilon: 0.2134648000090416
steps: 208    lr: 4e-05    evaluation reward: 4.24
episode: 2237  score: 5.0    memory length: 497526    epsilon:
0.21289654000903802    steps: 287    lr: 4e-05    evaluation reward: 4.25
episode: 2238  score: 2.0    memory length: 497707    epsilon:
0.21253816000903575    steps: 181    lr: 4e-05    evaluation reward: 4.23
episode: 2239  score: 6.0    memory length: 498061    epsilon: 0.2118372400090313
steps: 354    lr: 4e-05    evaluation reward: 4.18
episode: 2240  score: 8.0    memory length: 498499    epsilon:
0.21097000000902583    steps: 438    lr: 4e-05    evaluation reward: 4.2
episode: 2241  score: 5.0    memory length: 498807    epsilon:
0.21036016000902197    steps: 308    lr: 4e-05    evaluation reward: 4.22
episode: 2242  score: 5.0    memory length: 499121    epsilon:
0.20973844000901803    steps: 314    lr: 4e-05    evaluation reward: 4.2
episode: 2243  score: 4.0    memory length: 499379    epsilon: 0.2092276000090148
steps: 258    lr: 4e-05    evaluation reward: 4.17
episode: 2244  score: 6.0    memory length: 499721    epsilon:
0.20855044000901052    steps: 342    lr: 4e-05    evaluation reward: 4.2
episode: 2245  score: 6.0    memory length: 500063    epsilon:
0.20787328000900623    steps: 342    lr: 4e-05    evaluation reward: 4.19
episode: 2246  score: 5.0    memory length: 500351    epsilon:
0.20730304000900263    steps: 288    lr: 4e-05    evaluation reward: 4.19
episode: 2247  score: 5.0    memory length: 500656    epsilon: 0.2066991400089988
steps: 305    lr: 4e-05    evaluation reward: 4.2
episode: 2248  score: 3.0    memory length: 500884    epsilon:
0.20624770000899595    steps: 228    lr: 4e-05    evaluation reward: 4.18
episode: 2249  score: 6.0    memory length: 501209    epsilon:
0.20560420000899188    steps: 325    lr: 4e-05    evaluation reward: 4.21
episode: 2250  score: 4.0    memory length: 501468    epsilon:
0.20509138000898863    steps: 259    lr: 4e-05    evaluation reward: 4.2
episode: 2251  score: 3.0    memory length: 501717    epsilon: 0.2045983600089855
steps: 249    lr: 4e-05    evaluation reward: 4.21
episode: 2252  score: 4.0    memory length: 502011    epsilon:
0.20401624000898183    steps: 294    lr: 4e-05    evaluation reward: 4.18
episode: 2253  score: 4.0    memory length: 502254    epsilon: 0.2035351000089788
steps: 243    lr: 4e-05    evaluation reward: 4.19
episode: 2254  score: 2.0    memory length: 502435    epsilon:
0.20317672000897652    steps: 181    lr: 4e-05    evaluation reward: 4.18
episode: 2255  score: 10.0   memory length: 502943    epsilon:
0.20217088000897016    steps: 508    lr: 4e-05    evaluation reward: 4.23
episode: 2256  score: 2.0    memory length: 503140    epsilon: 0.2017808200089677
steps: 197    lr: 4e-05    evaluation reward: 4.22
episode: 2257  score: 6.0    memory length: 503482    epsilon: 0.2011036600089634

```

```

steps: 342    lr: 4e-05    evaluation reward: 4.23
episode: 2258  score: 3.0    memory length: 503713    epsilon: 0.2006462800089605
steps: 231    lr: 4e-05    evaluation reward: 4.22
episode: 2259  score: 5.0    memory length: 504018    epsilon: 0.2000423800089567
steps: 305    lr: 4e-05    evaluation reward: 4.19
episode: 2260  score: 6.0    memory length: 504344    epsilon: 0.1993969000089526
steps: 326    lr: 4e-05    evaluation reward: 4.21
episode: 2261  score: 7.0    memory length: 504718    epsilon:
0.19865638000894792    steps: 374    lr: 4e-05    evaluation reward: 4.21
episode: 2262  score: 4.0    memory length: 504992    epsilon: 0.1981138600089445
steps: 274    lr: 4e-05    evaluation reward: 4.22
episode: 2263  score: 3.0    memory length: 505241    epsilon:
0.19762084000894137    steps: 249    lr: 4e-05    evaluation reward: 4.2
episode: 2264  score: 5.0    memory length: 505529    epsilon:
0.19705060000893776    steps: 288    lr: 4e-05    evaluation reward: 4.23
episode: 2265  score: 4.0    memory length: 505808    epsilon:
0.19649818000893426    steps: 279    lr: 4e-05    evaluation reward: 4.19
episode: 2266  score: 5.0    memory length: 506112    epsilon:
0.19589626000893046    steps: 304    lr: 4e-05    evaluation reward: 4.18
episode: 2267  score: 2.0    memory length: 506293    epsilon: 0.1955378800089282
steps: 181    lr: 4e-05    evaluation reward: 4.12
episode: 2268  score: 4.0    memory length: 506534    epsilon:
0.19506070000892517    steps: 241    lr: 4e-05    evaluation reward: 4.11
episode: 2269  score: 5.0    memory length: 506860    epsilon:
0.19441522000892109    steps: 326    lr: 4e-05    evaluation reward: 4.14
episode: 2270  score: 3.0    memory length: 507085    epsilon:
0.19396972000891827    steps: 225    lr: 4e-05    evaluation reward: 4.12
episode: 2271  score: 4.0    memory length: 507341    epsilon:
0.19346284000891506    steps: 256    lr: 4e-05    evaluation reward: 4.11
episode: 2272  score: 2.0    memory length: 507522    epsilon: 0.1931044600089128
steps: 181    lr: 4e-05    evaluation reward: 4.11
episode: 2273  score: 2.0    memory length: 507704    epsilon: 0.1927441000089105
steps: 182    lr: 4e-05    evaluation reward: 4.09
episode: 2274  score: 2.0    memory length: 507904    epsilon: 0.192348100008908
steps: 200    lr: 4e-05    evaluation reward: 4.05
episode: 2275  score: 3.0    memory length: 508151    epsilon: 0.1918590400089049
steps: 247    lr: 4e-05    evaluation reward: 4.05
episode: 2276  score: 5.0    memory length: 508439    epsilon: 0.1912888000089013
steps: 288    lr: 4e-05    evaluation reward: 4.08
episode: 2277  score: 2.0    memory length: 508620    epsilon:
0.19093042000889904    steps: 181    lr: 4e-05    evaluation reward: 4.07
episode: 2278  score: 3.0    memory length: 508848    epsilon:
0.19047898000889618    steps: 228    lr: 4e-05    evaluation reward: 4.05
episode: 2279  score: 3.0    memory length: 509095    epsilon: 0.1899899200088931
steps: 247    lr: 4e-05    evaluation reward: 4.03
episode: 2280  score: 7.0    memory length: 509469    epsilon: 0.1892494000088884
steps: 374    lr: 4e-05    evaluation reward: 4.06
episode: 2281  score: 6.0    memory length: 509814    epsilon:

```

0.18856630000888408 steps: 345 lr: 4e-05 evaluation reward: 4.08
 episode: 2282 score: 5.0 memory length: 510094 epsilon:
 0.18801190000888057 steps: 280 lr: 4e-05 evaluation reward: 4.1
 episode: 2283 score: 2.0 memory length: 510274 epsilon:
 0.18765550000887832 steps: 180 lr: 4e-05 evaluation reward: 4.1
 episode: 2284 score: 8.0 memory length: 510697 epsilon:
 0.18681796000887302 steps: 423 lr: 4e-05 evaluation reward: 4.14
 episode: 2285 score: 5.0 memory length: 511002 epsilon: 0.1862140600088692
 steps: 305 lr: 4e-05 evaluation reward: 4.15
 episode: 2286 score: 6.0 memory length: 511342 epsilon:
 0.18554086000886494 steps: 340 lr: 4e-05 evaluation reward: 4.17
 episode: 2287 score: 4.0 memory length: 511583 epsilon:
 0.18506368000886192 steps: 241 lr: 4e-05 evaluation reward: 4.18
 episode: 2288 score: 7.0 memory length: 511959 epsilon: 0.1843192000088572
 steps: 376 lr: 4e-05 evaluation reward: 4.21
 episode: 2289 score: 4.0 memory length: 512218 epsilon:
 0.18380638000885396 steps: 259 lr: 4e-05 evaluation reward: 4.21
 episode: 2290 score: 4.0 memory length: 512492 epsilon:
 0.18326386000885053 steps: 274 lr: 4e-05 evaluation reward: 4.22
 episode: 2291 score: 9.0 memory length: 512960 epsilon:
 0.18233722000884467 steps: 468 lr: 4e-05 evaluation reward: 4.27
 episode: 2292 score: 6.0 memory length: 513299 epsilon:
 0.18166600000884042 steps: 339 lr: 4e-05 evaluation reward: 4.31
 episode: 2293 score: 6.0 memory length: 513624 epsilon:
 0.18102250000883635 steps: 325 lr: 4e-05 evaluation reward: 4.32
 episode: 2294 score: 3.0 memory length: 513855 epsilon:
 0.18056512000883346 steps: 231 lr: 4e-05 evaluation reward: 4.32
 episode: 2295 score: 4.0 memory length: 514113 epsilon:
 0.18005428000883023 steps: 258 lr: 4e-05 evaluation reward: 4.32
 episode: 2296 score: 2.0 memory length: 514294 epsilon:
 0.17969590000882796 steps: 181 lr: 4e-05 evaluation reward: 4.31
 episode: 2297 score: 6.0 memory length: 514629 epsilon:
 0.17903260000882376 steps: 335 lr: 4e-05 evaluation reward: 4.33
 episode: 2298 score: 5.0 memory length: 514919 epsilon:
 0.17845840000882013 steps: 290 lr: 4e-05 evaluation reward: 4.3
 episode: 2299 score: 4.0 memory length: 515162 epsilon:
 0.17797726000881708 steps: 243 lr: 4e-05 evaluation reward: 4.3
 episode: 2300 score: 4.0 memory length: 515419 epsilon:
 0.17746840000881386 steps: 257 lr: 4e-05 evaluation reward: 4.31
 episode: 2301 score: 5.0 memory length: 515708 epsilon:
 0.17689618000881024 steps: 289 lr: 4e-05 evaluation reward: 4.32
 episode: 2302 score: 2.0 memory length: 515889 epsilon:
 0.17653780000880798 steps: 181 lr: 4e-05 evaluation reward: 4.32
 episode: 2303 score: 2.0 memory length: 516071 epsilon: 0.1761774400088057
 steps: 182 lr: 4e-05 evaluation reward: 4.32
 episode: 2304 score: 8.0 memory length: 516467 epsilon:
 0.17539336000880074 steps: 396 lr: 4e-05 evaluation reward: 4.37
 episode: 2305 score: 4.0 memory length: 516743 epsilon:

0.17484688000879728 steps: 276 lr: 4e-05 evaluation reward: 4.37
 episode: 2306 score: 2.0 memory length: 516924 epsilon: 0.174488500008795
 steps: 181 lr: 4e-05 evaluation reward: 4.34
 episode: 2307 score: 4.0 memory length: 517183 epsilon:
 0.17397568000879177 steps: 259 lr: 4e-05 evaluation reward: 4.33
 episode: 2308 score: 5.0 memory length: 517473 epsilon:
 0.17340148000878813 steps: 290 lr: 4e-05 evaluation reward: 4.32
 episode: 2309 score: 5.0 memory length: 517779 epsilon: 0.1727956000087843
 steps: 306 lr: 4e-05 evaluation reward: 4.34
 episode: 2310 score: 4.0 memory length: 518040 epsilon:
 0.17227882000878103 steps: 261 lr: 4e-05 evaluation reward: 4.35
 episode: 2311 score: 6.0 memory length: 518385 epsilon: 0.1715957200087767
 steps: 345 lr: 4e-05 evaluation reward: 4.34
 episode: 2312 score: 6.0 memory length: 518690 epsilon: 0.1709918200087729
 steps: 305 lr: 4e-05 evaluation reward: 4.38
 episode: 2313 score: 3.0 memory length: 518900 epsilon:
 0.17057602000877026 steps: 210 lr: 4e-05 evaluation reward: 4.39
 episode: 2314 score: 2.0 memory length: 519080 epsilon: 0.170219620008768
 steps: 180 lr: 4e-05 evaluation reward: 4.37
 episode: 2315 score: 6.0 memory length: 519385 epsilon:
 0.16961572000876418 steps: 305 lr: 4e-05 evaluation reward: 4.39
 episode: 2316 score: 9.0 memory length: 519831 epsilon: 0.1687326400087586
 steps: 446 lr: 4e-05 evaluation reward: 4.42
 episode: 2317 score: 2.0 memory length: 520012 epsilon:
 0.16837426000875633 steps: 181 lr: 4e-05 evaluation reward: 4.38
 episode: 2318 score: 4.0 memory length: 520255 epsilon:
 0.16789312000875328 steps: 243 lr: 4e-05 evaluation reward: 4.4
 episode: 2319 score: 5.0 memory length: 520562 epsilon:
 0.16728526000874944 steps: 307 lr: 4e-05 evaluation reward: 4.41
 episode: 2320 score: 6.0 memory length: 520920 epsilon:
 0.16657642000874495 steps: 358 lr: 4e-05 evaluation reward: 4.44
 episode: 2321 score: 4.0 memory length: 521196 epsilon: 0.1660299400087415
 steps: 276 lr: 4e-05 evaluation reward: 4.46
 episode: 2322 score: 5.0 memory length: 521506 epsilon: 0.1654161400087376
 steps: 310 lr: 4e-05 evaluation reward: 4.47
 episode: 2323 score: 7.0 memory length: 521890 epsilon: 0.1646558200087328
 steps: 384 lr: 4e-05 evaluation reward: 4.46
 episode: 2324 score: 4.0 memory length: 522149 epsilon:
 0.16414300000872956 steps: 259 lr: 4e-05 evaluation reward: 4.47
 episode: 2325 score: 2.0 memory length: 522330 epsilon: 0.1637846200087273
 steps: 181 lr: 4e-05 evaluation reward: 4.46
 episode: 2326 score: 8.0 memory length: 522755 epsilon:
 0.16294312000872196 steps: 425 lr: 4e-05 evaluation reward: 4.48
 episode: 2327 score: 6.0 memory length: 523097 epsilon:
 0.16226596000871768 steps: 342 lr: 4e-05 evaluation reward: 4.5
 episode: 2328 score: 7.0 memory length: 523488 epsilon:
 0.16149178000871278 steps: 391 lr: 4e-05 evaluation reward: 4.53
 episode: 2329 score: 5.0 memory length: 523784 epsilon:

0.16090570000870907 steps: 296 lr: 4e-05 evaluation reward: 4.53
 episode: 2330 score: 5.0 memory length: 524080 epsilon:
 0.16031962000870537 steps: 296 lr: 4e-05 evaluation reward: 4.56
 episode: 2331 score: 2.0 memory length: 524261 epsilon: 0.1599612400087031
 steps: 181 lr: 4e-05 evaluation reward: 4.54
 episode: 2332 score: 12.0 memory length: 524787 epsilon:
 0.1589197600086965 steps: 526 lr: 4e-05 evaluation reward: 4.62
 episode: 2333 score: 2.0 memory length: 524966 epsilon:
 0.15856534000869427 steps: 179 lr: 4e-05 evaluation reward: 4.6
 episode: 2334 score: 4.0 memory length: 525244 epsilon:
 0.15801490000869078 steps: 278 lr: 4e-05 evaluation reward: 4.58
 episode: 2335 score: 5.0 memory length: 525540 epsilon:
 0.15742882000868708 steps: 296 lr: 4e-05 evaluation reward: 4.61
 episode: 2336 score: 5.0 memory length: 525831 epsilon:
 0.15685264000868343 steps: 291 lr: 4e-05 evaluation reward: 4.63
 episode: 2337 score: 3.0 memory length: 526081 epsilon: 0.1563576400086803
 steps: 250 lr: 4e-05 evaluation reward: 4.61
 episode: 2338 score: 5.0 memory length: 526392 epsilon: 0.1557418600086764
 steps: 311 lr: 4e-05 evaluation reward: 4.64
 episode: 2339 score: 4.0 memory length: 526673 epsilon:
 0.15518548000867288 steps: 281 lr: 4e-05 evaluation reward: 4.62
 episode: 2340 score: 4.0 memory length: 526934 epsilon: 0.1546687000086696
 steps: 261 lr: 4e-05 evaluation reward: 4.58
 episode: 2341 score: 2.0 memory length: 527115 epsilon:
 0.15431032000866735 steps: 181 lr: 4e-05 evaluation reward: 4.55
 episode: 2342 score: 6.0 memory length: 527432 epsilon:
 0.15368266000866337 steps: 317 lr: 4e-05 evaluation reward: 4.56
 episode: 2343 score: 2.0 memory length: 527613 epsilon: 0.1533242800086611
 steps: 181 lr: 4e-05 evaluation reward: 4.54
 episode: 2344 score: 4.0 memory length: 527892 epsilon: 0.1527718600086576
 steps: 279 lr: 4e-05 evaluation reward: 4.52
 episode: 2345 score: 7.0 memory length: 528254 epsilon:
 0.15205510000865308 steps: 362 lr: 4e-05 evaluation reward: 4.53
 episode: 2346 score: 3.0 memory length: 528479 epsilon:
 0.15160960000865026 steps: 225 lr: 4e-05 evaluation reward: 4.51
 episode: 2347 score: 4.0 memory length: 528755 epsilon: 0.1510631200086468
 steps: 276 lr: 4e-05 evaluation reward: 4.5
 episode: 2348 score: 3.0 memory length: 528967 epsilon:
 0.15064336000864414 steps: 212 lr: 4e-05 evaluation reward: 4.5
 episode: 2349 score: 8.0 memory length: 529364 epsilon:
 0.14985730000863917 steps: 397 lr: 4e-05 evaluation reward: 4.52
 episode: 2350 score: 5.0 memory length: 529644 epsilon:
 0.14930290000863566 steps: 280 lr: 4e-05 evaluation reward: 4.53
 episode: 2351 score: 2.0 memory length: 529825 epsilon: 0.1489445200086334
 steps: 181 lr: 4e-05 evaluation reward: 4.52
 episode: 2352 score: 6.0 memory length: 530138 epsilon:
 0.14832478000862948 steps: 313 lr: 4e-05 evaluation reward: 4.54
 episode: 2353 score: 5.0 memory length: 530431 epsilon: 0.1477446400086258

```

steps: 293    lr: 4e-05    evaluation reward: 4.55
episode: 2354  score: 4.0  memory length: 530672  epsilon:
0.14726746000862279  steps: 241    lr: 4e-05    evaluation reward: 4.57
episode: 2355  score: 5.0  memory length: 530980  epsilon:
0.14665762000861893  steps: 308    lr: 4e-05    evaluation reward: 4.52
episode: 2356  score: 4.0  memory length: 531258  epsilon:
0.14610718000861544  steps: 278    lr: 4e-05    evaluation reward: 4.54
episode: 2357  score: 4.0  memory length: 531521  epsilon:
0.14558644000861215  steps: 263    lr: 4e-05    evaluation reward: 4.52
episode: 2358  score: 4.0  memory length: 531782  epsilon:
0.14506966000860888  steps: 261    lr: 4e-05    evaluation reward: 4.53
episode: 2359  score: 4.0  memory length: 532043  epsilon: 0.1445528800086056
steps: 261    lr: 4e-05    evaluation reward: 4.52
episode: 2360  score: 5.0  memory length: 532366  epsilon:
0.14391334000860156  steps: 323    lr: 4e-05    evaluation reward: 4.51
episode: 2361  score: 6.0  memory length: 532708  epsilon:
0.14323618000859728  steps: 342    lr: 4e-05    evaluation reward: 4.5
episode: 2362  score: 4.0  memory length: 532984  epsilon:
0.14268970000859382  steps: 276    lr: 4e-05    evaluation reward: 4.5
episode: 2363  score: 2.0  memory length: 533165  epsilon:
0.14233132000859156  steps: 181    lr: 4e-05    evaluation reward: 4.49
episode: 2364  score: 2.0  memory length: 533346  epsilon: 0.1419729400085893
steps: 181    lr: 4e-05    evaluation reward: 4.46
episode: 2365  score: 6.0  memory length: 533688  epsilon: 0.141295780008585
steps: 342    lr: 4e-05    evaluation reward: 4.48
episode: 2366  score: 6.0  memory length: 534044  epsilon:
0.14059090000858054  steps: 356    lr: 4e-05    evaluation reward: 4.49
episode: 2367  score: 6.0  memory length: 534371  epsilon:
0.13994344000857645  steps: 327    lr: 4e-05    evaluation reward: 4.53
episode: 2368  score: 6.0  memory length: 534713  epsilon:
0.13926628000857216  steps: 342    lr: 4e-05    evaluation reward: 4.55
episode: 2369  score: 8.0  memory length: 535134  epsilon: 0.1384327000085669
steps: 421    lr: 4e-05    evaluation reward: 4.58
episode: 2370  score: 7.0  memory length: 535487  epsilon:
0.13773376000856247  steps: 353    lr: 4e-05    evaluation reward: 4.62
episode: 2371  score: 3.0  memory length: 535696  epsilon:
0.13731994000855985  steps: 209    lr: 4e-05    evaluation reward: 4.61
episode: 2372  score: 2.0  memory length: 535877  epsilon:
0.13696156000855758  steps: 181    lr: 4e-05    evaluation reward: 4.61
episode: 2373  score: 3.0  memory length: 536105  epsilon:
0.13651012000855473  steps: 228    lr: 4e-05    evaluation reward: 4.62
episode: 2374  score: 4.0  memory length: 536362  epsilon: 0.1360012600085515
steps: 257    lr: 4e-05    evaluation reward: 4.64
episode: 2375  score: 6.0  memory length: 536704  epsilon:
0.13532410000854722  steps: 342    lr: 4e-05    evaluation reward: 4.67
episode: 2376  score: 2.0  memory length: 536883  epsilon:
0.13496968000854498  steps: 179    lr: 4e-05    evaluation reward: 4.64
episode: 2377  score: 5.0  memory length: 537158  epsilon:

```

0.13442518000854153 steps: 275 lr: 4e-05 evaluation reward: 4.67
 episode: 2378 score: 4.0 memory length: 537455 epsilon: 0.1338371200085378
 steps: 297 lr: 4e-05 evaluation reward: 4.68
 episode: 2379 score: 2.0 memory length: 537653 epsilon:
 0.13344508000853533 steps: 198 lr: 4e-05 evaluation reward: 4.67
 episode: 2380 score: 5.0 memory length: 537950 epsilon: 0.1328570200085316
 steps: 297 lr: 4e-05 evaluation reward: 4.65
 episode: 2381 score: 3.0 memory length: 538178 epsilon:
 0.13240558000852876 steps: 228 lr: 4e-05 evaluation reward: 4.62
 episode: 2382 score: 4.0 memory length: 538437 epsilon: 0.1318927600085255
 steps: 259 lr: 4e-05 evaluation reward: 4.61
 episode: 2383 score: 2.0 memory length: 538635 epsilon:
 0.13150072000852303 steps: 198 lr: 4e-05 evaluation reward: 4.61
 episode: 2384 score: 2.0 memory length: 538816 epsilon:
 0.13114234000852076 steps: 181 lr: 4e-05 evaluation reward: 4.55
 episode: 2385 score: 5.0 memory length: 539140 epsilon: 0.1305008200085167
 steps: 324 lr: 4e-05 evaluation reward: 4.55
 episode: 2386 score: 4.0 memory length: 539418 epsilon:
 0.12995038000851322 steps: 278 lr: 4e-05 evaluation reward: 4.53
 episode: 2387 score: 5.0 memory length: 539698 epsilon:
 0.12939598000850971 steps: 280 lr: 4e-05 evaluation reward: 4.54
 episode: 2388 score: 4.0 memory length: 539954 epsilon: 0.1288891000085065
 steps: 256 lr: 4e-05 evaluation reward: 4.51
 episode: 2389 score: 6.0 memory length: 540296 epsilon:
 0.12821194000850222 steps: 342 lr: 4e-05 evaluation reward: 4.53
 episode: 2390 score: 5.0 memory length: 540576 epsilon:
 0.12765754000849872 steps: 280 lr: 4e-05 evaluation reward: 4.54
 episode: 2391 score: 2.0 memory length: 540757 epsilon:
 0.12729916000849645 steps: 181 lr: 4e-05 evaluation reward: 4.47
 episode: 2392 score: 2.0 memory length: 540938 epsilon:
 0.12694078000849418 steps: 181 lr: 4e-05 evaluation reward: 4.43
 episode: 2393 score: 4.0 memory length: 541200 epsilon: 0.1264220200084909
 steps: 262 lr: 4e-05 evaluation reward: 4.41
 episode: 2394 score: 6.0 memory length: 541542 epsilon:
 0.12574486000848661 steps: 342 lr: 4e-05 evaluation reward: 4.44
 episode: 2395 score: 4.0 memory length: 541822 epsilon: 0.1251904600084831
 steps: 280 lr: 4e-05 evaluation reward: 4.44
 episode: 2396 score: 6.0 memory length: 542164 epsilon:
 0.12451330000848224 steps: 342 lr: 4e-05 evaluation reward: 4.48
 episode: 2397 score: 4.0 memory length: 542420 epsilon:
 0.12400642000848258 steps: 256 lr: 4e-05 evaluation reward: 4.46
 episode: 2398 score: 5.0 memory length: 542701 epsilon:
 0.12345004000848296 steps: 281 lr: 4e-05 evaluation reward: 4.46
 episode: 2399 score: 9.0 memory length: 543176 epsilon: 0.1225095400084836
 steps: 475 lr: 4e-05 evaluation reward: 4.51
 episode: 2400 score: 4.0 memory length: 543414 epsilon:
 0.12203830000848392 steps: 238 lr: 4e-05 evaluation reward: 4.51
 episode: 2401 score: 5.0 memory length: 543704 epsilon:

0.12146410000848432 steps: 290 lr: 4e-05 evaluation reward: 4.51
 episode: 2402 score: 5.0 memory length: 544012 epsilon:
 0.12085426000848473 steps: 308 lr: 4e-05 evaluation reward: 4.54
 episode: 2403 score: 2.0 memory length: 544193 epsilon:
 0.12049588000848498 steps: 181 lr: 4e-05 evaluation reward: 4.54
 episode: 2404 score: 2.0 memory length: 544374 epsilon:
 0.12013750000848522 steps: 181 lr: 4e-05 evaluation reward: 4.48
 episode: 2405 score: 2.0 memory length: 544555 epsilon:
 0.11977912000848547 steps: 181 lr: 4e-05 evaluation reward: 4.46
 episode: 2406 score: 2.0 memory length: 544736 epsilon:
 0.11942074000848571 steps: 181 lr: 4e-05 evaluation reward: 4.46
 episode: 2407 score: 7.0 memory length: 545127 epsilon:
 0.11864656000848624 steps: 391 lr: 4e-05 evaluation reward: 4.49
 episode: 2408 score: 6.0 memory length: 545454 epsilon:
 0.11799910000848668 steps: 327 lr: 4e-05 evaluation reward: 4.5
 episode: 2409 score: 5.0 memory length: 545800 epsilon:
 0.11731402000848715 steps: 346 lr: 4e-05 evaluation reward: 4.5
 episode: 2410 score: 2.0 memory length: 545981 epsilon:
 0.11695564000848739 steps: 181 lr: 4e-05 evaluation reward: 4.48
 episode: 2411 score: 4.0 memory length: 546257 epsilon:
 0.11640916000848776 steps: 276 lr: 4e-05 evaluation reward: 4.46
 episode: 2412 score: 3.0 memory length: 546506 epsilon: 0.1159161400084881
 steps: 249 lr: 4e-05 evaluation reward: 4.43
 episode: 2413 score: 2.0 memory length: 546688 epsilon:
 0.11555578000848835 steps: 182 lr: 4e-05 evaluation reward: 4.42
 episode: 2414 score: 6.0 memory length: 547033 epsilon:
 0.11487268000848881 steps: 345 lr: 4e-05 evaluation reward: 4.46
 episode: 2415 score: 7.0 memory length: 547409 epsilon:
 0.11412820000848932 steps: 376 lr: 4e-05 evaluation reward: 4.47
 episode: 2416 score: 7.0 memory length: 547834 epsilon: 0.1132867000084899
 steps: 425 lr: 4e-05 evaluation reward: 4.45
 episode: 2417 score: 4.0 memory length: 548108 epsilon:
 0.11274418000849026 steps: 274 lr: 4e-05 evaluation reward: 4.47
 episode: 2418 score: 3.0 memory length: 548337 epsilon:
 0.11229076000849057 steps: 229 lr: 4e-05 evaluation reward: 4.46
 episode: 2419 score: 2.0 memory length: 548518 epsilon:
 0.11193238000849082 steps: 181 lr: 4e-05 evaluation reward: 4.43
 episode: 2420 score: 2.0 memory length: 548699 epsilon:
 0.11157400000849106 steps: 181 lr: 4e-05 evaluation reward: 4.39
 episode: 2421 score: 5.0 memory length: 549008 epsilon:
 0.11096218000849148 steps: 309 lr: 4e-05 evaluation reward: 4.4
 episode: 2422 score: 4.0 memory length: 549286 epsilon:
 0.11041174000849185 steps: 278 lr: 4e-05 evaluation reward: 4.39
 episode: 2423 score: 3.0 memory length: 549514 epsilon:
 0.10996030000849216 steps: 228 lr: 4e-05 evaluation reward: 4.35
 episode: 2424 score: 4.0 memory length: 549757 epsilon:
 0.10947916000849249 steps: 243 lr: 4e-05 evaluation reward: 4.35
 episode: 2425 score: 7.0 memory length: 550131 epsilon: 0.108738640008493


```

steps: 374    lr: 4e-05    evaluation reward: 4.4
episode: 2426  score: 2.0  memory length: 550312  epsilon:
0.10838026000849324  steps: 181  lr: 4e-05  evaluation reward: 4.34
episode: 2427  score: 8.0  memory length: 550719  epsilon:
0.10757440000849379  steps: 407  lr: 4e-05  evaluation reward: 4.36
episode: 2428  score: 4.0  memory length: 550995  epsilon:
0.10702792000849416  steps: 276  lr: 4e-05  evaluation reward: 4.33
episode: 2429  score: 5.0  memory length: 551275  epsilon:
0.10647352000849454  steps: 280  lr: 4e-05  evaluation reward: 4.33
episode: 2430  score: 3.0  memory length: 551485  epsilon:
0.10605772000849482  steps: 210  lr: 4e-05  evaluation reward: 4.31
episode: 2431  score: 6.0  memory length: 551810  epsilon:
0.10541422000849526  steps: 325  lr: 4e-05  evaluation reward: 4.35
episode: 2432  score: 2.0  memory length: 551991  epsilon:
0.10505584000849551  steps: 181  lr: 4e-05  evaluation reward: 4.25
episode: 2433  score: 2.0  memory length: 552172  epsilon:
0.10469746000849575  steps: 181  lr: 4e-05  evaluation reward: 4.25
episode: 2434  score: 3.0  memory length: 552401  epsilon:
0.10424404000849606  steps: 229  lr: 4e-05  evaluation reward: 4.24
episode: 2435  score: 4.0  memory length: 552675  epsilon:
0.10370152000849643  steps: 274  lr: 4e-05  evaluation reward: 4.23
episode: 2436  score: 7.0  memory length: 553032  epsilon:
0.10299466000849691  steps: 357  lr: 4e-05  evaluation reward: 4.25
episode: 2437  score: 4.0  memory length: 553292  epsilon:
0.10247986000849726  steps: 260  lr: 4e-05  evaluation reward: 4.26
episode: 2438  score: 2.0  memory length: 553473  epsilon:
0.10212148000849751  steps: 181  lr: 4e-05  evaluation reward: 4.23
episode: 2439  score: 4.0  memory length: 553731  epsilon:
0.10161064000849786  steps: 258  lr: 4e-05  evaluation reward: 4.23
episode: 2440  score: 5.0  memory length: 554057  epsilon: 0.1009651600084983
steps: 326  lr: 4e-05  evaluation reward: 4.24
episode: 2441  score: 5.0  memory length: 554361  epsilon:
0.10036324000849871  steps: 304  lr: 4e-05  evaluation reward: 4.27
episode: 2442  score: 4.0  memory length: 554639  epsilon:
0.09981280000849908  steps: 278  lr: 4e-05  evaluation reward: 4.25
episode: 2443  score: 6.0  memory length: 554981  epsilon:
0.09913564000849955  steps: 342  lr: 4e-05  evaluation reward: 4.29
episode: 2444  score: 4.0  memory length: 555239  epsilon: 0.0986248000084999
steps: 258  lr: 4e-05  evaluation reward: 4.29
episode: 2445  score: 3.0  memory length: 555469  epsilon: 0.0981694000085002
steps: 230  lr: 4e-05  evaluation reward: 4.25
episode: 2446  score: 8.0  memory length: 555873  epsilon:
0.09736948000850075  steps: 404  lr: 4e-05  evaluation reward: 4.3
episode: 2447  score: 4.0  memory length: 556149  epsilon:
0.09682300000850112  steps: 276  lr: 4e-05  evaluation reward: 4.3
episode: 2448  score: 8.0  memory length: 556544  epsilon:
0.09604090000850166  steps: 395  lr: 4e-05  evaluation reward: 4.35
episode: 2449  score: 4.0  memory length: 556822  epsilon:

```

0.09549046000850203 steps: 278 lr: 4e-05 evaluation reward: 4.31
 episode: 2450 score: 6.0 memory length: 557196 epsilon:
 0.09474994000850254 steps: 374 lr: 4e-05 evaluation reward: 4.32
 episode: 2451 score: 11.0 memory length: 557690 epsilon:
 0.0937718200085032 steps: 494 lr: 4e-05 evaluation reward: 4.41
 episode: 2452 score: 2.0 memory length: 557870 epsilon:
 0.09341542000850345 steps: 180 lr: 4e-05 evaluation reward: 4.37
 episode: 2453 score: 6.0 memory length: 558216 epsilon:
 0.09273034000850391 steps: 346 lr: 4e-05 evaluation reward: 4.38
 episode: 2454 score: 2.0 memory length: 558397 epsilon:
 0.09237196000850416 steps: 181 lr: 4e-05 evaluation reward: 4.36
 episode: 2455 score: 7.0 memory length: 558800 epsilon: 0.0915740200085047
 steps: 403 lr: 4e-05 evaluation reward: 4.38
 episode: 2456 score: 5.0 memory length: 559125 epsilon:
 0.09093052000850514 steps: 325 lr: 4e-05 evaluation reward: 4.39
 episode: 2457 score: 8.0 memory length: 559524 epsilon:
 0.09014050000850568 steps: 399 lr: 4e-05 evaluation reward: 4.43
 episode: 2458 score: 3.0 memory length: 559736 epsilon:
 0.08972074000850597 steps: 212 lr: 4e-05 evaluation reward: 4.42
 episode: 2459 score: 3.0 memory length: 559965 epsilon:
 0.08926732000850628 steps: 229 lr: 4e-05 evaluation reward: 4.41
 episode: 2460 score: 2.0 memory length: 560146 epsilon:
 0.08890894000850652 steps: 181 lr: 4e-05 evaluation reward: 4.38
 episode: 2461 score: 6.0 memory length: 560473 epsilon:
 0.08826148000850696 steps: 327 lr: 4e-05 evaluation reward: 4.38
 episode: 2462 score: 4.0 memory length: 560734 epsilon:
 0.08774470000850731 steps: 261 lr: 4e-05 evaluation reward: 4.38
 episode: 2463 score: 5.0 memory length: 561044 epsilon:
 0.08713090000850773 steps: 310 lr: 4e-05 evaluation reward: 4.41
 episode: 2464 score: 3.0 memory length: 561251 epsilon:
 0.08672104000850801 steps: 207 lr: 4e-05 evaluation reward: 4.42
 episode: 2465 score: 7.0 memory length: 561643 epsilon:
 0.08594488000850854 steps: 392 lr: 4e-05 evaluation reward: 4.43
 episode: 2466 score: 2.0 memory length: 561840 epsilon:
 0.08555482000850881 steps: 197 lr: 4e-05 evaluation reward: 4.39
 episode: 2467 score: 3.0 memory length: 562068 epsilon:
 0.08510338000850912 steps: 228 lr: 4e-05 evaluation reward: 4.36
 episode: 2468 score: 2.0 memory length: 562249 epsilon:
 0.08474500000850936 steps: 181 lr: 4e-05 evaluation reward: 4.32
 episode: 2469 score: 3.0 memory length: 562479 epsilon:
 0.08428960000850967 steps: 230 lr: 4e-05 evaluation reward: 4.27
 episode: 2470 score: 5.0 memory length: 562825 epsilon:
 0.08360452000851014 steps: 346 lr: 4e-05 evaluation reward: 4.25
 episode: 2471 score: 4.0 memory length: 563084 epsilon:
 0.08309170000851049 steps: 259 lr: 4e-05 evaluation reward: 4.26
 episode: 2472 score: 10.0 memory length: 563546 epsilon:
 0.08217694000851111 steps: 462 lr: 4e-05 evaluation reward: 4.34
 episode: 2473 score: 4.0 memory length: 563820 epsilon:

0.08163442000851148 steps: 274 lr: 4e-05 evaluation reward: 4.35
 episode: 2474 score: 2.0 memory length: 564001 epsilon:
 0.08127604000851173 steps: 181 lr: 4e-05 evaluation reward: 4.33
 episode: 2475 score: 2.0 memory length: 564198 epsilon:
 0.08088598000851199 steps: 197 lr: 4e-05 evaluation reward: 4.29
 episode: 2476 score: 2.0 memory length: 564379 epsilon:
 0.08052760000851224 steps: 181 lr: 4e-05 evaluation reward: 4.29
 episode: 2477 score: 7.0 memory length: 564757 epsilon:
 0.07977916000851275 steps: 378 lr: 4e-05 evaluation reward: 4.31
 episode: 2478 score: 3.0 memory length: 564988 epsilon:
 0.07932178000851306 steps: 231 lr: 4e-05 evaluation reward: 4.3
 episode: 2479 score: 7.0 memory length: 565364 epsilon:
 0.07857730000851357 steps: 376 lr: 4e-05 evaluation reward: 4.35
 episode: 2480 score: 5.0 memory length: 565636 epsilon:
 0.07803874000851393 steps: 272 lr: 4e-05 evaluation reward: 4.35
 episode: 2481 score: 2.0 memory length: 565817 epsilon:
 0.07768036000851418 steps: 181 lr: 4e-05 evaluation reward: 4.34
 episode: 2482 score: 5.0 memory length: 566108 epsilon:
 0.07710418000851457 steps: 291 lr: 4e-05 evaluation reward: 4.35
 episode: 2483 score: 2.0 memory length: 566305 epsilon:
 0.07671412000851484 steps: 197 lr: 4e-05 evaluation reward: 4.35
 episode: 2484 score: 2.0 memory length: 566486 epsilon:
 0.07635574000851508 steps: 181 lr: 4e-05 evaluation reward: 4.35
 episode: 2485 score: 4.0 memory length: 566764 epsilon:
 0.07580530000851546 steps: 278 lr: 4e-05 evaluation reward: 4.34
 episode: 2486 score: 4.0 memory length: 567023 epsilon:
 0.07529248000851581 steps: 259 lr: 4e-05 evaluation reward: 4.34
 episode: 2487 score: 4.0 memory length: 567266 epsilon:
 0.07481134000851614 steps: 243 lr: 4e-05 evaluation reward: 4.33
 episode: 2488 score: 2.0 memory length: 567447 epsilon:
 0.07445296000851638 steps: 181 lr: 4e-05 evaluation reward: 4.31
 episode: 2489 score: 8.0 memory length: 567838 epsilon:
 0.07367878000851691 steps: 391 lr: 4e-05 evaluation reward: 4.33
 episode: 2490 score: 3.0 memory length: 568067 epsilon:
 0.07322536000851722 steps: 229 lr: 4e-05 evaluation reward: 4.31
 episode: 2491 score: 4.0 memory length: 568326 epsilon:
 0.07271254000851757 steps: 259 lr: 4e-05 evaluation reward: 4.33
 episode: 2492 score: 2.0 memory length: 568507 epsilon:
 0.07235416000851781 steps: 181 lr: 4e-05 evaluation reward: 4.33
 episode: 2493 score: 3.0 memory length: 568738 epsilon:
 0.07189678000851812 steps: 231 lr: 4e-05 evaluation reward: 4.32
 episode: 2494 score: 8.0 memory length: 569169 epsilon: 0.0710434000085187
 steps: 431 lr: 4e-05 evaluation reward: 4.34
 episode: 2495 score: 5.0 memory length: 569449 epsilon:
 0.07048900000851908 steps: 280 lr: 4e-05 evaluation reward: 4.35
 episode: 2496 score: 7.0 memory length: 569807 epsilon:
 0.06978016000851957 steps: 358 lr: 4e-05 evaluation reward: 4.36
 episode: 2497 score: 4.0 memory length: 570050 epsilon: 0.0692990200085199

```

steps: 243    lr: 4e-05    evaluation reward: 4.36
episode: 2498  score: 2.0  memory length: 570231  epsilon:
0.06894064000852014  steps: 181  lr: 4e-05    evaluation reward: 4.33
episode: 2499  score: 4.0  memory length: 570490  epsilon:
0.06842782000852049  steps: 259  lr: 4e-05    evaluation reward: 4.28
episode: 2500  score: 3.0  memory length: 570721  epsilon: 0.0679704400085208
steps: 231    lr: 4e-05    evaluation reward: 4.27
episode: 2501  score: 7.0  memory length: 571097  epsilon:
0.06722596000852131  steps: 376  lr: 4e-05    evaluation reward: 4.29
episode: 2502  score: 2.0  memory length: 571278  epsilon:
0.06686758000852155  steps: 181  lr: 4e-05    evaluation reward: 4.26
episode: 2503  score: 2.0  memory length: 571459  epsilon: 0.0665092000085218
steps: 181    lr: 4e-05    evaluation reward: 4.26
episode: 2504  score: 7.0  memory length: 571863  epsilon:
0.06570928000852234  steps: 404  lr: 4e-05    evaluation reward: 4.31
episode: 2505  score: 5.0  memory length: 572143  epsilon:
0.06515488000852272  steps: 280  lr: 4e-05    evaluation reward: 4.34
episode: 2506  score: 2.0  memory length: 572324  epsilon:
0.06479650000852297  steps: 181  lr: 4e-05    evaluation reward: 4.34
episode: 2507  score: 4.0  memory length: 572567  epsilon: 0.0643153600085233
steps: 243    lr: 4e-05    evaluation reward: 4.31
episode: 2508  score: 3.0  memory length: 572792  epsilon: 0.0638698600085236
steps: 225    lr: 4e-05    evaluation reward: 4.28
episode: 2509  score: 6.0  memory length: 573134  epsilon:
0.06319270000852406  steps: 342  lr: 4e-05    evaluation reward: 4.29
episode: 2510  score: 6.0  memory length: 573479  epsilon:
0.06250960000852453  steps: 345  lr: 4e-05    evaluation reward: 4.33
episode: 2511  score: 5.0  memory length: 573773  epsilon:
0.061927480008524924  steps: 294  lr: 4e-05    evaluation reward: 4.34
episode: 2512  score: 4.0  memory length: 574047  epsilon:
0.061384960008525294  steps: 274  lr: 4e-05    evaluation reward: 4.35
episode: 2513  score: 5.0  memory length: 574337  epsilon:
0.060810760008525686  steps: 290  lr: 4e-05    evaluation reward: 4.38
episode: 2514  score: 5.0  memory length: 574644  epsilon: 0.0602029000085261
steps: 307    lr: 4e-05    evaluation reward: 4.37
episode: 2515  score: 7.0  memory length: 575032  epsilon:
0.059434660008526624  steps: 388  lr: 4e-05    evaluation reward: 4.37
episode: 2516  score: 4.0  memory length: 575293  epsilon:
0.05891788000852698  steps: 261  lr: 4e-05    evaluation reward: 4.34
episode: 2517  score: 5.0  memory length: 575616  epsilon:
0.05827834000852741  steps: 323  lr: 4e-05    evaluation reward: 4.35
episode: 2518  score: 2.0  memory length: 575795  epsilon:
0.057923920008527655  steps: 179  lr: 4e-05    evaluation reward: 4.34
episode: 2519  score: 3.0  memory length: 576026  epsilon:
0.057466540008527967  steps: 231  lr: 4e-05    evaluation reward: 4.35
episode: 2520  score: 6.0  memory length: 576324  epsilon:
0.05687650000852837  steps: 298  lr: 4e-05    evaluation reward: 4.39
episode: 2521  score: 7.0  memory length: 576731  epsilon:

```

0.05607064000852892 steps: 407 lr: 4e-05 evaluation reward: 4.41
 episode: 2522 score: 6.0 memory length: 577036 epsilon:
 0.05546674000852933 steps: 305 lr: 4e-05 evaluation reward: 4.43
 episode: 2523 score: 4.0 memory length: 577292 epsilon:
 0.054959860008529676 steps: 256 lr: 4e-05 evaluation reward: 4.44
 episode: 2524 score: 2.0 memory length: 577473 epsilon:
 0.05460148000852992 steps: 181 lr: 4e-05 evaluation reward: 4.42
 episode: 2525 score: 5.0 memory length: 577801 epsilon:
 0.053952040008530364 steps: 328 lr: 4e-05 evaluation reward: 4.4
 episode: 2526 score: 4.0 memory length: 578095 epsilon:
 0.05336992000853076 steps: 294 lr: 4e-05 evaluation reward: 4.42
 episode: 2527 score: 7.0 memory length: 578449 epsilon:
 0.05266900000853124 steps: 354 lr: 4e-05 evaluation reward: 4.41
 episode: 2528 score: 5.0 memory length: 578729 epsilon:
 0.05211460000853162 steps: 280 lr: 4e-05 evaluation reward: 4.42
 episode: 2529 score: 2.0 memory length: 578928 epsilon:
 0.051720580008531886 steps: 199 lr: 4e-05 evaluation reward: 4.39
 episode: 2530 score: 2.0 memory length: 579107 epsilon:
 0.05136616000853213 steps: 179 lr: 4e-05 evaluation reward: 4.38
 episode: 2531 score: 3.0 memory length: 579317 epsilon:
 0.05095036000853241 steps: 210 lr: 4e-05 evaluation reward: 4.35
 episode: 2532 score: 3.0 memory length: 579566 epsilon:
 0.05045734000853275 steps: 249 lr: 4e-05 evaluation reward: 4.36
 episode: 2533 score: 3.0 memory length: 579795 epsilon:
 0.050003920008533057 steps: 229 lr: 4e-05 evaluation reward: 4.37
 episode: 2534 score: 2.0 memory length: 579976 epsilon: 0.0496455400085333
 steps: 181 lr: 4e-05 evaluation reward: 4.36
 episode: 2535 score: 4.0 memory length: 580257 epsilon:
 0.04908916000853368 steps: 281 lr: 4e-05 evaluation reward: 4.36
 episode: 2536 score: 2.0 memory length: 580438 epsilon:
 0.048730780008533925 steps: 181 lr: 4e-05 evaluation reward: 4.31
 episode: 2537 score: 3.0 memory length: 580685 epsilon:
 0.04824172000853426 steps: 247 lr: 4e-05 evaluation reward: 4.3
 episode: 2538 score: 4.0 memory length: 580945 epsilon:
 0.04772692000853461 steps: 260 lr: 4e-05 evaluation reward: 4.32
 episode: 2539 score: 3.0 memory length: 581172 epsilon:
 0.047277460008534916 steps: 227 lr: 4e-05 evaluation reward: 4.31
 episode: 2540 score: 3.0 memory length: 581421 epsilon:
 0.04678444000853525 steps: 249 lr: 4e-05 evaluation reward: 4.29
 episode: 2541 score: 3.0 memory length: 581668 epsilon:
 0.046295380008535586 steps: 247 lr: 4e-05 evaluation reward: 4.27
 episode: 2542 score: 3.0 memory length: 581877 epsilon:
 0.04588156000853587 steps: 209 lr: 4e-05 evaluation reward: 4.26
 episode: 2543 score: 4.0 memory length: 582120 epsilon:
 0.045400420008536196 steps: 243 lr: 4e-05 evaluation reward: 4.24
 episode: 2544 score: 9.0 memory length: 582579 epsilon:
 0.044491600008536816 steps: 459 lr: 4e-05 evaluation reward: 4.29
 episode: 2545 score: 4.0 memory length: 582855 epsilon:

0.04394512000853719 steps: 276 lr: 4e-05 evaluation reward: 4.3
 episode: 2546 score: 3.0 memory length: 583086 epsilon: 0.0434877400085375
 steps: 231 lr: 4e-05 evaluation reward: 4.25
 episode: 2547 score: 2.0 memory length: 583267 epsilon:
 0.043129360008537745 steps: 181 lr: 4e-05 evaluation reward: 4.23
 episode: 2548 score: 3.0 memory length: 583495 epsilon:
 0.04267792000853805 steps: 228 lr: 4e-05 evaluation reward: 4.18
 episode: 2549 score: 2.0 memory length: 583676 epsilon: 0.0423195400085383
 steps: 181 lr: 4e-05 evaluation reward: 4.16
 episode: 2550 score: 2.0 memory length: 583857 epsilon:
 0.04196116000853854 steps: 181 lr: 4e-05 evaluation reward: 4.12
 episode: 2551 score: 4.0 memory length: 584135 epsilon:
 0.04141072000853892 steps: 278 lr: 4e-05 evaluation reward: 4.05
 episode: 2552 score: 3.0 memory length: 584364 epsilon:
 0.04095730000853923 steps: 229 lr: 4e-05 evaluation reward: 4.06
 episode: 2553 score: 2.0 memory length: 584544 epsilon:
 0.04060090000853947 steps: 180 lr: 4e-05 evaluation reward: 4.02
 episode: 2554 score: 3.0 memory length: 584793 epsilon:
 0.040107880008539806 steps: 249 lr: 4e-05 evaluation reward: 4.03
 episode: 2555 score: 5.0 memory length: 585073 epsilon:
 0.039553480008540184 steps: 280 lr: 4e-05 evaluation reward: 4.01
 episode: 2556 score: 4.0 memory length: 585331 epsilon:
 0.03904264000854053 steps: 258 lr: 4e-05 evaluation reward: 4.0
 episode: 2557 score: 6.0 memory length: 585636 epsilon:
 0.038438740008540945 steps: 305 lr: 4e-05 evaluation reward: 3.98
 episode: 2558 score: 2.0 memory length: 585817 epsilon:
 0.03808036000854119 steps: 181 lr: 4e-05 evaluation reward: 3.97
 episode: 2559 score: 8.0 memory length: 586252 epsilon:
 0.03721906000854178 steps: 435 lr: 4e-05 evaluation reward: 4.02
 episode: 2560 score: 6.0 memory length: 586573 epsilon:
 0.03658348000854221 steps: 321 lr: 4e-05 evaluation reward: 4.06
 episode: 2561 score: 7.0 memory length: 586925 epsilon:
 0.035886520008542686 steps: 352 lr: 4e-05 evaluation reward: 4.07
 episode: 2562 score: 2.0 memory length: 587106 epsilon:
 0.03552814000854293 steps: 181 lr: 4e-05 evaluation reward: 4.05
 episode: 2563 score: 2.0 memory length: 587287 epsilon:
 0.035169760008543174 steps: 181 lr: 4e-05 evaluation reward: 4.02
 episode: 2564 score: 3.0 memory length: 587515 epsilon:
 0.03471832000854348 steps: 228 lr: 4e-05 evaluation reward: 4.02
 episode: 2565 score: 5.0 memory length: 587795 epsilon:
 0.03416392000854386 steps: 280 lr: 4e-05 evaluation reward: 4.0
 episode: 2566 score: 5.0 memory length: 588075 epsilon:
 0.03360952000854424 steps: 280 lr: 4e-05 evaluation reward: 4.03
 episode: 2567 score: 8.0 memory length: 588460 epsilon:
 0.03284722000854476 steps: 385 lr: 4e-05 evaluation reward: 4.08
 episode: 2568 score: 5.0 memory length: 588740 epsilon:
 0.03229282000854514 steps: 280 lr: 4e-05 evaluation reward: 4.11
 episode: 2569 score: 2.0 memory length: 588921 epsilon:

0.03193444000854538 steps: 181 lr: 4e-05 evaluation reward: 4.1
 episode: 2570 score: 2.0 memory length: 589102 epsilon:
 0.031576060008545626 steps: 181 lr: 4e-05 evaluation reward: 4.07
 episode: 2571 score: 6.0 memory length: 589429 epsilon:
 0.030928600008546067 steps: 327 lr: 4e-05 evaluation reward: 4.09
 episode: 2572 score: 5.0 memory length: 589741 epsilon:
 0.03031084000854649 steps: 312 lr: 4e-05 evaluation reward: 4.04
 episode: 2573 score: 5.0 memory length: 590021 epsilon:
 0.029756440008546867 steps: 280 lr: 4e-05 evaluation reward: 4.05
 episode: 2574 score: 3.0 memory length: 590231 epsilon:
 0.02934064000854715 steps: 210 lr: 4e-05 evaluation reward: 4.06
 episode: 2575 score: 7.0 memory length: 590591 epsilon:
 0.028627840008547636 steps: 360 lr: 4e-05 evaluation reward: 4.11
 episode: 2576 score: 7.0 memory length: 590943 epsilon:
 0.027930880008548112 steps: 352 lr: 4e-05 evaluation reward: 4.16
 episode: 2577 score: 2.0 memory length: 591124 epsilon:
 0.027572500008548356 steps: 181 lr: 4e-05 evaluation reward: 4.11
 episode: 2578 score: 5.0 memory length: 591420 epsilon:
 0.026986420008548756 steps: 296 lr: 4e-05 evaluation reward: 4.13
 episode: 2579 score: 3.0 memory length: 591669 epsilon:
 0.026493400008549092 steps: 249 lr: 4e-05 evaluation reward: 4.09
 episode: 2580 score: 3.0 memory length: 591918 epsilon:
 0.02600038000854943 steps: 249 lr: 4e-05 evaluation reward: 4.07
 episode: 2581 score: 6.0 memory length: 592247 epsilon:
 0.025348960008549873 steps: 329 lr: 4e-05 evaluation reward: 4.11
 episode: 2582 score: 5.0 memory length: 592527 epsilon:
 0.02479456000855025 steps: 280 lr: 4e-05 evaluation reward: 4.11
 episode: 2583 score: 9.0 memory length: 592945 epsilon:
 0.023966920008550816 steps: 418 lr: 4e-05 evaluation reward: 4.18
 episode: 2584 score: 7.0 memory length: 593307 epsilon:
 0.023250160008551304 steps: 362 lr: 4e-05 evaluation reward: 4.23
 episode: 2585 score: 5.0 memory length: 593606 epsilon:
 0.022658140008551708 steps: 299 lr: 4e-05 evaluation reward: 4.24
 episode: 2586 score: 4.0 memory length: 593868 epsilon:
 0.022139380008552062 steps: 262 lr: 4e-05 evaluation reward: 4.24
 episode: 2587 score: 4.0 memory length: 594164 epsilon:
 0.021553300008552462 steps: 296 lr: 4e-05 evaluation reward: 4.24
 episode: 2588 score: 5.0 memory length: 594452 epsilon:
 0.02098306000855285 steps: 288 lr: 4e-05 evaluation reward: 4.27
 episode: 2589 score: 3.0 memory length: 594681 epsilon:
 0.02052964000855316 steps: 229 lr: 4e-05 evaluation reward: 4.22
 episode: 2590 score: 6.0 memory length: 595008 epsilon: 0.0198821800085536
 steps: 327 lr: 4e-05 evaluation reward: 4.25
 episode: 2591 score: 3.0 memory length: 595239 epsilon:
 0.019424800008553914 steps: 231 lr: 4e-05 evaluation reward: 4.24
 episode: 2592 score: 7.0 memory length: 595597 epsilon:
 0.018715960008554397 steps: 358 lr: 4e-05 evaluation reward: 4.29
 episode: 2593 score: 7.0 memory length: 595975 epsilon:

0.017967520008554907 steps: 378 lr: 4e-05 evaluation reward: 4.33
 episode: 2594 score: 6.0 memory length: 596333 epsilon:
 0.01725868000855539 steps: 358 lr: 4e-05 evaluation reward: 4.31
 episode: 2595 score: 6.0 memory length: 596660 epsilon:
 0.016611220008555833 steps: 327 lr: 4e-05 evaluation reward: 4.32
 episode: 2596 score: 4.0 memory length: 596934 epsilon:
 0.016068700008556203 steps: 274 lr: 4e-05 evaluation reward: 4.29
 episode: 2597 score: 4.0 memory length: 597197 epsilon:
 0.01554796000855649 steps: 263 lr: 4e-05 evaluation reward: 4.29
 episode: 2598 score: 4.0 memory length: 597456 epsilon:
 0.01503514000855639 steps: 259 lr: 4e-05 evaluation reward: 4.31
 episode: 2599 score: 10.0 memory length: 597958 epsilon:
 0.014041180008556198 steps: 502 lr: 4e-05 evaluation reward: 4.37
 episode: 2600 score: 7.0 memory length: 598313 epsilon:
 0.013338280008556061 steps: 355 lr: 4e-05 evaluation reward: 4.41
 episode: 2601 score: 5.0 memory length: 598593 epsilon:
 0.012783880008555954 steps: 280 lr: 4e-05 evaluation reward: 4.39
 episode: 2602 score: 8.0 memory length: 598952 epsilon:
 0.012073060008555816 steps: 359 lr: 4e-05 evaluation reward: 4.45
 episode: 2603 score: 6.0 memory length: 599257 epsilon:
 0.011469160008555699 steps: 305 lr: 4e-05 evaluation reward: 4.49
 episode: 2604 score: 5.0 memory length: 599553 epsilon:
 0.010883080008555585 steps: 296 lr: 4e-05 evaluation reward: 4.47
 episode: 2605 score: 2.0 memory length: 599734 epsilon:
 0.010524700008555515 steps: 181 lr: 4e-05 evaluation reward: 4.44
 episode: 2606 score: 3.0 memory length: 599964 epsilon:
 0.010069300008555427 steps: 230 lr: 4e-05 evaluation reward: 4.45
 episode: 2607 score: 6.0 memory length: 600318 epsilon:
 0.009998020008555413 steps: 354 lr: 4e-05 evaluation reward: 4.47
 episode: 2608 score: 3.0 memory length: 600567 epsilon:
 0.009998020008555413 steps: 249 lr: 4e-05 evaluation reward: 4.47
 episode: 2609 score: 6.0 memory length: 600894 epsilon:
 0.009998020008555413 steps: 327 lr: 4e-05 evaluation reward: 4.47
 episode: 2610 score: 3.0 memory length: 601122 epsilon:
 0.009998020008555413 steps: 228 lr: 4e-05 evaluation reward: 4.44
 episode: 2611 score: 6.0 memory length: 601482 epsilon:
 0.009998020008555413 steps: 360 lr: 4e-05 evaluation reward: 4.45
 episode: 2612 score: 4.0 memory length: 601744 epsilon:
 0.009998020008555413 steps: 262 lr: 4e-05 evaluation reward: 4.45
 episode: 2613 score: 4.0 memory length: 602002 epsilon:
 0.009998020008555413 steps: 258 lr: 4e-05 evaluation reward: 4.44
 episode: 2614 score: 4.0 memory length: 602279 epsilon:
 0.009998020008555413 steps: 277 lr: 4e-05 evaluation reward: 4.43
 episode: 2615 score: 4.0 memory length: 602535 epsilon:
 0.009998020008555413 steps: 256 lr: 4e-05 evaluation reward: 4.4
 episode: 2616 score: 3.0 memory length: 602766 epsilon:
 0.009998020008555413 steps: 231 lr: 4e-05 evaluation reward: 4.39
 episode: 2617 score: 3.0 memory length: 602978 epsilon:

| | | | |
|----------------------|------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 212 | lr: 4e-05 | evaluation reward: 4.37 |
| episode: 2618 | score: 6.0 | memory length: 603312 | epsilon: |
| 0.009998020008555413 | steps: 334 | lr: 4e-05 | evaluation reward: 4.41 |
| episode: 2619 | score: 5.0 | memory length: 603622 | epsilon: |
| 0.009998020008555413 | steps: 310 | lr: 4e-05 | evaluation reward: 4.43 |
| episode: 2620 | score: 3.0 | memory length: 603850 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.4 |
| episode: 2621 | score: 4.0 | memory length: 604128 | epsilon: |
| 0.009998020008555413 | steps: 278 | lr: 4e-05 | evaluation reward: 4.37 |
| episode: 2622 | score: 6.0 | memory length: 604457 | epsilon: |
| 0.009998020008555413 | steps: 329 | lr: 4e-05 | evaluation reward: 4.37 |
| episode: 2623 | score: 7.0 | memory length: 604814 | epsilon: |
| 0.009998020008555413 | steps: 357 | lr: 4e-05 | evaluation reward: 4.4 |
| episode: 2624 | score: 5.0 | memory length: 605126 | epsilon: |
| 0.009998020008555413 | steps: 312 | lr: 4e-05 | evaluation reward: 4.43 |
| episode: 2625 | score: 4.0 | memory length: 605400 | epsilon: |
| 0.009998020008555413 | steps: 274 | lr: 4e-05 | evaluation reward: 4.42 |
| episode: 2626 | score: 3.0 | memory length: 605647 | epsilon: |
| 0.009998020008555413 | steps: 247 | lr: 4e-05 | evaluation reward: 4.41 |
| episode: 2627 | score: 3.0 | memory length: 605878 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.37 |
| episode: 2628 | score: 6.0 | memory length: 606235 | epsilon: |
| 0.009998020008555413 | steps: 357 | lr: 4e-05 | evaluation reward: 4.38 |
| episode: 2629 | score: 6.0 | memory length: 606578 | epsilon: |
| 0.009998020008555413 | steps: 343 | lr: 4e-05 | evaluation reward: 4.42 |
| episode: 2630 | score: 4.0 | memory length: 606852 | epsilon: |
| 0.009998020008555413 | steps: 274 | lr: 4e-05 | evaluation reward: 4.44 |
| episode: 2631 | score: 6.0 | memory length: 607157 | epsilon: |
| 0.009998020008555413 | steps: 305 | lr: 4e-05 | evaluation reward: 4.47 |
| episode: 2632 | score: 5.0 | memory length: 607437 | epsilon: |
| 0.009998020008555413 | steps: 280 | lr: 4e-05 | evaluation reward: 4.49 |
| episode: 2633 | score: 6.0 | memory length: 607764 | epsilon: |
| 0.009998020008555413 | steps: 327 | lr: 4e-05 | evaluation reward: 4.52 |
| episode: 2634 | score: 9.0 | memory length: 608266 | epsilon: |
| 0.009998020008555413 | steps: 502 | lr: 4e-05 | evaluation reward: 4.59 |
| episode: 2635 | score: 3.0 | memory length: 608494 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.58 |
| episode: 2636 | score: 7.0 | memory length: 608865 | epsilon: |
| 0.009998020008555413 | steps: 371 | lr: 4e-05 | evaluation reward: 4.63 |
| episode: 2637 | score: 3.0 | memory length: 609093 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.63 |
| episode: 2638 | score: 3.0 | memory length: 609303 | epsilon: |
| 0.009998020008555413 | steps: 210 | lr: 4e-05 | evaluation reward: 4.62 |
| episode: 2639 | score: 4.0 | memory length: 609577 | epsilon: |
| 0.009998020008555413 | steps: 274 | lr: 4e-05 | evaluation reward: 4.63 |
| episode: 2640 | score: 4.0 | memory length: 609838 | epsilon: |
| 0.009998020008555413 | steps: 261 | lr: 4e-05 | evaluation reward: 4.64 |
| episode: 2641 | score: 4.0 | memory length: 610098 | epsilon: |

| | | | |
|----------------------|-------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 260 | lr: 4e-05 | evaluation reward: 4.65 |
| episode: 2642 | score: 6.0 | memory length: 610440 | epsilon: |
| 0.009998020008555413 | steps: 342 | lr: 4e-05 | evaluation reward: 4.68 |
| episode: 2643 | score: 6.0 | memory length: 610794 | epsilon: |
| 0.009998020008555413 | steps: 354 | lr: 4e-05 | evaluation reward: 4.7 |
| episode: 2644 | score: 3.0 | memory length: 611041 | epsilon: |
| 0.009998020008555413 | steps: 247 | lr: 4e-05 | evaluation reward: 4.64 |
| episode: 2645 | score: 6.0 | memory length: 611383 | epsilon: |
| 0.009998020008555413 | steps: 342 | lr: 4e-05 | evaluation reward: 4.66 |
| episode: 2646 | score: 5.0 | memory length: 611710 | epsilon: |
| 0.009998020008555413 | steps: 327 | lr: 4e-05 | evaluation reward: 4.68 |
| episode: 2647 | score: 4.0 | memory length: 611988 | epsilon: |
| 0.009998020008555413 | steps: 278 | lr: 4e-05 | evaluation reward: 4.7 |
| episode: 2648 | score: 10.0 | memory length: 612368 | epsilon: |
| 0.009998020008555413 | steps: 380 | lr: 4e-05 | evaluation reward: 4.77 |
| episode: 2649 | score: 3.0 | memory length: 612599 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.78 |
| episode: 2650 | score: 3.0 | memory length: 612830 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.79 |
| episode: 2651 | score: 4.0 | memory length: 613091 | epsilon: |
| 0.009998020008555413 | steps: 261 | lr: 4e-05 | evaluation reward: 4.79 |
| episode: 2652 | score: 4.0 | memory length: 613389 | epsilon: |
| 0.009998020008555413 | steps: 298 | lr: 4e-05 | evaluation reward: 4.8 |
| episode: 2653 | score: 6.0 | memory length: 613731 | epsilon: |
| 0.009998020008555413 | steps: 342 | lr: 4e-05 | evaluation reward: 4.84 |
| episode: 2654 | score: 3.0 | memory length: 613960 | epsilon: |
| 0.009998020008555413 | steps: 229 | lr: 4e-05 | evaluation reward: 4.84 |
| episode: 2655 | score: 4.0 | memory length: 614234 | epsilon: |
| 0.009998020008555413 | steps: 274 | lr: 4e-05 | evaluation reward: 4.83 |
| episode: 2656 | score: 6.0 | memory length: 614576 | epsilon: |
| 0.009998020008555413 | steps: 342 | lr: 4e-05 | evaluation reward: 4.85 |
| episode: 2657 | score: 6.0 | memory length: 614934 | epsilon: |
| 0.009998020008555413 | steps: 358 | lr: 4e-05 | evaluation reward: 4.85 |
| episode: 2658 | score: 2.0 | memory length: 615115 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.85 |
| episode: 2659 | score: 3.0 | memory length: 615327 | epsilon: |
| 0.009998020008555413 | steps: 212 | lr: 4e-05 | evaluation reward: 4.8 |
| episode: 2660 | score: 4.0 | memory length: 615581 | epsilon: |
| 0.009998020008555413 | steps: 254 | lr: 4e-05 | evaluation reward: 4.78 |
| episode: 2661 | score: 4.0 | memory length: 615859 | epsilon: |
| 0.009998020008555413 | steps: 278 | lr: 4e-05 | evaluation reward: 4.75 |
| episode: 2662 | score: 2.0 | memory length: 616040 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.75 |
| episode: 2663 | score: 6.0 | memory length: 616367 | epsilon: |
| 0.009998020008555413 | steps: 327 | lr: 4e-05 | evaluation reward: 4.79 |
| episode: 2664 | score: 2.0 | memory length: 616548 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.78 |
| episode: 2665 | score: 3.0 | memory length: 616776 | epsilon: |

| | | | |
|----------------------|------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.76 |
| episode: 2666 | score: 4.0 | memory length: 617074 | epsilon: |
| 0.009998020008555413 | steps: 298 | lr: 4e-05 | evaluation reward: 4.75 |
| episode: 2667 | score: 2.0 | memory length: 617255 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.69 |
| episode: 2668 | score: 5.0 | memory length: 617582 | epsilon: |
| 0.009998020008555413 | steps: 327 | lr: 4e-05 | evaluation reward: 4.69 |
| episode: 2669 | score: 6.0 | memory length: 617927 | epsilon: |
| 0.009998020008555413 | steps: 345 | lr: 4e-05 | evaluation reward: 4.73 |
| episode: 2670 | score: 5.0 | memory length: 618253 | epsilon: |
| 0.009998020008555413 | steps: 326 | lr: 4e-05 | evaluation reward: 4.76 |
| episode: 2671 | score: 7.0 | memory length: 618636 | epsilon: |
| 0.009998020008555413 | steps: 383 | lr: 4e-05 | evaluation reward: 4.77 |
| episode: 2672 | score: 6.0 | memory length: 618947 | epsilon: |
| 0.009998020008555413 | steps: 311 | lr: 4e-05 | evaluation reward: 4.78 |
| episode: 2673 | score: 2.0 | memory length: 619128 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.75 |
| episode: 2674 | score: 6.0 | memory length: 619486 | epsilon: |
| 0.009998020008555413 | steps: 358 | lr: 4e-05 | evaluation reward: 4.78 |
| episode: 2675 | score: 6.0 | memory length: 619846 | epsilon: |
| 0.009998020008555413 | steps: 360 | lr: 4e-05 | evaluation reward: 4.77 |
| episode: 2676 | score: 7.0 | memory length: 620236 | epsilon: |
| 0.009998020008555413 | steps: 390 | lr: 4e-05 | evaluation reward: 4.77 |
| episode: 2677 | score: 4.0 | memory length: 620495 | epsilon: |
| 0.009998020008555413 | steps: 259 | lr: 4e-05 | evaluation reward: 4.79 |
| episode: 2678 | score: 2.0 | memory length: 620676 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.76 |
| episode: 2679 | score: 8.0 | memory length: 621056 | epsilon: |
| 0.009998020008555413 | steps: 380 | lr: 4e-05 | evaluation reward: 4.81 |
| episode: 2680 | score: 4.0 | memory length: 621336 | epsilon: |
| 0.009998020008555413 | steps: 280 | lr: 4e-05 | evaluation reward: 4.82 |
| episode: 2681 | score: 5.0 | memory length: 621663 | epsilon: |
| 0.009998020008555413 | steps: 327 | lr: 4e-05 | evaluation reward: 4.81 |
| episode: 2682 | score: 5.0 | memory length: 621987 | epsilon: |
| 0.009998020008555413 | steps: 324 | lr: 4e-05 | evaluation reward: 4.81 |
| episode: 2683 | score: 7.0 | memory length: 622400 | epsilon: |
| 0.009998020008555413 | steps: 413 | lr: 4e-05 | evaluation reward: 4.79 |
| episode: 2684 | score: 4.0 | memory length: 622678 | epsilon: |
| 0.009998020008555413 | steps: 278 | lr: 4e-05 | evaluation reward: 4.76 |
| episode: 2685 | score: 3.0 | memory length: 622890 | epsilon: |
| 0.009998020008555413 | steps: 212 | lr: 4e-05 | evaluation reward: 4.74 |
| episode: 2686 | score: 5.0 | memory length: 623220 | epsilon: |
| 0.009998020008555413 | steps: 330 | lr: 4e-05 | evaluation reward: 4.75 |
| episode: 2687 | score: 4.0 | memory length: 623480 | epsilon: |
| 0.009998020008555413 | steps: 260 | lr: 4e-05 | evaluation reward: 4.75 |
| episode: 2688 | score: 3.0 | memory length: 623708 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.73 |
| episode: 2689 | score: 3.0 | memory length: 623957 | epsilon: |

| | | | |
|----------------------|------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 249 | lr: 4e-05 | evaluation reward: 4.73 |
| episode: 2690 | score: 5.0 | memory length: 624237 | epsilon: |
| 0.009998020008555413 | steps: 280 | lr: 4e-05 | evaluation reward: 4.72 |
| episode: 2691 | score: 3.0 | memory length: 624465 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.72 |
| episode: 2692 | score: 4.0 | memory length: 624743 | epsilon: |
| 0.009998020008555413 | steps: 278 | lr: 4e-05 | evaluation reward: 4.69 |
| episode: 2693 | score: 4.0 | memory length: 625003 | epsilon: |
| 0.009998020008555413 | steps: 260 | lr: 4e-05 | evaluation reward: 4.66 |
| episode: 2694 | score: 5.0 | memory length: 625331 | epsilon: |
| 0.009998020008555413 | steps: 328 | lr: 4e-05 | evaluation reward: 4.65 |
| episode: 2695 | score: 2.0 | memory length: 625512 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.61 |
| episode: 2696 | score: 4.0 | memory length: 625791 | epsilon: |
| 0.009998020008555413 | steps: 279 | lr: 4e-05 | evaluation reward: 4.61 |
| episode: 2697 | score: 3.0 | memory length: 626022 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.6 |
| episode: 2698 | score: 5.0 | memory length: 626368 | epsilon: |
| 0.009998020008555413 | steps: 346 | lr: 4e-05 | evaluation reward: 4.61 |
| episode: 2699 | score: 3.0 | memory length: 626599 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.54 |
| episode: 2700 | score: 2.0 | memory length: 626780 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.49 |
| episode: 2701 | score: 4.0 | memory length: 627078 | epsilon: |
| 0.009998020008555413 | steps: 298 | lr: 4e-05 | evaluation reward: 4.48 |
| episode: 2702 | score: 4.0 | memory length: 627375 | epsilon: |
| 0.009998020008555413 | steps: 297 | lr: 4e-05 | evaluation reward: 4.44 |
| episode: 2703 | score: 8.0 | memory length: 627790 | epsilon: |
| 0.009998020008555413 | steps: 415 | lr: 4e-05 | evaluation reward: 4.46 |
| episode: 2704 | score: 4.0 | memory length: 628086 | epsilon: |
| 0.009998020008555413 | steps: 296 | lr: 4e-05 | evaluation reward: 4.45 |
| episode: 2705 | score: 2.0 | memory length: 628267 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.45 |
| episode: 2706 | score: 4.0 | memory length: 628564 | epsilon: |
| 0.009998020008555413 | steps: 297 | lr: 4e-05 | evaluation reward: 4.46 |
| episode: 2707 | score: 4.0 | memory length: 628861 | epsilon: |
| 0.009998020008555413 | steps: 297 | lr: 4e-05 | evaluation reward: 4.44 |
| episode: 2708 | score: 3.0 | memory length: 629089 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.44 |
| episode: 2709 | score: 5.0 | memory length: 629397 | epsilon: |
| 0.009998020008555413 | steps: 308 | lr: 4e-05 | evaluation reward: 4.43 |
| episode: 2710 | score: 4.0 | memory length: 629674 | epsilon: |
| 0.009998020008555413 | steps: 277 | lr: 4e-05 | evaluation reward: 4.44 |
| episode: 2711 | score: 2.0 | memory length: 629855 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.4 |
| episode: 2712 | score: 2.0 | memory length: 630036 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.38 |
| episode: 2713 | score: 5.0 | memory length: 630361 | epsilon: |

| | | | |
|----------------------|-------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 325 | lr: 4e-05 | evaluation reward: 4.39 |
| episode: 2714 | score: 3.0 | memory length: 630591 | epsilon: |
| 0.009998020008555413 | steps: 230 | lr: 4e-05 | evaluation reward: 4.38 |
| episode: 2715 | score: 3.0 | memory length: 630819 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.37 |
| episode: 2716 | score: 4.0 | memory length: 631080 | epsilon: |
| 0.009998020008555413 | steps: 261 | lr: 4e-05 | evaluation reward: 4.38 |
| episode: 2717 | score: 6.0 | memory length: 631439 | epsilon: |
| 0.009998020008555413 | steps: 359 | lr: 4e-05 | evaluation reward: 4.41 |
| episode: 2718 | score: 2.0 | memory length: 631620 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.37 |
| episode: 2719 | score: 2.0 | memory length: 631801 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.34 |
| episode: 2720 | score: 2.0 | memory length: 631982 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.33 |
| episode: 2721 | score: 4.0 | memory length: 632236 | epsilon: |
| 0.009998020008555413 | steps: 254 | lr: 4e-05 | evaluation reward: 4.33 |
| episode: 2722 | score: 2.0 | memory length: 632417 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.29 |
| episode: 2723 | score: 5.0 | memory length: 632704 | epsilon: |
| 0.009998020008555413 | steps: 287 | lr: 4e-05 | evaluation reward: 4.27 |
| episode: 2724 | score: 5.0 | memory length: 633013 | epsilon: |
| 0.009998020008555413 | steps: 309 | lr: 4e-05 | evaluation reward: 4.27 |
| episode: 2725 | score: 10.0 | memory length: 633454 | epsilon: |
| 0.009998020008555413 | steps: 441 | lr: 4e-05 | evaluation reward: 4.33 |
| episode: 2726 | score: 2.0 | memory length: 633635 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.32 |
| episode: 2727 | score: 3.0 | memory length: 633866 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.32 |
| episode: 2728 | score: 3.0 | memory length: 634097 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.29 |
| episode: 2729 | score: 7.0 | memory length: 634490 | epsilon: |
| 0.009998020008555413 | steps: 393 | lr: 4e-05 | evaluation reward: 4.3 |
| episode: 2730 | score: 3.0 | memory length: 634738 | epsilon: |
| 0.009998020008555413 | steps: 248 | lr: 4e-05 | evaluation reward: 4.29 |
| episode: 2731 | score: 9.0 | memory length: 635208 | epsilon: |
| 0.009998020008555413 | steps: 470 | lr: 4e-05 | evaluation reward: 4.32 |
| episode: 2732 | score: 3.0 | memory length: 635439 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.3 |
| episode: 2733 | score: 9.0 | memory length: 635884 | epsilon: |
| 0.009998020008555413 | steps: 445 | lr: 4e-05 | evaluation reward: 4.33 |
| episode: 2734 | score: 5.0 | memory length: 636182 | epsilon: |
| 0.009998020008555413 | steps: 298 | lr: 4e-05 | evaluation reward: 4.29 |
| episode: 2735 | score: 5.0 | memory length: 636510 | epsilon: |
| 0.009998020008555413 | steps: 328 | lr: 4e-05 | evaluation reward: 4.31 |
| episode: 2736 | score: 3.0 | memory length: 636757 | epsilon: |
| 0.009998020008555413 | steps: 247 | lr: 4e-05 | evaluation reward: 4.27 |
| episode: 2737 | score: 7.0 | memory length: 637141 | epsilon: |

| | | | |
|----------------------|------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 384 | lr: 4e-05 | evaluation reward: 4.31 |
| episode: 2738 | score: 3.0 | memory length: 637351 | epsilon: |
| 0.009998020008555413 | steps: 210 | lr: 4e-05 | evaluation reward: 4.31 |
| episode: 2739 | score: 5.0 | memory length: 637642 | epsilon: |
| 0.009998020008555413 | steps: 291 | lr: 4e-05 | evaluation reward: 4.32 |
| episode: 2740 | score: 2.0 | memory length: 637823 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.3 |
| episode: 2741 | score: 2.0 | memory length: 638004 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.28 |
| episode: 2742 | score: 2.0 | memory length: 638185 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.24 |
| episode: 2743 | score: 3.0 | memory length: 638414 | epsilon: |
| 0.009998020008555413 | steps: 229 | lr: 4e-05 | evaluation reward: 4.21 |
| episode: 2744 | score: 2.0 | memory length: 638595 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.2 |
| episode: 2745 | score: 3.0 | memory length: 638822 | epsilon: |
| 0.009998020008555413 | steps: 227 | lr: 4e-05 | evaluation reward: 4.17 |
| episode: 2746 | score: 7.0 | memory length: 639212 | epsilon: |
| 0.009998020008555413 | steps: 390 | lr: 4e-05 | evaluation reward: 4.19 |
| episode: 2747 | score: 3.0 | memory length: 639443 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.18 |
| episode: 2748 | score: 3.0 | memory length: 639674 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.11 |
| episode: 2749 | score: 3.0 | memory length: 639901 | epsilon: |
| 0.009998020008555413 | steps: 227 | lr: 4e-05 | evaluation reward: 4.11 |
| episode: 2750 | score: 3.0 | memory length: 640132 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.11 |
| episode: 2751 | score: 4.0 | memory length: 640406 | epsilon: |
| 0.009998020008555413 | steps: 274 | lr: 4e-05 | evaluation reward: 4.11 |
| episode: 2752 | score: 6.0 | memory length: 640766 | epsilon: |
| 0.009998020008555413 | steps: 360 | lr: 4e-05 | evaluation reward: 4.13 |
| episode: 2753 | score: 3.0 | memory length: 640997 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.1 |
| episode: 2754 | score: 7.0 | memory length: 641370 | epsilon: |
| 0.009998020008555413 | steps: 373 | lr: 4e-05 | evaluation reward: 4.14 |
| episode: 2755 | score: 4.0 | memory length: 641650 | epsilon: |
| 0.009998020008555413 | steps: 280 | lr: 4e-05 | evaluation reward: 4.14 |
| episode: 2756 | score: 3.0 | memory length: 641881 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.11 |
| episode: 2757 | score: 4.0 | memory length: 642135 | epsilon: |
| 0.009998020008555413 | steps: 254 | lr: 4e-05 | evaluation reward: 4.09 |
| episode: 2758 | score: 8.0 | memory length: 642592 | epsilon: |
| 0.009998020008555413 | steps: 457 | lr: 4e-05 | evaluation reward: 4.15 |
| episode: 2759 | score: 6.0 | memory length: 642934 | epsilon: |
| 0.009998020008555413 | steps: 342 | lr: 4e-05 | evaluation reward: 4.18 |
| episode: 2760 | score: 3.0 | memory length: 643164 | epsilon: |
| 0.009998020008555413 | steps: 230 | lr: 4e-05 | evaluation reward: 4.17 |
| episode: 2761 | score: 3.0 | memory length: 643395 | epsilon: |

| | | | |
|----------------------|-------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.16 |
| episode: 2762 | score: 3.0 | memory length: 643624 | epsilon: |
| 0.009998020008555413 | steps: 229 | lr: 4e-05 | evaluation reward: 4.17 |
| episode: 2763 | score: 4.0 | memory length: 643885 | epsilon: |
| 0.009998020008555413 | steps: 261 | lr: 4e-05 | evaluation reward: 4.15 |
| episode: 2764 | score: 3.0 | memory length: 644113 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.16 |
| episode: 2765 | score: 5.0 | memory length: 644421 | epsilon: |
| 0.009998020008555413 | steps: 308 | lr: 4e-05 | evaluation reward: 4.18 |
| episode: 2766 | score: 3.0 | memory length: 644670 | epsilon: |
| 0.009998020008555413 | steps: 249 | lr: 4e-05 | evaluation reward: 4.17 |
| episode: 2767 | score: 8.0 | memory length: 644983 | epsilon: |
| 0.009998020008555413 | steps: 313 | lr: 4e-05 | evaluation reward: 4.23 |
| episode: 2768 | score: 6.0 | memory length: 645336 | epsilon: |
| 0.009998020008555413 | steps: 353 | lr: 4e-05 | evaluation reward: 4.24 |
| episode: 2769 | score: 2.0 | memory length: 645517 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.2 |
| episode: 2770 | score: 4.0 | memory length: 645797 | epsilon: |
| 0.009998020008555413 | steps: 280 | lr: 4e-05 | evaluation reward: 4.19 |
| episode: 2771 | score: 2.0 | memory length: 645978 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.14 |
| episode: 2772 | score: 2.0 | memory length: 646159 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.1 |
| episode: 2773 | score: 3.0 | memory length: 646389 | epsilon: |
| 0.009998020008555413 | steps: 230 | lr: 4e-05 | evaluation reward: 4.11 |
| episode: 2774 | score: 4.0 | memory length: 646648 | epsilon: |
| 0.009998020008555413 | steps: 259 | lr: 4e-05 | evaluation reward: 4.09 |
| episode: 2775 | score: 3.0 | memory length: 646875 | epsilon: |
| 0.009998020008555413 | steps: 227 | lr: 4e-05 | evaluation reward: 4.06 |
| episode: 2776 | score: 2.0 | memory length: 647056 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.01 |
| episode: 2777 | score: 3.0 | memory length: 647287 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.0 |
| episode: 2778 | score: 4.0 | memory length: 647547 | epsilon: |
| 0.009998020008555413 | steps: 260 | lr: 4e-05 | evaluation reward: 4.02 |
| episode: 2779 | score: 16.0 | memory length: 648197 | epsilon: |
| 0.009998020008555413 | steps: 650 | lr: 4e-05 | evaluation reward: 4.1 |
| episode: 2780 | score: 4.0 | memory length: 648474 | epsilon: |
| 0.009998020008555413 | steps: 277 | lr: 4e-05 | evaluation reward: 4.1 |
| episode: 2781 | score: 4.0 | memory length: 648733 | epsilon: |
| 0.009998020008555413 | steps: 259 | lr: 4e-05 | evaluation reward: 4.09 |
| episode: 2782 | score: 3.0 | memory length: 648962 | epsilon: |
| 0.009998020008555413 | steps: 229 | lr: 4e-05 | evaluation reward: 4.07 |
| episode: 2783 | score: 2.0 | memory length: 649143 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.02 |
| episode: 2784 | score: 5.0 | memory length: 649450 | epsilon: |
| 0.009998020008555413 | steps: 307 | lr: 4e-05 | evaluation reward: 4.03 |
| episode: 2785 | score: 5.0 | memory length: 649760 | epsilon: |

| | | | |
|----------------------|-------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 310 | lr: 4e-05 | evaluation reward: 4.05 |
| episode: 2786 | score: 8.0 | memory length: 650184 | epsilon: |
| 0.009998020008555413 | steps: 424 | lr: 4e-05 | evaluation reward: 4.08 |
| episode: 2787 | score: 4.0 | memory length: 650483 | epsilon: |
| 0.009998020008555413 | steps: 299 | lr: 4e-05 | evaluation reward: 4.08 |
| episode: 2788 | score: 3.0 | memory length: 650732 | epsilon: |
| 0.009998020008555413 | steps: 249 | lr: 4e-05 | evaluation reward: 4.08 |
| episode: 2789 | score: 4.0 | memory length: 650988 | epsilon: |
| 0.009998020008555413 | steps: 256 | lr: 4e-05 | evaluation reward: 4.09 |
| episode: 2790 | score: 4.0 | memory length: 651247 | epsilon: |
| 0.009998020008555413 | steps: 259 | lr: 4e-05 | evaluation reward: 4.08 |
| episode: 2791 | score: 6.0 | memory length: 651601 | epsilon: |
| 0.009998020008555413 | steps: 354 | lr: 4e-05 | evaluation reward: 4.11 |
| episode: 2792 | score: 3.0 | memory length: 651829 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.1 |
| episode: 2793 | score: 3.0 | memory length: 652060 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.09 |
| episode: 2794 | score: 3.0 | memory length: 652288 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.07 |
| episode: 2795 | score: 3.0 | memory length: 652516 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.08 |
| episode: 2796 | score: 4.0 | memory length: 652775 | epsilon: |
| 0.009998020008555413 | steps: 259 | lr: 4e-05 | evaluation reward: 4.08 |
| episode: 2797 | score: 4.0 | memory length: 653053 | epsilon: |
| 0.009998020008555413 | steps: 278 | lr: 4e-05 | evaluation reward: 4.09 |
| episode: 2798 | score: 2.0 | memory length: 653234 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.06 |
| episode: 2799 | score: 4.0 | memory length: 653495 | epsilon: |
| 0.009998020008555413 | steps: 261 | lr: 4e-05 | evaluation reward: 4.07 |
| episode: 2800 | score: 8.0 | memory length: 653935 | epsilon: |
| 0.009998020008555413 | steps: 440 | lr: 4e-05 | evaluation reward: 4.13 |
| episode: 2801 | score: 4.0 | memory length: 654211 | epsilon: |
| 0.009998020008555413 | steps: 276 | lr: 4e-05 | evaluation reward: 4.13 |
| episode: 2802 | score: 3.0 | memory length: 654442 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.12 |
| episode: 2803 | score: 7.0 | memory length: 654799 | epsilon: |
| 0.009998020008555413 | steps: 357 | lr: 4e-05 | evaluation reward: 4.11 |
| episode: 2804 | score: 4.0 | memory length: 655060 | epsilon: |
| 0.009998020008555413 | steps: 261 | lr: 4e-05 | evaluation reward: 4.11 |
| episode: 2805 | score: 3.0 | memory length: 655309 | epsilon: |
| 0.009998020008555413 | steps: 249 | lr: 4e-05 | evaluation reward: 4.12 |
| episode: 2806 | score: 7.0 | memory length: 655678 | epsilon: |
| 0.009998020008555413 | steps: 369 | lr: 4e-05 | evaluation reward: 4.15 |
| episode: 2807 | score: 11.0 | memory length: 656214 | epsilon: |
| 0.009998020008555413 | steps: 536 | lr: 4e-05 | evaluation reward: 4.22 |
| episode: 2808 | score: 3.0 | memory length: 656463 | epsilon: |
| 0.009998020008555413 | steps: 249 | lr: 4e-05 | evaluation reward: 4.22 |
| episode: 2809 | score: 5.0 | memory length: 656754 | epsilon: |

| | | | |
|----------------------|------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 291 | lr: 4e-05 | evaluation reward: 4.22 |
| episode: 2810 | score: 6.0 | memory length: 657088 | epsilon: |
| 0.009998020008555413 | steps: 334 | lr: 4e-05 | evaluation reward: 4.24 |
| episode: 2811 | score: 7.0 | memory length: 657449 | epsilon: |
| 0.009998020008555413 | steps: 361 | lr: 4e-05 | evaluation reward: 4.29 |
| episode: 2812 | score: 6.0 | memory length: 657795 | epsilon: |
| 0.009998020008555413 | steps: 346 | lr: 4e-05 | evaluation reward: 4.33 |
| episode: 2813 | score: 5.0 | memory length: 658088 | epsilon: |
| 0.009998020008555413 | steps: 293 | lr: 4e-05 | evaluation reward: 4.33 |
| episode: 2814 | score: 6.0 | memory length: 658436 | epsilon: |
| 0.009998020008555413 | steps: 348 | lr: 4e-05 | evaluation reward: 4.36 |
| episode: 2815 | score: 2.0 | memory length: 658617 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.35 |
| episode: 2816 | score: 5.0 | memory length: 658908 | epsilon: |
| 0.009998020008555413 | steps: 291 | lr: 4e-05 | evaluation reward: 4.36 |
| episode: 2817 | score: 4.0 | memory length: 659201 | epsilon: |
| 0.009998020008555413 | steps: 293 | lr: 4e-05 | evaluation reward: 4.34 |
| episode: 2818 | score: 4.0 | memory length: 659460 | epsilon: |
| 0.009998020008555413 | steps: 259 | lr: 4e-05 | evaluation reward: 4.36 |
| episode: 2819 | score: 5.0 | memory length: 659750 | epsilon: |
| 0.009998020008555413 | steps: 290 | lr: 4e-05 | evaluation reward: 4.39 |
| episode: 2820 | score: 4.0 | memory length: 660009 | epsilon: |
| 0.009998020008555413 | steps: 259 | lr: 4e-05 | evaluation reward: 4.41 |
| episode: 2821 | score: 4.0 | memory length: 660266 | epsilon: |
| 0.009998020008555413 | steps: 257 | lr: 4e-05 | evaluation reward: 4.41 |
| episode: 2822 | score: 4.0 | memory length: 660526 | epsilon: |
| 0.009998020008555413 | steps: 260 | lr: 4e-05 | evaluation reward: 4.43 |
| episode: 2823 | score: 4.0 | memory length: 660823 | epsilon: |
| 0.009998020008555413 | steps: 297 | lr: 4e-05 | evaluation reward: 4.42 |
| episode: 2824 | score: 7.0 | memory length: 661200 | epsilon: |
| 0.009998020008555413 | steps: 377 | lr: 4e-05 | evaluation reward: 4.44 |
| episode: 2825 | score: 5.0 | memory length: 661529 | epsilon: |
| 0.009998020008555413 | steps: 329 | lr: 4e-05 | evaluation reward: 4.39 |
| episode: 2826 | score: 4.0 | memory length: 661805 | epsilon: |
| 0.009998020008555413 | steps: 276 | lr: 4e-05 | evaluation reward: 4.41 |
| episode: 2827 | score: 5.0 | memory length: 662085 | epsilon: |
| 0.009998020008555413 | steps: 280 | lr: 4e-05 | evaluation reward: 4.43 |
| episode: 2828 | score: 6.0 | memory length: 662457 | epsilon: |
| 0.009998020008555413 | steps: 372 | lr: 4e-05 | evaluation reward: 4.46 |
| episode: 2829 | score: 5.0 | memory length: 662750 | epsilon: |
| 0.009998020008555413 | steps: 293 | lr: 4e-05 | evaluation reward: 4.44 |
| episode: 2830 | score: 8.0 | memory length: 663172 | epsilon: |
| 0.009998020008555413 | steps: 422 | lr: 4e-05 | evaluation reward: 4.49 |
| episode: 2831 | score: 4.0 | memory length: 663429 | epsilon: |
| 0.009998020008555413 | steps: 257 | lr: 4e-05 | evaluation reward: 4.44 |
| episode: 2832 | score: 5.0 | memory length: 663739 | epsilon: |
| 0.009998020008555413 | steps: 310 | lr: 4e-05 | evaluation reward: 4.46 |
| episode: 2833 | score: 7.0 | memory length: 664123 | epsilon: |

| | | | |
|----------------------|-------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 384 | lr: 4e-05 | evaluation reward: 4.44 |
| episode: 2834 | score: 6.0 | memory length: 664444 | epsilon: |
| 0.009998020008555413 | steps: 321 | lr: 4e-05 | evaluation reward: 4.45 |
| episode: 2835 | score: 6.0 | memory length: 664796 | epsilon: |
| 0.009998020008555413 | steps: 352 | lr: 4e-05 | evaluation reward: 4.46 |
| episode: 2836 | score: 6.0 | memory length: 665130 | epsilon: |
| 0.009998020008555413 | steps: 334 | lr: 4e-05 | evaluation reward: 4.49 |
| episode: 2837 | score: 4.0 | memory length: 665388 | epsilon: |
| 0.009998020008555413 | steps: 258 | lr: 4e-05 | evaluation reward: 4.46 |
| episode: 2838 | score: 5.0 | memory length: 665699 | epsilon: |
| 0.009998020008555413 | steps: 311 | lr: 4e-05 | evaluation reward: 4.48 |
| episode: 2839 | score: 7.0 | memory length: 666099 | epsilon: |
| 0.009998020008555413 | steps: 400 | lr: 4e-05 | evaluation reward: 4.5 |
| episode: 2840 | score: 2.0 | memory length: 666280 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.5 |
| episode: 2841 | score: 13.0 | memory length: 666785 | epsilon: |
| 0.009998020008555413 | steps: 505 | lr: 4e-05 | evaluation reward: 4.61 |
| episode: 2842 | score: 5.0 | memory length: 667115 | epsilon: |
| 0.009998020008555413 | steps: 330 | lr: 4e-05 | evaluation reward: 4.64 |
| episode: 2843 | score: 3.0 | memory length: 667342 | epsilon: |
| 0.009998020008555413 | steps: 227 | lr: 4e-05 | evaluation reward: 4.64 |
| episode: 2844 | score: 9.0 | memory length: 667813 | epsilon: |
| 0.009998020008555413 | steps: 471 | lr: 4e-05 | evaluation reward: 4.71 |
| episode: 2845 | score: 6.0 | memory length: 668171 | epsilon: |
| 0.009998020008555413 | steps: 358 | lr: 4e-05 | evaluation reward: 4.74 |
| episode: 2846 | score: 9.0 | memory length: 668671 | epsilon: |
| 0.009998020008555413 | steps: 500 | lr: 4e-05 | evaluation reward: 4.76 |
| episode: 2847 | score: 5.0 | memory length: 668956 | epsilon: |
| 0.009998020008555413 | steps: 285 | lr: 4e-05 | evaluation reward: 4.78 |
| episode: 2848 | score: 3.0 | memory length: 669183 | epsilon: |
| 0.009998020008555413 | steps: 227 | lr: 4e-05 | evaluation reward: 4.78 |
| episode: 2849 | score: 4.0 | memory length: 669461 | epsilon: |
| 0.009998020008555413 | steps: 278 | lr: 4e-05 | evaluation reward: 4.79 |
| episode: 2850 | score: 11.0 | memory length: 669938 | epsilon: |
| 0.009998020008555413 | steps: 477 | lr: 4e-05 | evaluation reward: 4.87 |
| episode: 2851 | score: 3.0 | memory length: 670166 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.86 |
| episode: 2852 | score: 4.0 | memory length: 670440 | epsilon: |
| 0.009998020008555413 | steps: 274 | lr: 4e-05 | evaluation reward: 4.84 |
| episode: 2853 | score: 2.0 | memory length: 670637 | epsilon: |
| 0.009998020008555413 | steps: 197 | lr: 4e-05 | evaluation reward: 4.83 |
| episode: 2854 | score: 5.0 | memory length: 670948 | epsilon: |
| 0.009998020008555413 | steps: 311 | lr: 4e-05 | evaluation reward: 4.81 |
| episode: 2855 | score: 5.0 | memory length: 671256 | epsilon: |
| 0.009998020008555413 | steps: 308 | lr: 4e-05 | evaluation reward: 4.82 |
| episode: 2856 | score: 4.0 | memory length: 671530 | epsilon: |
| 0.009998020008555413 | steps: 274 | lr: 4e-05 | evaluation reward: 4.83 |
| episode: 2857 | score: 8.0 | memory length: 671928 | epsilon: |

| | | | |
|----------------------|------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 398 | lr: 4e-05 | evaluation reward: 4.87 |
| episode: 2858 | score: 6.0 | memory length: 672270 | epsilon: |
| 0.009998020008555413 | steps: 342 | lr: 4e-05 | evaluation reward: 4.85 |
| episode: 2859 | score: 4.0 | memory length: 672528 | epsilon: |
| 0.009998020008555413 | steps: 258 | lr: 4e-05 | evaluation reward: 4.83 |
| episode: 2860 | score: 6.0 | memory length: 672863 | epsilon: |
| 0.009998020008555413 | steps: 335 | lr: 4e-05 | evaluation reward: 4.86 |
| episode: 2861 | score: 4.0 | memory length: 673120 | epsilon: |
| 0.009998020008555413 | steps: 257 | lr: 4e-05 | evaluation reward: 4.87 |
| episode: 2862 | score: 7.0 | memory length: 673474 | epsilon: |
| 0.009998020008555413 | steps: 354 | lr: 4e-05 | evaluation reward: 4.91 |
| episode: 2863 | score: 2.0 | memory length: 673655 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.89 |
| episode: 2864 | score: 5.0 | memory length: 673960 | epsilon: |
| 0.009998020008555413 | steps: 305 | lr: 4e-05 | evaluation reward: 4.91 |
| episode: 2865 | score: 3.0 | memory length: 674188 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.89 |
| episode: 2866 | score: 6.0 | memory length: 674543 | epsilon: |
| 0.009998020008555413 | steps: 355 | lr: 4e-05 | evaluation reward: 4.92 |
| episode: 2867 | score: 4.0 | memory length: 674821 | epsilon: |
| 0.009998020008555413 | steps: 278 | lr: 4e-05 | evaluation reward: 4.88 |
| episode: 2868 | score: 3.0 | memory length: 675052 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.85 |
| episode: 2869 | score: 6.0 | memory length: 675391 | epsilon: |
| 0.009998020008555413 | steps: 339 | lr: 4e-05 | evaluation reward: 4.89 |
| episode: 2870 | score: 6.0 | memory length: 675733 | epsilon: |
| 0.009998020008555413 | steps: 342 | lr: 4e-05 | evaluation reward: 4.91 |
| episode: 2871 | score: 6.0 | memory length: 676075 | epsilon: |
| 0.009998020008555413 | steps: 342 | lr: 4e-05 | evaluation reward: 4.95 |
| episode: 2872 | score: 5.0 | memory length: 676384 | epsilon: |
| 0.009998020008555413 | steps: 309 | lr: 4e-05 | evaluation reward: 4.98 |
| episode: 2873 | score: 3.0 | memory length: 676615 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.98 |
| episode: 2874 | score: 8.0 | memory length: 677103 | epsilon: |
| 0.009998020008555413 | steps: 488 | lr: 4e-05 | evaluation reward: 5.02 |
| episode: 2875 | score: 4.0 | memory length: 677365 | epsilon: |
| 0.009998020008555413 | steps: 262 | lr: 4e-05 | evaluation reward: 5.03 |
| episode: 2876 | score: 4.0 | memory length: 677639 | epsilon: |
| 0.009998020008555413 | steps: 274 | lr: 4e-05 | evaluation reward: 5.05 |
| episode: 2877 | score: 4.0 | memory length: 677913 | epsilon: |
| 0.009998020008555413 | steps: 274 | lr: 4e-05 | evaluation reward: 5.06 |
| episode: 2878 | score: 3.0 | memory length: 678162 | epsilon: |
| 0.009998020008555413 | steps: 249 | lr: 4e-05 | evaluation reward: 5.05 |
| episode: 2879 | score: 2.0 | memory length: 678343 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.91 |
| episode: 2880 | score: 5.0 | memory length: 678668 | epsilon: |
| 0.009998020008555413 | steps: 325 | lr: 4e-05 | evaluation reward: 4.92 |
| episode: 2881 | score: 3.0 | memory length: 678896 | epsilon: |

| | | | |
|----------------------|-------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 228 | lr: 4e-05 | evaluation reward: 4.91 |
| episode: 2882 | score: 6.0 | memory length: 679229 | epsilon: |
| 0.009998020008555413 | steps: 333 | lr: 4e-05 | evaluation reward: 4.94 |
| episode: 2883 | score: 10.0 | memory length: 679730 | epsilon: |
| 0.009998020008555413 | steps: 501 | lr: 4e-05 | evaluation reward: 5.02 |
| episode: 2884 | score: 6.0 | memory length: 680074 | epsilon: |
| 0.009998020008555413 | steps: 344 | lr: 4e-05 | evaluation reward: 5.03 |
| episode: 2885 | score: 11.0 | memory length: 680578 | epsilon: |
| 0.009998020008555413 | steps: 504 | lr: 4e-05 | evaluation reward: 5.09 |
| episode: 2886 | score: 5.0 | memory length: 680858 | epsilon: |
| 0.009998020008555413 | steps: 280 | lr: 4e-05 | evaluation reward: 5.06 |
| episode: 2887 | score: 4.0 | memory length: 681121 | epsilon: |
| 0.009998020008555413 | steps: 263 | lr: 4e-05 | evaluation reward: 5.06 |
| episode: 2888 | score: 4.0 | memory length: 681397 | epsilon: |
| 0.009998020008555413 | steps: 276 | lr: 4e-05 | evaluation reward: 5.07 |
| episode: 2889 | score: 2.0 | memory length: 681577 | epsilon: |
| 0.009998020008555413 | steps: 180 | lr: 4e-05 | evaluation reward: 5.05 |
| episode: 2890 | score: 2.0 | memory length: 681758 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 5.03 |
| episode: 2891 | score: 5.0 | memory length: 682083 | epsilon: |
| 0.009998020008555413 | steps: 325 | lr: 4e-05 | evaluation reward: 5.02 |
| episode: 2892 | score: 8.0 | memory length: 682502 | epsilon: |
| 0.009998020008555413 | steps: 419 | lr: 4e-05 | evaluation reward: 5.07 |
| episode: 2893 | score: 3.0 | memory length: 682733 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 5.07 |
| episode: 2894 | score: 6.0 | memory length: 683062 | epsilon: |
| 0.009998020008555413 | steps: 329 | lr: 4e-05 | evaluation reward: 5.1 |
| episode: 2895 | score: 4.0 | memory length: 683324 | epsilon: |
| 0.009998020008555413 | steps: 262 | lr: 4e-05 | evaluation reward: 5.11 |
| episode: 2896 | score: 5.0 | memory length: 683604 | epsilon: |
| 0.009998020008555413 | steps: 280 | lr: 4e-05 | evaluation reward: 5.12 |
| episode: 2897 | score: 5.0 | memory length: 683917 | epsilon: |
| 0.009998020008555413 | steps: 313 | lr: 4e-05 | evaluation reward: 5.13 |
| episode: 2898 | score: 4.0 | memory length: 684160 | epsilon: |
| 0.009998020008555413 | steps: 243 | lr: 4e-05 | evaluation reward: 5.15 |
| episode: 2899 | score: 3.0 | memory length: 684391 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 5.14 |
| episode: 2900 | score: 7.0 | memory length: 684770 | epsilon: |
| 0.009998020008555413 | steps: 379 | lr: 4e-05 | evaluation reward: 5.13 |
| episode: 2901 | score: 2.0 | memory length: 684951 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 5.11 |
| episode: 2902 | score: 4.0 | memory length: 685229 | epsilon: |
| 0.009998020008555413 | steps: 278 | lr: 4e-05 | evaluation reward: 5.12 |
| episode: 2903 | score: 3.0 | memory length: 685478 | epsilon: |
| 0.009998020008555413 | steps: 249 | lr: 4e-05 | evaluation reward: 5.08 |
| episode: 2904 | score: 8.0 | memory length: 685755 | epsilon: |
| 0.009998020008555413 | steps: 277 | lr: 4e-05 | evaluation reward: 5.12 |
| episode: 2905 | score: 5.0 | memory length: 686086 | epsilon: |

| | | | |
|----------------------|-------------|-----------------------|-------------------------|
| 0.009998020008555413 | steps: 331 | lr: 4e-05 | evaluation reward: 5.14 |
| episode: 2906 | score: 6.0 | memory length: 686407 | epsilon: |
| 0.009998020008555413 | steps: 321 | lr: 4e-05 | evaluation reward: 5.13 |
| episode: 2907 | score: 3.0 | memory length: 686655 | epsilon: |
| 0.009998020008555413 | steps: 248 | lr: 4e-05 | evaluation reward: 5.05 |
| episode: 2908 | score: 2.0 | memory length: 686852 | epsilon: |
| 0.009998020008555413 | steps: 197 | lr: 4e-05 | evaluation reward: 5.04 |
| episode: 2909 | score: 7.0 | memory length: 687228 | epsilon: |
| 0.009998020008555413 | steps: 376 | lr: 4e-05 | evaluation reward: 5.06 |
| episode: 2910 | score: 2.0 | memory length: 687409 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 5.02 |
| episode: 2911 | score: 5.0 | memory length: 687733 | epsilon: |
| 0.009998020008555413 | steps: 324 | lr: 4e-05 | evaluation reward: 5.0 |
| episode: 2912 | score: 4.0 | memory length: 688030 | epsilon: |
| 0.009998020008555413 | steps: 297 | lr: 4e-05 | evaluation reward: 4.98 |
| episode: 2913 | score: 5.0 | memory length: 688310 | epsilon: |
| 0.009998020008555413 | steps: 280 | lr: 4e-05 | evaluation reward: 4.98 |
| episode: 2914 | score: 7.0 | memory length: 688661 | epsilon: |
| 0.009998020008555413 | steps: 351 | lr: 4e-05 | evaluation reward: 4.99 |
| episode: 2915 | score: 4.0 | memory length: 688937 | epsilon: |
| 0.009998020008555413 | steps: 276 | lr: 4e-05 | evaluation reward: 5.01 |
| episode: 2916 | score: 6.0 | memory length: 689294 | epsilon: |
| 0.009998020008555413 | steps: 357 | lr: 4e-05 | evaluation reward: 5.02 |
| episode: 2917 | score: 8.0 | memory length: 689675 | epsilon: |
| 0.009998020008555413 | steps: 381 | lr: 4e-05 | evaluation reward: 5.06 |
| episode: 2918 | score: 8.0 | memory length: 690078 | epsilon: |
| 0.009998020008555413 | steps: 403 | lr: 4e-05 | evaluation reward: 5.1 |
| episode: 2919 | score: 4.0 | memory length: 690321 | epsilon: |
| 0.009998020008555413 | steps: 243 | lr: 4e-05 | evaluation reward: 5.09 |
| episode: 2920 | score: 5.0 | memory length: 690601 | epsilon: |
| 0.009998020008555413 | steps: 280 | lr: 4e-05 | evaluation reward: 5.1 |
| episode: 2921 | score: 4.0 | memory length: 690875 | epsilon: |
| 0.009998020008555413 | steps: 274 | lr: 4e-05 | evaluation reward: 5.1 |
| episode: 2922 | score: 2.0 | memory length: 691056 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 5.08 |
| episode: 2923 | score: 4.0 | memory length: 691315 | epsilon: |
| 0.009998020008555413 | steps: 259 | lr: 4e-05 | evaluation reward: 5.08 |
| episode: 2924 | score: 6.0 | memory length: 691643 | epsilon: |
| 0.009998020008555413 | steps: 328 | lr: 4e-05 | evaluation reward: 5.07 |
| episode: 2925 | score: 6.0 | memory length: 691988 | epsilon: |
| 0.009998020008555413 | steps: 345 | lr: 4e-05 | evaluation reward: 5.08 |
| episode: 2926 | score: 4.0 | memory length: 692267 | epsilon: |
| 0.009998020008555413 | steps: 279 | lr: 4e-05 | evaluation reward: 5.08 |
| episode: 2927 | score: 10.0 | memory length: 692762 | epsilon: |
| 0.009998020008555413 | steps: 495 | lr: 4e-05 | evaluation reward: 5.13 |
| episode: 2928 | score: 7.0 | memory length: 693186 | epsilon: |
| 0.009998020008555413 | steps: 424 | lr: 4e-05 | evaluation reward: 5.14 |
| episode: 2929 | score: 3.0 | memory length: 693435 | epsilon: |

| | | | |
|----------------------|-------------|----------------------------|-------------------------|
| 0.009998020008555413 | steps: 249 | lr: 4e-05 | evaluation reward: 5.12 |
| episode: 2930 | score: 4.0 | memory length: 693713 | epsilon: |
| 0.009998020008555413 | steps: 278 | lr: 4e-05 | evaluation reward: 5.08 |
| episode: 2931 | score: 8.0 | memory length: 694105 | epsilon: |
| 0.009998020008555413 | steps: 392 | lr: 4e-05 | evaluation reward: 5.12 |
| episode: 2932 | score: 4.0 | memory length: 694381 | epsilon: |
| 0.009998020008555413 | steps: 276 | lr: 4e-05 | evaluation reward: 5.11 |
| episode: 2933 | score: 4.0 | memory length: 694620 | epsilon: |
| 0.009998020008555413 | steps: 239 | lr: 4e-05 | evaluation reward: 5.08 |
| episode: 2934 | score: 6.0 | memory length: 695017 | epsilon: |
| 0.009998020008555413 | steps: 397 | lr: 4e-05 | evaluation reward: 5.08 |
| episode: 2935 | score: 2.0 | memory length: 695198 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 5.04 |
| episode: 2936 | score: 7.0 | memory length: 695593 | epsilon: |
| 0.009998020008555413 | steps: 395 | lr: 4e-05 | evaluation reward: 5.05 |
| episode: 2937 | score: 5.0 | memory length: 695873 | epsilon: |
| 0.009998020008555413 | steps: 280 | lr: 4e-05 | evaluation reward: 5.06 |
| episode: 2938 | score: 4.0 | memory length: 696170 | epsilon: |
| 0.009998020008555413 | steps: 297 | lr: 4e-05 | evaluation reward: 5.05 |
| episode: 2939 | score: 3.0 | memory length: 696401 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 5.01 |
| episode: 2940 | score: 4.0 | memory length: 696679 | epsilon: |
| 0.009998020008555413 | steps: 278 | lr: 4e-05 | evaluation reward: 5.03 |
| episode: 2941 | score: 3.0 | memory length: 696910 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.93 |
| episode: 2942 | score: 3.0 | memory length: 697141 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.91 |
| episode: 2943 | score: 7.0 | memory length: 697512 | epsilon: |
| 0.009998020008555413 | steps: 371 | lr: 4e-05 | evaluation reward: 4.95 |
| episode: 2944 | score: 4.0 | memory length: 697810 | epsilon: |
| 0.009998020008555413 | steps: 298 | lr: 4e-05 | evaluation reward: 4.9 |
| episode: 2945 | score: 5.0 | memory length: 698112 | epsilon: |
| 0.009998020008555413 | steps: 302 | lr: 4e-05 | evaluation reward: 4.89 |
| episode: 2946 | score: 2.0 | memory length: 698293 | epsilon: |
| 0.009998020008555413 | steps: 181 | lr: 4e-05 | evaluation reward: 4.82 |
| episode: 2947 | score: 3.0 | memory length: 698524 | epsilon: |
| 0.009998020008555413 | steps: 231 | lr: 4e-05 | evaluation reward: 4.8 |
| episode: 2948 | score: 6.0 | memory length: 698864 | epsilon: |
| 0.009998020008555413 | steps: 340 | lr: 4e-05 | evaluation reward: 4.83 |
| episode: 2949 | score: 6.0 | memory length: 699173 | epsilon: |
| 0.009998020008555413 | steps: 309 | lr: 4e-05 | evaluation reward: 4.85 |
| episode: 2950 | score: 4.0 | memory length: 699450 | epsilon: |
| 0.009998020008555413 | steps: 277 | lr: 4e-05 | evaluation reward: 4.78 |
| episode: 2951 | score: 12.0 | memory length: 699883 | epsilon: |
| 0.009998020008555413 | steps: 433 | lr: 4e-05 | evaluation reward: 4.87 |
| episode: 2952 | score: 3.0 | memory length: 700111 | epsilon: |
| 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation reward: 4.86 |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 2953 | score: 3.0 | memory length: 700342 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 2954 | score: 4.0 | memory length: 700604 | epsilon: 0.009998020008555413 | steps: 262 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 2955 | score: 6.0 | memory length: 700936 | epsilon: 0.009998020008555413 | steps: 332 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 2956 | score: 5.0 | memory length: 701243 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 2957 | score: 2.0 | memory length: 701440 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 2958 | score: 7.0 | memory length: 701774 | epsilon: 0.009998020008555413 | steps: 334 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 2959 | score: 9.0 | memory length: 702224 | epsilon: 0.009998020008555413 | steps: 450 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 2960 | score: 3.0 | memory length: 702468 | epsilon: 0.009998020008555413 | steps: 244 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 2961 | score: 5.0 | memory length: 702748 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 2962 | score: 4.0 | memory length: 703007 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 2963 | score: 7.0 | memory length: 703360 | epsilon: 0.009998020008555413 | steps: 353 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 2964 | score: 3.0 | memory length: 703572 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 2965 | score: 8.0 | memory length: 703849 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.91 | | | | | | |
| episode: 2966 | score: 13.0 | memory length: 704332 | epsilon: 0.009998020008555413 | steps: 483 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 2967 | score: 5.0 | memory length: 704658 | epsilon: 0.009998020008555413 | steps: 326 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 2968 | score: 5.0 | memory length: 704965 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 2969 | score: 4.0 | memory length: 705239 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 2970 | score: 6.0 | memory length: 705595 | epsilon: 0.009998020008555413 | steps: 356 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 2971 | score: 4.0 | memory length: 705854 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 2972 | score: 6.0 | memory length: 706172 | epsilon: 0.009998020008555413 | steps: 318 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 2973 | score: 5.0 | memory length: 706465 | epsilon: 0.009998020008555413 | steps: 293 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.0 | | | | | | |
| episode: 2974 | score: 4.0 | memory length: 706740 | epsilon: 0.009998020008555413 | steps: 275 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 2975 | score: 5.0 | memory length: 707047 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 2976 | score: 4.0 | memory length: 707290 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 2977 | score: 5.0 | memory length: 707600 | epsilon: 0.009998020008555413 | steps: 310 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 2978 | score: 4.0 | memory length: 707857 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 2979 | score: 2.0 | memory length: 708038 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 2980 | score: 10.0 | memory length: 708411 | epsilon: 0.009998020008555413 | steps: 373 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 2981 | score: 4.0 | memory length: 708654 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 2982 | score: 6.0 | memory length: 709006 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 2983 | score: 7.0 | memory length: 709411 | epsilon: 0.009998020008555413 | steps: 405 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 2984 | score: 6.0 | memory length: 709728 | epsilon: 0.009998020008555413 | steps: 317 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 2985 | score: 4.0 | memory length: 710007 | epsilon: 0.009998020008555413 | steps: 279 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |
| episode: 2986 | score: 7.0 | memory length: 710362 | epsilon: 0.009998020008555413 | steps: 355 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 2987 | score: 3.0 | memory length: 710611 | epsilon: 0.009998020008555413 | steps: 249 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 2988 | score: 6.0 | memory length: 710938 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 2989 | score: 3.0 | memory length: 711169 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 2990 | score: 6.0 | memory length: 711531 | epsilon: 0.009998020008555413 | steps: 362 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 2991 | score: 12.0 | memory length: 711992 | epsilon: 0.009998020008555413 | steps: 461 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.1 | | | | | | |
| episode: 2992 | score: 2.0 | memory length: 712173 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 2993 | score: 5.0 | memory length: 712464 | epsilon: 0.009998020008555413 | steps: 291 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.06 | | | | | | |
| episode: 2994 | score: 5.0 | memory length: 712755 | epsilon: 0.009998020008555413 | steps: 291 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 2995 | score: 5.0 | memory length: 713045 | epsilon: 0.009998020008555413 | steps: 290 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.06 | | | | | | |
| episode: 2996 | score: 6.0 | memory length: 713341 | epsilon: 0.009998020008555413 | steps: 296 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.07 | | | | | | |
| episode: 2997 | score: 4.0 | memory length: 713584 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.06 | | | | | | |
| episode: 2998 | score: 3.0 | memory length: 713811 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 2999 | score: 6.0 | memory length: 714152 | epsilon: 0.009998020008555413 | steps: 341 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.08 | | | | | | |
| episode: 3000 | score: 2.0 | memory length: 714332 | epsilon: 0.009998020008555413 | steps: 180 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3001 | score: 4.0 | memory length: 714592 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 3002 | score: 2.0 | memory length: 714773 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 3003 | score: 3.0 | memory length: 715005 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 3004 | score: 3.0 | memory length: 715234 | epsilon: 0.009998020008555413 | steps: 229 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 3005 | score: 9.0 | memory length: 715540 | epsilon: 0.009998020008555413 | steps: 306 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 3006 | score: 9.0 | memory length: 715846 | epsilon: 0.009998020008555413 | steps: 306 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 3007 | score: 7.0 | memory length: 716215 | epsilon: 0.009998020008555413 | steps: 369 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.09 | | | | | | |
| episode: 3008 | score: 4.0 | memory length: 716471 | epsilon: 0.009998020008555413 | steps: 256 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.11 | | | | | | |
| episode: 3009 | score: 2.0 | memory length: 716652 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.06 | | | | | | |
| episode: 3010 | score: 9.0 | memory length: 716979 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.13 | | | | | | |
| episode: 3011 | score: 5.0 | memory length: 717287 | epsilon: 0.009998020008555413 | steps: 308 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.13 | | | | | | |
| episode: 3012 | score: 2.0 | memory length: 717468 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.11 | | | | | | |
| episode: 3013 | score: 2.0 | memory length: 717665 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.08 | | | | | | |
| episode: 3014 | score: 6.0 | memory length: 717988 | epsilon: 0.009998020008555413 | steps: 323 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.07 | | | | | | |
| episode: 3015 | score: 3.0 | memory length: 718220 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.06 | | | | | | |
| episode: 3016 | score: 2.0 | memory length: 718417 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3017 | score: 5.0 | memory length: 718684 | epsilon: 0.009998020008555413 | steps: 267 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 3018 | score: 5.0 | memory length: 718994 | epsilon: 0.009998020008555413 | steps: 310 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 3019 | score: 9.0 | memory length: 719322 | epsilon: 0.009998020008555413 | steps: 328 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3020 | score: 5.0 | memory length: 719615 | epsilon: 0.009998020008555413 | steps: 293 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3021 | score: 4.0 | memory length: 719872 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3022 | score: 5.0 | memory length: 720175 | epsilon: 0.009998020008555413 | steps: 303 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3023 | score: 5.0 | memory length: 720482 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 3024 | score: 4.0 | memory length: 720760 | epsilon: 0.009998020008555413 | steps: 278 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 3025 | score: 2.0 | memory length: 720941 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 3026 | score: 9.0 | memory length: 721385 | epsilon: 0.009998020008555413 | steps: 444 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3027 | score: 5.0 | memory length: 721695 | epsilon: 0.009998020008555413 | steps: 310 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 3028 | score: 3.0 | memory length: 721942 | epsilon: 0.009998020008555413 | steps: 247 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |
| episode: 3029 | score: 3.0 | memory length: 722169 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |
| episode: 3030 | score: 5.0 | memory length: 722454 | epsilon: 0.009998020008555413 | steps: 285 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 3031 | score: 5.0 | memory length: 722775 | epsilon: 0.009998020008555413 | steps: 321 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3032 | score: 4.0 | memory length: 723048 | epsilon: 0.009998020008555413 | steps: 273 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3033 | score: 2.0 | memory length: 723229 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.91 | | | | | | |
| episode: 3034 | score: 5.0 | memory length: 723534 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3035 | score: 5.0 | memory length: 723844 | epsilon: 0.009998020008555413 | steps: 310 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3036 | score: 2.0 | memory length: 724025 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3037 | score: 3.0 | memory length: 724257 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3038 | score: 4.0 | memory length: 724518 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3039 | score: 2.0 | memory length: 724699 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3040 | score: 2.0 | memory length: 724880 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3041 | score: 9.0 | memory length: 725186 | epsilon: 0.009998020008555413 | steps: 306 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 3042 | score: 2.0 | memory length: 725367 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3043 | score: 2.0 | memory length: 725548 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3044 | score: 7.0 | memory length: 725918 | epsilon: 0.009998020008555413 | steps: 370 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3045 | score: 3.0 | memory length: 726146 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3046 | score: 5.0 | memory length: 726451 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3047 | score: 3.0 | memory length: 726680 | epsilon: 0.009998020008555413 | steps: 229 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3048 | score: 3.0 | memory length: 726911 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3049 | score: 5.0 | memory length: 727220 | epsilon: 0.009998020008555413 | steps: 309 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3050 | score: 2.0 | memory length: 727401 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3051 | score: 5.0 | memory length: 727708 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3052 | score: 5.0 | memory length: 728031 | epsilon: 0.009998020008555413 | steps: 323 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3053 | score: 6.0 | memory length: 728367 | epsilon: 0.009998020008555413 | steps: 336 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3054 | score: 4.0 | memory length: 728625 | epsilon: 0.009998020008555413 | steps: 258 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3055 | score: 5.0 | memory length: 728930 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3056 | score: 8.0 | memory length: 729207 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3057 | score: 4.0 | memory length: 729468 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3058 | score: 5.0 | memory length: 729762 | epsilon: 0.009998020008555413 | steps: 294 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3059 | score: 5.0 | memory length: 730088 | epsilon: 0.009998020008555413 | steps: 326 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3060 | score: 4.0 | memory length: 730366 | epsilon: 0.009998020008555413 | steps: 278 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3061 | score: 5.0 | memory length: 730676 | epsilon: 0.009998020008555413 | steps: 310 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3062 | score: 2.0 | memory length: 730857 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3063 | score: 5.0 | memory length: 731167 | epsilon: 0.009998020008555413 | steps: 310 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3064 | score: 3.0 | memory length: 731377 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3065 | score: 3.0 | memory length: 731604 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3066 | score: 3.0 | memory length: 731813 | epsilon: 0.009998020008555413 | steps: 209 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.59 | | | | | | |
| episode: 3067 | score: 2.0 | memory length: 731994 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.56 | | | | | | |
| episode: 3068 | score: 5.0 | memory length: 732302 | epsilon: 0.009998020008555413 | steps: 308 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.56 | | | | | | |
| episode: 3069 | score: 11.0 | memory length: 732792 | epsilon: 0.009998020008555413 | steps: 490 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.63 | | | | | | |
| episode: 3070 | score: 8.0 | memory length: 733152 | epsilon: 0.009998020008555413 | steps: 360 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.65 | | | | | | |
| episode: 3071 | score: 8.0 | memory length: 733512 | epsilon: 0.009998020008555413 | steps: 360 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3072 | score: 11.0 | memory length: 733971 | epsilon: 0.009998020008555413 | steps: 459 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3073 | score: 5.0 | memory length: 734251 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3074 | score: 10.0 | memory length: 734670 | epsilon: 0.009998020008555413 | steps: 419 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3075 | score: 8.0 | memory length: 735022 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3076 | score: 6.0 | memory length: 735349 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3077 | score: 8.0 | memory length: 735701 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3078 | score: 7.0 | memory length: 736046 | epsilon: 0.009998020008555413 | steps: 345 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.91 | | | | | | |
| episode: 3079 | score: 3.0 | memory length: 736275 | epsilon: 0.009998020008555413 | steps: 229 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3080 | score: 5.0 | memory length: 736550 | epsilon: 0.009998020008555413 | steps: 275 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3081 | score: 2.0 | memory length: 736731 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3082 | score: 6.0 | memory length: 737104 | epsilon: 0.009998020008555413 | steps: 373 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3083 | score: 6.0 | memory length: 737475 | epsilon: 0.009998020008555413 | steps: 371 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3084 | score: 9.0 | memory length: 737948 | epsilon: 0.009998020008555413 | steps: 473 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3085 | score: 5.0 | memory length: 738295 | epsilon: 0.009998020008555413 | steps: 347 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3086 | score: 6.0 | memory length: 738648 | epsilon: 0.009998020008555413 | steps: 353 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3087 | score: 4.0 | memory length: 738926 | epsilon: 0.009998020008555413 | steps: 278 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3088 | score: 4.0 | memory length: 739206 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3089 | score: 2.0 | memory length: 739407 | epsilon: 0.009998020008555413 | steps: 201 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3090 | score: 4.0 | memory length: 739683 | epsilon: 0.009998020008555413 | steps: 276 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3091 | score: 7.0 | memory length: 740035 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3092 | score: 4.0 | memory length: 740276 | epsilon: 0.009998020008555413 | steps: 241 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3093 | score: 3.0 | memory length: 740485 | epsilon: 0.009998020008555413 | steps: 209 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3094 | score: 12.0 | memory length: 740787 | epsilon: 0.009998020008555413 | steps: 302 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3095 | score: 3.0 | memory length: 741018 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3096 | score: 2.0 | memory length: 741199 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3097 | score: 5.0 | memory length: 741484 | epsilon: 0.009998020008555413 | steps: 285 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3098 | score: 8.0 | memory length: 741844 | epsilon: 0.009998020008555413 | steps: 360 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3099 | score: 5.0 | memory length: 742157 | epsilon: 0.009998020008555413 | steps: 313 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3100 | score: 2.0 | memory length: 742338 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3101 | score: 3.0 | memory length: 742583 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3102 | score: 3.0 | memory length: 742832 | epsilon: 0.009998020008555413 | steps: 249 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3103 | score: 4.0 | memory length: 743073 | epsilon: 0.009998020008555413 | steps: 241 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3104 | score: 5.0 | memory length: 743353 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3105 | score: 8.0 | memory length: 743630 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3106 | score: 2.0 | memory length: 743811 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3107 | score: 2.0 | memory length: 744008 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3108 | score: 3.0 | memory length: 744217 | epsilon: 0.009998020008555413 | steps: 209 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3109 | score: 5.0 | memory length: 744511 | epsilon: 0.009998020008555413 | steps: 294 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3110 | score: 4.0 | memory length: 744772 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3111 | score: 3.0 | memory length: 745000 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3112 | score: 5.0 | memory length: 745294 | epsilon: 0.009998020008555413 | steps: 294 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3113 | score: 4.0 | memory length: 745535 | epsilon: 0.009998020008555413 | steps: 241 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3114 | score: 10.0 | memory length: 745910 | epsilon: 0.009998020008555413 | steps: 375 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3115 | score: 3.0 | memory length: 746137 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3116 | score: 3.0 | memory length: 746383 | epsilon: 0.009998020008555413 | steps: 246 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3117 | score: 6.0 | memory length: 746706 | epsilon: 0.009998020008555413 | steps: 323 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3118 | score: 5.0 | memory length: 747047 | epsilon: 0.009998020008555413 | steps: 341 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3119 | score: 5.0 | memory length: 747388 | epsilon: 0.009998020008555413 | steps: 341 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3120 | score: 2.0 | memory length: 747569 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3121 | score: 2.0 | memory length: 747750 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3122 | score: 3.0 | memory length: 747960 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3123 | score: 5.0 | memory length: 748247 | epsilon: 0.009998020008555413 | steps: 287 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3124 | score: 3.0 | memory length: 748490 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.68 | | | | | | |
| episode: 3125 | score: 5.0 | memory length: 748813 | epsilon: 0.009998020008555413 | steps: 323 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3126 | score: 2.0 | memory length: 748994 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.64 | | | | | | |
| episode: 3127 | score: 8.0 | memory length: 749423 | epsilon: 0.009998020008555413 | steps: 429 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.67 | | | | | | |
| episode: 3128 | score: 4.0 | memory length: 749666 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.68 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3129 | score: 6.0 | memory length: 749969 | epsilon: 0.009998020008555413 | steps: 303 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3130 | score: 3.0 | memory length: 750200 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3131 | score: 3.0 | memory length: 750432 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.67 | | | | | | |
| episode: 3132 | score: 5.0 | memory length: 750722 | epsilon: 0.009998020008555413 | steps: 290 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.68 | | | | | | |
| episode: 3133 | score: 5.0 | memory length: 751016 | epsilon: 0.009998020008555413 | steps: 294 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3134 | score: 8.0 | memory length: 751368 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3135 | score: 2.0 | memory length: 751549 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3136 | score: 6.0 | memory length: 751942 | epsilon: 0.009998020008555413 | steps: 393 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3137 | score: 3.0 | memory length: 752150 | epsilon: 0.009998020008555413 | steps: 208 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3138 | score: 9.0 | memory length: 752559 | epsilon: 0.009998020008555413 | steps: 409 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3139 | score: 5.0 | memory length: 752866 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3140 | score: 5.0 | memory length: 753173 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3141 | score: 4.0 | memory length: 753451 | epsilon: 0.009998020008555413 | steps: 278 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3142 | score: 4.0 | memory length: 753710 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3143 | score: 9.0 | memory length: 754117 | epsilon: 0.009998020008555413 | steps: 407 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3144 | score: 3.0 | memory length: 754344 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3145 | score: 4.0 | memory length: 754600 | epsilon: 0.009998020008555413 | steps: 256 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3146 | score: 4.0 | memory length: 754861 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3147 | score: 6.0 | memory length: 755206 | epsilon: 0.009998020008555413 | steps: 345 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 3148 | score: 3.0 | memory length: 755436 | epsilon: 0.009998020008555413 | steps: 230 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 3149 | score: 3.0 | memory length: 755646 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3150 | score: 7.0 | memory length: 756041 | epsilon: 0.009998020008555413 | steps: 395 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3151 | score: 5.0 | memory length: 756368 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3152 | score: 5.0 | memory length: 756710 | epsilon: 0.009998020008555413 | steps: 342 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3153 | score: 7.0 | memory length: 757114 | epsilon: 0.009998020008555413 | steps: 404 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3154 | score: 7.0 | memory length: 757466 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 3155 | score: 5.0 | memory length: 757759 | epsilon: 0.009998020008555413 | steps: 293 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 3156 | score: 7.0 | memory length: 758152 | epsilon: 0.009998020008555413 | steps: 393 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |
| episode: 3157 | score: 3.0 | memory length: 758381 | epsilon: 0.009998020008555413 | steps: 229 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 3158 | score: 4.0 | memory length: 758624 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3159 | score: 4.0 | memory length: 758883 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3160 | score: 8.0 | memory length: 759285 | epsilon: 0.009998020008555413 | steps: 402 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3161 | score: 4.0 | memory length: 759560 | epsilon: 0.009998020008555413 | steps: 275 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |
| episode: 3162 | score: 6.0 | memory length: 759887 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 3163 | score: 7.0 | memory length: 760254 | epsilon: 0.009998020008555413 | steps: 367 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3164 | score: 4.0 | memory length: 760497 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 3165 | score: 6.0 | memory length: 760820 | epsilon: 0.009998020008555413 | steps: 323 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 3166 | score: 2.0 | memory length: 761001 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3167 | score: 4.0 | memory length: 761259 | epsilon: 0.009998020008555413 | steps: 258 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.06 | | | | | | |
| episode: 3168 | score: 5.0 | memory length: 761552 | epsilon: 0.009998020008555413 | steps: 293 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.06 | | | | | | |
| episode: 3169 | score: 7.0 | memory length: 761939 | epsilon: 0.009998020008555413 | steps: 387 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 3170 | score: 4.0 | memory length: 762196 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 3171 | score: 6.0 | memory length: 762507 | epsilon: 0.009998020008555413 | steps: 311 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 3172 | score: 3.0 | memory length: 762754 | epsilon: 0.009998020008555413 | steps: 247 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3173 | score: 7.0 | memory length: 763120 | epsilon: 0.009998020008555413 | steps: 366 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3174 | score: 4.0 | memory length: 763398 | epsilon: 0.009998020008555413 | steps: 278 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3175 | score: 6.0 | memory length: 763742 | epsilon: 0.009998020008555413 | steps: 344 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3176 | score: 7.0 | memory length: 764088 | epsilon: 0.009998020008555413 | steps: 346 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3177 | score: 4.0 | memory length: 764365 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3178 | score: 2.0 | memory length: 764544 | epsilon: 0.009998020008555413 | steps: 179 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3179 | score: 3.0 | memory length: 764787 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3180 | score: 7.0 | memory length: 765128 | epsilon: 0.009998020008555413 | steps: 341 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3181 | score: 2.0 | memory length: 765309 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3182 | score: 2.0 | memory length: 765490 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |
| episode: 3183 | score: 2.0 | memory length: 765671 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.68 | | | | | | |
| episode: 3184 | score: 7.0 | memory length: 766065 | epsilon: 0.009998020008555413 | steps: 394 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.66 | | | | | | |
| episode: 3185 | score: 6.0 | memory length: 766421 | epsilon: 0.009998020008555413 | steps: 356 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.67 | | | | | | |
| episode: 3186 | score: 8.0 | memory length: 766781 | epsilon: 0.009998020008555413 | steps: 360 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3187 | score: 5.0 | memory length: 767075 | epsilon: 0.009998020008555413 | steps: 294 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.7 | | | | | | |
| episode: 3188 | score: 5.0 | memory length: 767381 | epsilon: 0.009998020008555413 | steps: 306 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3189 | score: 6.0 | memory length: 767708 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3190 | score: 5.0 | memory length: 768002 | epsilon: 0.009998020008555413 | steps: 294 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3191 | score: 6.0 | memory length: 768345 | epsilon: 0.009998020008555413 | steps: 343 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3192 | score: 2.0 | memory length: 768526 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3193 | score: 6.0 | memory length: 768854 | epsilon: 0.009998020008555413 | steps: 328 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3194 | score: 7.0 | memory length: 769231 | epsilon: 0.009998020008555413 | steps: 377 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3195 | score: 2.0 | memory length: 769412 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.7 | | | | | | |
| episode: 3196 | score: 5.0 | memory length: 769737 | epsilon: 0.009998020008555413 | steps: 325 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3197 | score: 7.0 | memory length: 770114 | epsilon: 0.009998020008555413 | steps: 377 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3198 | score: 5.0 | memory length: 770408 | epsilon: 0.009998020008555413 | steps: 294 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |
| episode: 3199 | score: 2.0 | memory length: 770589 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3200 | score: 6.0 | memory length: 770931 | epsilon: 0.009998020008555413 | steps: 342 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3201 | score: 5.0 | memory length: 771241 | epsilon: 0.009998020008555413 | steps: 310 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3202 | score: 4.0 | memory length: 771501 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3203 | score: 4.0 | memory length: 771775 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3204 | score: 4.0 | memory length: 772018 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3205 | score: 4.0 | memory length: 772296 | epsilon: 0.009998020008555413 | steps: 278 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3206 | score: 4.0 | memory length: 772570 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3207 | score: 2.0 | memory length: 772751 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3208 | score: 4.0 | memory length: 773029 | epsilon: 0.009998020008555413 | steps: 278 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3209 | score: 5.0 | memory length: 773356 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3210 | score: 7.0 | memory length: 773710 | epsilon: 0.009998020008555413 | steps: 354 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3211 | score: 3.0 | memory length: 773938 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3212 | score: 2.0 | memory length: 774119 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3213 | score: 4.0 | memory length: 774393 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3214 | score: 2.0 | memory length: 774574 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.66 | | | | | | |
| episode: 3215 | score: 6.0 | memory length: 774891 | epsilon: 0.009998020008555413 | steps: 317 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3216 | score: 8.0 | memory length: 775331 | epsilon: 0.009998020008555413 | steps: 440 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3217 | score: 4.0 | memory length: 775605 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |
| episode: 3218 | score: 7.0 | memory length: 775999 | epsilon: 0.009998020008555413 | steps: 394 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3219 | score: 7.0 | memory length: 776344 | epsilon: 0.009998020008555413 | steps: 345 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3220 | score: 5.0 | memory length: 776634 | epsilon: 0.009998020008555413 | steps: 290 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3221 | score: 2.0 | memory length: 776815 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3222 | score: 6.0 | memory length: 777136 | epsilon: 0.009998020008555413 | steps: 321 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3223 | score: 3.0 | memory length: 777367 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3224 | score: 6.0 | memory length: 777711 | epsilon: 0.009998020008555413 | steps: 344 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3225 | score: 5.0 | memory length: 778005 | epsilon: 0.009998020008555413 | steps: 294 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3226 | score: 4.0 | memory length: 778265 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3227 | score: 4.0 | memory length: 778522 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3228 | score: 3.0 | memory length: 778766 | epsilon: 0.009998020008555413 | steps: 244 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3229 | score: 6.0 | memory length: 779104 | epsilon: 0.009998020008555413 | steps: 338 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3230 | score: 3.0 | memory length: 779334 | epsilon: 0.009998020008555413 | steps: 230 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3231 | score: 3.0 | memory length: 779544 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3232 | score: 9.0 | memory length: 780006 | epsilon: 0.009998020008555413 | steps: 462 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3233 | score: 4.0 | memory length: 780266 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3234 | score: 7.0 | memory length: 780607 | epsilon: 0.009998020008555413 | steps: 341 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3235 | score: 7.0 | memory length: 780952 | epsilon: 0.009998020008555413 | steps: 345 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3236 | score: 6.0 | memory length: 781250 | epsilon: 0.009998020008555413 | steps: 298 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3237 | score: 8.0 | memory length: 781610 | epsilon: 0.009998020008555413 | steps: 360 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3238 | score: 7.0 | memory length: 781964 | epsilon: 0.009998020008555413 | steps: 354 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3239 | score: 7.0 | memory length: 782317 | epsilon: 0.009998020008555413 | steps: 353 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3240 | score: 8.0 | memory length: 782739 | epsilon: 0.009998020008555413 | steps: 422 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3241 | score: 3.0 | memory length: 782968 | epsilon: 0.009998020008555413 | steps: 229 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 3242 | score: 4.0 | memory length: 783222 | epsilon: 0.009998020008555413 | steps: 254 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 3243 | score: 4.0 | memory length: 783478 | epsilon: 0.009998020008555413 | steps: 256 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 3244 | score: 4.0 | memory length: 783736 | epsilon: 0.009998020008555413 | steps: 258 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3245 | score: 3.0 | memory length: 783961 | epsilon: 0.009998020008555413 | steps: 225 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 3246 | score: 4.0 | memory length: 784235 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 3247 | score: 2.0 | memory length: 784415 | epsilon: 0.009998020008555413 | steps: 180 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3248 | score: 4.0 | memory length: 784689 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3249 | score: 4.0 | memory length: 784946 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3250 | score: 5.0 | memory length: 785226 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3251 | score: 5.0 | memory length: 785517 | epsilon: 0.009998020008555413 | steps: 291 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3252 | score: 4.0 | memory length: 785795 | epsilon: 0.009998020008555413 | steps: 278 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3253 | score: 7.0 | memory length: 786140 | epsilon: 0.009998020008555413 | steps: 345 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3254 | score: 6.0 | memory length: 786438 | epsilon: 0.009998020008555413 | steps: 298 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3255 | score: 3.0 | memory length: 786669 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3256 | score: 4.0 | memory length: 786946 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3257 | score: 3.0 | memory length: 787156 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3258 | score: 4.0 | memory length: 787434 | epsilon: 0.009998020008555413 | steps: 278 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3259 | score: 5.0 | memory length: 787721 | epsilon: 0.009998020008555413 | steps: 287 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3260 | score: 3.0 | memory length: 787952 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3261 | score: 6.0 | memory length: 788257 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3262 | score: 6.0 | memory length: 788555 | epsilon: 0.009998020008555413 | steps: 298 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3263 | score: 4.0 | memory length: 788830 | epsilon: 0.009998020008555413 | steps: 275 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3264 | score: 5.0 | memory length: 789140 | epsilon: 0.009998020008555413 | steps: 310 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3265 | score: 3.0 | memory length: 789350 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3266 | score: 6.0 | memory length: 789675 | epsilon: 0.009998020008555413 | steps: 325 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3267 | score: 7.0 | memory length: 790027 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3268 | score: 6.0 | memory length: 790356 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3269 | score: 4.0 | memory length: 790635 | epsilon: 0.009998020008555413 | steps: 279 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3270 | score: 2.0 | memory length: 790816 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3271 | score: 9.0 | memory length: 791273 | epsilon: 0.009998020008555413 | steps: 457 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3272 | score: 10.0 | memory length: 791756 | epsilon: 0.009998020008555413 | steps: 483 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3273 | score: 4.0 | memory length: 792015 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3274 | score: 8.0 | memory length: 792411 | epsilon: 0.009998020008555413 | steps: 396 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3275 | score: 6.0 | memory length: 792716 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3276 | score: 5.0 | memory length: 792996 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3277 | score: 7.0 | memory length: 793333 | epsilon: 0.009998020008555413 | steps: 337 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3278 | score: 6.0 | memory length: 793660 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3279 | score: 6.0 | memory length: 793965 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3280 | score: 7.0 | memory length: 794358 | epsilon: 0.009998020008555413 | steps: 393 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3281 | score: 2.0 | memory length: 794539 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3282 | score: 3.0 | memory length: 794749 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 3283 | score: 2.0 | memory length: 794930 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 3284 | score: 2.0 | memory length: 795111 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 3285 | score: 2.0 | memory length: 795292 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3286 | score: 2.0 | memory length: 795473 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3287 | score: 2.0 | memory length: 795654 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3288 | score: 2.0 | memory length: 795851 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3289 | score: 8.0 | memory length: 796269 | epsilon: 0.009998020008555413 | steps: 418 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3290 | score: 2.0 | memory length: 796450 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |
| episode: 3291 | score: 6.0 | memory length: 796755 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |
| episode: 3292 | score: 2.0 | memory length: 796936 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |
| episode: 3293 | score: 3.0 | memory length: 797146 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3294 | score: 2.0 | memory length: 797327 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.64 | | | | | | |
| episode: 3295 | score: 4.0 | memory length: 797568 | epsilon: 0.009998020008555413 | steps: 241 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.66 | | | | | | |
| episode: 3296 | score: 2.0 | memory length: 797749 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.63 | | | | | | |
| episode: 3297 | score: 3.0 | memory length: 797959 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.59 | | | | | | |
| episode: 3298 | score: 5.0 | memory length: 798283 | epsilon: 0.009998020008555413 | steps: 324 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.59 | | | | | | |
| episode: 3299 | score: 6.0 | memory length: 798635 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.63 | | | | | | |
| episode: 3300 | score: 4.0 | memory length: 798889 | epsilon: 0.009998020008555413 | steps: 254 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.61 | | | | | | |
| episode: 3301 | score: 8.0 | memory length: 799310 | epsilon: 0.009998020008555413 | steps: 421 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.64 | | | | | | |
| episode: 3302 | score: 2.0 | memory length: 799491 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.62 | | | | | | |
| episode: 3303 | score: 5.0 | memory length: 799783 | epsilon: 0.009998020008555413 | steps: 292 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.63 | | | | | | |
| episode: 3304 | score: 4.0 | memory length: 800042 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.63 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3305 | score: 3.0 | memory length: 800270 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.62 | | | | | | |
| episode: 3306 | score: 4.0 | memory length: 800544 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.62 | | | | | | |
| episode: 3307 | score: 6.0 | memory length: 800900 | epsilon: 0.009998020008555413 | steps: 356 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.66 | | | | | | |
| episode: 3308 | score: 7.0 | memory length: 801276 | epsilon: 0.009998020008555413 | steps: 376 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3309 | score: 6.0 | memory length: 801603 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.7 | | | | | | |
| episode: 3310 | score: 7.0 | memory length: 802000 | epsilon: 0.009998020008555413 | steps: 397 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.7 | | | | | | |
| episode: 3311 | score: 6.0 | memory length: 802329 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3312 | score: 6.0 | memory length: 802658 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3313 | score: 6.0 | memory length: 802987 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3314 | score: 6.0 | memory length: 803316 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3315 | score: 7.0 | memory length: 803692 | epsilon: 0.009998020008555413 | steps: 376 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3316 | score: 6.0 | memory length: 804021 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3317 | score: 3.0 | memory length: 804268 | epsilon: 0.009998020008555413 | steps: 247 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3318 | score: 6.0 | memory length: 804597 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3319 | score: 3.0 | memory length: 804824 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3320 | score: 2.0 | memory length: 805005 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3321 | score: 5.0 | memory length: 805313 | epsilon: 0.009998020008555413 | steps: 308 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3322 | score: 5.0 | memory length: 805640 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3323 | score: 6.0 | memory length: 805993 | epsilon: 0.009998020008555413 | steps: 353 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3324 | score: 3.0 | memory length: 806223 | epsilon: 0.009998020008555413 | steps: 230 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3325 | score: 3.0 | memory length: 806454 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3326 | score: 5.0 | memory length: 806783 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3327 | score: 6.0 | memory length: 807128 | epsilon: 0.009998020008555413 | steps: 345 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3328 | score: 10.0 | memory length: 807520 | epsilon: 0.009998020008555413 | steps: 392 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3329 | score: 4.0 | memory length: 807779 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3330 | score: 3.0 | memory length: 808028 | epsilon: 0.009998020008555413 | steps: 249 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3331 | score: 6.0 | memory length: 808357 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3332 | score: 3.0 | memory length: 808567 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3333 | score: 3.0 | memory length: 808798 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3334 | score: 6.0 | memory length: 809127 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3335 | score: 5.0 | memory length: 809438 | epsilon: 0.009998020008555413 | steps: 311 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3336 | score: 4.0 | memory length: 809718 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3337 | score: 3.0 | memory length: 809967 | epsilon: 0.009998020008555413 | steps: 249 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.67 | | | | | | |
| episode: 3338 | score: 6.0 | memory length: 810322 | epsilon: 0.009998020008555413 | steps: 355 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.66 | | | | | | |
| episode: 3339 | score: 6.0 | memory length: 810651 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.65 | | | | | | |
| episode: 3340 | score: 3.0 | memory length: 810900 | epsilon: 0.009998020008555413 | steps: 249 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.6 | | | | | | |
| episode: 3341 | score: 7.0 | memory length: 811293 | epsilon: 0.009998020008555413 | steps: 393 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.64 | | | | | | |
| episode: 3342 | score: 3.0 | memory length: 811518 | epsilon: 0.009998020008555413 | steps: 225 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.63 | | | | | | |
| episode: 3343 | score: 4.0 | memory length: 811783 | epsilon: 0.009998020008555413 | steps: 265 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.63 | | | | | | |
| episode: 3344 | score: 3.0 | memory length: 811993 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.62 | | | | | | |
| episode: 3345 | score: 4.0 | memory length: 812253 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.63 | | | | | | |
| episode: 3346 | score: 6.0 | memory length: 812597 | epsilon: 0.009998020008555413 | steps: 344 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.65 | | | | | | |
| episode: 3347 | score: 11.0 | memory length: 813118 | epsilon: 0.009998020008555413 | steps: 521 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3348 | score: 5.0 | memory length: 813426 | epsilon: 0.009998020008555413 | steps: 308 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3349 | score: 6.0 | memory length: 813800 | epsilon: 0.009998020008555413 | steps: 374 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3350 | score: 6.0 | memory length: 814129 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3351 | score: 4.0 | memory length: 814407 | epsilon: 0.009998020008555413 | steps: 278 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3352 | score: 4.0 | memory length: 814683 | epsilon: 0.009998020008555413 | steps: 276 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3353 | score: 6.0 | memory length: 815012 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3354 | score: 5.0 | memory length: 815299 | epsilon: 0.009998020008555413 | steps: 287 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3355 | score: 4.0 | memory length: 815573 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3356 | score: 6.0 | memory length: 815881 | epsilon: 0.009998020008555413 | steps: 308 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3357 | score: 6.0 | memory length: 816208 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3358 | score: 6.0 | memory length: 816537 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3359 | score: 6.0 | memory length: 816866 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3360 | score: 3.0 | memory length: 817115 | epsilon: 0.009998020008555413 | steps: 249 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3361 | score: 6.0 | memory length: 817444 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3362 | score: 6.0 | memory length: 817773 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3363 | score: 6.0 | memory length: 818102 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3364 | score: 5.0 | memory length: 818426 | epsilon: 0.009998020008555413 | steps: 324 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3365 | score: 7.0 | memory length: 818834 | epsilon: 0.009998020008555413 | steps: 408 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3366 | score: 5.0 | memory length: 819125 | epsilon: 0.009998020008555413 | steps: 291 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 3367 | score: 3.0 | memory length: 819356 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3368 | score: 3.0 | memory length: 819603 | epsilon: 0.009998020008555413 | steps: 247 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3369 | score: 3.0 | memory length: 819847 | epsilon: 0.009998020008555413 | steps: 244 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3370 | score: 5.0 | memory length: 820156 | epsilon: 0.009998020008555413 | steps: 309 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3371 | score: 10.0 | memory length: 820611 | epsilon: 0.009998020008555413 | steps: 455 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3372 | score: 6.0 | memory length: 820940 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3373 | score: 6.0 | memory length: 821269 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3374 | score: 3.0 | memory length: 821500 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3375 | score: 5.0 | memory length: 821780 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3376 | score: 5.0 | memory length: 822108 | epsilon: 0.009998020008555413 | steps: 328 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3377 | score: 6.0 | memory length: 822437 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3378 | score: 3.0 | memory length: 822662 | epsilon: 0.009998020008555413 | steps: 225 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3379 | score: 6.0 | memory length: 822991 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3380 | score: 7.0 | memory length: 823367 | epsilon: 0.009998020008555413 | steps: 376 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3381 | score: 6.0 | memory length: 823694 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3382 | score: 5.0 | memory length: 823974 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3383 | score: 6.0 | memory length: 824303 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3384 | score: 2.0 | memory length: 824484 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3385 | score: 7.0 | memory length: 824861 | epsilon: 0.009998020008555413 | steps: 377 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3386 | score: 2.0 | memory length: 825058 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3387 | score: 2.0 | memory length: 825239 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3388 | score: 3.0 | memory length: 825470 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 3389 | score: 3.0 | memory length: 825701 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3390 | score: 3.0 | memory length: 825932 | epsilon: 0.009998020008555413 | steps: 231 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3391 | score: 3.0 | memory length: 826142 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3392 | score: 2.0 | memory length: 826339 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3393 | score: 4.0 | memory length: 826598 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3394 | score: 4.0 | memory length: 826860 | epsilon: 0.009998020008555413 | steps: 262 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3395 | score: 6.0 | memory length: 827189 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3396 | score: 6.0 | memory length: 827518 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.91 | | | | | | |
| episode: 3397 | score: 6.0 | memory length: 827847 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 3398 | score: 5.0 | memory length: 828175 | epsilon: 0.009998020008555413 | steps: 328 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 3399 | score: 5.0 | memory length: 828455 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3400 | score: 6.0 | memory length: 828782 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3401 | score: 6.0 | memory length: 829125 | epsilon: 0.009998020008555413 | steps: 343 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3402 | score: 5.0 | memory length: 829409 | epsilon: 0.009998020008555413 | steps: 284 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 3403 | score: 6.0 | memory length: 829738 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 3404 | score: 5.0 | memory length: 830018 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 3405 | score: 6.0 | memory length: 830347 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3406 | score: 7.0 | memory length: 830743 | epsilon: 0.009998020008555413 | steps: 396 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3407 | score: 5.0 | memory length: 831034 | epsilon: 0.009998020008555413 | steps: 291 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 3408 | score: 6.0 | memory length: 831363 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 3409 | score: 9.0 | memory length: 831802 | epsilon: 0.009998020008555413 | steps: 439 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 3410 | score: 11.0 | memory length: 832370 | epsilon: 0.009998020008555413 | steps: 568 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.09 | | | | | | |
| episode: 3411 | score: 8.0 | memory length: 832743 | epsilon: 0.009998020008555413 | steps: 373 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.11 | | | | | | |
| episode: 3412 | score: 11.0 | memory length: 833291 | epsilon: 0.009998020008555413 | steps: 548 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.16 | | | | | | |
| episode: 3413 | score: 3.0 | memory length: 833519 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.13 | | | | | | |
| episode: 3414 | score: 5.0 | memory length: 833803 | epsilon: 0.009998020008555413 | steps: 284 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.12 | | | | | | |
| episode: 3415 | score: 9.0 | memory length: 834222 | epsilon: 0.009998020008555413 | steps: 419 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.14 | | | | | | |
| episode: 3416 | score: 6.0 | memory length: 834551 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.14 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3417 | score: 4.0 | memory length: 834808 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.15 | | | | | | |
| episode: 3418 | score: 8.0 | memory length: 835213 | epsilon: 0.009998020008555413 | steps: 405 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.17 | | | | | | |
| episode: 3419 | score: 6.0 | memory length: 835559 | epsilon: 0.009998020008555413 | steps: 346 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.2 | | | | | | |
| episode: 3420 | score: 2.0 | memory length: 835740 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.2 | | | | | | |
| episode: 3421 | score: 2.0 | memory length: 835921 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.17 | | | | | | |
| episode: 3422 | score: 4.0 | memory length: 836182 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.16 | | | | | | |
| episode: 3423 | score: 8.0 | memory length: 836630 | epsilon: 0.009998020008555413 | steps: 448 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.18 | | | | | | |
| episode: 3424 | score: 12.0 | memory length: 836932 | epsilon: 0.009998020008555413 | steps: 302 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 3425 | score: 4.0 | memory length: 837173 | epsilon: 0.009998020008555413 | steps: 241 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.28 | | | | | | |
| episode: 3426 | score: 4.0 | memory length: 837416 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 3427 | score: 3.0 | memory length: 837628 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.24 | | | | | | |
| episode: 3428 | score: 9.0 | memory length: 837954 | epsilon: 0.009998020008555413 | steps: 326 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.23 | | | | | | |
| episode: 3429 | score: 3.0 | memory length: 838166 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.22 | | | | | | |
| episode: 3430 | score: 4.0 | memory length: 838427 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.23 | | | | | | |
| episode: 3431 | score: 3.0 | memory length: 838656 | epsilon: 0.009998020008555413 | steps: 229 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.2 | | | | | | |
| episode: 3432 | score: 3.0 | memory length: 838883 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.2 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3433 | score: 3.0 | memory length: 839093 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.2 | | | | | | |
| episode: 3434 | score: 2.0 | memory length: 839290 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.16 | | | | | | |
| episode: 3435 | score: 4.0 | memory length: 839551 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.15 | | | | | | |
| episode: 3436 | score: 5.0 | memory length: 839897 | epsilon: 0.009998020008555413 | steps: 346 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.16 | | | | | | |
| episode: 3437 | score: 7.0 | memory length: 840292 | epsilon: 0.009998020008555413 | steps: 395 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.2 | | | | | | |
| episode: 3438 | score: 4.0 | memory length: 840546 | epsilon: 0.009998020008555413 | steps: 254 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.18 | | | | | | |
| episode: 3439 | score: 3.0 | memory length: 840758 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.15 | | | | | | |
| episode: 3440 | score: 3.0 | memory length: 840970 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.15 | | | | | | |
| episode: 3441 | score: 12.0 | memory length: 841522 | epsilon: 0.009998020008555413 | steps: 552 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.2 | | | | | | |
| episode: 3442 | score: 3.0 | memory length: 841734 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.2 | | | | | | |
| episode: 3443 | score: 5.0 | memory length: 842033 | epsilon: 0.009998020008555413 | steps: 299 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.21 | | | | | | |
| episode: 3444 | score: 4.0 | memory length: 842287 | epsilon: 0.009998020008555413 | steps: 254 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.22 | | | | | | |
| episode: 3445 | score: 4.0 | memory length: 842543 | epsilon: 0.009998020008555413 | steps: 256 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.22 | | | | | | |
| episode: 3446 | score: 2.0 | memory length: 842724 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.18 | | | | | | |
| episode: 3447 | score: 4.0 | memory length: 842983 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.11 | | | | | | |
| episode: 3448 | score: 3.0 | memory length: 843210 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.09 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3449 | score: 6.0 | memory length: 843546 | epsilon: 0.009998020008555413 | steps: 336 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.09 | | | | | | |
| episode: 3450 | score: 4.0 | memory length: 843805 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.07 | | | | | | |
| episode: 3451 | score: 5.0 | memory length: 844109 | epsilon: 0.009998020008555413 | steps: 304 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.08 | | | | | | |
| episode: 3452 | score: 2.0 | memory length: 844290 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.06 | | | | | | |
| episode: 3453 | score: 4.0 | memory length: 844544 | epsilon: 0.009998020008555413 | steps: 254 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3454 | score: 5.0 | memory length: 844867 | epsilon: 0.009998020008555413 | steps: 323 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3455 | score: 5.0 | memory length: 845195 | epsilon: 0.009998020008555413 | steps: 328 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 3456 | score: 5.0 | memory length: 845485 | epsilon: 0.009998020008555413 | steps: 290 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3457 | score: 5.0 | memory length: 845770 | epsilon: 0.009998020008555413 | steps: 285 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 3458 | score: 5.0 | memory length: 846097 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 3459 | score: 6.0 | memory length: 846451 | epsilon: 0.009998020008555413 | steps: 354 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 3460 | score: 5.0 | memory length: 846741 | epsilon: 0.009998020008555413 | steps: 290 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3461 | score: 2.0 | memory length: 846922 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.0 | | | | | | |
| episode: 3462 | score: 6.0 | memory length: 847280 | epsilon: 0.009998020008555413 | steps: 358 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.0 | | | | | | |
| episode: 3463 | score: 3.0 | memory length: 847492 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 3464 | score: 3.0 | memory length: 847704 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3465 | score: 4.0 | memory length: 847961 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3466 | score: 7.0 | memory length: 848354 | epsilon: 0.009998020008555413 | steps: 393 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 3467 | score: 4.0 | memory length: 848615 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |
| episode: 3468 | score: 9.0 | memory length: 849042 | epsilon: 0.009998020008555413 | steps: 427 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3469 | score: 4.0 | memory length: 849318 | epsilon: 0.009998020008555413 | steps: 276 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 3470 | score: 4.0 | memory length: 849579 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3471 | score: 5.0 | memory length: 849883 | epsilon: 0.009998020008555413 | steps: 304 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 3472 | score: 2.0 | memory length: 850082 | epsilon: 0.009998020008555413 | steps: 199 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3473 | score: 3.0 | memory length: 850309 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 3474 | score: 2.0 | memory length: 850506 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3475 | score: 9.0 | memory length: 850955 | epsilon: 0.009998020008555413 | steps: 449 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3476 | score: 2.0 | memory length: 851136 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 3477 | score: 10.0 | memory length: 851608 | epsilon: 0.009998020008555413 | steps: 472 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3478 | score: 9.0 | memory length: 852046 | epsilon: 0.009998020008555413 | steps: 438 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 3479 | score: 6.0 | memory length: 852367 | epsilon: 0.009998020008555413 | steps: 321 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 3480 | score: 3.0 | memory length: 852579 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3481 | score: 3.0 | memory length: 852788 | epsilon: 0.009998020008555413 | steps: 209 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3482 | score: 3.0 | memory length: 852997 | epsilon: 0.009998020008555413 | steps: 209 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3483 | score: 4.0 | memory length: 853256 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3484 | score: 7.0 | memory length: 853641 | epsilon: 0.009998020008555413 | steps: 385 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3485 | score: 4.0 | memory length: 853902 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3486 | score: 3.0 | memory length: 854114 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.91 | | | | | | |
| episode: 3487 | score: 8.0 | memory length: 854384 | epsilon: 0.009998020008555413 | steps: 270 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 3488 | score: 3.0 | memory length: 854596 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 3489 | score: 3.0 | memory length: 854805 | epsilon: 0.009998020008555413 | steps: 209 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 3490 | score: 3.0 | memory length: 855032 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 3491 | score: 4.0 | memory length: 855286 | epsilon: 0.009998020008555413 | steps: 254 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 3492 | score: 6.0 | memory length: 855626 | epsilon: 0.009998020008555413 | steps: 340 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 3493 | score: 4.0 | memory length: 855887 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 3494 | score: 7.0 | memory length: 856266 | epsilon: 0.009998020008555413 | steps: 379 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 3495 | score: 4.0 | memory length: 856507 | epsilon: 0.009998020008555413 | steps: 241 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 3496 | score: 11.0 | memory length: 856874 | epsilon: 0.009998020008555413 | steps: 367 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.08 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3497 | score: 3.0 | memory length: 857102 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.05 | | | | | | |
| episode: 3498 | score: 4.0 | memory length: 857362 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3499 | score: 4.0 | memory length: 857623 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 3500 | score: 7.0 | memory length: 857999 | epsilon: 0.009998020008555413 | steps: 376 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3501 | score: 3.0 | memory length: 858209 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3502 | score: 4.0 | memory length: 858463 | epsilon: 0.009998020008555413 | steps: 254 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.0 | | | | | | |
| episode: 3503 | score: 4.0 | memory length: 858704 | epsilon: 0.009998020008555413 | steps: 241 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 3504 | score: 4.0 | memory length: 858978 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 3505 | score: 4.0 | memory length: 859220 | epsilon: 0.009998020008555413 | steps: 242 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |
| episode: 3506 | score: 4.0 | memory length: 859481 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3507 | score: 6.0 | memory length: 859786 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3508 | score: 9.0 | memory length: 860216 | epsilon: 0.009998020008555413 | steps: 430 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 3509 | score: 3.0 | memory length: 860428 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3510 | score: 6.0 | memory length: 860757 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3511 | score: 8.0 | memory length: 861109 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3512 | score: 7.0 | memory length: 861446 | epsilon: 0.009998020008555413 | steps: 337 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3513 | score: 7.0 | memory length: 861822 | epsilon: 0.009998020008555413 | steps: 376 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3514 | score: 5.0 | memory length: 862089 | epsilon: 0.009998020008555413 | steps: 267 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3515 | score: 3.0 | memory length: 862317 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3516 | score: 3.0 | memory length: 862526 | epsilon: 0.009998020008555413 | steps: 209 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3517 | score: 3.0 | memory length: 862738 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3518 | score: 8.0 | memory length: 863144 | epsilon: 0.009998020008555413 | steps: 406 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3519 | score: 8.0 | memory length: 863539 | epsilon: 0.009998020008555413 | steps: 395 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3520 | score: 3.0 | memory length: 863751 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3521 | score: 4.0 | memory length: 864013 | epsilon: 0.009998020008555413 | steps: 262 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3522 | score: 10.0 | memory length: 864484 | epsilon: 0.009998020008555413 | steps: 471 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3523 | score: 4.0 | memory length: 864745 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3524 | score: 8.0 | memory length: 865166 | epsilon: 0.009998020008555413 | steps: 421 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3525 | score: 6.0 | memory length: 865488 | epsilon: 0.009998020008555413 | steps: 322 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3526 | score: 5.0 | memory length: 865786 | epsilon: 0.009998020008555413 | steps: 298 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3527 | score: 4.0 | memory length: 866031 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3528 | score: 4.0 | memory length: 866288 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3529 | score: 3.0 | memory length: 866500 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3530 | score: 3.0 | memory length: 866710 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3531 | score: 4.0 | memory length: 866967 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3532 | score: 4.0 | memory length: 867226 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3533 | score: 3.0 | memory length: 867438 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3534 | score: 5.0 | memory length: 867729 | epsilon: 0.009998020008555413 | steps: 291 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3535 | score: 6.0 | memory length: 868070 | epsilon: 0.009998020008555413 | steps: 341 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3536 | score: 3.0 | memory length: 868282 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3537 | score: 4.0 | memory length: 868556 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.78 | | | | | | |
| episode: 3538 | score: 3.0 | memory length: 868784 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3539 | score: 3.0 | memory length: 869014 | epsilon: 0.009998020008555413 | steps: 230 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3540 | score: 3.0 | memory length: 869226 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3541 | score: 3.0 | memory length: 869438 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.68 | | | | | | |
| episode: 3542 | score: 5.0 | memory length: 869745 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.7 | | | | | | |
| episode: 3543 | score: 3.0 | memory length: 869957 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.68 | | | | | | |
| episode: 3544 | score: 4.0 | memory length: 870218 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.68 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3545 | score: 7.0 | memory length: 870601 | epsilon: 0.009998020008555413 | steps: 383 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3546 | score: 4.0 | memory length: 870857 | epsilon: 0.009998020008555413 | steps: 256 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3547 | score: 4.0 | memory length: 871116 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3548 | score: 4.0 | memory length: 871377 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3549 | score: 4.0 | memory length: 871634 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |
| episode: 3550 | score: 3.0 | memory length: 871844 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3551 | score: 3.0 | memory length: 872056 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3552 | score: 3.0 | memory length: 872268 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.7 | | | | | | |
| episode: 3553 | score: 3.0 | memory length: 872480 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3554 | score: 9.0 | memory length: 872804 | epsilon: 0.009998020008555413 | steps: 324 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3555 | score: 3.0 | memory length: 873014 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3556 | score: 3.0 | memory length: 873226 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3557 | score: 3.0 | memory length: 873436 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.67 | | | | | | |
| episode: 3558 | score: 6.0 | memory length: 873749 | epsilon: 0.009998020008555413 | steps: 313 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.68 | | | | | | |
| episode: 3559 | score: 3.0 | memory length: 873961 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.65 | | | | | | |
| episode: 3560 | score: 6.0 | memory length: 874280 | epsilon: 0.009998020008555413 | steps: 319 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.66 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3561 | score: 5.0 | memory length: 874587 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3562 | score: 11.0 | memory length: 874966 | epsilon: 0.009998020008555413 | steps: 379 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3563 | score: 6.0 | memory length: 875288 | epsilon: 0.009998020008555413 | steps: 322 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3564 | score: 6.0 | memory length: 875601 | epsilon: 0.009998020008555413 | steps: 313 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3565 | score: 6.0 | memory length: 875957 | epsilon: 0.009998020008555413 | steps: 356 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3566 | score: 4.0 | memory length: 876207 | epsilon: 0.009998020008555413 | steps: 250 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3574 | score: 5.0 | memory length: 878487 | epsilon: 0.009998020008555413 | steps: 268 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3575 | score: 3.0 | memory length: 878699 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3576 | score: 3.0 | memory length: 878911 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3577 | score: 5.0 | memory length: 879199 | epsilon: 0.009998020008555413 | steps: 288 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.75 | | | | | | |
| episode: 3578 | score: 3.0 | memory length: 879411 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3579 | score: 3.0 | memory length: 879643 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.66 | | | | | | |
| episode: 3580 | score: 3.0 | memory length: 879856 | epsilon: 0.009998020008555413 | steps: 213 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.66 | | | | | | |
| episode: 3581 | score: 6.0 | memory length: 880155 | epsilon: 0.009998020008555413 | steps: 299 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3582 | score: 3.0 | memory length: 880367 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3583 | score: 7.0 | memory length: 880720 | epsilon: 0.009998020008555413 | steps: 353 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3584 | score: 8.0 | memory length: 881144 | epsilon: 0.009998020008555413 | steps: 424 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3585 | score: 7.0 | memory length: 881515 | epsilon: 0.009998020008555413 | steps: 371 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3586 | score: 7.0 | memory length: 881911 | epsilon: 0.009998020008555413 | steps: 396 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3587 | score: 9.0 | memory length: 882327 | epsilon: 0.009998020008555413 | steps: 416 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3588 | score: 6.0 | memory length: 882674 | epsilon: 0.009998020008555413 | steps: 347 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3589 | score: 3.0 | memory length: 882884 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3590 | score: 12.0 | memory length: 883186 | epsilon: 0.009998020008555413 | steps: 302 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3591 | score: 3.0 | memory length: 883398 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3592 | score: 7.0 | memory length: 883774 | epsilon: 0.009998020008555413 | steps: 376 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3593 | score: 3.0 | memory length: 884001 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3594 | score: 5.0 | memory length: 884307 | epsilon: 0.009998020008555413 | steps: 306 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3595 | score: 6.0 | memory length: 884663 | epsilon: 0.009998020008555413 | steps: 356 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3596 | score: 3.0 | memory length: 884875 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3597 | score: 4.0 | memory length: 885135 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3598 | score: 4.0 | memory length: 885377 | epsilon: 0.009998020008555413 | steps: 242 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3599 | score: 8.0 | memory length: 885654 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.89 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3600 | score: 3.0 | memory length: 885866 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3601 | score: 4.0 | memory length: 886127 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3602 | score: 3.0 | memory length: 886339 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3603 | score: 12.0 | memory length: 886641 | epsilon: 0.009998020008555413 | steps: 302 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3604 | score: 6.0 | memory length: 886988 | epsilon: 0.009998020008555413 | steps: 347 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.95 | | | | | | |
| episode: 3605 | score: 6.0 | memory length: 887361 | epsilon: 0.009998020008555413 | steps: 373 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 3606 | score: 8.0 | memory length: 887638 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3607 | score: 4.0 | memory length: 887888 | epsilon: 0.009998020008555413 | steps: 250 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 3608 | score: 6.0 | memory length: 888202 | epsilon: 0.009998020008555413 | steps: 314 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 3609 | score: 5.0 | memory length: 888482 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 3610 | score: 2.0 | memory length: 888662 | epsilon: 0.009998020008555413 | steps: 180 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 3611 | score: 7.0 | memory length: 889035 | epsilon: 0.009998020008555413 | steps: 373 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3612 | score: 4.0 | memory length: 889273 | epsilon: 0.009998020008555413 | steps: 238 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3613 | score: 3.0 | memory length: 889483 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3614 | score: 3.0 | memory length: 889709 | epsilon: 0.009998020008555413 | steps: 226 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3615 | score: 4.0 | memory length: 889969 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3616 | score: 5.0 | memory length: 890237 | epsilon: 0.009998020008555413 | steps: 268 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3617 | score: 3.0 | memory length: 890449 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3618 | score: 4.0 | memory length: 890723 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3619 | score: 5.0 | memory length: 891002 | epsilon: 0.009998020008555413 | steps: 279 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3620 | score: 4.0 | memory length: 891262 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3621 | score: 5.0 | memory length: 891569 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3622 | score: 5.0 | memory length: 891874 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3623 | score: 3.0 | memory length: 892086 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3624 | score: 4.0 | memory length: 892336 | epsilon: 0.009998020008555413 | steps: 250 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |
| episode: 3625 | score: 4.0 | memory length: 892592 | epsilon: 0.009998020008555413 | steps: 256 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.7 | | | | | | |
| episode: 3626 | score: 6.0 | memory length: 892933 | epsilon: 0.009998020008555413 | steps: 341 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3627 | score: 7.0 | memory length: 893295 | epsilon: 0.009998020008555413 | steps: 362 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3628 | score: 6.0 | memory length: 893624 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3629 | score: 6.0 | memory length: 893953 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |
| episode: 3630 | score: 5.0 | memory length: 894265 | epsilon: 0.009998020008555413 | steps: 312 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3631 | score: 3.0 | memory length: 894493 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3632 | score: 6.0 | memory length: 894804 | epsilon: 0.009998020008555413 | steps: 311 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3633 | score: 7.0 | memory length: 895180 | epsilon: 0.009998020008555413 | steps: 376 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3634 | score: 5.0 | memory length: 895452 | epsilon: 0.009998020008555413 | steps: 272 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3635 | score: 5.0 | memory length: 895724 | epsilon: 0.009998020008555413 | steps: 272 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3636 | score: 3.0 | memory length: 895936 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3637 | score: 5.0 | memory length: 896232 | epsilon: 0.009998020008555413 | steps: 296 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3638 | score: 2.0 | memory length: 896413 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3639 | score: 2.0 | memory length: 896594 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3640 | score: 4.0 | memory length: 896854 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3641 | score: 3.0 | memory length: 897066 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3642 | score: 4.0 | memory length: 897345 | epsilon: 0.009998020008555413 | steps: 279 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3643 | score: 3.0 | memory length: 897573 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.84 | | | | | | |
| episode: 3644 | score: 6.0 | memory length: 897930 | epsilon: 0.009998020008555413 | steps: 357 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.86 | | | | | | |
| episode: 3645 | score: 2.0 | memory length: 898111 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3646 | score: 3.0 | memory length: 898323 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3647 | score: 3.0 | memory length: 898535 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.79 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3648 | score: 2.0 | memory length: 898716 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3649 | score: 3.0 | memory length: 898928 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3650 | score: 3.0 | memory length: 899140 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3651 | score: 4.0 | memory length: 899385 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3652 | score: 3.0 | memory length: 899597 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3653 | score: 8.0 | memory length: 899874 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3654 | score: 8.0 | memory length: 900151 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3655 | score: 3.0 | memory length: 900380 | epsilon: 0.009998020008555413 | steps: 229 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3656 | score: 3.0 | memory length: 900605 | epsilon: 0.009998020008555413 | steps: 225 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3657 | score: 5.0 | memory length: 900910 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 3658 | score: 3.0 | memory length: 901122 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3659 | score: 4.0 | memory length: 901382 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3660 | score: 6.0 | memory length: 901703 | epsilon: 0.009998020008555413 | steps: 321 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3661 | score: 9.0 | memory length: 902009 | epsilon: 0.009998020008555413 | steps: 306 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 3662 | score: 3.0 | memory length: 902221 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.77 | | | | | | |
| episode: 3663 | score: 3.0 | memory length: 902433 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3664 | score: 3.0 | memory length: 902645 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3665 | score: 4.0 | memory length: 902890 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3666 | score: 3.0 | memory length: 903102 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.68 | | | | | | |
| episode: 3667 | score: 4.0 | memory length: 903347 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.66 | | | | | | |
| episode: 3668 | score: 3.0 | memory length: 903557 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.65 | | | | | | |
| episode: 3669 | score: 3.0 | memory length: 903769 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.65 | | | | | | |
| episode: 3670 | score: 3.0 | memory length: 903981 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.6 | | | | | | |
| episode: 3671 | score: 3.0 | memory length: 904208 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.58 | | | | | | |
| episode: 3672 | score: 6.0 | memory length: 904513 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.61 | | | | | | |
| episode: 3673 | score: 3.0 | memory length: 904725 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.59 | | | | | | |
| episode: 3674 | score: 10.0 | memory length: 905073 | epsilon: 0.009998020008555413 | steps: 348 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.64 | | | | | | |
| episode: 3675 | score: 5.0 | memory length: 905362 | epsilon: 0.009998020008555413 | steps: 289 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.66 | | | | | | |
| episode: 3676 | score: 4.0 | memory length: 905607 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.67 | | | | | | |
| episode: 3677 | score: 3.0 | memory length: 905833 | epsilon: 0.009998020008555413 | steps: 226 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.65 | | | | | | |
| episode: 3678 | score: 7.0 | memory length: 906190 | epsilon: 0.009998020008555413 | steps: 357 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3679 | score: 6.0 | memory length: 906503 | epsilon: 0.009998020008555413 | steps: 313 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.72 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3680 | score: 7.0 | memory length: 906874 | epsilon: 0.009998020008555413 | steps: 371 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.76 | | | | | | |
| episode: 3681 | score: 4.0 | memory length: 907119 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3682 | score: 3.0 | memory length: 907331 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3683 | score: 7.0 | memory length: 907665 | epsilon: 0.009998020008555413 | steps: 334 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3684 | score: 8.0 | memory length: 908088 | epsilon: 0.009998020008555413 | steps: 423 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 3685 | score: 4.0 | memory length: 908367 | epsilon: 0.009998020008555413 | steps: 279 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3686 | score: 7.0 | memory length: 908721 | epsilon: 0.009998020008555413 | steps: 354 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.71 | | | | | | |
| episode: 3687 | score: 3.0 | memory length: 908948 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.65 | | | | | | |
| episode: 3688 | score: 4.0 | memory length: 909205 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.63 | | | | | | |
| episode: 3689 | score: 3.0 | memory length: 909417 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.63 | | | | | | |
| episode: 3690 | score: 3.0 | memory length: 909649 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.54 | | | | | | |
| episode: 3691 | score: 5.0 | memory length: 909953 | epsilon: 0.009998020008555413 | steps: 304 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.56 | | | | | | |
| episode: 3692 | score: 3.0 | memory length: 910165 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.52 | | | | | | |
| episode: 3693 | score: 3.0 | memory length: 910377 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.52 | | | | | | |
| episode: 3694 | score: 4.0 | memory length: 910633 | epsilon: 0.009998020008555413 | steps: 256 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.51 | | | | | | |
| episode: 3695 | score: 3.0 | memory length: 910843 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.48 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3696 | score: 3.0 | memory length: 911053 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.48 | | | | | | |
| episode: 3697 | score: 5.0 | memory length: 911394 | epsilon: 0.009998020008555413 | steps: 341 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.49 | | | | | | |
| episode: 3698 | score: 6.0 | memory length: 911752 | epsilon: 0.009998020008555413 | steps: 358 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.51 | | | | | | |
| episode: 3699 | score: 3.0 | memory length: 911962 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.46 | | | | | | |
| episode: 3700 | score: 4.0 | memory length: 912236 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.47 | | | | | | |
| episode: 3701 | score: 3.0 | memory length: 912465 | epsilon: 0.009998020008555413 | steps: 229 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.46 | | | | | | |
| episode: 3702 | score: 6.0 | memory length: 912763 | epsilon: 0.009998020008555413 | steps: 298 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.49 | | | | | | |
| episode: 3703 | score: 5.0 | memory length: 913086 | epsilon: 0.009998020008555413 | steps: 323 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.42 | | | | | | |
| episode: 3704 | score: 3.0 | memory length: 913298 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.39 | | | | | | |
| episode: 3705 | score: 3.0 | memory length: 913510 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.36 | | | | | | |
| episode: 3706 | score: 6.0 | memory length: 913837 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.34 | | | | | | |
| episode: 3707 | score: 4.0 | memory length: 914111 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.34 | | | | | | |
| episode: 3708 | score: 5.0 | memory length: 914418 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.33 | | | | | | |
| episode: 3709 | score: 4.0 | memory length: 914679 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.32 | | | | | | |
| episode: 3710 | score: 7.0 | memory length: 915053 | epsilon: 0.009998020008555413 | steps: 374 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.37 | | | | | | |
| episode: 3711 | score: 3.0 | memory length: 915263 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.33 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3712 | score: 3.0 | memory length: 915495 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.32 | | | | | | |
| episode: 3713 | score: 8.0 | memory length: 915772 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.37 | | | | | | |
| episode: 3714 | score: 4.0 | memory length: 916046 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.38 | | | | | | |
| episode: 3715 | score: 3.0 | memory length: 916278 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.37 | | | | | | |
| episode: 3716 | score: 3.0 | memory length: 916488 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.35 | | | | | | |
| episode: 3717 | score: 6.0 | memory length: 916817 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.38 | | | | | | |
| episode: 3718 | score: 7.0 | memory length: 917177 | epsilon: 0.009998020008555413 | steps: 360 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.41 | | | | | | |
| episode: 3719 | score: 6.0 | memory length: 917486 | epsilon: 0.009998020008555413 | steps: 309 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.42 | | | | | | |
| episode: 3720 | score: 7.0 | memory length: 917861 | epsilon: 0.009998020008555413 | steps: 375 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.45 | | | | | | |
| episode: 3721 | score: 7.0 | memory length: 918197 | epsilon: 0.009998020008555413 | steps: 336 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.47 | | | | | | |
| episode: 3722 | score: 9.0 | memory length: 918676 | epsilon: 0.009998020008555413 | steps: 479 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.51 | | | | | | |
| episode: 3723 | score: 8.0 | memory length: 919114 | epsilon: 0.009998020008555413 | steps: 438 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.56 | | | | | | |
| episode: 3724 | score: 3.0 | memory length: 919346 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.55 | | | | | | |
| episode: 3725 | score: 4.0 | memory length: 919589 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.55 | | | | | | |
| episode: 3726 | score: 4.0 | memory length: 919834 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.53 | | | | | | |
| episode: 3727 | score: 3.0 | memory length: 920046 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.49 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3728 | score: 3.0 | memory length: 920278 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.46 | | | | | | |
| episode: 3729 | score: 4.0 | memory length: 920536 | epsilon: 0.009998020008555413 | steps: 258 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.44 | | | | | | |
| episode: 3730 | score: 4.0 | memory length: 920798 | epsilon: 0.009998020008555413 | steps: 262 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.43 | | | | | | |
| episode: 3731 | score: 4.0 | memory length: 921072 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.44 | | | | | | |
| episode: 3732 | score: 4.0 | memory length: 921346 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.42 | | | | | | |
| episode: 3733 | score: 7.0 | memory length: 921698 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.42 | | | | | | |
| episode: 3734 | score: 10.0 | memory length: 922185 | epsilon: 0.009998020008555413 | steps: 487 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.47 | | | | | | |
| episode: 3735 | score: 5.0 | memory length: 922478 | epsilon: 0.009998020008555413 | steps: 293 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.47 | | | | | | |
| episode: 3736 | score: 12.0 | memory length: 922984 | epsilon: 0.009998020008555413 | steps: 506 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.56 | | | | | | |
| episode: 3737 | score: 9.0 | memory length: 923428 | epsilon: 0.009998020008555413 | steps: 444 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.6 | | | | | | |
| episode: 3738 | score: 8.0 | memory length: 923836 | epsilon: 0.009998020008555413 | steps: 408 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.66 | | | | | | |
| episode: 3739 | score: 5.0 | memory length: 924116 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.69 | | | | | | |
| episode: 3740 | score: 8.0 | memory length: 924522 | epsilon: 0.009998020008555413 | steps: 406 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3741 | score: 3.0 | memory length: 924734 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.73 | | | | | | |
| episode: 3742 | score: 12.0 | memory length: 925036 | epsilon: 0.009998020008555413 | steps: 302 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.81 | | | | | | |
| episode: 3743 | score: 4.0 | memory length: 925315 | epsilon: 0.009998020008555413 | steps: 279 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3744 | score: 4.0 | memory length: 925589 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.8 | | | | | | |
| episode: 3745 | score: 4.0 | memory length: 925849 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3746 | score: 3.0 | memory length: 926059 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 3747 | score: 9.0 | memory length: 926464 | epsilon: 0.009998020008555413 | steps: 405 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3748 | score: 5.0 | memory length: 926748 | epsilon: 0.009998020008555413 | steps: 284 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.91 | | | | | | |
| episode: 3749 | score: 3.0 | memory length: 926958 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.91 | | | | | | |
| episode: 3750 | score: 2.0 | memory length: 927139 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 3751 | score: 7.0 | memory length: 927491 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3752 | score: 7.0 | memory length: 927867 | epsilon: 0.009998020008555413 | steps: 376 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 3753 | score: 3.0 | memory length: 928077 | epsilon: 0.009998020008555413 | steps: 210 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3754 | score: 7.0 | memory length: 928436 | epsilon: 0.009998020008555413 | steps: 359 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.91 | | | | | | |
| episode: 3755 | score: 4.0 | memory length: 928690 | epsilon: 0.009998020008555413 | steps: 254 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.92 | | | | | | |
| episode: 3756 | score: 7.0 | memory length: 929042 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 3757 | score: 5.0 | memory length: 929347 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 3758 | score: 4.0 | memory length: 929607 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 3759 | score: 4.0 | memory length: 929864 | epsilon: 0.009998020008555413 | steps: 257 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.97 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3760 | score: 2.0 | memory length: 930045 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 3761 | score: 3.0 | memory length: 930273 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 3762 | score: 4.0 | memory length: 930531 | epsilon: 0.009998020008555413 | steps: 258 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.88 | | | | | | |
| episode: 3763 | score: 9.0 | memory length: 930955 | epsilon: 0.009998020008555413 | steps: 424 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 3764 | score: 7.0 | memory length: 931336 | epsilon: 0.009998020008555413 | steps: 381 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 3765 | score: 7.0 | memory length: 931672 | epsilon: 0.009998020008555413 | steps: 336 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3766 | score: 6.0 | memory length: 932001 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3767 | score: 6.0 | memory length: 932330 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.06 | | | | | | |
| episode: 3768 | score: 5.0 | memory length: 932610 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.08 | | | | | | |
| episode: 3769 | score: 7.0 | memory length: 932986 | epsilon: 0.009998020008555413 | steps: 376 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.12 | | | | | | |
| episode: 3770 | score: 3.0 | memory length: 933214 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.12 | | | | | | |
| episode: 3771 | score: 9.0 | memory length: 933620 | epsilon: 0.009998020008555413 | steps: 406 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.18 | | | | | | |
| episode: 3772 | score: 6.0 | memory length: 933943 | epsilon: 0.009998020008555413 | steps: 323 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.18 | | | | | | |
| episode: 3773 | score: 5.0 | memory length: 934242 | epsilon: 0.009998020008555413 | steps: 299 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.2 | | | | | | |
| episode: 3774 | score: 8.0 | memory length: 934669 | epsilon: 0.009998020008555413 | steps: 427 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.18 | | | | | | |
| episode: 3775 | score: 7.0 | memory length: 935032 | epsilon: 0.009998020008555413 | steps: 363 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.2 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3776 | score: 5.0 | memory length: 935312 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.21 | | | | | | |
| episode: 3777 | score: 7.0 | memory length: 935664 | epsilon: 0.009998020008555413 | steps: 352 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.25 | | | | | | |
| episode: 3778 | score: 7.0 | memory length: 936048 | epsilon: 0.009998020008555413 | steps: 384 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.25 | | | | | | |
| episode: 3779 | score: 6.0 | memory length: 936371 | epsilon: 0.009998020008555413 | steps: 323 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.25 | | | | | | |
| episode: 3780 | score: 11.0 | memory length: 936882 | epsilon: 0.009998020008555413 | steps: 511 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.29 | | | | | | |
| episode: 3781 | score: 5.0 | memory length: 937207 | epsilon: 0.009998020008555413 | steps: 325 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.3 | | | | | | |
| episode: 3782 | score: 7.0 | memory length: 937602 | epsilon: 0.009998020008555413 | steps: 395 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.34 | | | | | | |
| episode: 3783 | score: 9.0 | memory length: 938025 | epsilon: 0.009998020008555413 | steps: 423 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3784 | score: 2.0 | memory length: 938206 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.3 | | | | | | |
| episode: 3785 | score: 7.0 | memory length: 938560 | epsilon: 0.009998020008555413 | steps: 354 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.33 | | | | | | |
| episode: 3786 | score: 11.0 | memory length: 939095 | epsilon: 0.009998020008555413 | steps: 535 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.37 | | | | | | |
| episode: 3787 | score: 2.0 | memory length: 939292 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3788 | score: 5.0 | memory length: 939620 | epsilon: 0.009998020008555413 | steps: 328 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.37 | | | | | | |
| episode: 3789 | score: 6.0 | memory length: 939934 | epsilon: 0.009998020008555413 | steps: 314 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.4 | | | | | | |
| episode: 3790 | score: 5.0 | memory length: 940241 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.42 | | | | | | |
| episode: 3791 | score: 6.0 | memory length: 940546 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.43 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3792 | score: 6.0 | memory length: 940885 | epsilon: 0.009998020008555413 | steps: 339 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.46 | | | | | | |
| episode: 3793 | score: 7.0 | memory length: 941312 | epsilon: 0.009998020008555413 | steps: 427 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.5 | | | | | | |
| episode: 3794 | score: 7.0 | memory length: 941707 | epsilon: 0.009998020008555413 | steps: 395 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.53 | | | | | | |
| episode: 3795 | score: 7.0 | memory length: 942099 | epsilon: 0.009998020008555413 | steps: 392 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.57 | | | | | | |
| episode: 3796 | score: 4.0 | memory length: 942373 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.58 | | | | | | |
| episode: 3797 | score: 4.0 | memory length: 942646 | epsilon: 0.009998020008555413 | steps: 273 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.57 | | | | | | |
| episode: 3798 | score: 2.0 | memory length: 942827 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.53 | | | | | | |
| episode: 3799 | score: 5.0 | memory length: 943099 | epsilon: 0.009998020008555413 | steps: 272 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.55 | | | | | | |
| episode: 3800 | score: 5.0 | memory length: 943409 | epsilon: 0.009998020008555413 | steps: 310 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.56 | | | | | | |
| episode: 3801 | score: 4.0 | memory length: 943683 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.57 | | | | | | |
| episode: 3802 | score: 8.0 | memory length: 944127 | epsilon: 0.009998020008555413 | steps: 444 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.59 | | | | | | |
| episode: 3803 | score: 14.0 | memory length: 944739 | epsilon: 0.009998020008555413 | steps: 612 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.68 | | | | | | |
| episode: 3804 | score: 5.0 | memory length: 945019 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.7 | | | | | | |
| episode: 3805 | score: 7.0 | memory length: 945373 | epsilon: 0.009998020008555413 | steps: 354 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.74 | | | | | | |
| episode: 3806 | score: 5.0 | memory length: 945653 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.73 | | | | | | |
| episode: 3807 | score: 3.0 | memory length: 945865 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.72 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3808 | score: 4.0 | memory length: 946125 | epsilon: 0.009998020008555413 | steps: 260 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.71 | | | | | | |
| episode: 3809 | score: 3.0 | memory length: 946355 | epsilon: 0.009998020008555413 | steps: 230 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.7 | | | | | | |
| episode: 3810 | score: 5.0 | memory length: 946662 | epsilon: 0.009998020008555413 | steps: 307 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.68 | | | | | | |
| episode: 3811 | score: 3.0 | memory length: 946892 | epsilon: 0.009998020008555413 | steps: 230 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.68 | | | | | | |
| episode: 3812 | score: 3.0 | memory length: 947122 | epsilon: 0.009998020008555413 | steps: 230 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.68 | | | | | | |
| episode: 3813 | score: 4.0 | memory length: 947395 | epsilon: 0.009998020008555413 | steps: 273 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.64 | | | | | | |
| episode: 3814 | score: 3.0 | memory length: 947625 | epsilon: 0.009998020008555413 | steps: 230 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.63 | | | | | | |
| episode: 3815 | score: 5.0 | memory length: 947905 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.65 | | | | | | |
| episode: 3816 | score: 4.0 | memory length: 948164 | epsilon: 0.009998020008555413 | steps: 259 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.66 | | | | | | |
| episode: 3817 | score: 6.0 | memory length: 948476 | epsilon: 0.009998020008555413 | steps: 312 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.66 | | | | | | |
| episode: 3818 | score: 4.0 | memory length: 948717 | epsilon: 0.009998020008555413 | steps: 241 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.63 | | | | | | |
| episode: 3819 | score: 5.0 | memory length: 948997 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.62 | | | | | | |
| episode: 3820 | score: 6.0 | memory length: 949302 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.61 | | | | | | |
| episode: 3821 | score: 8.0 | memory length: 949579 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.62 | | | | | | |
| episode: 3822 | score: 5.0 | memory length: 949859 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.58 | | | | | | |
| episode: 3823 | score: 5.0 | memory length: 950140 | epsilon: 0.009998020008555413 | steps: 281 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.55 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3824 | score: 8.0 | memory length: 950417 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.6 | | | | | | |
| episode: 3825 | score: 6.0 | memory length: 950757 | epsilon: 0.009998020008555413 | steps: 340 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.62 | | | | | | |
| episode: 3826 | score: 5.0 | memory length: 951047 | epsilon: 0.009998020008555413 | steps: 290 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.63 | | | | | | |
| episode: 3827 | score: 9.0 | memory length: 951487 | epsilon: 0.009998020008555413 | steps: 440 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.69 | | | | | | |
| episode: 3828 | score: 3.0 | memory length: 951699 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.69 | | | | | | |
| episode: 3829 | score: 5.0 | memory length: 951979 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.7 | | | | | | |
| episode: 3830 | score: 2.0 | memory length: 952160 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.68 | | | | | | |
| episode: 3831 | score: 5.0 | memory length: 952432 | epsilon: 0.009998020008555413 | steps: 272 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.69 | | | | | | |
| episode: 3832 | score: 9.0 | memory length: 952731 | epsilon: 0.009998020008555413 | steps: 299 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.74 | | | | | | |
| episode: 3833 | score: 2.0 | memory length: 952913 | epsilon: 0.009998020008555413 | steps: 182 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.69 | | | | | | |
| episode: 3834 | score: 5.0 | memory length: 953193 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.64 | | | | | | |
| episode: 3835 | score: 2.0 | memory length: 953374 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.61 | | | | | | |
| episode: 3836 | score: 4.0 | memory length: 953617 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.53 | | | | | | |
| episode: 3837 | score: 3.0 | memory length: 953847 | epsilon: 0.009998020008555413 | steps: 230 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.47 | | | | | | |
| episode: 3838 | score: 9.0 | memory length: 954146 | epsilon: 0.009998020008555413 | steps: 299 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.48 | | | | | | |
| episode: 3839 | score: 5.0 | memory length: 954426 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.48 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3840 | score: 5.0 | memory length: 954716 | epsilon: 0.009998020008555413 | steps: 290 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.45 | | | | | | |
| episode: 3841 | score: 8.0 | memory length: 954986 | epsilon: 0.009998020008555413 | steps: 270 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.5 | | | | | | |
| episode: 3842 | score: 6.0 | memory length: 955297 | epsilon: 0.009998020008555413 | steps: 311 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.44 | | | | | | |
| episode: 3843 | score: 4.0 | memory length: 955572 | epsilon: 0.009998020008555413 | steps: 275 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.44 | | | | | | |
| episode: 3844 | score: 8.0 | memory length: 955962 | epsilon: 0.009998020008555413 | steps: 390 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.48 | | | | | | |
| episode: 3845 | score: 3.0 | memory length: 956174 | epsilon: 0.009998020008555413 | steps: 212 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.47 | | | | | | |
| episode: 3846 | score: 9.0 | memory length: 956473 | epsilon: 0.009998020008555413 | steps: 299 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.53 | | | | | | |
| episode: 3847 | score: 2.0 | memory length: 956654 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.46 | | | | | | |
| episode: 3848 | score: 2.0 | memory length: 956835 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.43 | | | | | | |
| episode: 3849 | score: 2.0 | memory length: 957016 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.42 | | | | | | |
| episode: 3850 | score: 4.0 | memory length: 957261 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.44 | | | | | | |
| episode: 3851 | score: 5.0 | memory length: 957541 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.42 | | | | | | |
| episode: 3852 | score: 6.0 | memory length: 957850 | epsilon: 0.009998020008555413 | steps: 309 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 3853 | score: 6.0 | memory length: 958159 | epsilon: 0.009998020008555413 | steps: 309 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.44 | | | | | | |
| episode: 3854 | score: 7.0 | memory length: 958537 | epsilon: 0.009998020008555413 | steps: 378 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.44 | | | | | | |
| episode: 3855 | score: 6.0 | memory length: 958846 | epsilon: 0.009998020008555413 | steps: 309 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.46 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3856 | score: 2.0 | memory length: 959027 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 3857 | score: 12.0 | memory length: 959451 | epsilon: 0.009998020008555413 | steps: 424 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.48 | | | | | | |
| episode: 3858 | score: 6.0 | memory length: 959780 | epsilon: 0.009998020008555413 | steps: 329 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.5 | | | | | | |
| episode: 3859 | score: 3.0 | memory length: 960025 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.49 | | | | | | |
| episode: 3860 | score: 5.0 | memory length: 960305 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.52 | | | | | | |
| episode: 3861 | score: 4.0 | memory length: 960567 | epsilon: 0.009998020008555413 | steps: 262 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.53 | | | | | | |
| episode: 3862 | score: 5.0 | memory length: 960849 | epsilon: 0.009998020008555413 | steps: 282 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.54 | | | | | | |
| episode: 3863 | score: 6.0 | memory length: 961158 | epsilon: 0.009998020008555413 | steps: 309 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.51 | | | | | | |
| episode: 3864 | score: 9.0 | memory length: 961582 | epsilon: 0.009998020008555413 | steps: 424 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.53 | | | | | | |
| episode: 3865 | score: 6.0 | memory length: 961894 | epsilon: 0.009998020008555413 | steps: 312 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.52 | | | | | | |
| episode: 3866 | score: 5.0 | memory length: 962174 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.51 | | | | | | |
| episode: 3867 | score: 2.0 | memory length: 962355 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.47 | | | | | | |
| episode: 3868 | score: 9.0 | memory length: 962770 | epsilon: 0.009998020008555413 | steps: 415 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.51 | | | | | | |
| episode: 3869 | score: 2.0 | memory length: 962951 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.46 | | | | | | |
| episode: 3870 | score: 4.0 | memory length: 963196 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.47 | | | | | | |
| episode: 3871 | score: 5.0 | memory length: 963476 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.43 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3872 | score: 8.0 | memory length: 963871 | epsilon: 0.009998020008555413 | steps: 395 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.45 | | | | | | |
| episode: 3873 | score: 7.0 | memory length: 964244 | epsilon: 0.009998020008555413 | steps: 373 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.47 | | | | | | |
| episode: 3874 | score: 2.0 | memory length: 964425 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 3875 | score: 5.0 | memory length: 964705 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.39 | | | | | | |
| episode: 3876 | score: 2.0 | memory length: 964886 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3877 | score: 2.0 | memory length: 965067 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.31 | | | | | | |
| episode: 3878 | score: 4.0 | memory length: 965328 | epsilon: 0.009998020008555413 | steps: 261 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.28 | | | | | | |
| episode: 3879 | score: 2.0 | memory length: 965509 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.24 | | | | | | |
| episode: 3880 | score: 2.0 | memory length: 965690 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.15 | | | | | | |
| episode: 3881 | score: 7.0 | memory length: 966066 | epsilon: 0.009998020008555413 | steps: 376 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.17 | | | | | | |
| episode: 3882 | score: 3.0 | memory length: 966291 | epsilon: 0.009998020008555413 | steps: 225 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.13 | | | | | | |
| episode: 3883 | score: 4.0 | memory length: 966534 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.08 | | | | | | |
| episode: 3884 | score: 2.0 | memory length: 966715 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.08 | | | | | | |
| episode: 3885 | score: 6.0 | memory length: 967042 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.07 | | | | | | |
| episode: 3886 | score: 2.0 | memory length: 967223 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 3887 | score: 5.0 | memory length: 967550 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3888 | score: 3.0 | memory length: 967779 | epsilon: 0.009998020008555413 | steps: 229 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 3889 | score: 8.0 | memory length: 968188 | epsilon: 0.009998020008555413 | steps: 409 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3890 | score: 4.0 | memory length: 968438 | epsilon: 0.009998020008555413 | steps: 250 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.0 | | | | | | |
| episode: 3891 | score: 6.0 | memory length: 968783 | epsilon: 0.009998020008555413 | steps: 345 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.0 | | | | | | |
| episode: 3892 | score: 9.0 | memory length: 969210 | epsilon: 0.009998020008555413 | steps: 427 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 3893 | score: 6.0 | memory length: 969519 | epsilon: 0.009998020008555413 | steps: 309 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 3894 | score: 5.0 | memory length: 969817 | epsilon: 0.009998020008555413 | steps: 298 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.0 | | | | | | |
| episode: 3895 | score: 8.0 | memory length: 970224 | epsilon: 0.009998020008555413 | steps: 407 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 3896 | score: 5.0 | memory length: 970504 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 3897 | score: 2.0 | memory length: 970685 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.0 | | | | | | |
| episode: 3898 | score: 5.0 | memory length: 970965 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 3899 | score: 15.0 | memory length: 971367 | epsilon: 0.009998020008555413 | steps: 402 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.13 | | | | | | |
| episode: 3900 | score: 6.0 | memory length: 971714 | epsilon: 0.009998020008555413 | steps: 347 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.14 | | | | | | |
| episode: 3901 | score: 5.0 | memory length: 971993 | epsilon: 0.009998020008555413 | steps: 279 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.15 | | | | | | |
| episode: 3902 | score: 4.0 | memory length: 972236 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.11 | | | | | | |
| episode: 3903 | score: 4.0 | memory length: 972479 | epsilon: 0.009998020008555413 | steps: 243 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.01 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3904 | score: 7.0 | memory length: 972885 | epsilon: 0.009998020008555413 | steps: 406 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 3905 | score: 7.0 | memory length: 973244 | epsilon: 0.009998020008555413 | steps: 359 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 3906 | score: 6.0 | memory length: 973589 | epsilon: 0.009998020008555413 | steps: 345 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 3907 | score: 6.0 | memory length: 973914 | epsilon: 0.009998020008555413 | steps: 325 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.07 | | | | | | |
| episode: 3908 | score: 6.0 | memory length: 974227 | epsilon: 0.009998020008555413 | steps: 313 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.09 | | | | | | |
| episode: 3909 | score: 6.0 | memory length: 974536 | epsilon: 0.009998020008555413 | steps: 309 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.12 | | | | | | |
| episode: 3910 | score: 8.0 | memory length: 974980 | epsilon: 0.009998020008555413 | steps: 444 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.15 | | | | | | |
| episode: 3911 | score: 5.0 | memory length: 975260 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.17 | | | | | | |
| episode: 3912 | score: 11.0 | memory length: 975734 | epsilon: 0.009998020008555413 | steps: 474 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.25 | | | | | | |
| episode: 3913 | score: 6.0 | memory length: 976032 | epsilon: 0.009998020008555413 | steps: 298 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 3914 | score: 7.0 | memory length: 976368 | epsilon: 0.009998020008555413 | steps: 336 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.31 | | | | | | |
| episode: 3915 | score: 7.0 | memory length: 976704 | epsilon: 0.009998020008555413 | steps: 336 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.33 | | | | | | |
| episode: 3916 | score: 6.0 | memory length: 977013 | epsilon: 0.009998020008555413 | steps: 309 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.35 | | | | | | |
| episode: 3917 | score: 7.0 | memory length: 977366 | epsilon: 0.009998020008555413 | steps: 353 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3918 | score: 3.0 | memory length: 977611 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.35 | | | | | | |
| episode: 3919 | score: 8.0 | memory length: 977888 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.38 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3920 | score: 3.0 | memory length: 978133 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.35 | | | | | | |
| episode: 3921 | score: 6.0 | memory length: 978445 | epsilon: 0.009998020008555413 | steps: 312 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.33 | | | | | | |
| episode: 3922 | score: 8.0 | memory length: 978722 | epsilon: 0.009998020008555413 | steps: 277 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3923 | score: 7.0 | memory length: 979063 | epsilon: 0.009998020008555413 | steps: 341 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.38 | | | | | | |
| episode: 3924 | score: 3.0 | memory length: 979290 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.33 | | | | | | |
| episode: 3925 | score: 9.0 | memory length: 979694 | epsilon: 0.009998020008555413 | steps: 404 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3926 | score: 7.0 | memory length: 980054 | epsilon: 0.009998020008555413 | steps: 360 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.38 | | | | | | |
| episode: 3927 | score: 6.0 | memory length: 980363 | epsilon: 0.009998020008555413 | steps: 309 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.35 | | | | | | |
| episode: 3928 | score: 4.0 | memory length: 980613 | epsilon: 0.009998020008555413 | steps: 250 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3929 | score: 5.0 | memory length: 980893 | epsilon: 0.009998020008555413 | steps: 280 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3930 | score: 2.0 | memory length: 981074 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3931 | score: 10.0 | memory length: 981430 | epsilon: 0.009998020008555413 | steps: 356 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 3932 | score: 4.0 | memory length: 981688 | epsilon: 0.009998020008555413 | steps: 258 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3933 | score: 15.0 | memory length: 982234 | epsilon: 0.009998020008555413 | steps: 546 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.49 | | | | | | |
| episode: 3934 | score: 6.0 | memory length: 982590 | epsilon: 0.009998020008555413 | steps: 356 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.5 | | | | | | |
| episode: 3935 | score: 6.0 | memory length: 982895 | epsilon: 0.009998020008555413 | steps: 305 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.54 | | | | | | |

| | | | | | | |
|---------------|------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3936 | score: 4.0 | memory length: 983169 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.54 | | | | | | |
| episode: 3937 | score: 3.0 | memory length: 983401 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.54 | | | | | | |
| episode: 3938 | score: 5.0 | memory length: 983668 | epsilon: 0.009998020008555413 | steps: 267 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.5 | | | | | | |
| episode: 3939 | score: 2.0 | memory length: 983849 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.47 | | | | | | |
| episode: 3940 | score: 2.0 | memory length: 984030 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.44 | | | | | | |
| episode: 3941 | score: 2.0 | memory length: 984227 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.38 | | | | | | |
| episode: 3942 | score: 3.0 | memory length: 984453 | epsilon: 0.009998020008555413 | steps: 226 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.35 | | | | | | |
| episode: 3943 | score: 2.0 | memory length: 984634 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.33 | | | | | | |
| episode: 3944 | score: 4.0 | memory length: 984884 | epsilon: 0.009998020008555413 | steps: 250 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.29 | | | | | | |
| episode: 3945 | score: 4.0 | memory length: 985160 | epsilon: 0.009998020008555413 | steps: 276 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.3 | | | | | | |
| episode: 3946 | score: 5.0 | memory length: 985458 | epsilon: 0.009998020008555413 | steps: 298 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.26 | | | | | | |
| episode: 3947 | score: 3.0 | memory length: 985686 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 3948 | score: 4.0 | memory length: 985931 | epsilon: 0.009998020008555413 | steps: 245 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.29 | | | | | | |
| episode: 3949 | score: 7.0 | memory length: 986327 | epsilon: 0.009998020008555413 | steps: 396 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.34 | | | | | | |
| episode: 3950 | score: 7.0 | memory length: 986691 | epsilon: 0.009998020008555413 | steps: 364 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.37 | | | | | | |
| episode: 3951 | score: 3.0 | memory length: 986923 | epsilon: 0.009998020008555413 | steps: 232 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.35 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3952 | score: 12.0 | memory length: 987430 | epsilon: 0.009998020008555413 | steps: 507 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 3953 | score: 6.0 | memory length: 987755 | epsilon: 0.009998020008555413 | steps: 325 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 3954 | score: 2.0 | memory length: 987936 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3955 | score: 4.0 | memory length: 988210 | epsilon: 0.009998020008555413 | steps: 274 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.34 | | | | | | |
| episode: 3956 | score: 5.0 | memory length: 988494 | epsilon: 0.009998020008555413 | steps: 284 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.37 | | | | | | |
| episode: 3957 | score: 7.0 | memory length: 988862 | epsilon: 0.009998020008555413 | steps: 368 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.32 | | | | | | |
| episode: 3958 | score: 2.0 | memory length: 989043 | epsilon: 0.009998020008555413 | steps: 181 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.28 | | | | | | |
| episode: 3959 | score: 9.0 | memory length: 989467 | epsilon: 0.009998020008555413 | steps: 424 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.34 | | | | | | |
| episode: 3960 | score: 2.0 | memory length: 989664 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.31 | | | | | | |
| episode: 3961 | score: 3.0 | memory length: 989891 | epsilon: 0.009998020008555413 | steps: 227 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.3 | | | | | | |
| episode: 3962 | score: 2.0 | memory length: 990088 | epsilon: 0.009998020008555413 | steps: 197 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 3963 | score: 9.0 | memory length: 990512 | epsilon: 0.009998020008555413 | steps: 424 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.3 | | | | | | |
| episode: 3964 | score: 6.0 | memory length: 990810 | epsilon: 0.009998020008555413 | steps: 298 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 3965 | score: 7.0 | memory length: 991215 | epsilon: 0.009998020008555413 | steps: 405 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.28 | | | | | | |
| episode: 3966 | score: 13.0 | memory length: 991781 | epsilon: 0.009998020008555413 | steps: 566 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 3967 | score: 10.0 | memory length: 992273 | epsilon: 0.009998020008555413 | steps: 492 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.44 | | | | | | |

| | | | | | | |
|---------------|-------------|-----------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3968 | score: 6.0 | memory length: 992600 | epsilon: 0.009998020008555413 | steps: 327 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 3969 | score: 6.0 | memory length: 992975 | epsilon: 0.009998020008555413 | steps: 375 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.45 | | | | | | |
| episode: 3970 | score: 9.0 | memory length: 993381 | epsilon: 0.009998020008555413 | steps: 406 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.5 | | | | | | |
| episode: 3971 | score: 8.0 | memory length: 993754 | epsilon: 0.009998020008555413 | steps: 373 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.53 | | | | | | |
| episode: 3972 | score: 7.0 | memory length: 994108 | epsilon: 0.009998020008555413 | steps: 354 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.52 | | | | | | |
| episode: 3973 | score: 7.0 | memory length: 994501 | epsilon: 0.009998020008555413 | steps: 393 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.52 | | | | | | |
| episode: 3974 | score: 3.0 | memory length: 994729 | epsilon: 0.009998020008555413 | steps: 228 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.53 | | | | | | |
| episode: 3975 | score: 12.0 | memory length: 995244 | epsilon: 0.009998020008555413 | steps: 515 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.6 | | | | | | |
| episode: 3976 | score: 9.0 | memory length: 995692 | epsilon: 0.009998020008555413 | steps: 448 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.67 | | | | | | |
| episode: 3977 | score: 6.0 | memory length: 996035 | epsilon: 0.009998020008555413 | steps: 343 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.71 | | | | | | |
| episode: 3978 | score: 8.0 | memory length: 996441 | epsilon: 0.009998020008555413 | steps: 406 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.75 | | | | | | |
| episode: 3979 | score: 8.0 | memory length: 996848 | epsilon: 0.009998020008555413 | steps: 407 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.81 | | | | | | |
| episode: 3980 | score: 7.0 | memory length: 997204 | epsilon: 0.009998020008555413 | steps: 356 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.86 | | | | | | |
| episode: 3981 | score: 7.0 | memory length: 997562 | epsilon: 0.009998020008555413 | steps: 358 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.86 | | | | | | |
| episode: 3982 | score: 7.0 | memory length: 997902 | epsilon: 0.009998020008555413 | steps: 340 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.9 | | | | | | |
| episode: 3983 | score: 11.0 | memory length: 998373 | epsilon: 0.009998020008555413 | steps: 471 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 5.97 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 3984 | score: 6.0 | memory length: 998686 | epsilon: 0.009998020008555413 | steps: 313 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 6.01 | | | | | | |
| episode: 3985 | score: 9.0 | memory length: 999092 | epsilon: 0.009998020008555413 | steps: 406 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 6.04 | | | | | | |
| episode: 3986 | score: 6.0 | memory length: 999405 | epsilon: 0.009998020008555413 | steps: 313 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 6.08 | | | | | | |
| episode: 3987 | score: 6.0 | memory length: 999728 | epsilon: 0.009998020008555413 | steps: 323 | lr: 1.6000000000000003e-05 | evaluation |
| reward: 6.09 | | | | | | |
| episode: 3988 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.13 | | | | | | |
| episode: 3989 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.08 | | | | | | |
| episode: 3990 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 323 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.1 | | | | | | |
| episode: 3991 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 305 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.1 | | | | | | |
| episode: 3992 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 454 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.11 | | | | | | |
| episode: 3993 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 454 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.15 | | | | | | |
| episode: 3994 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 352 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.17 | | | | | | |
| episode: 3995 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.19 | | | | | | |
| episode: 3996 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 316 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.2 | | | | | | |
| episode: 3997 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.22 | | | | | | |
| episode: 3998 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.23 | | | | | | |
| episode: 3999 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 276 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.12 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4000 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.12 | | | | | | |
| episode: 4001 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.13 | | | | | | |
| episode: 4002 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.18 | | | | | | |
| episode: 4003 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.2 | | | | | | |
| episode: 4004 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 356 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.19 | | | | | | |
| episode: 4005 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 473 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.22 | | | | | | |
| episode: 4006 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 310 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.21 | | | | | | |
| episode: 4007 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.17 | | | | | | |
| episode: 4008 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 323 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.17 | | | | | | |
| episode: 4009 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 451 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.19 | | | | | | |
| episode: 4010 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 439 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.19 | | | | | | |
| episode: 4011 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 381 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.22 | | | | | | |
| episode: 4012 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.17 | | | | | | |
| episode: 4013 | score: 12.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 498 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.23 | | | | | | |
| episode: 4014 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.25 | | | | | | |
| episode: 4015 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 442 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.26 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4016 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.29 | | | | | | |
| episode: 4017 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 381 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.3 | | | | | | |
| episode: 4018 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 334 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.34 | | | | | | |
| episode: 4019 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 305 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.32 | | | | | | |
| episode: 4020 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.38 | | | | | | |
| episode: 4021 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 451 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.42 | | | | | | |
| episode: 4022 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 323 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.4 | | | | | | |
| episode: 4023 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.42 | | | | | | |
| episode: 4024 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 279 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.43 | | | | | | |
| episode: 4025 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.43 | | | | | | |
| episode: 4026 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.42 | | | | | | |
| episode: 4027 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.46 | | | | | | |
| episode: 4028 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 249 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.45 | | | | | | |
| episode: 4029 | score: 14.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 676 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.54 | | | | | | |
| episode: 4030 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 403 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.61 | | | | | | |
| episode: 4031 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.57 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4032 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 330 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.59 | | | | | | |
| episode: 4033 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.49 | | | | | | |
| episode: 4034 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 409 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.5 | | | | | | |
| episode: 4035 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.47 | | | | | | |
| episode: 4036 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.52 | | | | | | |
| episode: 4037 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.55 | | | | | | |
| episode: 4038 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.56 | | | | | | |
| episode: 4039 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 245 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.57 | | | | | | |
| episode: 4040 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.57 | | | | | | |
| episode: 4041 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 210 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.58 | | | | | | |
| episode: 4042 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.57 | | | | | | |
| episode: 4043 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.57 | | | | | | |
| episode: 4044 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.55 | | | | | | |
| episode: 4045 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 259 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.55 | | | | | | |
| episode: 4046 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.52 | | | | | | |
| episode: 4047 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.51 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4048 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 245 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.5 | | | | | | |
| episode: 4049 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.49 | | | | | | |
| episode: 4050 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 336 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.48 | | | | | | |
| episode: 4051 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 276 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.49 | | | | | | |
| episode: 4052 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 197 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.39 | | | | | | |
| episode: 4053 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 198 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.35 | | | | | | |
| episode: 4054 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 210 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.36 | | | | | | |
| episode: 4055 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.39 | | | | | | |
| episode: 4056 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 261 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.38 | | | | | | |
| episode: 4057 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 241 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.35 | | | | | | |
| episode: 4058 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 443 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.41 | | | | | | |
| episode: 4059 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.35 | | | | | | |
| episode: 4060 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 285 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.38 | | | | | | |
| episode: 4061 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.42 | | | | | | |
| episode: 4062 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 355 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.47 | | | | | | |
| episode: 4063 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 405 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.46 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|---------------------------|------------|
| episode: 4064 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.45 | | | | | | |
| episode: 4065 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 334 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.45 | | | | | | |
| episode: 4066 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.38 | | | | | | |
| episode: 4067 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 328 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.38 | | | | | | |
| episode: 4068 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 197 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.34 | | | | | | |
| episode: 4069 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 245 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.31 | | | | | | |
| episode: 4070 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 318 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.28 | | | | | | |
| episode: 4071 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 389 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.28 | | | | | | |
| episode: 4072 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.23 | | | | | | |
| episode: 4073 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 199 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.18 | | | | | | |
| episode: 4074 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 258 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.19 | | | | | | |
| episode: 4075 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 291 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.12 | | | | | | |
| episode: 4076 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 336 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.09 | | | | | | |
| episode: 4077 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 291 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.08 | | | | | | |
| episode: 4078 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.07 | | | | | | |
| episode: 4079 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.400000000000001e-06 | evaluation |
| reward: 6.06 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4080 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 228 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.02 | | | | | | |
| episode: 4081 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.0 | | | | | | |
| episode: 4082 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 355 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.0 | | | | | | |
| episode: 4083 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 249 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.92 | | | | | | |
| episode: 4084 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.95 | | | | | | |
| episode: 4085 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 343 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.92 | | | | | | |
| episode: 4086 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.92 | | | | | | |
| episode: 4087 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.93 | | | | | | |
| episode: 4088 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.91 | | | | | | |
| episode: 4089 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.97 | | | | | | |
| episode: 4094 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 391 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.92 | | | | | | |
| episode: 4095 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.92 | | | | | | |
| episode: 4096 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.95 | | | | | | |
| episode: 4097 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.0 | | | | | | |
| episode: 4098 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.03 | | | | | | |
| episode: 4099 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 259 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.03 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4100 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.06 | | | | | | |
| episode: 4101 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 277 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.08 | | | | | | |
| episode: 4102 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 258 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.03 | | | | | | |
| episode: 4103 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 357 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.03 | | | | | | |
| episode: 4104 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 272 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.02 | | | | | | |
| episode: 4105 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 229 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.95 | | | | | | |
| episode: 4106 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 306 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.96 | | | | | | |
| episode: 4107 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 275 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.99 | | | | | | |
| episode: 4108 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 290 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.98 | | | | | | |
| episode: 4109 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 473 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.0 | | | | | | |
| episode: 4110 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.02 | | | | | | |
| episode: 4111 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.03 | | | | | | |
| episode: 4112 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.07 | | | | | | |
| episode: 4113 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 490 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.04 | | | | | | |
| episode: 4114 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.0 | | | | | | |
| episode: 4115 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.98 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4116 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.96 | | | | | | |
| episode: 4117 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.93 | | | | | | |
| episode: 4118 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 342 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.92 | | | | | | |
| episode: 4119 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 543 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.97 | | | | | | |
| episode: 4120 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 373 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.94 | | | | | | |
| episode: 4121 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.9 | | | | | | |
| episode: 4122 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 383 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.92 | | | | | | |
| episode: 4123 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.88 | | | | | | |
| episode: 4124 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 336 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.9 | | | | | | |
| episode: 4125 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.86 | | | | | | |
| episode: 4126 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.84 | | | | | | |
| episode: 4127 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.79 | | | | | | |
| episode: 4128 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.81 | | | | | | |
| episode: 4129 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.71 | | | | | | |
| episode: 4130 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.67 | | | | | | |
| episode: 4131 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.67 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4132 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 380 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.68 | | | | | | |
| episode: 4133 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 467 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.73 | | | | | | |
| episode: 4134 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 382 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.74 | | | | | | |
| episode: 4135 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.76 | | | | | | |
| episode: 4136 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.72 | | | | | | |
| episode: 4137 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 397 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.73 | | | | | | |
| episode: 4138 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 410 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.75 | | | | | | |
| episode: 4139 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 303 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.78 | | | | | | |
| episode: 4140 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.82 | | | | | | |
| episode: 4141 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.85 | | | | | | |
| episode: 4142 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.86 | | | | | | |
| episode: 4143 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.9 | | | | | | |
| episode: 4144 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 356 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.95 | | | | | | |
| episode: 4145 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 254 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.95 | | | | | | |
| episode: 4146 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 321 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.98 | | | | | | |
| episode: 4147 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 249 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.99 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4148 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.03 | | | | | | |
| episode: 4149 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 376 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.04 | | | | | | |
| episode: 4150 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 403 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.06 | | | | | | |
| episode: 4151 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 348 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.08 | | | | | | |
| episode: 4152 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.12 | | | | | | |
| episode: 4153 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 327 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.16 | | | | | | |
| episode: 4154 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.18 | | | | | | |
| episode: 4155 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.17 | | | | | | |
| episode: 4156 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 412 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.21 | | | | | | |
| episode: 4157 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 513 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.28 | | | | | | |
| episode: 4158 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 180 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.22 | | | | | | |
| episode: 4159 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 357 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.25 | | | | | | |
| episode: 4160 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 413 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.28 | | | | | | |
| episode: 4161 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 358 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.28 | | | | | | |
| episode: 4162 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 340 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.28 | | | | | | |
| episode: 4163 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.26 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4164 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 278 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.25 | | | | | | |
| episode: 4165 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.24 | | | | | | |
| episode: 4166 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.23 | | | | | | |
| episode: 4167 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 416 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.22 | | | | | | |
| episode: 4168 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.26 | | | | | | |
| episode: 4169 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 397 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.3 | | | | | | |
| episode: 4170 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.31 | | | | | | |
| episode: 4171 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.3 | | | | | | |
| episode: 4172 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 479 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.38 | | | | | | |
| episode: 4173 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.41 | | | | | | |
| episode: 4174 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 323 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.43 | | | | | | |
| episode: 4175 | score: 16.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 524 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.54 | | | | | | |
| episode: 4176 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 270 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.56 | | | | | | |
| episode: 4177 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.6 | | | | | | |
| episode: 4178 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 294 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.57 | | | | | | |
| episode: 4179 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 342 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.56 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4180 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 349 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.59 | | | | | | |
| episode: 4181 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 344 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.61 | | | | | | |
| episode: 4182 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 346 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.6 | | | | | | |
| episode: 4183 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 343 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.63 | | | | | | |
| episode: 4184 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 305 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.6 | | | | | | |
| episode: 4185 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.61 | | | | | | |
| episode: 4186 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.61 | | | | | | |
| episode: 4187 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 321 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.6 | | | | | | |
| episode: 4188 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.6 | | | | | | |
| episode: 4189 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 399 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.59 | | | | | | |
| episode: 4190 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 385 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.64 | | | | | | |
| episode: 4191 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.64 | | | | | | |
| episode: 4192 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 379 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.62 | | | | | | |
| episode: 4193 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.63 | | | | | | |
| episode: 4194 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 399 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.62 | | | | | | |
| episode: 4195 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.59 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4196 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.6 | | | | | | |
| episode: 4197 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.57 | | | | | | |
| episode: 4198 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 397 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.55 | | | | | | |
| episode: 4199 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.57 | | | | | | |
| episode: 4200 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.54 | | | | | | |
| episode: 4201 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 511 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.57 | | | | | | |
| episode: 4202 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.59 | | | | | | |
| episode: 4203 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 444 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.62 | | | | | | |
| episode: 4204 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.66 | | | | | | |
| episode: 4205 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 435 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.72 | | | | | | |
| episode: 4206 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.71 | | | | | | |
| episode: 4207 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.72 | | | | | | |
| episode: 4208 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.76 | | | | | | |
| episode: 4209 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.73 | | | | | | |
| episode: 4210 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.68 | | | | | | |
| episode: 4211 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 387 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.67 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4212 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.63 | | | | | | |
| episode: 4213 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.59 | | | | | | |
| episode: 4214 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 349 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.6 | | | | | | |
| episode: 4215 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.6 | | | | | | |
| episode: 4216 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.62 | | | | | | |
| episode: 4217 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.66 | | | | | | |
| episode: 4218 | score: 12.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 492 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.72 | | | | | | |
| episode: 4219 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.71 | | | | | | |
| episode: 4220 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.7 | | | | | | |
| episode: 4221 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.71 | | | | | | |
| episode: 4222 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.68 | | | | | | |
| episode: 4223 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 380 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.7 | | | | | | |
| episode: 4224 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 474 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.75 | | | | | | |
| episode: 4225 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 407 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.78 | | | | | | |
| episode: 4226 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.81 | | | | | | |
| episode: 4227 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.83 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4228 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.84 | | | | | | |
| episode: 4229 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 422 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.88 | | | | | | |
| episode: 4230 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 471 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.93 | | | | | | |
| episode: 4231 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 473 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.97 | | | | | | |
| episode: 4232 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 473 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.0 | | | | | | |
| episode: 4233 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.94 | | | | | | |
| episode: 4234 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.96 | | | | | | |
| episode: 4235 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.0 | | | | | | |
| episode: 4236 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 395 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.03 | | | | | | |
| episode: 4237 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.98 | | | | | | |
| episode: 4238 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.96 | | | | | | |
| episode: 4239 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 350 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.97 | | | | | | |
| episode: 4240 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 360 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.99 | | | | | | |
| episode: 4241 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 473 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.03 | | | | | | |
| episode: 4242 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 473 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.1 | | | | | | |
| episode: 4243 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 460 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.14 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4244 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.16 | | | | | | |
| episode: 4245 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.19 | | | | | | |
| episode: 4246 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 395 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.22 | | | | | | |
| episode: 4247 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.28 | | | | | | |
| episode: 4248 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 390 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.3 | | | | | | |
| episode: 4249 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.28 | | | | | | |
| episode: 4250 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 425 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.28 | | | | | | |
| episode: 4251 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.28 | | | | | | |
| episode: 4252 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 499 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.31 | | | | | | |
| episode: 4253 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.31 | | | | | | |
| episode: 4254 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 210 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.29 | | | | | | |
| episode: 4255 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 332 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.29 | | | | | | |
| episode: 4256 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 393 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.29 | | | | | | |
| episode: 4257 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 382 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.25 | | | | | | |
| episode: 4258 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.33 | | | | | | |
| episode: 4259 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.33 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4260 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 398 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.33 | | | | | | |
| episode: 4261 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.31 | | | | | | |
| episode: 4262 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.3 | | | | | | |
| episode: 4263 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.33 | | | | | | |
| episode: 4264 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.31 | | | | | | |
| episode: 4265 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.3 | | | | | | |
| episode: 4266 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.34 | | | | | | |
| episode: 4267 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.34 | | | | | | |
| episode: 4268 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 383 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.36 | | | | | | |
| episode: 4269 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.36 | | | | | | |
| episode: 4270 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 393 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.38 | | | | | | |
| episode: 4271 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.4 | | | | | | |
| episode: 4272 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.4 | | | | | | |
| episode: 4273 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.45 | | | | | | |
| episode: 4274 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 409 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.48 | | | | | | |
| episode: 4275 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 378 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.39 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4276 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 373 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.38 | | | | | | |
| episode: 4277 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 344 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.36 | | | | | | |
| episode: 4278 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.37 | | | | | | |
| episode: 4279 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.36 | | | | | | |
| episode: 4280 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 437 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.39 | | | | | | |
| episode: 4281 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 460 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.41 | | | | | | |
| episode: 4282 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 357 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.41 | | | | | | |
| episode: 4283 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 459 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.44 | | | | | | |
| episode: 4284 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 355 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.44 | | | | | | |
| episode: 4285 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 425 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.45 | | | | | | |
| episode: 4286 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 344 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.46 | | | | | | |
| episode: 4287 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 356 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.47 | | | | | | |
| episode: 4288 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 454 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.51 | | | | | | |
| episode: 4289 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 343 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.49 | | | | | | |
| episode: 4290 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 310 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.46 | | | | | | |
| episode: 4291 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.47 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4292 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 421 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.5 | | | | | | |
| episode: 4293 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 396 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.49 | | | | | | |
| episode: 4294 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 321 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.47 | | | | | | |
| episode: 4295 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 382 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.48 | | | | | | |
| episode: 4296 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.44 | | | | | | |
| episode: 4297 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.47 | | | | | | |
| episode: 4298 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.5 | | | | | | |
| episode: 4299 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.54 | | | | | | |
| episode: 4300 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.57 | | | | | | |
| episode: 4301 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 310 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.51 | | | | | | |
| episode: 4302 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 444 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.55 | | | | | | |
| episode: 4303 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 354 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.54 | | | | | | |
| episode: 4304 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 360 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.53 | | | | | | |
| episode: 4305 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 356 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.51 | | | | | | |
| episode: 4306 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.52 | | | | | | |
| episode: 4307 | score: 13.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 448 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.59 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4308 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.6 | | | | | | |
| episode: 4309 | score: 15.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 501 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.68 | | | | | | |
| episode: 4310 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 332 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.7 | | | | | | |
| episode: 4311 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 452 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.71 | | | | | | |
| episode: 4312 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.74 | | | | | | |
| episode: 4313 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 405 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.77 | | | | | | |
| episode: 4314 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.75 | | | | | | |
| episode: 4315 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.76 | | | | | | |
| episode: 4316 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.74 | | | | | | |
| episode: 4317 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 356 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.72 | | | | | | |
| episode: 4318 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 393 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.69 | | | | | | |
| episode: 4319 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 542 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.7 | | | | | | |
| episode: 4320 | score: 12.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 523 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.77 | | | | | | |
| episode: 4321 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.8 | | | | | | |
| episode: 4322 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 474 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.86 | | | | | | |
| episode: 4323 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.86 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4324 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 542 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.86 | | | | | | |
| episode: 4325 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.81 | | | | | | |
| episode: 4326 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.8 | | | | | | |
| episode: 4327 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.78 | | | | | | |
| episode: 4328 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 344 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.78 | | | | | | |
| episode: 4329 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.77 | | | | | | |
| episode: 4330 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 360 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.74 | | | | | | |
| episode: 4331 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.71 | | | | | | |
| episode: 4332 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.67 | | | | | | |
| episode: 4333 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.69 | | | | | | |
| episode: 4334 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.66 | | | | | | |
| episode: 4335 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 296 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.62 | | | | | | |
| episode: 4336 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 356 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.61 | | | | | | |
| episode: 4337 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 477 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.68 | | | | | | |
| episode: 4338 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.68 | | | | | | |
| episode: 4339 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.67 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4340 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.64 | | | | | | |
| episode: 4341 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.6 | | | | | | |
| episode: 4342 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.57 | | | | | | |
| episode: 4343 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 357 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.54 | | | | | | |
| episode: 4344 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.52 | | | | | | |
| episode: 4345 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 409 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.53 | | | | | | |
| episode: 4346 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 343 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.51 | | | | | | |
| episode: 4347 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 399 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.53 | | | | | | |
| episode: 4348 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.51 | | | | | | |
| episode: 4349 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.5 | | | | | | |
| episode: 4350 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 352 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.49 | | | | | | |
| episode: 4351 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.48 | | | | | | |
| episode: 4352 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.45 | | | | | | |
| episode: 4353 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.44 | | | | | | |
| episode: 4354 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.48 | | | | | | |
| episode: 4355 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 304 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.47 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4356 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 303 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.45 | | | | | | |
| episode: 4357 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 312 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.44 | | | | | | |
| episode: 4358 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 376 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.41 | | | | | | |
| episode: 4359 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.4 | | | | | | |
| episode: 4360 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.39 | | | | | | |
| episode: 4361 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.41 | | | | | | |
| episode: 4362 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 242 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.39 | | | | | | |
| episode: 4363 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.36 | | | | | | |
| episode: 4364 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 343 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.4 | | | | | | |
| episode: 4365 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.4 | | | | | | |
| episode: 4366 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.38 | | | | | | |
| episode: 4367 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.36 | | | | | | |
| episode: 4368 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 352 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.35 | | | | | | |
| episode: 4369 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 397 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.35 | | | | | | |
| episode: 4370 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 350 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.33 | | | | | | |
| episode: 4371 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 357 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.31 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4372 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 305 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.27 | | | | | | |
| episode: 4373 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.26 | | | | | | |
| episode: 4374 | score: 13.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 544 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.3 | | | | | | |
| episode: 4375 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.32 | | | | | | |
| episode: 4376 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.34 | | | | | | |
| episode: 4377 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.37 | | | | | | |
| episode: 4378 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 346 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.39 | | | | | | |
| episode: 4379 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 436 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.44 | | | | | | |
| episode: 4380 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 405 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.43 | | | | | | |
| episode: 4381 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.41 | | | | | | |
| episode: 4382 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.42 | | | | | | |
| episode: 4383 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 427 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.4 | | | | | | |
| episode: 4384 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.41 | | | | | | |
| episode: 4385 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 426 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.41 | | | | | | |
| episode: 4386 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.41 | | | | | | |
| episode: 4387 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 409 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.41 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4388 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 343 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.38 | | | | | | |
| episode: 4389 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 246 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.35 | | | | | | |
| episode: 4390 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 355 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.36 | | | | | | |
| episode: 4391 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.35 | | | | | | |
| episode: 4392 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 354 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.33 | | | | | | |
| episode: 4393 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 328 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.31 | | | | | | |
| episode: 4394 | score: 13.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 544 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.38 | | | | | | |
| episode: 4395 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.39 | | | | | | |
| episode: 4396 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.39 | | | | | | |
| episode: 4397 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.39 | | | | | | |
| episode: 4398 | score: 12.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 503 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.41 | | | | | | |
| episode: 4399 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.34 | | | | | | |
| episode: 4400 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.32 | | | | | | |
| episode: 4401 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 495 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.37 | | | | | | |
| episode: 4402 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.34 | | | | | | |
| episode: 4403 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 432 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.35 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4404 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 477 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.38 | | | | | | |
| episode: 4405 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.38 | | | | | | |
| episode: 4406 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 359 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.39 | | | | | | |
| episode: 4407 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 474 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.37 | | | | | | |
| episode: 4408 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 420 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.34 | | | | | | |
| episode: 4409 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 467 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.29 | | | | | | |
| episode: 4410 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.27 | | | | | | |
| episode: 4411 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 290 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.23 | | | | | | |
| episode: 4412 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 243 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.18 | | | | | | |
| episode: 4413 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 391 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.16 | | | | | | |
| episode: 4414 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 421 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.2 | | | | | | |
| episode: 4415 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.18 | | | | | | |
| episode: 4416 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.2 | | | | | | |
| episode: 4417 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 473 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.23 | | | | | | |
| episode: 4418 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.19 | | | | | | |
| episode: 4419 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 357 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.15 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4420 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.06 | | | | | | |
| episode: 4421 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.99 | | | | | | |
| episode: 4422 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.91 | | | | | | |
| episode: 4423 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.87 | | | | | | |
| episode: 4424 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.79 | | | | | | |
| episode: 4425 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.79 | | | | | | |
| episode: 4426 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.76 | | | | | | |
| episode: 4427 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.74 | | | | | | |
| episode: 4428 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 325 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.74 | | | | | | |
| episode: 4429 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.72 | | | | | | |
| episode: 4430 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 390 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.74 | | | | | | |
| episode: 4431 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.73 | | | | | | |
| episode: 4432 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 457 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.76 | | | | | | |
| episode: 4433 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 352 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.77 | | | | | | |
| episode: 4434 | score: 13.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 544 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.83 | | | | | | |
| episode: 4435 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 440 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.89 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4436 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 352 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.89 | | | | | | |
| episode: 4437 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 447 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.9 | | | | | | |
| episode: 4438 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 350 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.91 | | | | | | |
| episode: 4439 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.88 | | | | | | |
| episode: 4440 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.87 | | | | | | |
| episode: 4441 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 444 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.91 | | | | | | |
| episode: 4442 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.87 | | | | | | |
| episode: 4443 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 459 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.9 | | | | | | |
| episode: 4444 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 303 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.89 | | | | | | |
| episode: 4445 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.86 | | | | | | |
| episode: 4446 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 352 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.87 | | | | | | |
| episode: 4447 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 350 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.83 | | | | | | |
| episode: 4448 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.85 | | | | | | |
| episode: 4449 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.9 | | | | | | |
| episode: 4450 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.92 | | | | | | |
| episode: 4451 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 230 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.9 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4452 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 230 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.87 | | | | | | |
| episode: 4453 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 344 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.89 | | | | | | |
| episode: 4454 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.88 | | | | | | |
| episode: 4455 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.86 | | | | | | |
| episode: 4456 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 352 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.87 | | | | | | |
| episode: 4457 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.9 | | | | | | |
| episode: 4458 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.92 | | | | | | |
| episode: 4459 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 305 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.93 | | | | | | |
| episode: 4460 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 268 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.91 | | | | | | |
| episode: 4461 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 449 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.93 | | | | | | |
| episode: 4462 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.98 | | | | | | |
| episode: 4463 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.98 | | | | | | |
| episode: 4464 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.01 | | | | | | |
| episode: 4465 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.0 | | | | | | |
| episode: 4466 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.97 | | | | | | |
| episode: 4467 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.93 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4468 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 230 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.89 | | | | | | |
| episode: 4469 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.87 | | | | | | |
| episode: 4470 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.83 | | | | | | |
| episode: 4471 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.79 | | | | | | |
| episode: 4472 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.77 | | | | | | |
| episode: 4473 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 294 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.73 | | | | | | |
| episode: 4474 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 230 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.63 | | | | | | |
| episode: 4475 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 276 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.58 | | | | | | |
| episode: 4476 | score: 12.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 498 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.61 | | | | | | |
| episode: 4477 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.54 | | | | | | |
| episode: 4478 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.56 | | | | | | |
| episode: 4479 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.49 | | | | | | |
| episode: 4480 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 258 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.45 | | | | | | |
| episode: 4481 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 294 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.43 | | | | | | |
| episode: 4482 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.39 | | | | | | |
| episode: 4483 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 262 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.36 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4484 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 373 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.37 | | | | | | |
| episode: 4485 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 307 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.34 | | | | | | |
| episode: 4486 | score: 16.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 584 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.43 | | | | | | |
| episode: 4487 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 476 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.47 | | | | | | |
| episode: 4488 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.43 | | | | | | |
| episode: 4489 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 420 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.49 | | | | | | |
| episode: 4490 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.46 | | | | | | |
| episode: 4491 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 499 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.52 | | | | | | |
| episode: 4492 | score: 14.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 530 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.58 | | | | | | |
| episode: 4493 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.57 | | | | | | |
| episode: 4494 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 238 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.48 | | | | | | |
| episode: 4495 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.42 | | | | | | |
| episode: 4496 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.39 | | | | | | |
| episode: 4497 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.39 | | | | | | |
| episode: 4498 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 281 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.31 | | | | | | |
| episode: 4499 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 354 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.34 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4500 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 245 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.31 | | | | | | |
| episode: 4501 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.25 | | | | | | |
| episode: 4502 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.21 | | | | | | |
| episode: 4503 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.2 | | | | | | |
| episode: 4504 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.12 | | | | | | |
| episode: 4505 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.08 | | | | | | |
| episode: 4506 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.05 | | | | | | |
| episode: 4507 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.97 | | | | | | |
| episode: 4508 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.96 | | | | | | |
| episode: 4509 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.89 | | | | | | |
| episode: 4510 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.87 | | | | | | |
| episode: 4511 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 279 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.86 | | | | | | |
| episode: 4512 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 340 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.88 | | | | | | |
| episode: 4513 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 308 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.87 | | | | | | |
| episode: 4514 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 244 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.82 | | | | | | |
| episode: 4515 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 245 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.81 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4516 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.75 | | | | | | |
| episode: 4517 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 276 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.69 | | | | | | |
| episode: 4518 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 274 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.68 | | | | | | |
| episode: 4519 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.64 | | | | | | |
| episode: 4520 | score: 18.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 623 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.79 | | | | | | |
| episode: 4521 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 230 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.79 | | | | | | |
| episode: 4522 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 358 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.82 | | | | | | |
| episode: 4523 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.88 | | | | | | |
| episode: 4524 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.88 | | | | | | |
| episode: 4525 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 246 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.89 | | | | | | |
| episode: 4526 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 347 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.92 | | | | | | |
| episode: 4527 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 376 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.96 | | | | | | |
| episode: 4528 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.92 | | | | | | |
| episode: 4529 | score: 17.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 598 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.04 | | | | | | |
| episode: 4530 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.98 | | | | | | |
| episode: 4531 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.95 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4532 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 343 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.92 | | | | | | |
| episode: 4533 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 506 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.95 | | | | | | |
| episode: 4534 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 213 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.85 | | | | | | |
| episode: 4535 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.83 | | | | | | |
| episode: 4536 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.85 | | | | | | |
| episode: 4537 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.84 | | | | | | |
| episode: 4538 | score: 12.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 520 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.89 | | | | | | |
| episode: 4539 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 446 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.96 | | | | | | |
| episode: 4540 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 473 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.02 | | | | | | |
| episode: 4541 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.97 | | | | | | |
| episode: 4542 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.03 | | | | | | |
| episode: 4543 | score: 15.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 524 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.08 | | | | | | |
| episode: 4544 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.08 | | | | | | |
| episode: 4545 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 473 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.13 | | | | | | |
| episode: 4546 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.15 | | | | | | |
| episode: 4547 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.17 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4548 | score: 12.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 534 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.2 | | | | | | |
| episode: 4549 | score: 12.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 511 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.23 | | | | | | |
| episode: 4550 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 226 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.17 | | | | | | |
| episode: 4551 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 364 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.21 | | | | | | |
| episode: 4552 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 393 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.24 | | | | | | |
| episode: 4553 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.26 | | | | | | |
| episode: 4554 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 376 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.27 | | | | | | |
| episode: 4555 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.27 | | | | | | |
| episode: 4556 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 238 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.24 | | | | | | |
| episode: 4557 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 278 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.2 | | | | | | |
| episode: 4558 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.17 | | | | | | |
| episode: 4559 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.2 | | | | | | |
| episode: 4560 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 506 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.26 | | | | | | |
| episode: 4561 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.2 | | | | | | |
| episode: 4562 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.14 | | | | | | |
| episode: 4563 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 278 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.13 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4564 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 226 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.07 | | | | | | |
| episode: 4565 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 362 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.1 | | | | | | |
| episode: 4566 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 449 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.16 | | | | | | |
| episode: 4567 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.16 | | | | | | |
| episode: 4568 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.16 | | | | | | |
| episode: 4569 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.14 | | | | | | |
| episode: 4570 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 418 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.2 | | | | | | |
| episode: 4571 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.26 | | | | | | |
| episode: 4572 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.31 | | | | | | |
| episode: 4573 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 500 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.37 | | | | | | |
| episode: 4574 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.43 | | | | | | |
| episode: 4575 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 227 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.42 | | | | | | |
| episode: 4576 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 433 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.4 | | | | | | |
| episode: 4577 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 375 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.45 | | | | | | |
| episode: 4578 | score: 15.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 501 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.51 | | | | | | |
| episode: 4579 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 360 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.56 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4580 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 433 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.62 | | | | | | |
| episode: 4581 | score: 14.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 612 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.71 | | | | | | |
| episode: 4582 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.73 | | | | | | |
| episode: 4583 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.75 | | | | | | |
| episode: 4584 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 275 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.71 | | | | | | |
| episode: 4585 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 344 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.73 | | | | | | |
| episode: 4586 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 381 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.65 | | | | | | |
| episode: 4587 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.59 | | | | | | |
| episode: 4588 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.63 | | | | | | |
| episode: 4589 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.59 | | | | | | |
| episode: 4590 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.64 | | | | | | |
| episode: 4591 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 296 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.58 | | | | | | |
| episode: 4592 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.49 | | | | | | |
| episode: 4593 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 327 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.5 | | | | | | |
| episode: 4594 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 272 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.54 | | | | | | |
| episode: 4595 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 486 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.61 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4596 | score: 12.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 394 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.7 | | | | | | |
| episode: 4597 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 375 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.68 | | | | | | |
| episode: 4598 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 332 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.7 | | | | | | |
| episode: 4599 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.69 | | | | | | |
| episode: 4600 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 457 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.74 | | | | | | |
| episode: 4601 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.74 | | | | | | |
| episode: 4602 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 331 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.77 | | | | | | |
| episode: 4603 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 499 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.8 | | | | | | |
| episode: 4604 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.86 | | | | | | |
| episode: 4605 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 332 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.89 | | | | | | |
| episode: 4606 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 500 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.95 | | | | | | |
| episode: 4607 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 360 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.99 | | | | | | |
| episode: 4608 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 429 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.02 | | | | | | |
| episode: 4609 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.04 | | | | | | |
| episode: 4610 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.06 | | | | | | |
| episode: 4611 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 348 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.08 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4612 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 327 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.08 | | | | | | |
| episode: 4613 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.1 | | | | | | |
| episode: 4614 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 327 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.13 | | | | | | |
| episode: 4615 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 333 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.15 | | | | | | |
| episode: 4616 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.17 | | | | | | |
| episode: 4617 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.18 | | | | | | |
| episode: 4618 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 327 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.2 | | | | | | |
| episode: 4619 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 405 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.25 | | | | | | |
| episode: 4620 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 331 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.13 | | | | | | |
| episode: 4621 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 327 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.16 | | | | | | |
| episode: 4622 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 314 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.16 | | | | | | |
| episode: 4623 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 334 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.13 | | | | | | |
| episode: 4624 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 327 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.16 | | | | | | |
| episode: 4625 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 426 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.2 | | | | | | |
| episode: 4626 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.22 | | | | | | |
| episode: 4627 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.22 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4628 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 274 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.24 | | | | | | |
| episode: 4629 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 367 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.15 | | | | | | |
| episode: 4630 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 378 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.19 | | | | | | |
| episode: 4631 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.24 | | | | | | |
| episode: 4632 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.25 | | | | | | |
| episode: 4633 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 230 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.18 | | | | | | |
| episode: 4634 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 379 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.22 | | | | | | |
| episode: 4635 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 437 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.22 | | | | | | |
| episode: 4636 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 396 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.2 | | | | | | |
| episode: 4637 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 471 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.21 | | | | | | |
| episode: 4638 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 396 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.16 | | | | | | |
| episode: 4639 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 379 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.14 | | | | | | |
| episode: 4640 | score: 13.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 544 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.17 | | | | | | |
| episode: 4641 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 476 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.21 | | | | | | |
| episode: 4642 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.19 | | | | | | |
| episode: 4643 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 501 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.14 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4644 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 479 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.17 | | | | | | |
| episode: 4645 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.11 | | | | | | |
| episode: 4646 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 381 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.09 | | | | | | |
| episode: 4647 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 453 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.09 | | | | | | |
| episode: 4648 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.05 | | | | | | |
| episode: 4649 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 357 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.0 | | | | | | |
| episode: 4650 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 397 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.04 | | | | | | |
| episode: 4651 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 376 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.04 | | | | | | |
| episode: 4652 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 449 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.07 | | | | | | |
| episode: 4653 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.05 | | | | | | |
| episode: 4654 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.03 | | | | | | |
| episode: 4655 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.02 | | | | | | |
| episode: 4656 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.05 | | | | | | |
| episode: 4657 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.07 | | | | | | |
| episode: 4658 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 292 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.05 | | | | | | |
| episode: 4659 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 476 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.05 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4660 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 350 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.0 | | | | | | |
| episode: 4661 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.0 | | | | | | |
| episode: 4662 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 453 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.06 | | | | | | |
| episode: 4663 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 396 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.08 | | | | | | |
| episode: 4664 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.08 | | | | | | |
| episode: 4665 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 371 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.08 | | | | | | |
| episode: 4666 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.01 | | | | | | |
| episode: 4667 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 414 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.07 | | | | | | |
| episode: 4668 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.09 | | | | | | |
| episode: 4669 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 503 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.16 | | | | | | |
| episode: 4670 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 396 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.14 | | | | | | |
| episode: 4671 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.1 | | | | | | |
| episode: 4672 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 311 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.07 | | | | | | |
| episode: 4673 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 7.01 | | | | | | |
| episode: 4674 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.95 | | | | | | |
| episode: 4675 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.97 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4676 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 441 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.96 | | | | | | |
| episode: 4677 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 374 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.96 | | | | | | |
| episode: 4678 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.84 | | | | | | |
| episode: 4679 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 308 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.82 | | | | | | |
| episode: 4680 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.75 | | | | | | |
| episode: 4681 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.66 | | | | | | |
| episode: 4682 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.66 | | | | | | |
| episode: 4683 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.65 | | | | | | |
| episode: 4684 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 327 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.67 | | | | | | |
| episode: 4685 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.66 | | | | | | |
| episode: 4686 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.62 | | | | | | |
| episode: 4687 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.63 | | | | | | |
| episode: 4688 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.62 | | | | | | |
| episode: 4689 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 274 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.61 | | | | | | |
| episode: 4690 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.58 | | | | | | |
| episode: 4691 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 278 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.58 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4692 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 317 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.59 | | | | | | |
| episode: 4693 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 360 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.6 | | | | | | |
| episode: 4694 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.57 | | | | | | |
| episode: 4695 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.53 | | | | | | |
| episode: 4696 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.45 | | | | | | |
| episode: 4697 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.42 | | | | | | |
| episode: 4698 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.4 | | | | | | |
| episode: 4699 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 260 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.39 | | | | | | |
| episode: 4700 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.35 | | | | | | |
| episode: 4701 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.34 | | | | | | |
| episode: 4702 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 358 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.35 | | | | | | |
| episode: 4703 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.3 | | | | | | |
| episode: 4704 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 378 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.28 | | | | | | |
| episode: 4705 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 210 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.25 | | | | | | |
| episode: 4706 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.18 | | | | | | |
| episode: 4707 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.17 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4708 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 378 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.15 | | | | | | |
| episode: 4709 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 397 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.17 | | | | | | |
| episode: 4710 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.16 | | | | | | |
| episode: 4711 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.15 | | | | | | |
| episode: 4712 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 245 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.12 | | | | | | |
| episode: 4713 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 197 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.07 | | | | | | |
| episode: 4714 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 392 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.08 | | | | | | |
| episode: 4715 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 294 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.07 | | | | | | |
| episode: 4716 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.08 | | | | | | |
| episode: 4717 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 294 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.08 | | | | | | |
| episode: 4718 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.07 | | | | | | |
| episode: 4719 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 278 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.04 | | | | | | |
| episode: 4720 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 311 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.04 | | | | | | |
| episode: 4721 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.02 | | | | | | |
| episode: 4722 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 6.02 | | | | | | |
| episode: 4723 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.97 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4724 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 423 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.99 | | | | | | |
| episode: 4725 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.94 | | | | | | |
| episode: 4726 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 425 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.94 | | | | | | |
| episode: 4727 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.91 | | | | | | |
| episode: 4728 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 292 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.91 | | | | | | |
| episode: 4729 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.86 | | | | | | |
| episode: 4730 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.82 | | | | | | |
| episode: 4731 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.79 | | | | | | |
| episode: 4732 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 260 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.76 | | | | | | |
| episode: 4733 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 468 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.83 | | | | | | |
| episode: 4734 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 228 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.79 | | | | | | |
| episode: 4735 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 353 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.76 | | | | | | |
| episode: 4736 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 257 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.73 | | | | | | |
| episode: 4737 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 458 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.72 | | | | | | |
| episode: 4738 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 358 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.72 | | | | | | |
| episode: 4739 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.69 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4740 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.59 | | | | | | |
| episode: 4741 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 274 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.54 | | | | | | |
| episode: 4742 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.5 | | | | | | |
| episode: 4743 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 376 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.47 | | | | | | |
| episode: 4744 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 358 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.45 | | | | | | |
| episode: 4745 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.46 | | | | | | |
| episode: 4746 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 260 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.43 | | | | | | |
| episode: 4747 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.38 | | | | | | |
| episode: 4748 | score: 12.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 443 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.42 | | | | | | |
| episode: 4749 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 343 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 4750 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 364 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 4751 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 210 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.37 | | | | | | |
| episode: 4752 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 409 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 4753 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 420 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.37 | | | | | | |
| episode: 4754 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.37 | | | | | | |
| episode: 4755 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 358 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.42 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4756 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.38 | | | | | | |
| episode: 4757 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.34 | | | | | | |
| episode: 4758 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 343 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.37 | | | | | | |
| episode: 4759 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.31 | | | | | | |
| episode: 4760 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 535 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.36 | | | | | | |
| episode: 4761 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 293 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.37 | | | | | | |
| episode: 4762 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 365 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.35 | | | | | | |
| episode: 4763 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 453 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.38 | | | | | | |
| episode: 4764 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 4765 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 378 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 4766 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.41 | | | | | | |
| episode: 4767 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 280 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.37 | | | | | | |
| episode: 4768 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 197 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.34 | | | | | | |
| episode: 4769 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 4770 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 292 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.25 | | | | | | |
| episode: 4771 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 409 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.28 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4772 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 298 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 4773 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 438 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.32 | | | | | | |
| episode: 4774 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.31 | | | | | | |
| episode: 4775 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 362 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.33 | | | | | | |
| episode: 4776 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 326 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.29 | | | | | | |
| episode: 4777 | score: 14.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 546 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.35 | | | | | | |
| episode: 4778 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.35 | | | | | | |
| episode: 4779 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.35 | | | | | | |
| episode: 4780 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.35 | | | | | | |
| episode: 4781 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.34 | | | | | | |
| episode: 4782 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.32 | | | | | | |
| episode: 4783 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 228 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.3 | | | | | | |
| episode: 4784 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 228 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 4785 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.24 | | | | | | |
| episode: 4786 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.23 | | | | | | |
| episode: 4787 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 387 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.25 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4788 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 373 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 4789 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 4790 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.25 | | | | | | |
| episode: 4791 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.23 | | | | | | |
| episode: 4792 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 327 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.22 | | | | | | |
| episode: 4793 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.18 | | | | | | |
| episode: 4794 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.16 | | | | | | |
| episode: 4795 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 308 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.15 | | | | | | |
| episode: 4796 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.12 | | | | | | |
| episode: 4797 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.11 | | | | | | |
| episode: 4798 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 343 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.13 | | | | | | |
| episode: 4799 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 365 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.15 | | | | | | |
| episode: 4800 | score: 14.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 484 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.24 | | | | | | |
| episode: 4801 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.27 | | | | | | |
| episode: 4802 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.22 | | | | | | |
| episode: 4803 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.17 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|---------------------------|------------|
| episode: 4804 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.400000000000001e-06 | evaluation |
| reward: 5.12 | | | | | | |
| episode: 4805 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.400000000000001e-06 | evaluation |
| reward: 5.11 | | | | | | |
| episode: 4806 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.400000000000001e-06 | evaluation |
| reward: 5.1 | | | | | | |
| episode: 4807 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.400000000000001e-06 | evaluation |
| reward: 5.06 | | | | | | |
| episode: 4808 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.400000000000001e-06 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 4809 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 327 | lr: 6.400000000000001e-06 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 4810 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 4811 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 4812 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 358 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.97 | | | | | | |
| episode: 4813 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 4814 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 408 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 4815 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 296 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 4816 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 402 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 4817 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 439 | lr: 6.400000000000001e-06 | evaluation |
| reward: 5.03 | | | | | | |
| episode: 4818 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 245 | lr: 6.400000000000001e-06 | evaluation |
| reward: 5.02 | | | | | | |
| episode: 4819 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 6.400000000000001e-06 | evaluation |
| reward: 5.03 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4820 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.98 | | | | | | |
| episode: 4821 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.95 | | | | | | |
| episode: 4822 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 4823 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 362 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 4824 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.89 | | | | | | |
| episode: 4825 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 255 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 4826 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.83 | | | | | | |
| episode: 4827 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 335 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.85 | | | | | | |
| episode: 4828 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.9 | | | | | | |
| episode: 4829 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 354 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.93 | | | | | | |
| episode: 4830 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 255 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.94 | | | | | | |
| episode: 4831 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 360 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.96 | | | | | | |
| episode: 4832 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 378 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.99 | | | | | | |
| episode: 4833 | score: 12.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 469 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 4834 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 465 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.09 | | | | | | |
| episode: 4835 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 447 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.11 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4836 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 309 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.13 | | | | | | |
| episode: 4837 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 397 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.11 | | | | | | |
| episode: 4838 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 424 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.12 | | | | | | |
| episode: 4839 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 329 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.13 | | | | | | |
| episode: 4840 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 419 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.18 | | | | | | |
| episode: 4841 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.15 | | | | | | |
| episode: 4842 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 308 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.17 | | | | | | |
| episode: 4843 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.11 | | | | | | |
| episode: 4844 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 197 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.06 | | | | | | |
| episode: 4845 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.04 | | | | | | |
| episode: 4846 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 4847 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 259 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 5.01 | | | | | | |
| episode: 4848 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 181 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.91 | | | | | | |
| episode: 4849 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 197 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.87 | | | | | | |
| episode: 4850 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.82 | | | | | | |
| episode: 4851 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.81 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4852 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.74 | | | | | | |
| episode: 4853 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.68 | | | | | | |
| episode: 4854 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.66 | | | | | | |
| episode: 4855 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.62 | | | | | | |
| episode: 4856 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.61 | | | | | | |
| episode: 4857 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.59 | | | | | | |
| episode: 4858 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 243 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.56 | | | | | | |
| episode: 4859 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 197 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.55 | | | | | | |
| episode: 4860 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.46 | | | | | | |
| episode: 4861 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 392 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.5 | | | | | | |
| episode: 4862 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.45 | | | | | | |
| episode: 4863 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.38 | | | | | | |
| episode: 4864 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 336 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.38 | | | | | | |
| episode: 4865 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 226 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.34 | | | | | | |
| episode: 4866 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.33 | | | | | | |
| episode: 4867 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.4000000000000001e-06 | evaluation |
| reward: 4.31 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|---------------------------|------------|
| episode: 4868 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 277 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.33 | | | | | | |
| episode: 4869 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.32 | | | | | | |
| episode: 4870 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 306 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.32 | | | | | | |
| episode: 4871 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 426 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.32 | | | | | | |
| episode: 4872 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.3 | | | | | | |
| episode: 4873 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 388 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.28 | | | | | | |
| episode: 4874 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 245 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.3 | | | | | | |
| episode: 4875 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.25 | | | | | | |
| episode: 4876 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 397 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.28 | | | | | | |
| episode: 4877 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.16 | | | | | | |
| episode: 4878 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.18 | | | | | | |
| episode: 4879 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 339 | lr: 6.400000000000001e-06 | evaluation |
| reward: 4.17 | | | | | | |
| episode: 4880 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.560000000000001e-06 | evaluation |
| reward: 4.19 | | | | | | |
| episode: 4881 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.560000000000001e-06 | evaluation |
| reward: 4.18 | | | | | | |
| episode: 4882 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 279 | lr: 2.560000000000001e-06 | evaluation |
| reward: 4.19 | | | | | | |
| episode: 4883 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 536 | lr: 2.560000000000001e-06 | evaluation |
| reward: 4.26 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4884 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.25 | | | | | | |
| episode: 4885 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.24 | | | | | | |
| episode: 4886 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.23 | | | | | | |
| episode: 4887 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 231 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.18 | | | | | | |
| episode: 4888 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 444 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.19 | | | | | | |
| episode: 4889 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 197 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.17 | | | | | | |
| episode: 4890 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.16 | | | | | | |
| episode: 4891 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.16 | | | | | | |
| episode: 4892 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.13 | | | | | | |
| episode: 4893 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.12 | | | | | | |
| episode: 4894 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.11 | | | | | | |
| episode: 4895 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 151 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.07 | | | | | | |
| episode: 4896 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.07 | | | | | | |
| episode: 4897 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 453 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.13 | | | | | | |
| episode: 4898 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 243 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.11 | | | | | | |
| episode: 4899 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 243 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 4.09 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4900 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.96 | | | | | | |
| episode: 4901 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.91 | | | | | | |
| episode: 4902 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.91 | | | | | | |
| episode: 4903 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 197 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.92 | | | | | | |
| episode: 4904 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.91 | | | | | | |
| episode: 4905 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.9 | | | | | | |
| episode: 4906 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.89 | | | | | | |
| episode: 4907 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.89 | | | | | | |
| episode: 4908 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 319 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.92 | | | | | | |
| episode: 4909 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.88 | | | | | | |
| episode: 4910 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.89 | | | | | | |
| episode: 4911 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.93 | | | | | | |
| episode: 4912 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.89 | | | | | | |
| episode: 4913 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.89 | | | | | | |
| episode: 4914 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.84 | | | | | | |
| episode: 4915 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.85 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4916 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.79 | | | | | | |
| episode: 4917 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.72 | | | | | | |
| episode: 4918 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.7 | | | | | | |
| episode: 4919 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 438 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.72 | | | | | | |
| episode: 4920 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.73 | | | | | | |
| episode: 4921 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.73 | | | | | | |
| episode: 4922 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.75 | | | | | | |
| episode: 4923 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.71 | | | | | | |
| episode: 4924 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.72 | | | | | | |
| episode: 4925 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.7 | | | | | | |
| episode: 4926 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.7 | | | | | | |
| episode: 4927 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.65 | | | | | | |
| episode: 4928 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.57 | | | | | | |
| episode: 4929 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.54 | | | | | | |
| episode: 4930 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.53 | | | | | | |
| episode: 4931 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.48 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4932 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 438 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.49 | | | | | | |
| episode: 4933 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.39 | | | | | | |
| episode: 4934 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 151 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.29 | | | | | | |
| episode: 4935 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.23 | | | | | | |
| episode: 4936 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.18 | | | | | | |
| episode: 4937 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.13 | | | | | | |
| episode: 4938 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 180 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.07 | | | | | | |
| episode: 4939 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.03 | | | | | | |
| episode: 4940 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.97 | | | | | | |
| episode: 4941 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 197 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.98 | | | | | | |
| episode: 4942 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 257 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.97 | | | | | | |
| episode: 4943 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.98 | | | | | | |
| episode: 4944 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 257 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.0 | | | | | | |
| episode: 4945 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.99 | | | | | | |
| episode: 4946 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.0 | | | | | | |
| episode: 4947 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 347 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.02 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4948 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 339 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.06 | | | | | | |
| episode: 4949 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 313 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.09 | | | | | | |
| episode: 4950 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 384 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.15 | | | | | | |
| episode: 4951 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 438 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.21 | | | | | | |
| episode: 4952 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 438 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.28 | | | | | | |
| episode: 4953 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 261 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.3 | | | | | | |
| episode: 4954 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.29 | | | | | | |
| episode: 4955 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.29 | | | | | | |
| episode: 4956 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 457 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.37 | | | | | | |
| episode: 4957 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.39 | | | | | | |
| episode: 4958 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.37 | | | | | | |
| episode: 4959 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.36 | | | | | | |
| episode: 4960 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.37 | | | | | | |
| episode: 4961 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.32 | | | | | | |
| episode: 4962 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.31 | | | | | | |
| episode: 4963 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 243 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.31 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4964 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.28 | | | | | | |
| episode: 4965 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.28 | | | | | | |
| episode: 4966 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.27 | | | | | | |
| episode: 4967 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.26 | | | | | | |
| episode: 4968 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.23 | | | | | | |
| episode: 4969 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.24 | | | | | | |
| episode: 4970 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.22 | | | | | | |
| episode: 4971 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.15 | | | | | | |
| episode: 4972 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.15 | | | | | | |
| episode: 4973 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.08 | | | | | | |
| episode: 4974 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 226 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.07 | | | | | | |
| episode: 4975 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 226 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.08 | | | | | | |
| episode: 4976 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 383 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.07 | | | | | | |
| episode: 4977 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.08 | | | | | | |
| episode: 4978 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.06 | | | | | | |
| episode: 4979 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.04 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4980 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.01 | | | | | | |
| episode: 4981 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.0 | | | | | | |
| episode: 4982 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.99 | | | | | | |
| episode: 4983 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 439 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.97 | | | | | | |
| episode: 4984 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 434 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.04 | | | | | | |
| episode: 4985 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 406 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.09 | | | | | | |
| episode: 4986 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.09 | | | | | | |
| episode: 4987 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.07 | | | | | | |
| episode: 4988 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.01 | | | | | | |
| episode: 4989 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.02 | | | | | | |
| episode: 4990 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 434 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.09 | | | | | | |
| episode: 4991 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.08 | | | | | | |
| episode: 4992 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.08 | | | | | | |
| episode: 4993 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.09 | | | | | | |
| episode: 4994 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.1 | | | | | | |
| episode: 4995 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.1 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 4996 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.11 | | | | | | |
| episode: 4997 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.05 | | | | | | |
| episode: 4998 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.02 | | | | | | |
| episode: 4999 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.99 | | | | | | |
| episode: 5000 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.99 | | | | | | |
| episode: 5001 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.0 | | | | | | |
| episode: 5002 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.01 | | | | | | |
| episode: 5003 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.01 | | | | | | |
| episode: 5004 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 495 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.1 | | | | | | |
| episode: 5005 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 350 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.15 | | | | | | |
| episode: 5006 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.16 | | | | | | |
| episode: 5007 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 289 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.19 | | | | | | |
| episode: 5008 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 394 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.2 | | | | | | |
| episode: 5009 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.21 | | | | | | |
| episode: 5010 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 434 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.27 | | | | | | |
| episode: 5011 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 466 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.32 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5012 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.31 | | | | | | |
| episode: 5013 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 342 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.34 | | | | | | |
| episode: 5014 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 197 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.33 | | | | | | |
| episode: 5015 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.29 | | | | | | |
| episode: 5016 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.3 | | | | | | |
| episode: 5017 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.3 | | | | | | |
| episode: 5018 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 255 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.32 | | | | | | |
| episode: 5019 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 345 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.3 | | | | | | |
| episode: 5020 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 445 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.37 | | | | | | |
| episode: 5021 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.38 | | | | | | |
| episode: 5022 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.37 | | | | | | |
| episode: 5023 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.36 | | | | | | |
| episode: 5024 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.36 | | | | | | |
| episode: 5025 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.37 | | | | | | |
| episode: 5026 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.39 | | | | | | |
| episode: 5027 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.43 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5028 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.45 | | | | | | |
| episode: 5029 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 294 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.47 | | | | | | |
| episode: 5030 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.46 | | | | | | |
| episode: 5031 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.46 | | | | | | |
| episode: 5032 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.4 | | | | | | |
| episode: 5033 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.4 | | | | | | |
| episode: 5034 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.41 | | | | | | |
| episode: 5035 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.41 | | | | | | |
| episode: 5036 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 490 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.49 | | | | | | |
| episode: 5037 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.5 | | | | | | |
| episode: 5038 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.51 | | | | | | |
| episode: 5039 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.51 | | | | | | |
| episode: 5040 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.51 | | | | | | |
| episode: 5041 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.52 | | | | | | |
| episode: 5042 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.51 | | | | | | |
| episode: 5043 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.51 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5044 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.49 | | | | | | |
| episode: 5045 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.5 | | | | | | |
| episode: 5046 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.49 | | | | | | |
| episode: 5047 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.44 | | | | | | |
| episode: 5048 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.4 | | | | | | |
| episode: 5049 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.36 | | | | | | |
| episode: 5050 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.3 | | | | | | |
| episode: 5051 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.25 | | | | | | |
| episode: 5052 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 416 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.25 | | | | | | |
| episode: 5053 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.23 | | | | | | |
| episode: 5054 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.23 | | | | | | |
| episode: 5055 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.22 | | | | | | |
| episode: 5056 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.14 | | | | | | |
| episode: 5057 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.13 | | | | | | |
| episode: 5058 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 357 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.17 | | | | | | |
| episode: 5059 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.18 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5060 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.17 | | | | | | |
| episode: 5061 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.16 | | | | | | |
| episode: 5062 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.17 | | | | | | |
| episode: 5063 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.16 | | | | | | |
| episode: 5064 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 325 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.18 | | | | | | |
| episode: 5065 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.17 | | | | | | |
| episode: 5066 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.19 | | | | | | |
| episode: 5067 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.2 | | | | | | |
| episode: 5068 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.22 | | | | | | |
| episode: 5069 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.22 | | | | | | |
| episode: 5070 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 226 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.22 | | | | | | |
| episode: 5071 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.23 | | | | | | |
| episode: 5072 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.21 | | | | | | |
| episode: 5073 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.22 | | | | | | |
| episode: 5074 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.21 | | | | | | |
| episode: 5075 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 328 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.24 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5076 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.18 | | | | | | |
| episode: 5077 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.16 | | | | | | |
| episode: 5078 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.15 | | | | | | |
| episode: 5079 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.13 | | | | | | |
| episode: 5080 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.13 | | | | | | |
| episode: 5081 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.13 | | | | | | |
| episode: 5082 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.12 | | | | | | |
| episode: 5083 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 278 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.08 | | | | | | |
| episode: 5084 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.01 | | | | | | |
| episode: 5085 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.96 | | | | | | |
| episode: 5086 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.96 | | | | | | |
| episode: 5087 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.97 | | | | | | |
| episode: 5088 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 299 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.99 | | | | | | |
| episode: 5089 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.99 | | | | | | |
| episode: 5090 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.92 | | | | | | |
| episode: 5091 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.93 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5092 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.92 | | | | | | |
| episode: 5093 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.91 | | | | | | |
| episode: 5094 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.91 | | | | | | |
| episode: 5095 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.93 | | | | | | |
| episode: 5096 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.93 | | | | | | |
| episode: 5097 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.92 | | | | | | |
| episode: 5098 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.93 | | | | | | |
| episode: 5099 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.94 | | | | | | |
| episode: 5100 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.95 | | | | | | |
| episode: 5101 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.95 | | | | | | |
| episode: 5102 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.94 | | | | | | |
| episode: 5103 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.94 | | | | | | |
| episode: 5104 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.87 | | | | | | |
| episode: 5105 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.84 | | | | | | |
| episode: 5106 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.84 | | | | | | |
| episode: 5107 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.81 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5108 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.77 | | | | | | |
| episode: 5109 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 277 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.78 | | | | | | |
| episode: 5110 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.72 | | | | | | |
| episode: 5111 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.64 | | | | | | |
| episode: 5112 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.65 | | | | | | |
| episode: 5113 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.61 | | | | | | |
| episode: 5114 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 289 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.64 | | | | | | |
| episode: 5115 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.65 | | | | | | |
| episode: 5116 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.65 | | | | | | |
| episode: 5117 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.65 | | | | | | |
| episode: 5118 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.64 | | | | | | |
| episode: 5119 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.61 | | | | | | |
| episode: 5120 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.54 | | | | | | |
| episode: 5121 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.55 | | | | | | |
| episode: 5122 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.55 | | | | | | |
| episode: 5123 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.55 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5124 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.55 | | | | | | |
| episode: 5125 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.54 | | | | | | |
| episode: 5126 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.53 | | | | | | |
| episode: 5127 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.5 | | | | | | |
| episode: 5128 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.5 | | | | | | |
| episode: 5129 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.47 | | | | | | |
| episode: 5130 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 339 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.51 | | | | | | |
| episode: 5131 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.51 | | | | | | |
| episode: 5132 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.52 | | | | | | |
| episode: 5133 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.52 | | | | | | |
| episode: 5134 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.52 | | | | | | |
| episode: 5135 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.52 | | | | | | |
| episode: 5136 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.45 | | | | | | |
| episode: 5137 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.44 | | | | | | |
| episode: 5138 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.43 | | | | | | |
| episode: 5139 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.43 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5140 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.43 | | | | | | |
| episode: 5141 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.42 | | | | | | |
| episode: 5142 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.42 | | | | | | |
| episode: 5143 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 289 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.45 | | | | | | |
| episode: 5144 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 260 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.47 | | | | | | |
| episode: 5145 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 339 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.5 | | | | | | |
| episode: 5146 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.52 | | | | | | |
| episode: 5147 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.54 | | | | | | |
| episode: 5148 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 317 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.58 | | | | | | |
| episode: 5149 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.6 | | | | | | |
| episode: 5150 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.61 | | | | | | |
| episode: 5151 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 289 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.63 | | | | | | |
| episode: 5152 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.57 | | | | | | |
| episode: 5153 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.57 | | | | | | |
| episode: 5154 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.57 | | | | | | |
| episode: 5155 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.57 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5156 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.57 | | | | | | |
| episode: 5157 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.57 | | | | | | |
| episode: 5158 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.53 | | | | | | |
| episode: 5159 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.56 | | | | | | |
| episode: 5160 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.56 | | | | | | |
| episode: 5161 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.56 | | | | | | |
| episode: 5162 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 307 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.59 | | | | | | |
| episode: 5163 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.59 | | | | | | |
| episode: 5164 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.56 | | | | | | |
| episode: 5165 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 319 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.6 | | | | | | |
| episode: 5166 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.59 | | | | | | |
| episode: 5167 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.59 | | | | | | |
| episode: 5168 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.58 | | | | | | |
| episode: 5169 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.57 | | | | | | |
| episode: 5170 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.57 | | | | | | |
| episode: 5171 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.58 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5172 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.6 | | | | | | |
| episode: 5173 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.61 | | | | | | |
| episode: 5174 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 391 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.66 | | | | | | |
| episode: 5175 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 286 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.65 | | | | | | |
| episode: 5176 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.67 | | | | | | |
| episode: 5177 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 411 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.74 | | | | | | |
| episode: 5178 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.75 | | | | | | |
| episode: 5179 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.77 | | | | | | |
| episode: 5180 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 297 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.79 | | | | | | |
| episode: 5181 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.79 | | | | | | |
| episode: 5182 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.79 | | | | | | |
| episode: 5183 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.78 | | | | | | |
| episode: 5184 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.78 | | | | | | |
| episode: 5185 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 289 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.81 | | | | | | |
| episode: 5186 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.82 | | | | | | |
| episode: 5187 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.83 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5188 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.81 | | | | | | |
| episode: 5189 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.8 | | | | | | |
| episode: 5190 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.8 | | | | | | |
| episode: 5191 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.79 | | | | | | |
| episode: 5192 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.81 | | | | | | |
| episode: 5193 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.82 | | | | | | |
| episode: 5194 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.82 | | | | | | |
| episode: 5195 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.82 | | | | | | |
| episode: 5196 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 226 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.83 | | | | | | |
| episode: 5197 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.84 | | | | | | |
| episode: 5198 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.85 | | | | | | |
| episode: 5199 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.85 | | | | | | |
| episode: 5200 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.86 | | | | | | |
| episode: 5201 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.89 | | | | | | |
| episode: 5202 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.9 | | | | | | |
| episode: 5203 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.9 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5204 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.9 | | | | | | |
| episode: 5205 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 228 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.9 | | | | | | |
| episode: 5206 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 290 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.93 | | | | | | |
| episode: 5207 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.93 | | | | | | |
| episode: 5208 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 263 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.94 | | | | | | |
| episode: 5209 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.93 | | | | | | |
| episode: 5210 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.92 | | | | | | |
| episode: 5211 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 290 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.95 | | | | | | |
| episode: 5212 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 212 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.96 | | | | | | |
| episode: 5213 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.95 | | | | | | |
| episode: 5214 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.92 | | | | | | |
| episode: 5215 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.93 | | | | | | |
| episode: 5216 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 317 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.97 | | | | | | |
| episode: 5217 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 226 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.98 | | | | | | |
| episode: 5218 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.97 | | | | | | |
| episode: 5219 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 199 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.96 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5220 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 248 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.98 | | | | | | |
| episode: 5221 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 2.98 | | | | | | |
| episode: 5222 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 521 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.06 | | | | | | |
| episode: 5223 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.07 | | | | | | |
| episode: 5224 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.09 | | | | | | |
| episode: 5225 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.09 | | | | | | |
| episode: 5226 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 256 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.11 | | | | | | |
| episode: 5227 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.11 | | | | | | |
| episode: 5228 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.11 | | | | | | |
| episode: 5229 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 226 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.12 | | | | | | |
| episode: 5230 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 226 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.09 | | | | | | |
| episode: 5231 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.1 | | | | | | |
| episode: 5232 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.1 | | | | | | |
| episode: 5233 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.11 | | | | | | |
| episode: 5234 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 445 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.18 | | | | | | |
| episode: 5235 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.18 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5236 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 374 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.24 | | | | | | |
| episode: 5237 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 232 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.25 | | | | | | |
| episode: 5238 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.25 | | | | | | |
| episode: 5239 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.28 | | | | | | |
| episode: 5240 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.29 | | | | | | |
| episode: 5241 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.3 | | | | | | |
| episode: 5242 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.3 | | | | | | |
| episode: 5243 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.27 | | | | | | |
| episode: 5244 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.25 | | | | | | |
| episode: 5245 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.21 | | | | | | |
| episode: 5246 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.2 | | | | | | |
| episode: 5247 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.19 | | | | | | |
| episode: 5248 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.15 | | | | | | |
| episode: 5249 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.14 | | | | | | |
| episode: 5250 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.13 | | | | | | |
| episode: 5251 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.1 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5252 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.1 | | | | | | |
| episode: 5253 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.1 | | | | | | |
| episode: 5254 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.09 | | | | | | |
| episode: 5255 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.09 | | | | | | |
| episode: 5256 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.1 | | | | | | |
| episode: 5257 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.11 | | | | | | |
| episode: 5258 | score: 8.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 438 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.17 | | | | | | |
| episode: 5259 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 319 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.18 | | | | | | |
| episode: 5260 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.21 | | | | | | |
| episode: 5261 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.22 | | | | | | |
| episode: 5262 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.2 | | | | | | |
| episode: 5263 | score: 11.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 551 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.29 | | | | | | |
| episode: 5264 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.29 | | | | | | |
| episode: 5265 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.25 | | | | | | |
| episode: 5266 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.26 | | | | | | |
| episode: 5267 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.26 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5268 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.26 | | | | | | |
| episode: 5269 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.26 | | | | | | |
| episode: 5270 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.25 | | | | | | |
| episode: 5271 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.24 | | | | | | |
| episode: 5272 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.24 | | | | | | |
| episode: 5273 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.24 | | | | | | |
| episode: 5274 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.19 | | | | | | |
| episode: 5275 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.19 | | | | | | |
| episode: 5276 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.21 | | | | | | |
| episode: 5277 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 259 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.17 | | | | | | |
| episode: 5278 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 404 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.21 | | | | | | |
| episode: 5279 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.2 | | | | | | |
| episode: 5280 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.17 | | | | | | |
| episode: 5281 | score: 1.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 150 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.16 | | | | | | |
| episode: 5282 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.19 | | | | | | |
| episode: 5283 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.21 | | | | | | |

| | | | | | | |
|---------------|------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5284 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.22 | | | | | | |
| episode: 5285 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.19 | | | | | | |
| episode: 5286 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 246 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.19 | | | | | | |
| episode: 5287 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.19 | | | | | | |
| episode: 5288 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.2 | | | | | | |
| episode: 5289 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.23 | | | | | | |
| episode: 5290 | score: 4.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 259 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.25 | | | | | | |
| episode: 5291 | score: 7.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 370 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.3 | | | | | | |
| episode: 5292 | score: 5.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 295 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.32 | | | | | | |
| episode: 5293 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.31 | | | | | | |
| episode: 5294 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.31 | | | | | | |
| episode: 5295 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.3 | | | | | | |
| episode: 5296 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.3 | | | | | | |
| episode: 5297 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.3 | | | | | | |
| episode: 5298 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 318 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.33 | | | | | | |
| episode: 5299 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.34 | | | | | | |

| | | | | | | |
|---------------|-------------|------------------------|-------------------------------|------------|----------------------------|------------|
| episode: 5300 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.34 | | | | | | |
| episode: 5301 | score: 6.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 319 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.35 | | | | | | |
| episode: 5302 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.35 | | | | | | |
| episode: 5303 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.35 | | | | | | |
| episode: 5304 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.34 | | | | | | |
| episode: 5305 | score: 9.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 455 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.4 | | | | | | |
| episode: 5306 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 199 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.37 | | | | | | |
| episode: 5307 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.38 | | | | | | |
| episode: 5308 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.36 | | | | | | |
| episode: 5309 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.35 | | | | | | |
| episode: 5310 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.36 | | | | | | |
| episode: 5311 | score: 2.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 179 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.33 | | | | | | |
| episode: 5312 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.33 | | | | | | |
| episode: 5313 | score: 13.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 522 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.45 | | | | | | |
| episode: 5314 | score: 10.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 555 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.53 | | | | | | |
| episode: 5315 | score: 3.0 | memory length: 1000000 | epsilon: 0.009998020008555413 | steps: 225 | lr: 2.5600000000000001e-06 | evaluation |
| reward: 3.53 | | | | | | |

```

episode: 5316   score: 5.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 295   lr: 2.5600000000000001e-06   evaluation
reward: 3.52
episode: 5317   score: 3.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 225   lr: 2.5600000000000001e-06   evaluation
reward: 3.52
episode: 5318   score: 2.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 179   lr: 2.5600000000000001e-06   evaluation
reward: 3.52
episode: 5319   score: 3.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 225   lr: 2.5600000000000001e-06   evaluation
reward: 3.53
episode: 5320   score: 3.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 225   lr: 2.5600000000000001e-06   evaluation
reward: 3.52
episode: 5321   score: 3.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 225   lr: 2.5600000000000001e-06   evaluation
reward: 3.52
episode: 5322   score: 3.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 225   lr: 2.5600000000000001e-06   evaluation
reward: 3.45
episode: 5323   score: 3.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 225   lr: 2.5600000000000001e-06   evaluation
reward: 3.45
episode: 5324   score: 5.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 295   lr: 2.5600000000000001e-06   evaluation
reward: 3.46
episode: 5325   score: 3.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 225   lr: 2.5600000000000001e-06   evaluation
reward: 3.47
episode: 5326   score: 3.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 225   lr: 2.5600000000000001e-06   evaluation
reward: 3.46
episode: 5327   score: 9.0   memory length: 1000000   epsilon:
0.009998020008555413   steps: 498   lr: 2.5600000000000001e-06   evaluation
reward: 3.53

```

```

-----
KeyboardInterrupt                                Traceback (most recent call last)
Input In [8], in <cell line: 7>()
      42 # Start training after random sample generation
      43 if(frame >= train_frame):
----> 44     agent.train_policy_net(frame)
      45 score += reward
      46 history[:, :, :] = history[1:, :, :]

```

```

File ~/QA_generator/CS444/mp5/agent_lstm.py:52, in Agent_LSTM.
    ↪train_policy_net(self, frame)
      49 if self.epsilon > self.epsilon_min:
      50     self.epsilon -= self.epsilon_decay
---> 52 mini_batch = self.memory.sample_mini_batch(frame)
      53 mini_batch = np.array(mini_batch, dtype=object).transpose()
      55 history = np.stack(mini_batch[0], axis=0)

```

```

File ~/QA_generator/CS444/mp5/memory.py:65, in ReplayMemoryLSTM.
    ↪sample_mini_batch(self, frame)
      63 sample = []
      64 for j in range(lstm_seq_length + 1):
---> 65     sample.append(self.memory[i + j])
      66 sample = np.array(sample, dtype=object)
      67 mini_batch.append((np.stack(sample[:, 0], axis=0),
    ↪sample[lstm_seq_length - 1, 1], sample[lstm_seq_length - 1, 2],
    ↪sample[lstm_seq_length - 1, 3]))

```

KeyboardInterrupt:

Error in callback <function flush_figures at 0x14eeeda8c1f0> (for post_execute):

```

-----
KeyboardInterrupt                                Traceback (most recent call last)
File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/
    ↪matplotlib_inline/backend_inline.py:121, in flush_figures()
      118 if InlineBackend.instance().close_figures:
      119     # ignore the tracking, just draw and close all figures
      120     try:
--> 121         return show(True)
      122     except Exception as e:
      123         # safely show traceback if in IPython, else raise
      124         ip = get_ipython()

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/
    ↪matplotlib_inline/backend_inline.py:41, in show(close, block)
      39 try:
      40     for figure_manager in Gcf.get_all_fig_managers():
---> 41         display(
      42             figure_manager.canvas.figure,
      43             metadata=_fetch_figure_metadata(figure_manager.canvas.figure)
      44         )
      45 finally:
      46     show._to_draw = []

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/IPython/core
    ↪display_functions.py:298, in display(include, exclude, metadata, transient,
    ↪display_id, raw, clear, *objs, **kwargs)

```

```

296     publish_display_data(data=obj, metadata=metadata, **kwargs)
297 else:
--> 298     format_dict, md_dict = format(obj, include=include, exclude=exclude)
299     if not format_dict:
300         # nothing to display (e.g. _ipython_display_ took over)
301         continue

```

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/IPython/core/formatters.py:178, in DisplayFormatter.format(self, obj, include, exclude)

```

176 md = None
177 try:
--> 178     data = formatter(obj)
179 except:
180     # FIXME: log the exception
181     raise

```

File <decorator-gen-2>:2, in __call__(self, obj)

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/IPython/core/formatters.py:222, in catch_format_error(method, self, *args, **kwargs)

```

220 """show traceback on failed format call"""
221 try:
--> 222     r = method(self, *args, **kwargs)
223 except NotImplementedError:
224     # don't warn on NotImplementedError
225     return self._check_return(None, args[0])

```

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/IPython/core/formatters.py:339, in BaseFormatter.__call__(self, obj)

```

337     pass
338 else:
--> 339     return printer(obj)
340 # Finally look for special method names
341 method = get_real_method(obj, self.print_method)

```

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/IPython/core/pylabtools.py:151, in print_figure(fig, fmt, bbox_inches, base64, **kwargs)

```

148     from matplotlib.backend_bases import FigureCanvasBase
149     FigureCanvasBase(fig)
--> 151     fig.canvas.print_figure(bytes_io, **kw)
152     data = bytes_io.getvalue()
153     if fmt == 'svg':

```

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/backend_bases.py:2319, in FigureCanvasBase.print_figure(self, filename, dpi, facecolor, edgecolor, orientation, format, bbox_inches, pad_inches, bbox_extra_artists, backend, **kwargs)

```

2315 try:

```

```

2316     # _get_renderer may change the figure dpi (as vector formats
2317     # force the figure dpi to 72), so we need to set it again here.
2318     with cbook._setattr_cm(self.figure, dpi=dpi):
-> 2319         result = print_method(
2320             filename,
2321             facecolor=facecolor,
2322             edgecolor=edgecolor,
2323             orientation=orientation,
2324             bbox_inches_restore=_bbox_inches_restore,
2325             **kwargs)
2326     finally:
2327         if bbox_inches and restore_bbox:

```

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↳ backend_bases.py:1648, in _check_savefig_extra_args.<locals>.wrapper(*args,
↳ **kwargs)

```

1640     _api.warn_deprecated(
1641         '3.3', name=name, removal='3.6',
1642         message='%(name)s() got unexpected keyword argument "'
1643             + arg + '" which is no longer supported as of '
1644             + '%(since)s and will become an error '
1645             + '%(removal)s')
1646     kwargs.pop(arg)
-> 1648 return func(*args, **kwargs)

```

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↳ _api/deprecation.py:412, in delete_parameter.<locals>.wrapper(*inner_args,
↳ **inner_kwargs)

```

402     deprecation_addendum = (
403         f"If any parameter follows {name!r}, they should be passed as "
404         f"keyword, not positionally.")
405     warn_deprecated(
406         since,
407         name=repr(name),
408         (...))
409     else deprecation_addendum,
410     **kwargs)
--> 412 return func(*inner_args, **inner_kwargs)

```

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↳ backends/backend_agg.py:540, in FigureCanvasAgg.print_png(self,
↳ filename_or_obj, metadata, pil_kwargs, *args)

```

490 @_check_savefig_extra_args
491 @_api.delete_parameter("3.5", "args")
492 def print_png(self, filename_or_obj, *args,
493             metadata=None, pil_kwargs=None):
494     """
495     Write the figure to a PNG file.

```

```

496
(...)
538         *metadata*, including the default 'Software' key.
539         """
--> 540     FigureCanvasAgg.draw(self)
541     mpl.image.imsave(
542         filename_or_obj, self.buffer_rgba(), format="png",
↪origin="upper",
543         dpi=self.figure.dpi, metadata=metadata, pil_kwargs=pil_kwargs)

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↪backends/backend_agg.py:436, in FigureCanvasAgg.draw(self)
432 # Acquire a lock on the shared font cache.
433 with RendererAgg.lock, \
434     (self.toolbar._wait_cursor_for_draw_cm() if self.toolbar
435      else nullcontext()):
--> 436     self.figure.draw(self.renderer)
437     # A GUI class may need to update a window using this draw, so
438     # don't forget to call the superclass.
439     super().draw()

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↪artist.py:73, in _finalize_rasterization.<locals>.draw_wrapper(artist,
↪renderer, *args, **kwargs)
71 @wraps(draw)
72 def draw_wrapper(artist, renderer, *args, **kwargs):
---> 73     result = draw(artist, renderer, *args, **kwargs)
74     if renderer._rasterizing:
75         renderer.stop_rasterizing()

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↪artist.py:50, in allow_rasterization.<locals>.draw_wrapper(artist, renderer)
47     if artist.get_agg_filter() is not None:
48         renderer.start_filter()
---> 50     return draw(artist, renderer)
51 finally:
52     if artist.get_agg_filter() is not None:

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↪figure.py:2810, in Figure.draw(self, renderer)
2807         # ValueError can occur when resizing a window.
2809 self.patch.draw(renderer)
-> 2810 mimage._draw_list_compositing_images(
2811     renderer, self, artists, self.suppressComposite)
2813 for sfig in self.subfigs:
2814     sfig.draw(renderer)

```

```
File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↳image.py:132, in _draw_list_compositing_images(renderer, parent, artists,
↳suppress_composite)
```

```
    130 if not_composite or not has_images:
    131     for a in artists:
--> 132         a.draw(renderer)
    133 else:
    134     # Composite any adjacent images together
    135     image_group = []
```

```
File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↳artist.py:50, in allow_rasterization.<locals>.draw_wrapper(artist, renderer)
```

```
    47 if artist.get_agg_filter() is not None:
    48     renderer.start_filter()
---> 50     return draw(artist, renderer)
    51 finally:
    52     if artist.get_agg_filter() is not None:
```

```
File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↳axes/_base.py:3082, in _AxesBase.draw(self, renderer)
```

```
    3079 a.draw(renderer)
    3080 renderer.stop_rasterizing()
-> 3082 mimage._draw_list_compositing_images(
    3083     renderer, self, artists, self.figure.suppressComposite)
    3085 renderer.close_group('axes')
    3086 self.stale = False
```

```
File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↳image.py:132, in _draw_list_compositing_images(renderer, parent, artists,
↳suppress_composite)
```

```
    130 if not_composite or not has_images:
    131     for a in artists:
--> 132         a.draw(renderer)
    133 else:
    134     # Composite any adjacent images together
    135     image_group = []
```

```
File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↳artist.py:50, in allow_rasterization.<locals>.draw_wrapper(artist, renderer)
```

```
    47 if artist.get_agg_filter() is not None:
    48     renderer.start_filter()
---> 50     return draw(artist, renderer)
    51 finally:
    52     if artist.get_agg_filter() is not None:
```

```
File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/
↳lines.py:776, in Line2D.draw(self, renderer)
```

```
    773         gc.set_sketch_params(*self.get_sketch_params())
```

```

775         gc.set_dashes(self._dashOffset, self._dashSeq)
--> 776         renderer.draw_path(gc, tpath, affine.frozen())
777         gc.restore()
779 if self._marker and self._markersize > 0:

```

File /sw/external/python/anaconda3_gpu/lib/python3.9/site-packages/matplotlib/backends/backend_agg.py:170, in RendererAgg.draw_path(self, gc, path, transform, rgbFace)

```

168 else:
169     try:
--> 170         self._renderer.draw_path(gc, path, transform, rgbFace)
171     except OverflowError as err:
172         cant_chunk = ''

```

KeyboardInterrupt:

```

[ ]: pylab.plot(episodes, rewards, 'b')
pylab.xlabel('Episodes')
pylab.ylabel('Rewards')
pylab.title('Episodes vs Reward')
pylab.savefig("./save_graph/breakout_dqn_lstm.png") # save graph for training
    ↪ visualization

```



2 Visualize Agent Performance

```
[ ]: torch.save(agent.policy_net, "./save_model/breakout_dqn_lstm_latest.pth")

[ ]: from gym.wrappers import RecordVideo # If importing monitor raises issues, try_
    ↪ using `from gym.wrappers import RecordVideo`
import glob
import io
import base64

from IPython.display import HTML
from IPython import display as ipythondisplay

from pyvirtualdisplay import Display

env = gym.make('BreakoutDeterministic-v4', render_mode='rgb_array')
state = env.reset()
number_lives = find_max_lives(env)
state_size = env.observation_space.shape
action_size = 3 #fire, left, and right

# Displaying the game live
def show_state(env, step=0, info=""):
    plt.figure(3)
    plt.clf()
    plt.imshow(env.render(mode='rgb_array'))
    plt.title("%s | Step: %d %s" % ("Agent Playing", step, info))
    plt.axis('off')

    ipythondisplay.clear_output(wait=True)
    ipythondisplay.display(plt.gcf())

# Recording the game and replaying the game afterwards
def show_video():
    mp4list = glob.glob('video/*.mp4')
    if len(mp4list) > 0:
        mp4 = mp4list[0]
        video = io.open(mp4, 'r+b').read()
        encoded = base64.b64encode(video)
        ipythondisplay.display(HTML(data='''<video style="height: 400px;"_
    ↪controls="" loop="" autoplay="" alt="test">
        <source type="video/mp4" src="data:video/mp4;base64,{0}">
        </video>''' .format(encoded.decode('ascii'))))
    else:
        print("Could not find video")
```

```
def wrap_env(env):
    env = RecordVideo(env, './video')
    return env
```

```
[ ]: # display = Display(visible=0, size=(300, 200))
# display.start()
HISTORY_SIZE = 1

# Load agent
agent.load_policy_net("./save_model/breakout_dqn_lstm_latest.pth")
agent.epsilon = 0.0 # Set agent to only exploit the best action

env = gym.make('BreakoutDeterministic-v4', render_mode='rgb_array')
env = wrap_env(env)

done = False
score = 0
step = 0
state, _ = env.reset()
next_state = state
life = number_lives
history = np.zeros([HISTORY_SIZE + 1, 84, 84], dtype=np.uint8)
get_init_state(history, state, HISTORY_SIZE)
frame = 0
hidden = None

while not done:

    # Render breakout
    # env.render()
    # show_state(env, step) # uncommenting this provides another way to
    ↪ visualize the game

    step += 1
    frame += 1

    # Perform a fire action if ball is no longer on screen
    if step > 1 and len(np.unique(next_state[:189] == state[:189])) < 2:
        action = torch.tensor([[0]]).cuda()
    else:
        action, hidden = agent.get_action(np.float32(history[:1, :, :]) / 255.,
    ↪ hidden)
        state = next_state

    next_state, reward, done, _, info = env.step(action + 1)

    frame_next_state = get_frame(next_state)
```

```

history[1, :, :] = frame_next_state
terminal_state = check_live(life, info['lives'])

life = info['lives']
r = np.clip(reward, -1, 1)
r = reward

# Store the transition in memory
agent.memory.push(deepcopy(frame_next_state), action, r, terminal_state)
# Start training after random sample generation
score += reward

history[:1, :, :] = history[1:, :, :]
env.close()
show_video()
# display.stop()

```

3 Save as PDF

```

[ ]: %%capture

from google.colab import drive
drive.mount('/content/drive')
# install tex; first run may take several minutes
! apt-get install texlive-xetex
# file path and save location below are default; please change if they do not
→match yours
! jupyter nbconvert --output-dir='/content/drive/MyDrive/' '/content/drive/
→MyDrive/CS444/assignment5/MP5_DQNLSTM_EC.ipynb' --to pdf

```