Lexicographic Reinforcement Learning Benchmark on MuJoCo environment -Appendix-

Symbolic and Evolutionary Artificial Intelligence a.a. 2022/2023

Michele Lisi

September 2023

Appendix Guide

In this appendix, all the data regarding testing and training are showed.
All the plots are obtained using `tf_reader` and can be better visualized (and interacted with) simply by running the command `python3 main.py ./trainings`, assuming the trainings folder is in the same as tf_reader.

To gain a better understanding, here follows an explanation of the meaning of the model tags used:

- **[20S]** - a network with sample size of 20
- **[HHH,BBB]** - a network with layer size of HHH and Experience Replay Buffer's batch size of BBB
- **[NOB]** - the network does not use bias in its layers
- **[LT-X]** - the loss threshold used in the lexicographic network is 0.X
- **[SL-X]** - the slack used in the lexicographic network is 0.X

Tags not listed are not relevant to the purpose of the report.

The reward tags, instead, are defined as follows:

- **[HLT] Healthy Reward:** a fixed reward (1.0) given every timestamp that the ant is considered healthy (which is defined below).
- **[FWD] Forward Reward:** reward calculated as the difference between the x position of the Ant before and after the action, divided by `dt`, which is the time between frames. The reward is positive if the movement is in the positive direction of the x axis.
- **[CST] Control Cost:** a negative reward to penalize actions too large. It is calculated as the sum of the squared actions taken, multiplied by a weight factor
- **[ORG] Distance from Origin:** represents the distance from the spawn point and is calculated similarly to the Forward Reward, but using the xy position instead of the x position

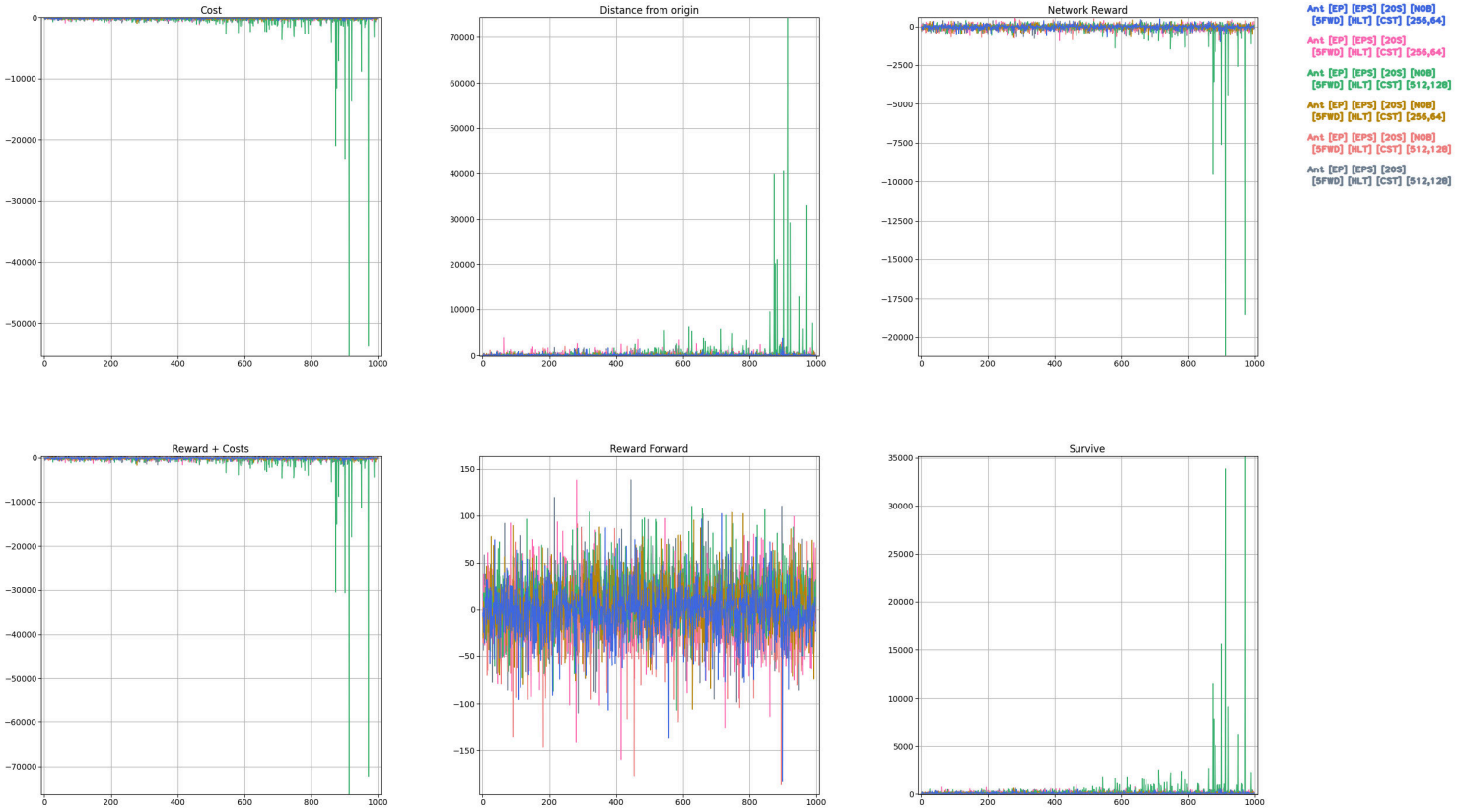
Finally, as for the name of the plots, the prefix dictates which metric they belong to:

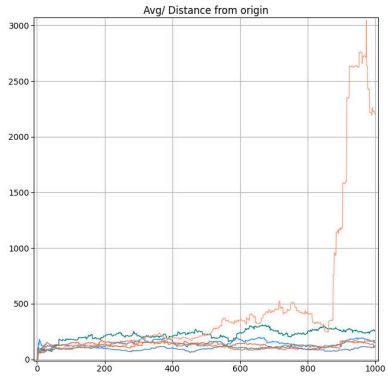
- **No prefix** - cumulative reward obtained during training in every episode

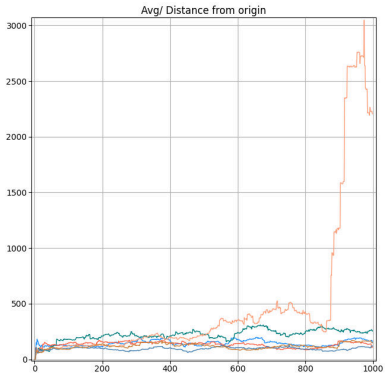
- **Avg/** - average cumulative reward obtained during training over 100 episodes
- **Test/** - cumulative reward obtained during test in every episode
- **Test Avg/** - average cumulative reward obtained during test over 100 episodes

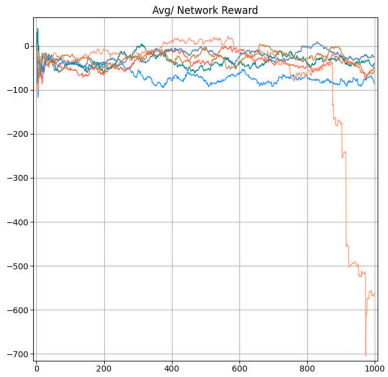
Appendix

Continuous DQN - [5FWD] [HLT] [CST]

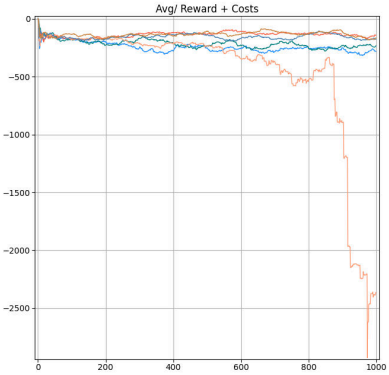


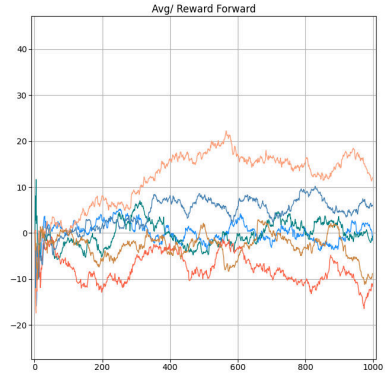

Avg/ Distance from origin
3000
2500
2000
1500
1000
500
0
0
200
400
600
800
1000

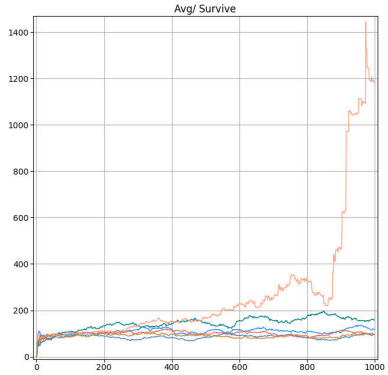

Avg/ Distance from origin
3000
2500
2000
1500
1000
500
0
0
200
400
600
800
1000

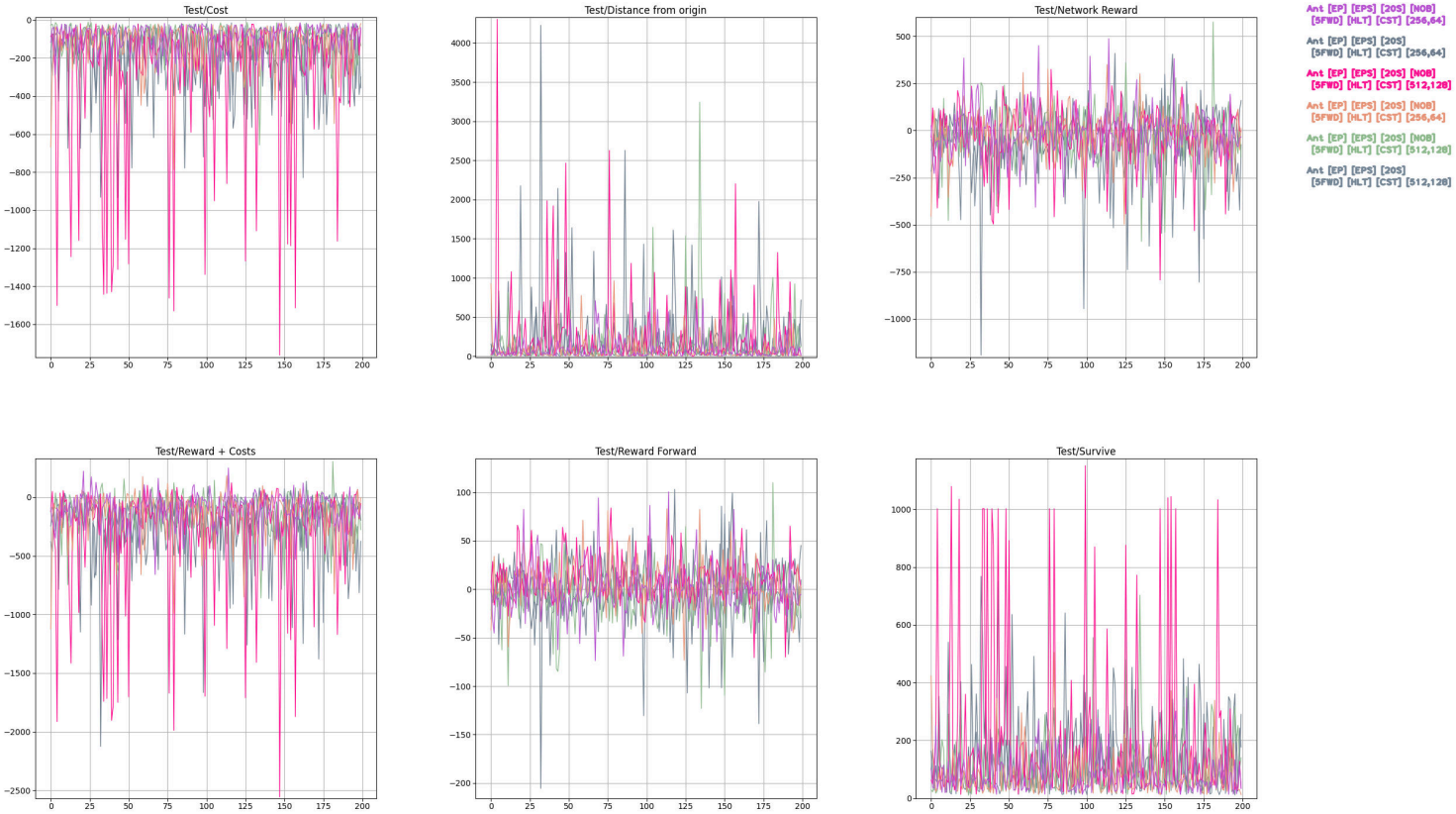

Avg/ Network Reward
0
−100
−200
−300
−400
−500
−600
−700
0
200
400
600
800
1000

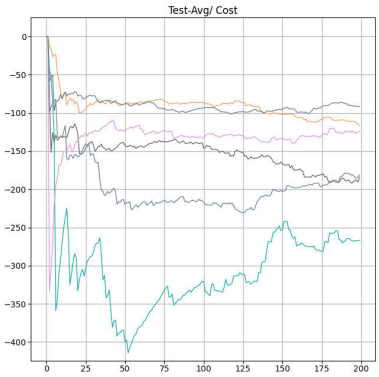

Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [SFWD] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [512,128]
Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [512,128]
Ant [EP] [EPS] [20S] [SFWD] [HLT] [CST] [512,128]

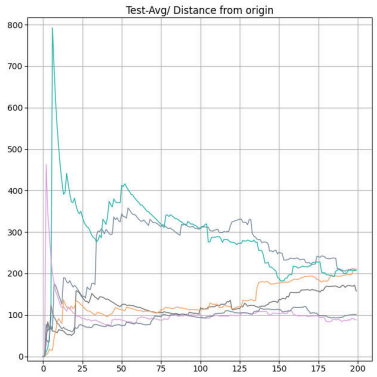

Avg/ Reward + Costs
0
−500
−1000
−1500
−2000
−2500
0
200
400
600
800
1000

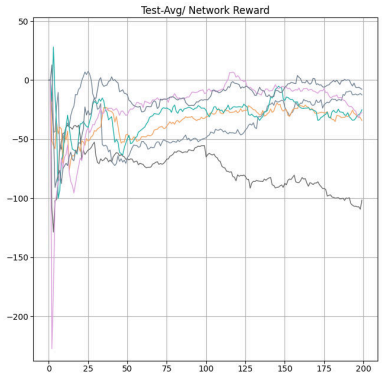

Avg/ Reward Forward
40
30
20
10
0
−10
−20
0
200
400
600
800
1000


Avg/ Survive
1400
1200
1000
800
600
400
200
0
0
200
400
600
800
1000

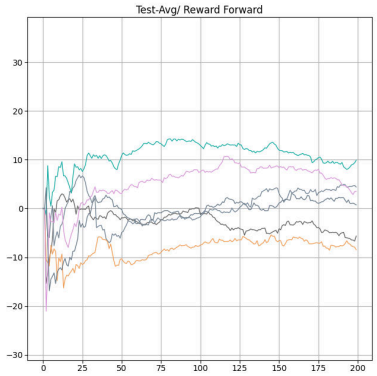

Test/Cost
Test/Distance from origin
Test/Network Reward
Test/Reward + Costs
Test/Reward Forward
Test/Survive
Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [SFWD] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [512,128]
Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [512,128]
Ant [EP] [EPS] [20S] [SFWD] [HLT] [CST] [512,128]

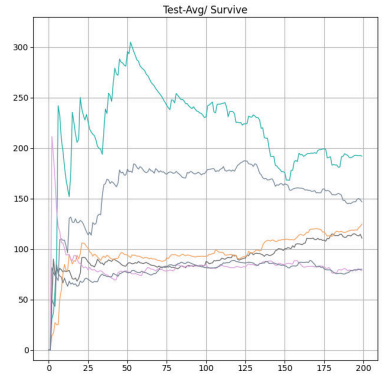

Test-Avg/ Cost


Test-Avg/ Distance from origin


Test-Avg/ Network Reward

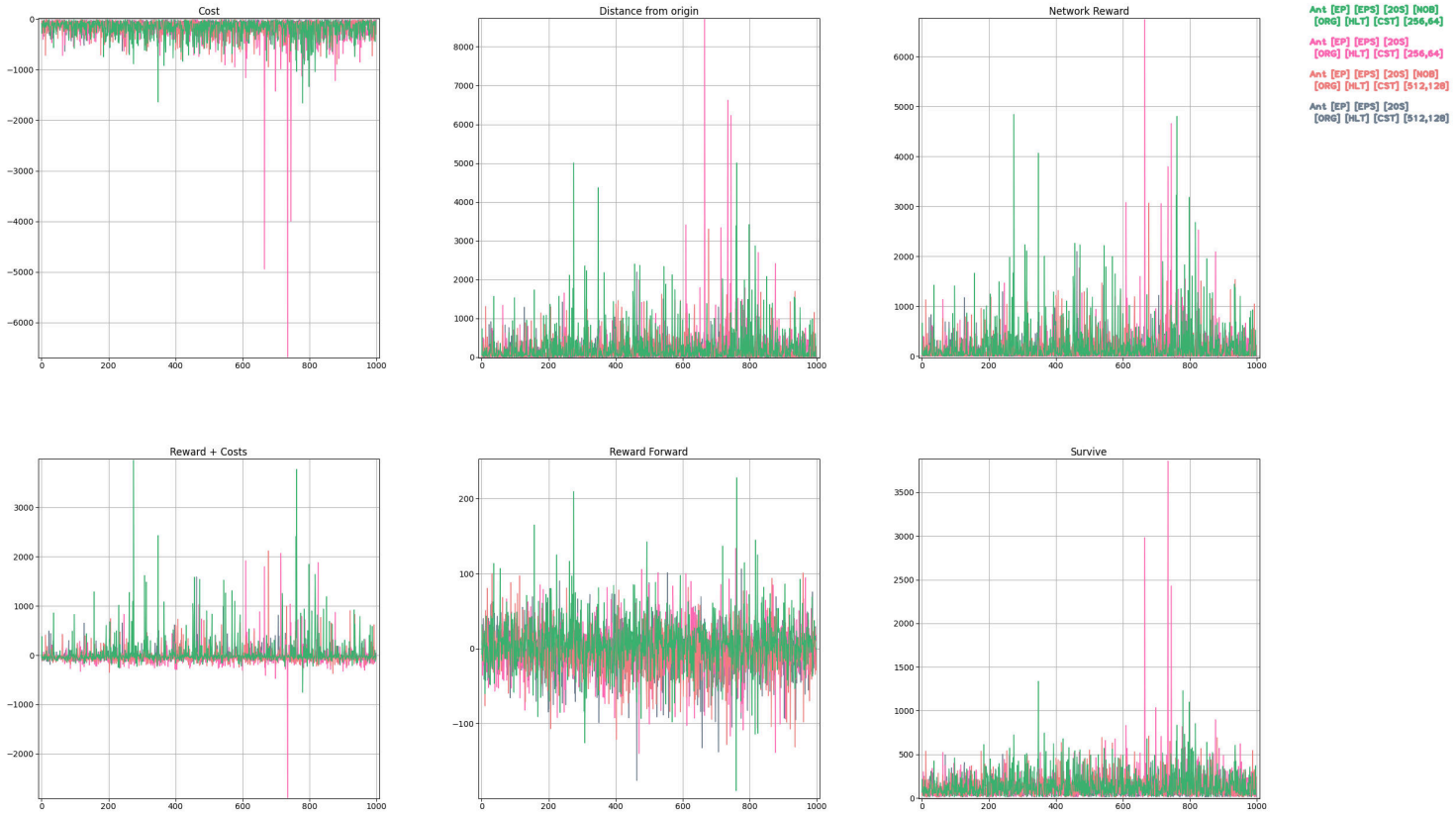

Test-Avg/ Reward + Costs

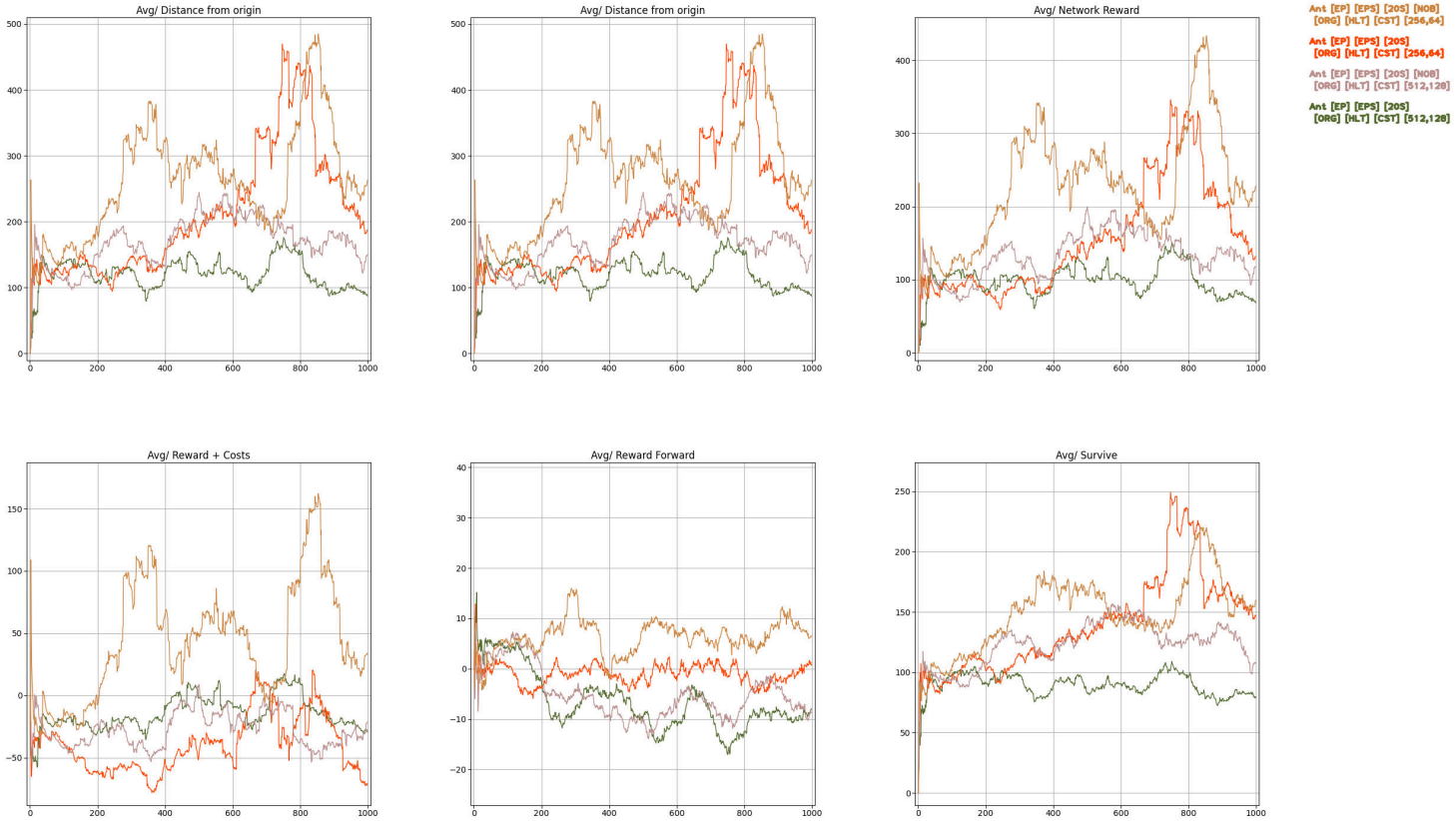

Test-Avg/ Reward Forward

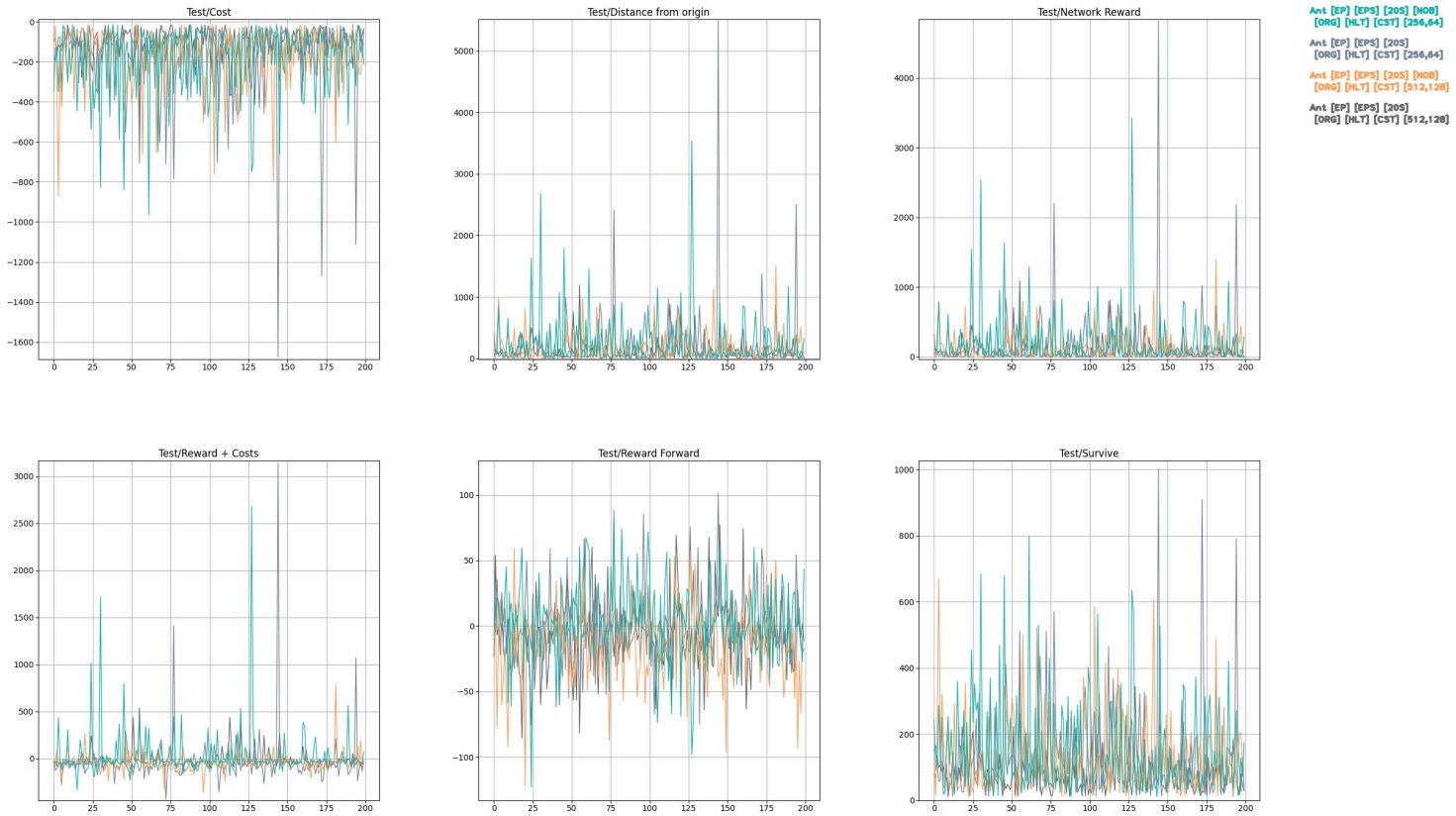

Test-Avg/ Survive


Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [SFWD] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [512,128]
Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [NOB] [SFWD] [HLT] [CST] [512,128]
Ant [EP] [EPS] [20S] [SFWD] [HLT] [CST] [512,128]

Continuous DQN - [ORG] [HLT] [CST]

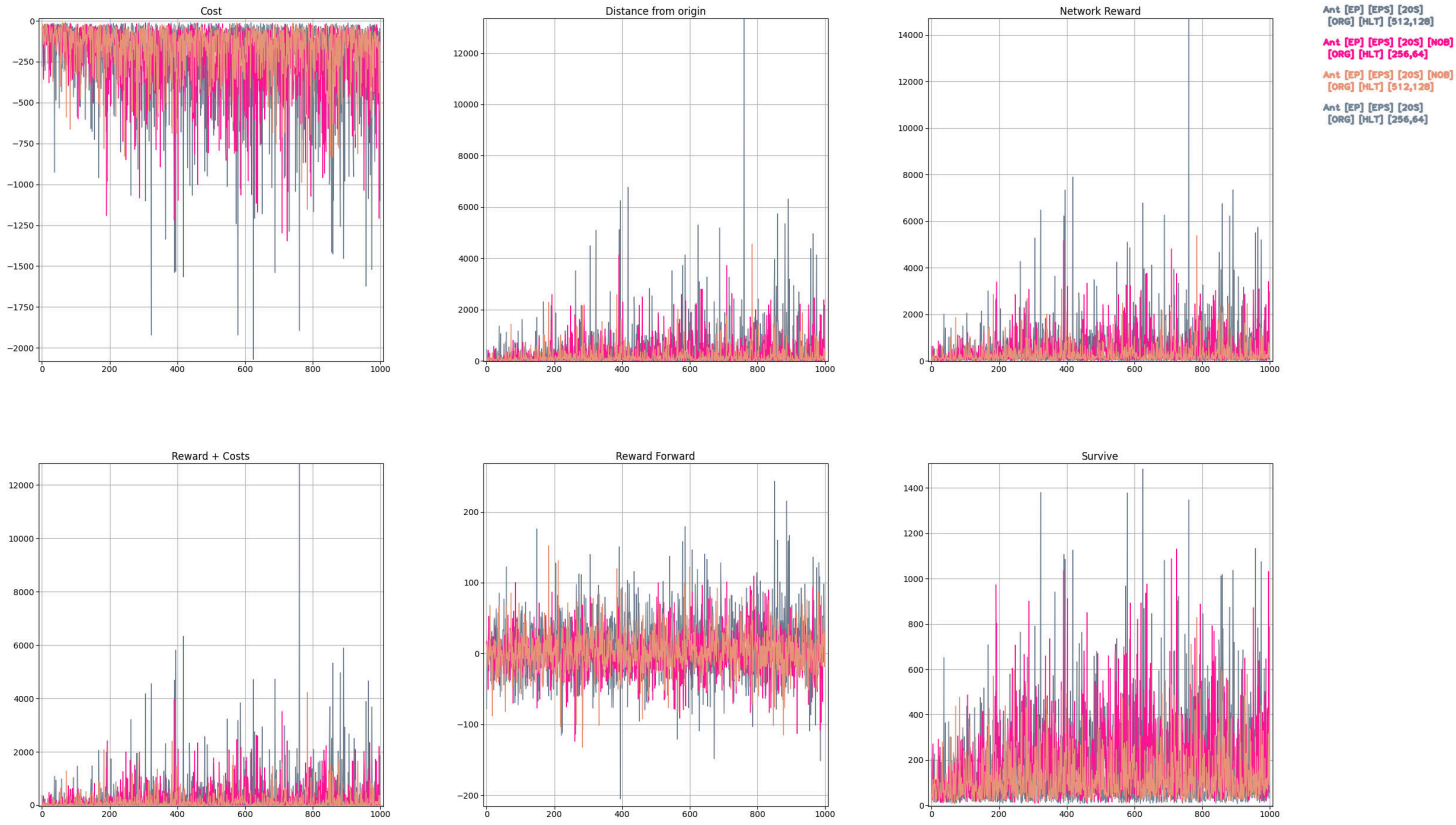


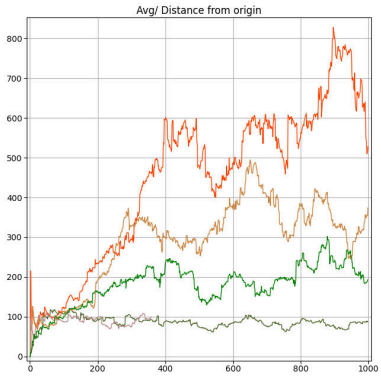

Avg/ Distance from origin
Avg/ Distance from origin
Avg/ Network Reward
Avg/ Reward + Costs
Avg/ Reward Forward
Avg/ Survive
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [ORG] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [CST] [512,128]
Ant [EP] [EPS] [20S] [ORG] [HLT] [CST] [512,128]

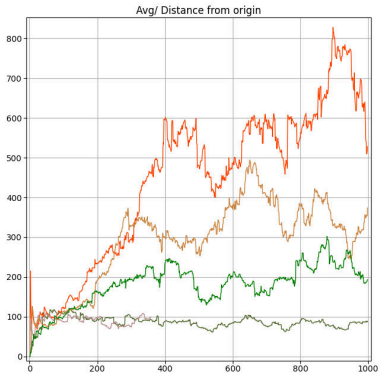

Test/Cost
Test/Distance from origin
Test/Network Reward
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [ORG] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [CST] [512,128]
Ant [EP] [EPS] [20S] [ORG] [HLT] [CST] [512,128]
Test/Reward + Costs
Test/Reward Forward
Test/Survive

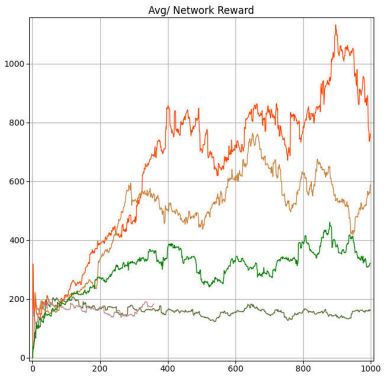

Test-Avg/ Cost
Test-Avg/ Distance from origin
Test-Avg/ Network Reward
Test-Avg/ Reward + Costs
Test-Avg/ Reward Forward
Test-Avg/ Survive
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [ORG] [HLT] [CST] [256,64]
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [CST] [512,128]
Ant [EP] [EPS] [20S] [ORG] [HLT] [CST] [512,128]

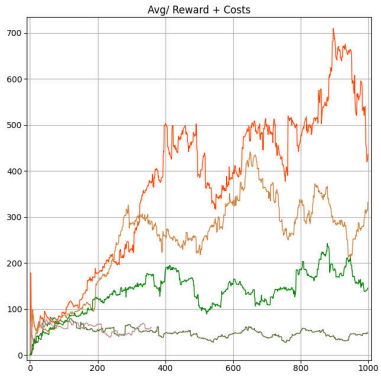
Continuous DQN - [ORG] [HLT]




Avg/ Distance from origin


Avg/ Distance from origin

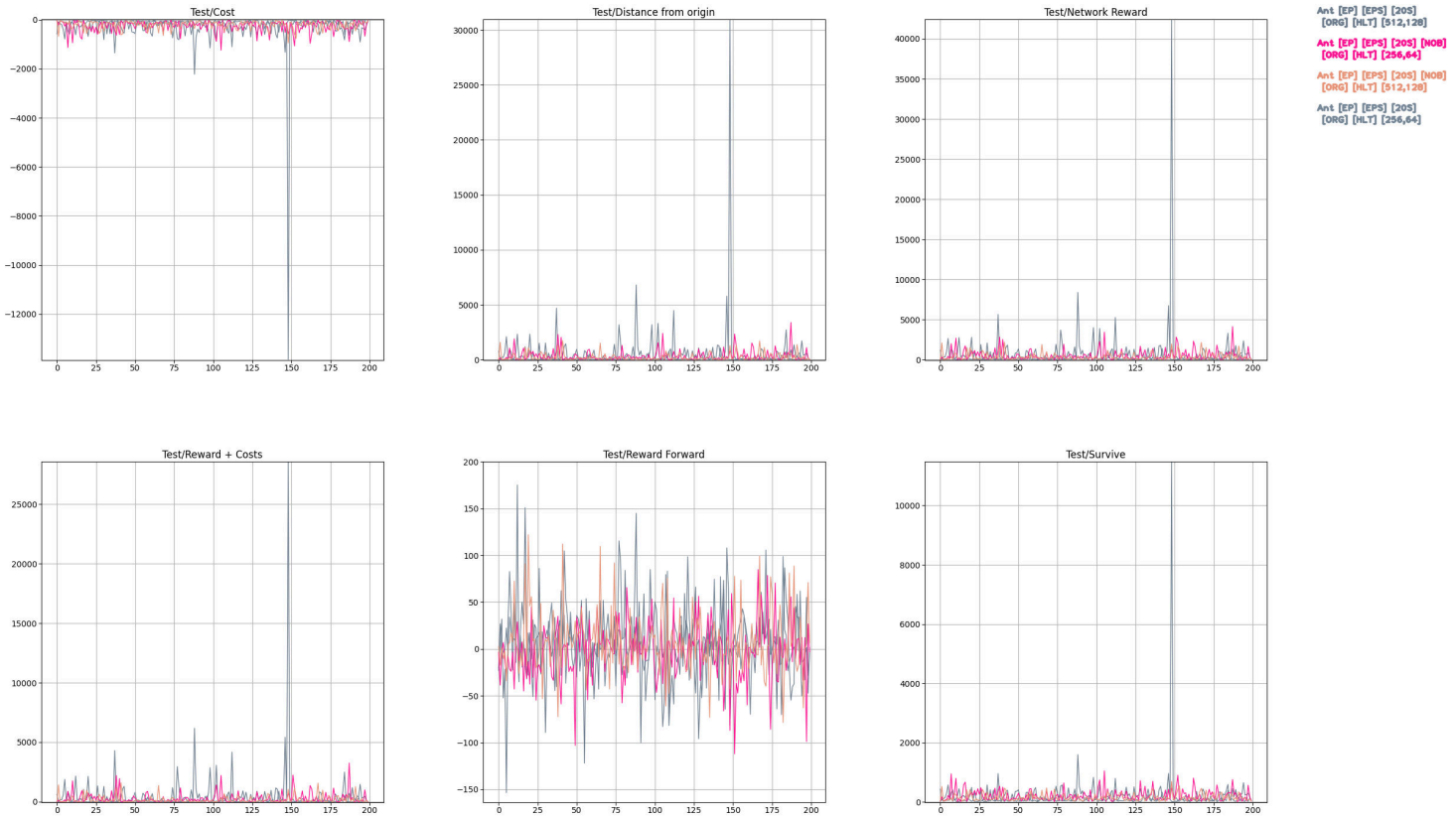

Avg/ Network Reward

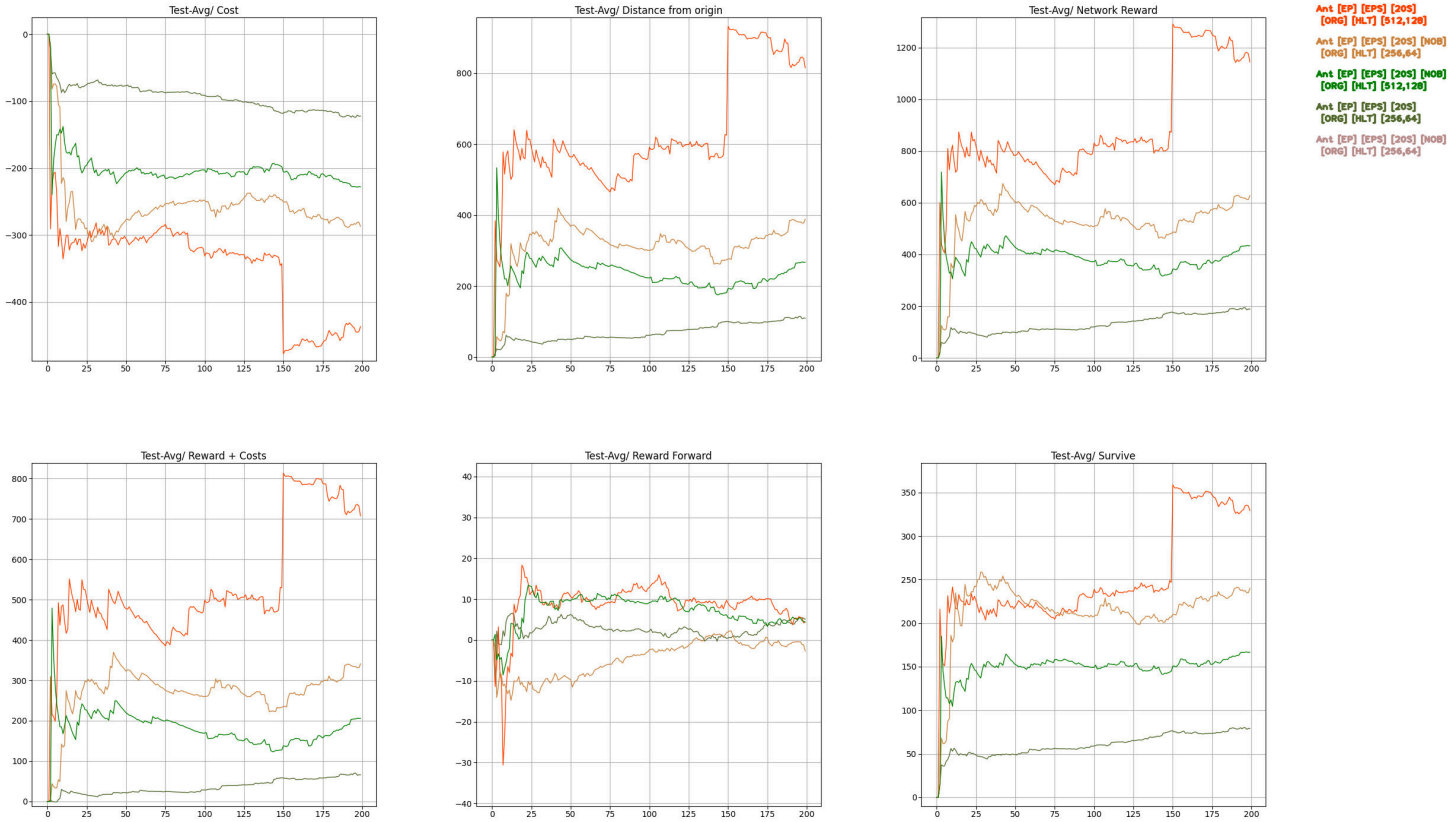

Avg/ Reward + Costs


Avg/ Reward Forward

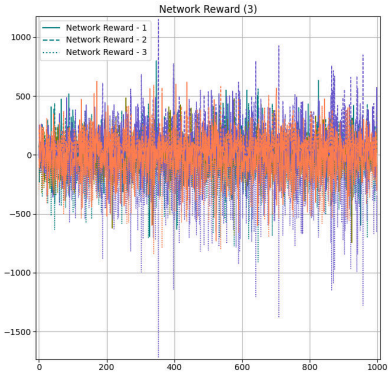

Avg/ Survive

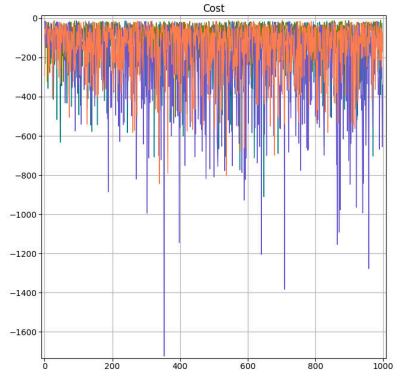

Ant [EP] [EPS] [20S] [ORG] [HLT] [512,128]
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [256,64]
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [512,128]
Ant [EP] [EPS] [20S] [ORG] [HLT] [256,64]
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [256,64]

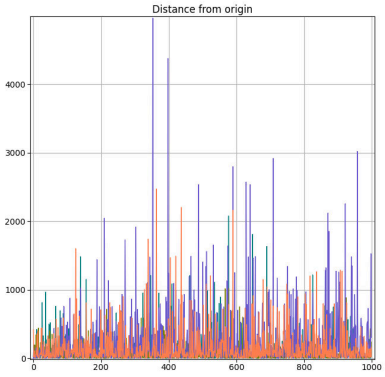

Test/Cost
Test/Distance from origin
Test/Network Reward
Ant [EP] [EPS] [20S] [ORG] [HLT] [512,128]
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [256,64]
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [512,128]
Ant [EP] [EPS] [20S] [ORG] [HLT] [256,64]
Test/Reward + Costs
Test/Reward Forward
Test/Survive

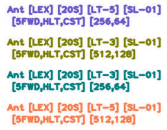

Test-Avg/ Cost
Test-Avg/ Distance from origin
Test-Avg/ Network Reward
Test-Avg/ Reward + Costs
Test-Avg/ Reward Forward
Test-Avg/ Survive
Ant [EP] [EPS] [20S] [ORG] [HLT] [512,128]
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [256,64]
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [512,128]
Ant [EP] [EPS] [20S] [ORG] [HLT] [256,64]
Ant [EP] [EPS] [20S] [NOB] [ORG] [HLT] [256,64]

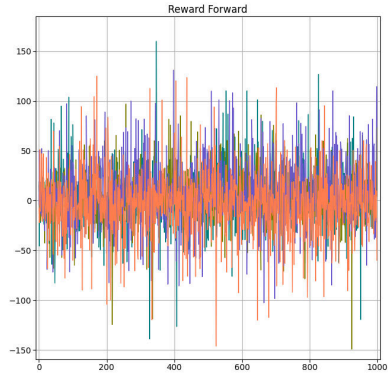
Lexicographic CDQN - [5FWD,HLT,CST]

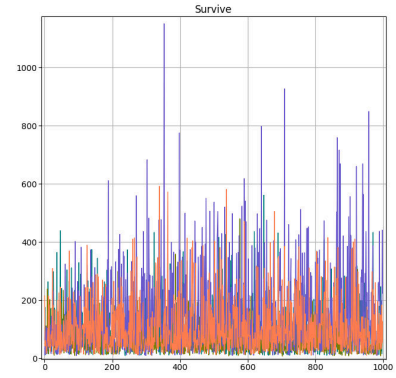


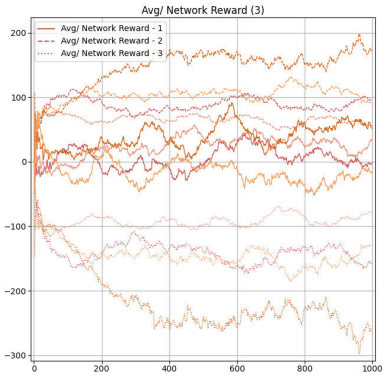


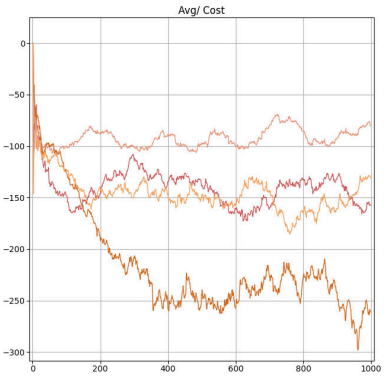


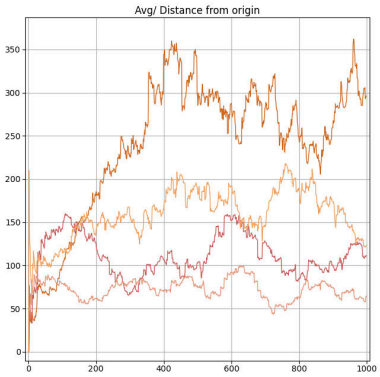










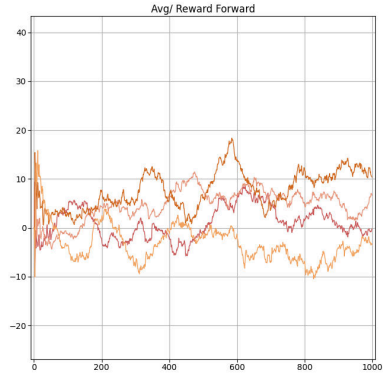


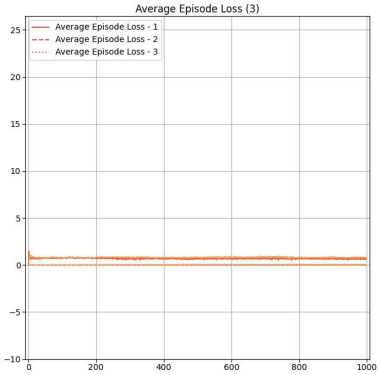
Ant [LEX] [20S] [LT-5] [SL-01] [SFWD,HLT,CST] [256,64]

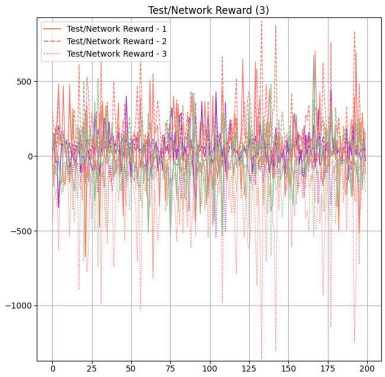
Ant [LEX] [20S] [LT-3] [SL-01] [SFWD,HLT,CST] [512,128]

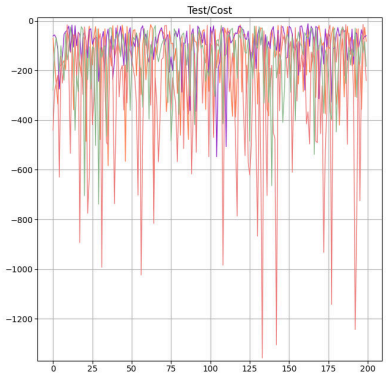
Ant [LEX] [20S] [LT-3] [SL-01] [SFWD,HLT,CST] [256,64]

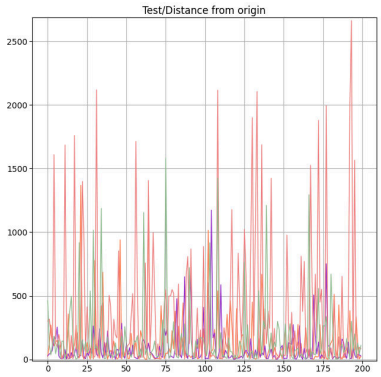
Ant [LEX] [20S] [LT-5] [SL-01] [SFWD,HLT,CST] [512,128]

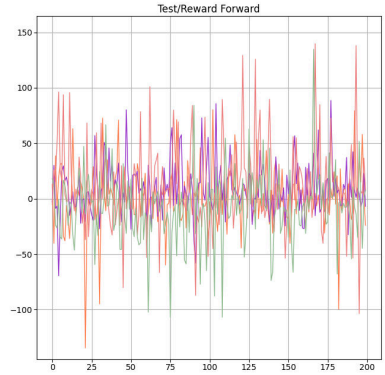


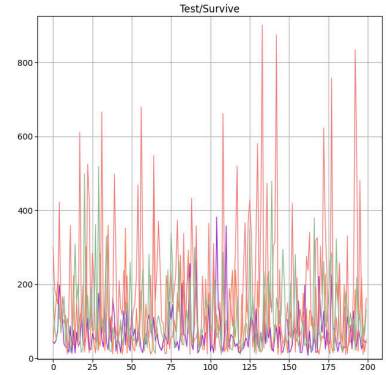


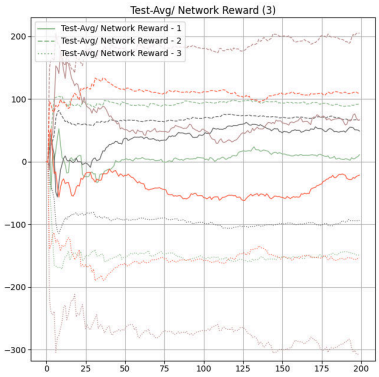

Test/Network Reward (3)
Test/Network Reward - 1
Test/Network Reward - 2
Test/Network Reward - 3

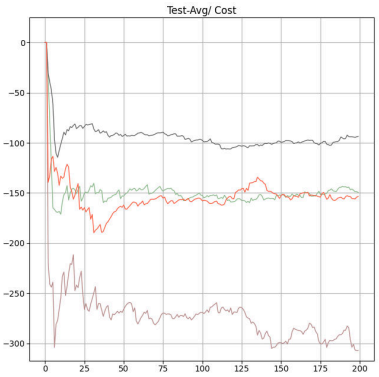

Test/Cost

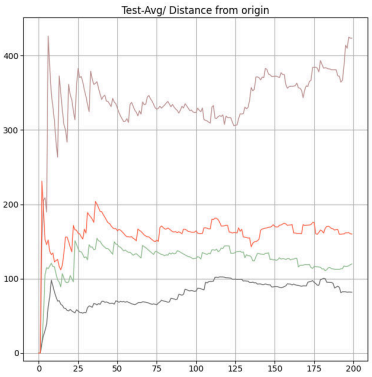

Test/Distance from origin


Test/Reward Forward


Test/Survive





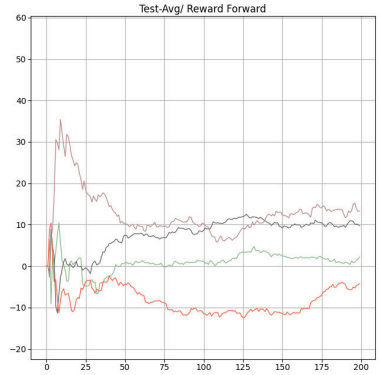


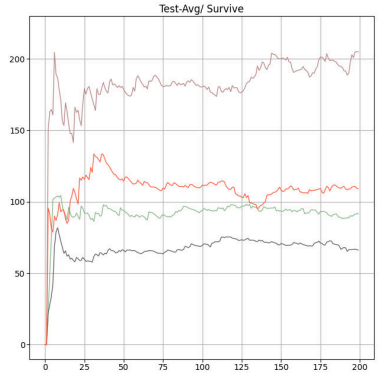
Ant [LEX] [20S] [LT-5] [SL-01] [5FWD,HLT,CST] [256,64]

Ant [LEX] [20S] [LT-3] [SL-01] [5FWD,HLT,CST] [512,128]

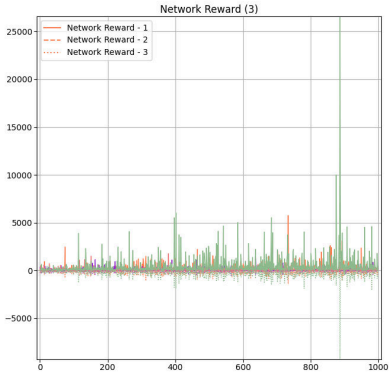
Ant [LEX] [20S] [LT-3] [SL-01] [5FWD,HLT,CST] [256,64]

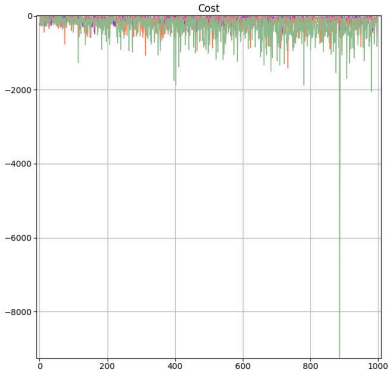
Ant [LEX] [20S] [LT-5] [SL-01] [5FWD,HLT,CST] [512,128]

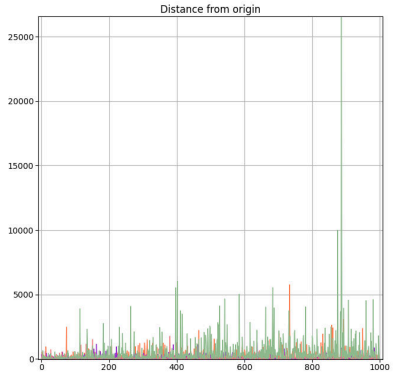


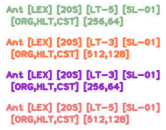


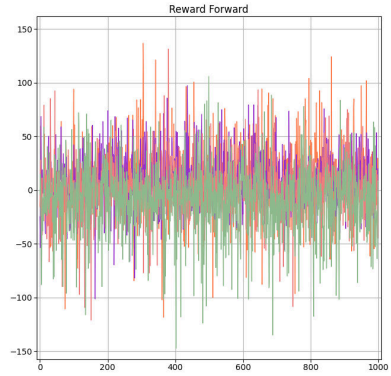
Lexicographic CDQN - [ORG,HLT,CST]

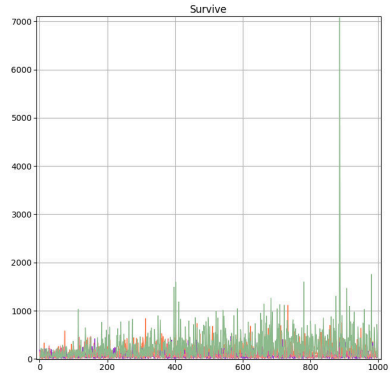


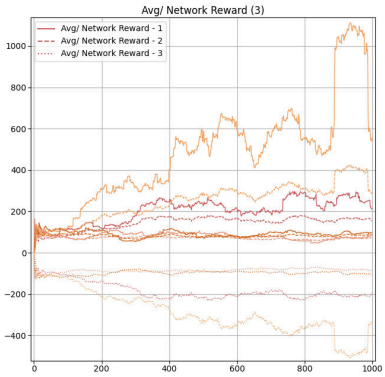


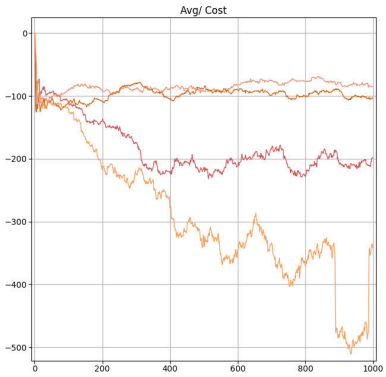










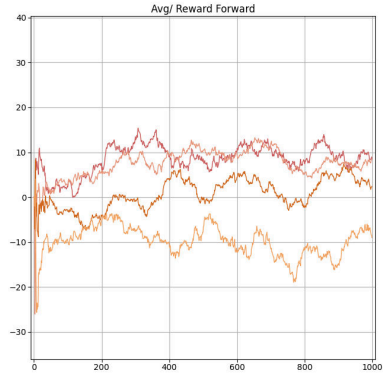


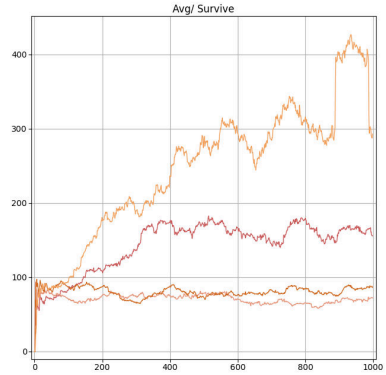
Ant [LEX] [20S] [LT-5] [SL-01] [ORG,HLT,CST] [256,64]

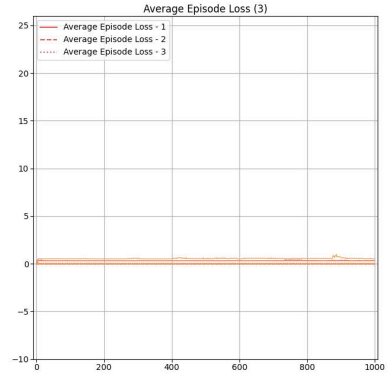
Ant [LEX] [20S] [LT-3] [SL-01] [ORG,HLT,CST] [512,128]

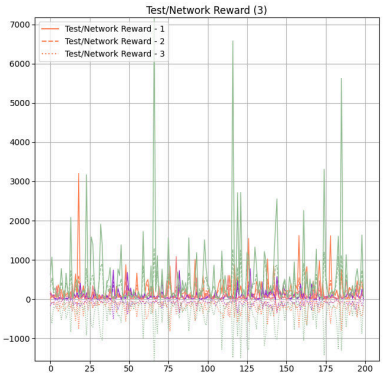
Ant [LEX] [20S] [LT-3] [SL-01] [ORG,HLT,CST] [256,64]

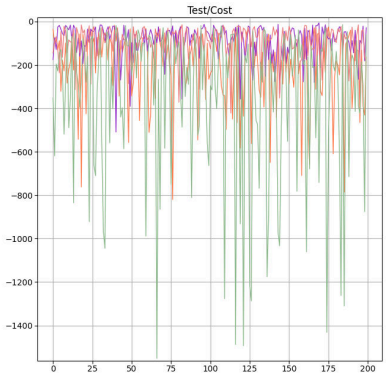
Ant [LEX] [20S] [LT-5] [SL-01] [ORG,HLT,CST] [512,128]

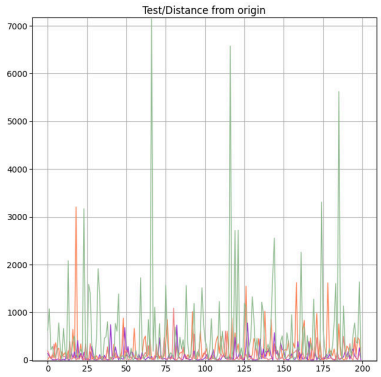




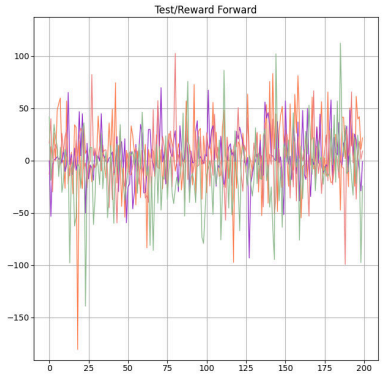


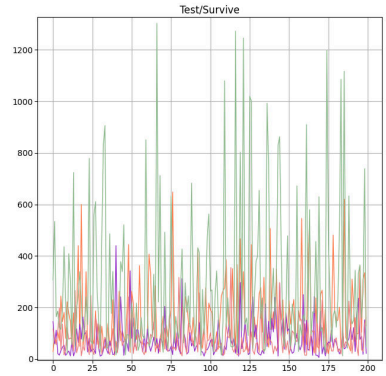

Test/Network Reward (3)
Test/Network Reward - 1
Test/Network Reward - 2
Test/Network Reward - 3

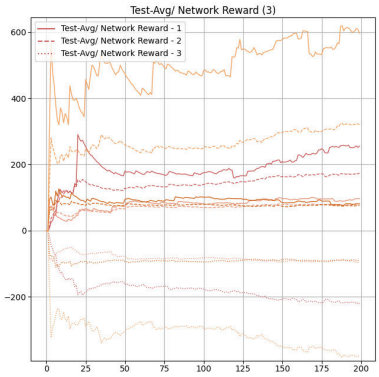

Test/Cost

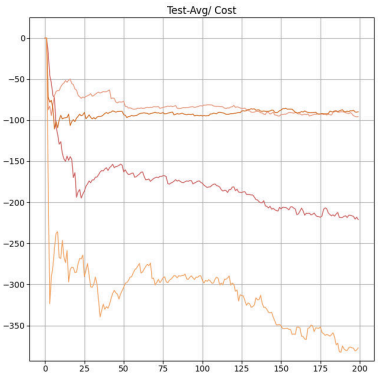

Test/Distance from origin

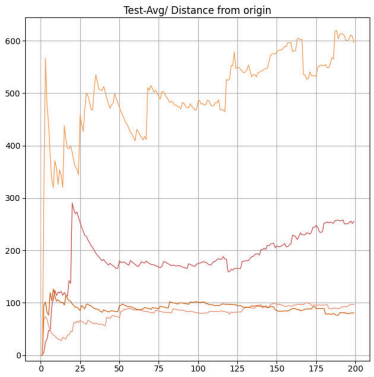

Ant [LEX] [20S] [LT-5] [SL-01] [ORG,HLT,CST] [256,64]
Ant [LEX] [20S] [LT-3] [SL-01] [ORG,HLT,CST] [512,128]
Ant [LEX] [20S] [LT-3] [SL-01] [ORG,HLT,CST] [256,64]
Ant [LEX] [20S] [LT-5] [SL-01] [ORG,HLT,CST] [512,128]

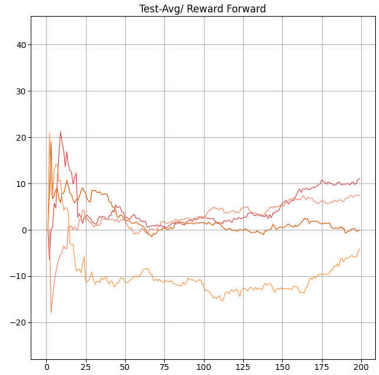

Test/Reward Forward

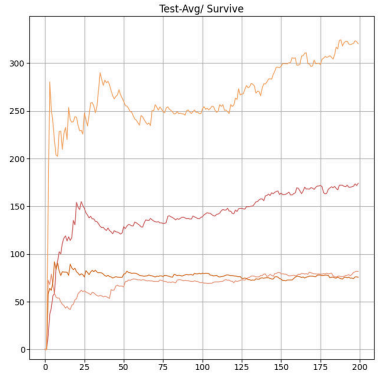

Test/Survive











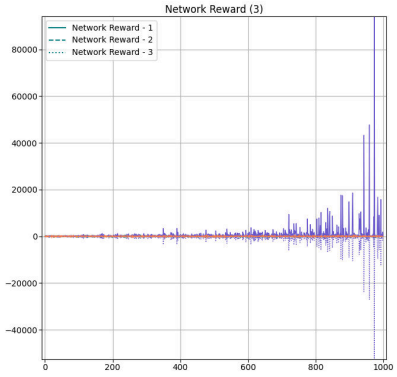
Ant [LEX] [20S] [LT-5] [SL-01] [ORG,HLT,CST] [256,64]

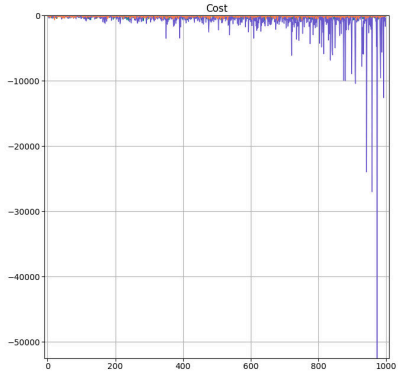
Ant [LEX] [20S] [LT-3] [SL-01] [ORG,HLT,CST] [512,128]

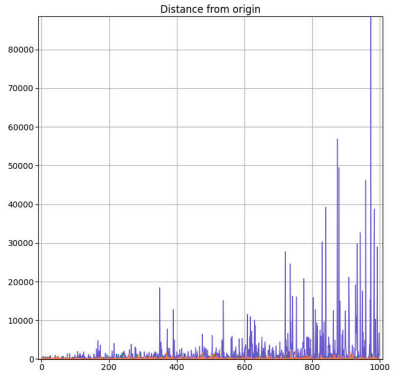
Ant [LEX] [20S] [LT-3] [SL-01] [ORG,HLT,CST] [256,64]

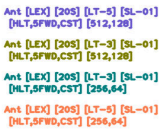
Ant [LEX] [20S] [LT-5] [SL-01] [ORG,HLT,CST] [512,128]

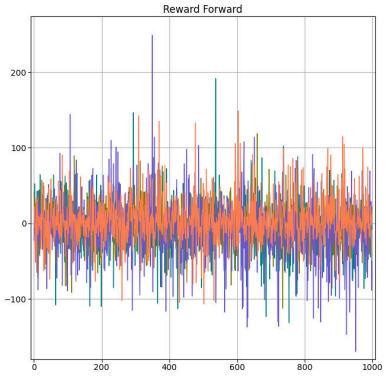
Lexicographic CDQN - [HLT,5FWD,CST]

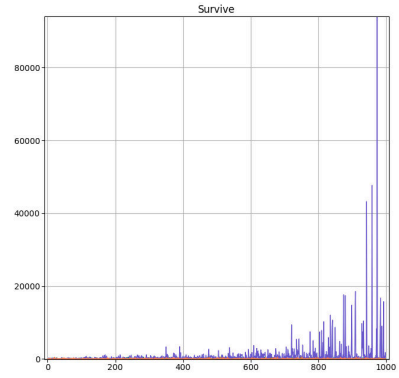


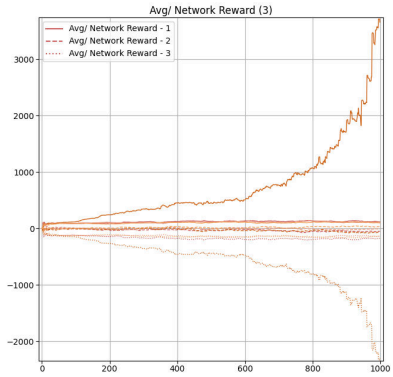


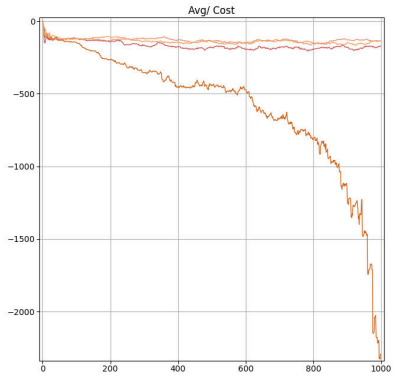


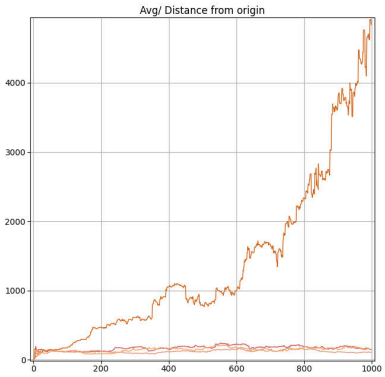


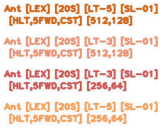


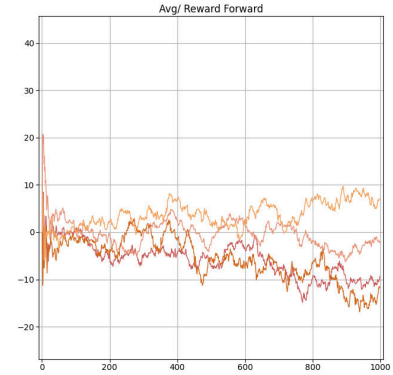


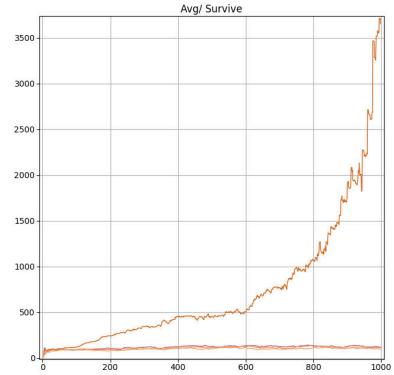

Avg/ Network Reward (3)
Avg/ Network Reward - 1
Avg/ Network Reward - 2
Avg/ Network Reward - 3

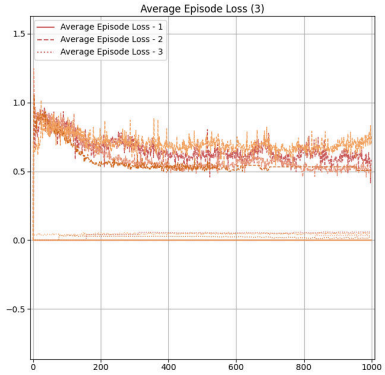

Avg/ Cost

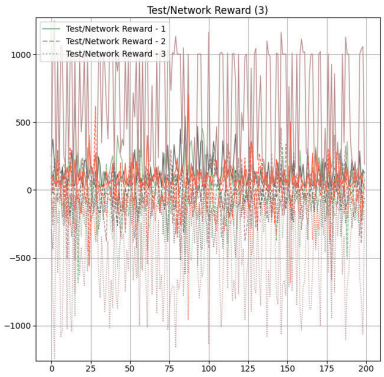

Avg/ Distance from origin

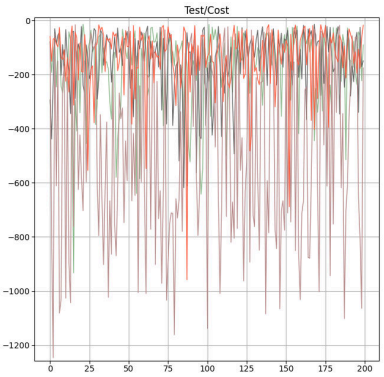

Ant [LEX] [20S] [LT-5] [SL-01] [HLT,5FWD,CST] [512,128]
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,5FWD,CST] [512,128]
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,5FWD,CST] [256,64]
Ant [LEX] [20S] [LT-5] [SL-01] [HLT,5FWD,CST] [256,64]

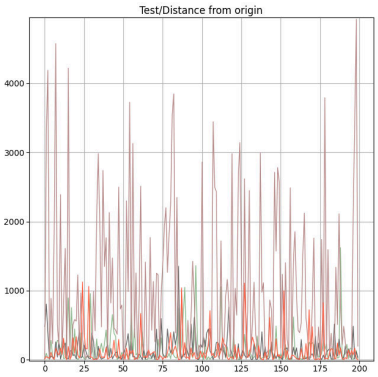

Avg/ Reward Forward


Avg/ Survive

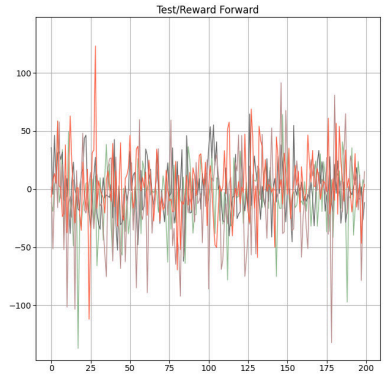

Average Episode Loss (3)
Average Episode Loss - 1
Average Episode Loss - 2
Average Episode Loss - 3

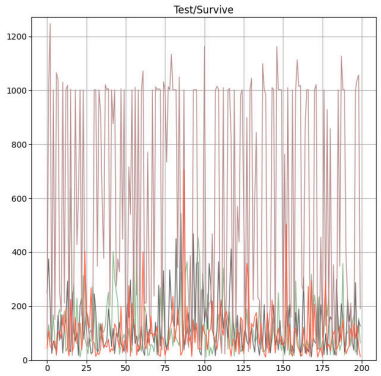

Test/Network Reward (3)
Test/Network Reward - 1
Test/Network Reward - 2
Test/Network Reward - 3

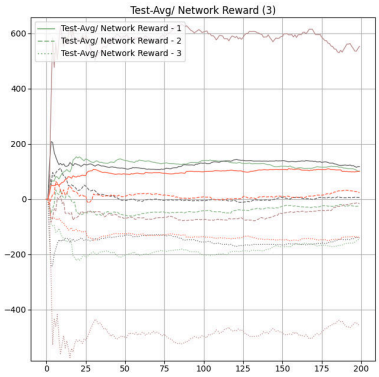

Test/Cost

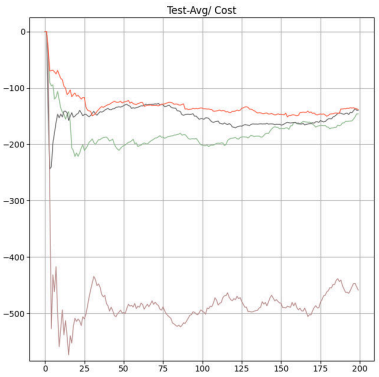

Test/Distance from origin

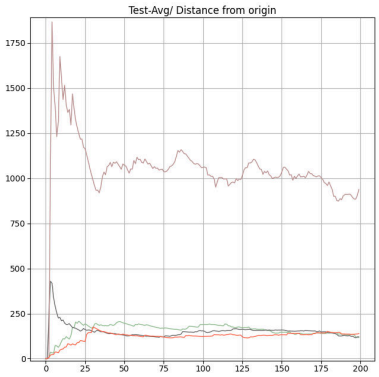

Ant [LEX] [20S] [LT-5] [SL-01] [HLT,5FWD,CST] [512,128]
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,5FWD,CST] [512,128]
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,5FWD,CST] [256,64]
Ant [LEX] [20S] [LT-5] [SL-01] [HLT,5FWD,CST] [256,64]


Test/Reward Forward

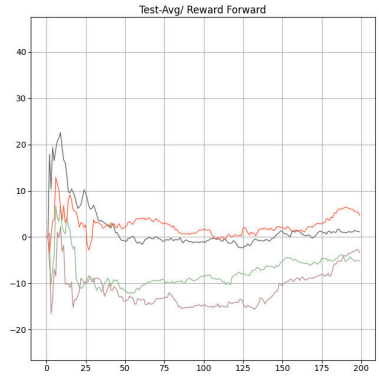

Test/Survive

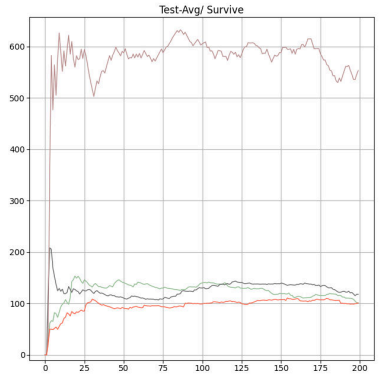

Test-Avg/ Network Reward (3)
Test-Avg/ Network Reward - 1
Test-Avg/ Network Reward - 2
Test-Avg/ Network Reward - 3


Test-Avg/ Cost

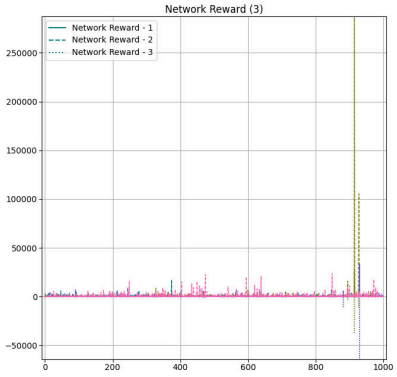

Test-Avg/ Distance from origin

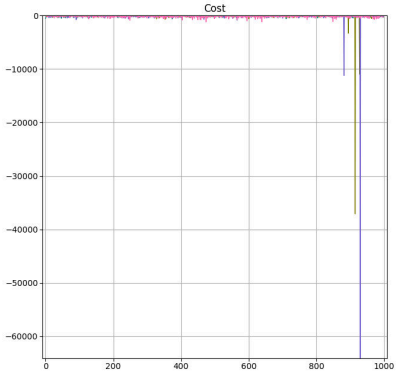

Ant [LEX] [20S] [LT-5] [SL-01] [HLT,5FWD,CST] [512,128]
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,5FWD,CST] [512,128]
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,5FWD,CST] [256,64]
Ant [LEX] [20S] [LT-5] [SL-01] [HLT,5FWD,CST] [256,64]

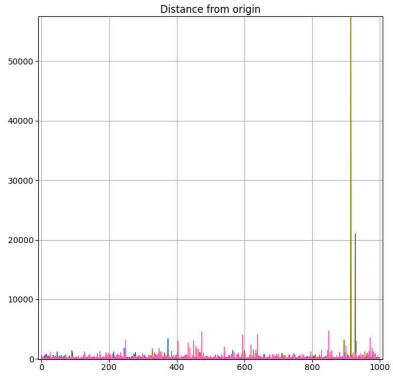

Test-Avg/ Reward Forward

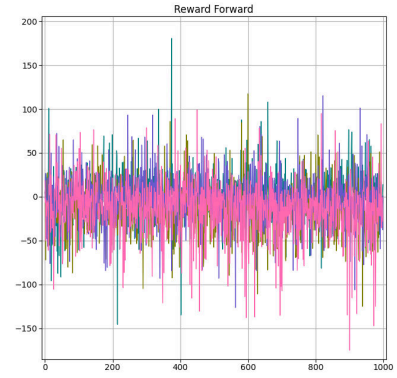

Test-Avg/ Survive

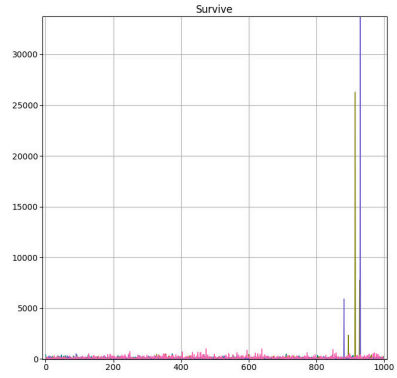
Lexicographic CDQN - [HLT,ORG,CST]

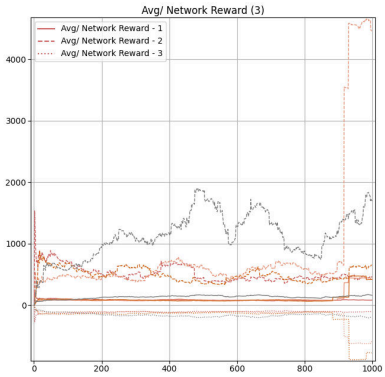


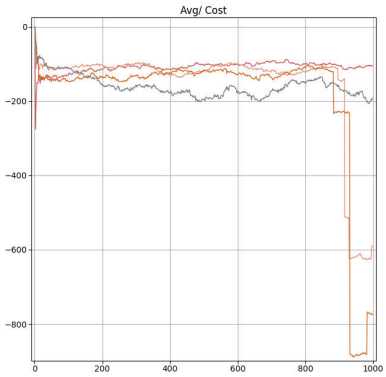


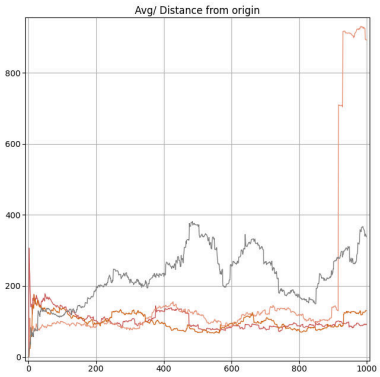










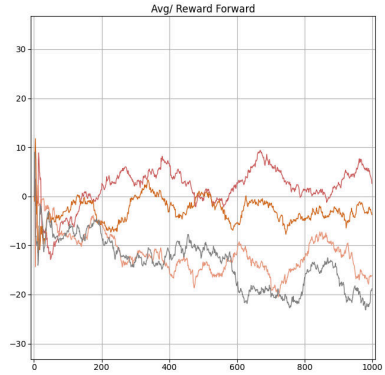


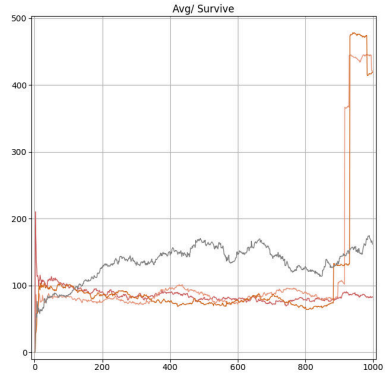
Ant [LEX] [20S] [LT-5] [SL-01] [HLT,ORG,CST] [256,64]

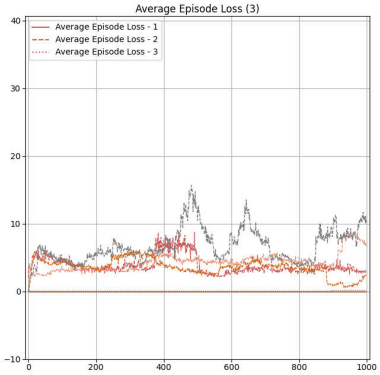
Ant [LEX] [20S] [LT-5] [SL-01] [HLT,ORG,CST] [512,128]

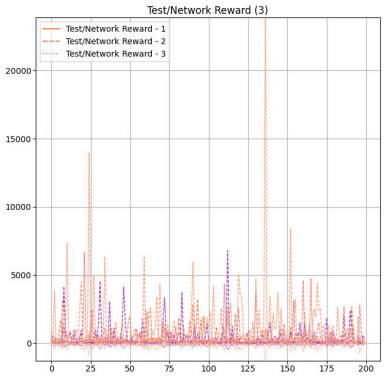
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,ORG,CST] [256,64]

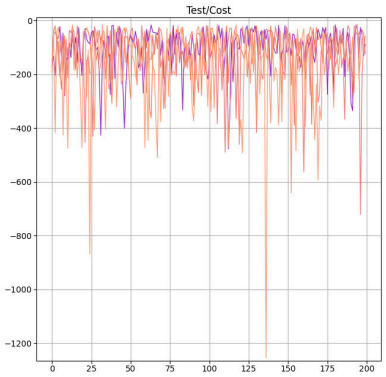
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,ORG,CST] [512,128]

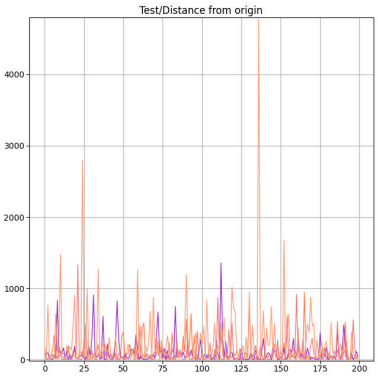










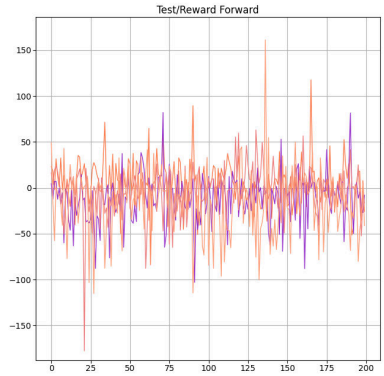


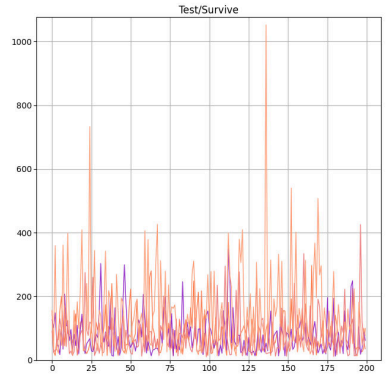
Ant [LEX] [20S] [LT-5] [SL-01] [HLT,ORG,CST] [256,64]

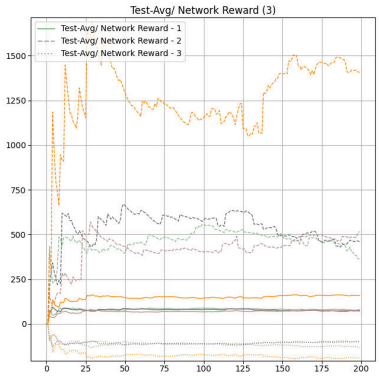
Ant [LEX] [20S] [LT-5] [SL-01] [HLT,ORG,CST] [512,128]

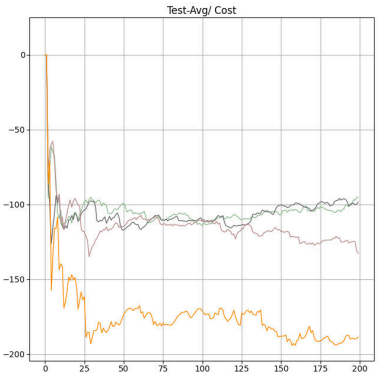
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,ORG,CST] [256,64]

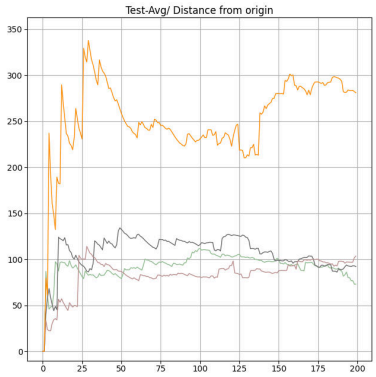
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,ORG,CST] [512,128]

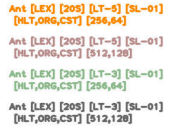


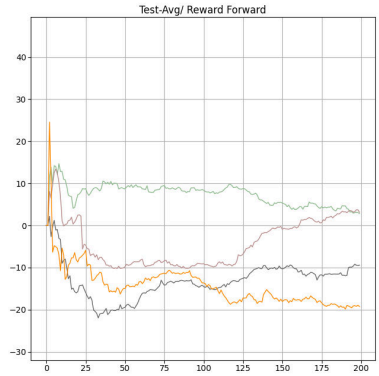


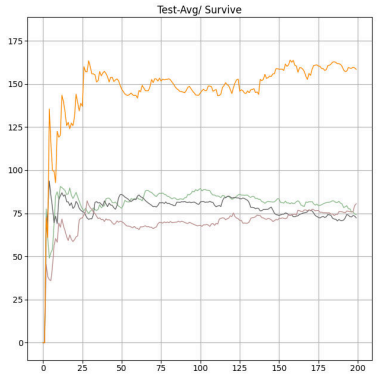

Test-Avg/ Network Reward (3)
Test-Avg/ Network Reward - 1
Test-Avg/ Network Reward - 2
Test-Avg/ Network Reward - 3


Test-Avg/ Cost

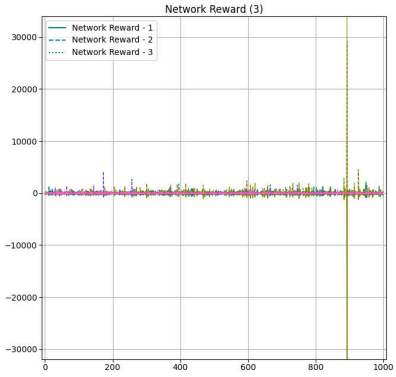

Test-Avg/ Distance from origin

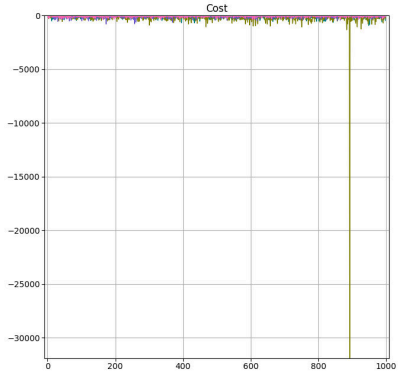

Ant [LEX] [20S] [LT-5] [SL-01] [HLT,ORG,CST] [256,64]
Ant [LEX] [20S] [LT-5] [SL-01] [HLT,ORG,CST] [512,128]
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,ORG,CST] [256,64]
Ant [LEX] [20S] [LT-3] [SL-01] [HLT,ORG,CST] [512,128]

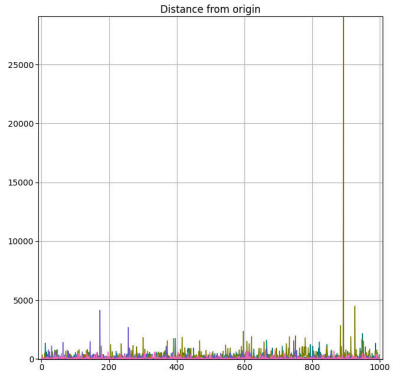

Test-Avg/ Reward Forward

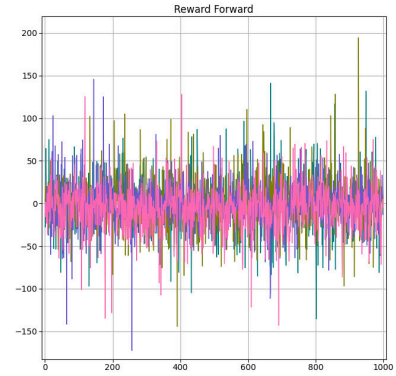

Test-Avg/ Survive

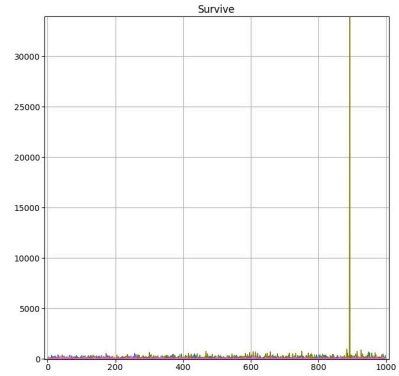
Lexicographic CDQN - [CST,ORG,HLT]

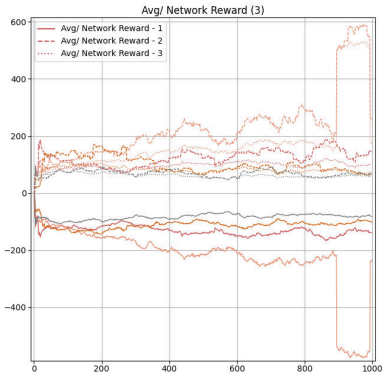


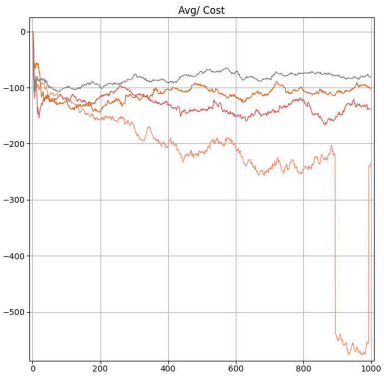


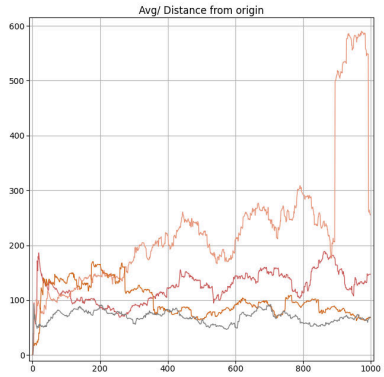










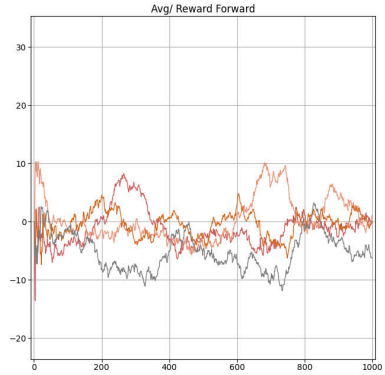


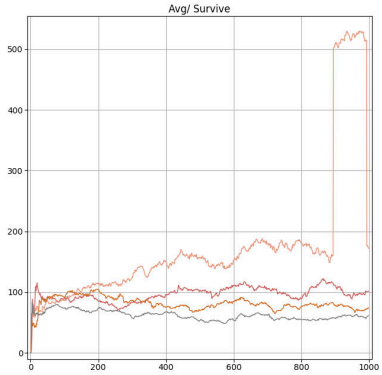
Ant [LEX] [20S] [LT-5] [SL-01] [CST,ORG,HLT] [512,128]

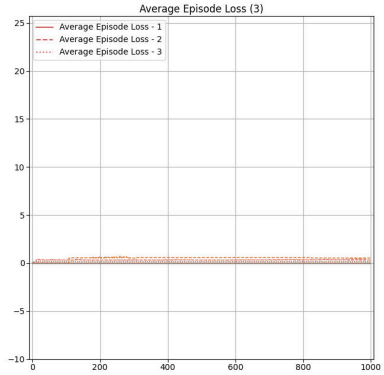
Ant [LEX] [20S] [LT-5] [SL-01] [CST,ORG,HLT] [256,64]

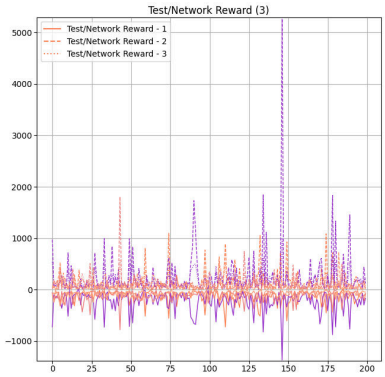
Ant [LEX] [20S] [LT-3] [SL-01] [CST,ORG,HLT] [512,128]

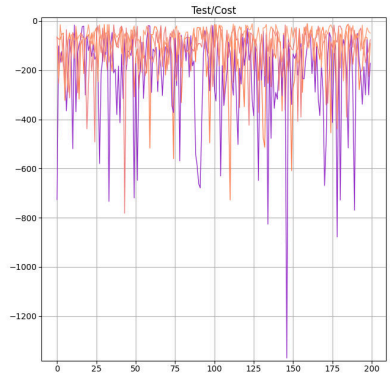
Ant [LEX] [20S] [LT-3] [SL-01] [CST,ORG,HLT] [256,64]

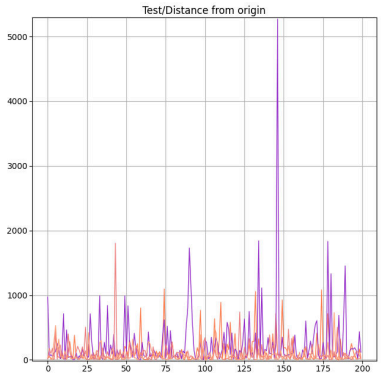










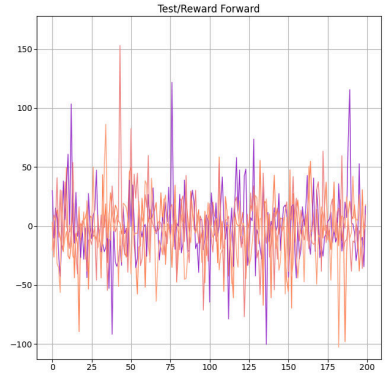


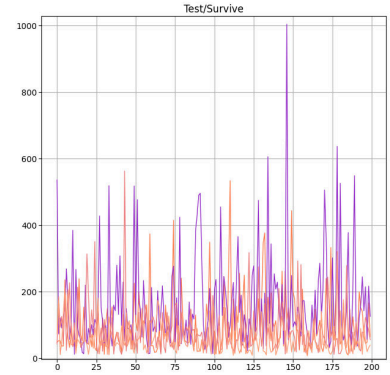
Ant [LEX] [20S] [LT-5] [SL-01] [CST,ORG,HLT] [512,128]

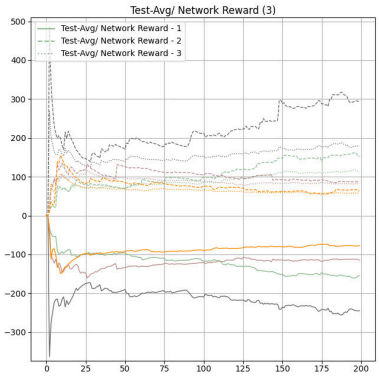
Ant [LEX] [20S] [LT-5] [SL-01] [CST,ORG,HLT] [256,64]

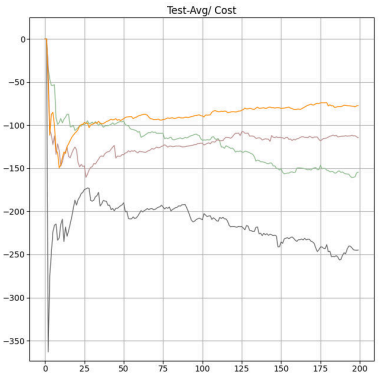
Ant [LEX] [20S] [LT-3] [SL-01] [CST,ORG,HLT] [512,128]

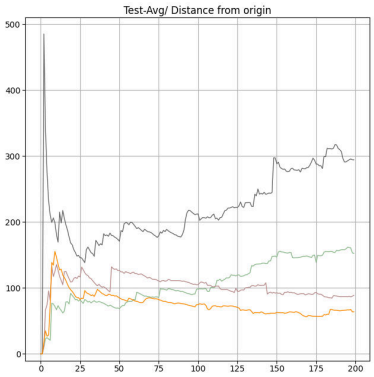
Ant [LEX] [20S] [LT-3] [SL-01] [CST,ORG,HLT] [256,64]

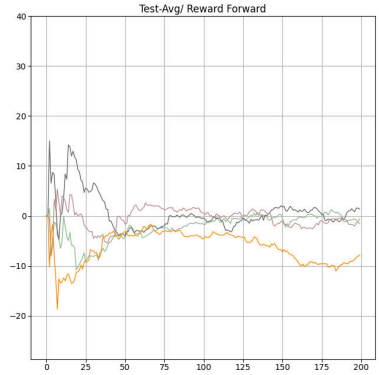


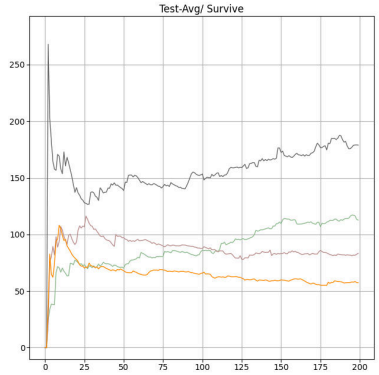












Ant [LEX] [20S] [LT-5] [SL-01] [CST,ORG,HLT] [512,128]

Ant [LEX] [20S] [LT-5] [SL-01] [CST,ORG,HLT] [256,64]

Ant [LEX] [20S] [LT-3] [SL-01] [CST,ORG,HLT] [512,128]

Ant [LEX] [20S] [LT-3] [SL-01] [CST,ORG,HLT] [256,64]