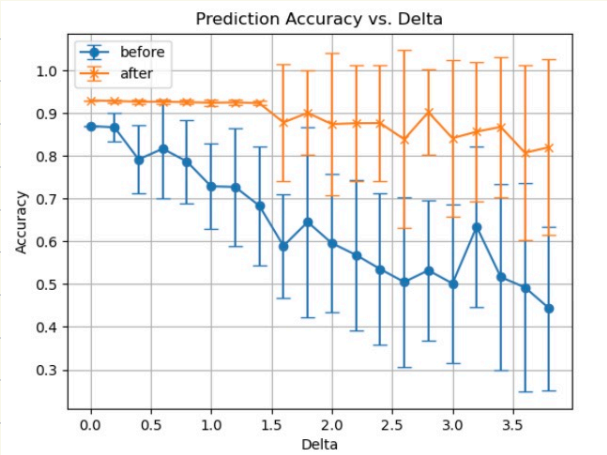
Q1


Prediction Accuracy vs. Delta
before
after
Accuracy
Delta
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5

Q2

(a)

number represent the order of cell visited

| | | | | |
|---|---|---|---|---|
| 12 | 13 | 14 | 15 | 16 |
| 11 | 2 | 3 | 4 | 17 |
| 10 | 1 | 0 | 5 | 18 |
| 9 | 8 | 7 | 6 | 19 |
| 24 | 23 | 22 | 21 | 20 |

start at here

$= \gamma^0 \times 1$ $= \gamma^2 \times 1$ reward

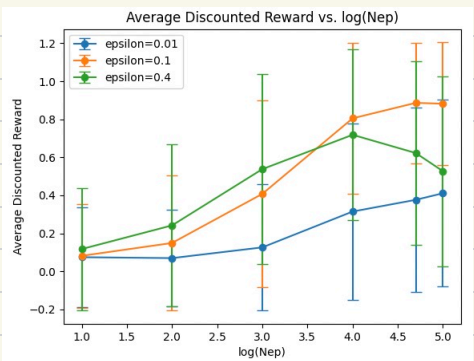
start location

as prize can at any cell except middle

$$\text{average discounted reward} = \frac{(\gamma^0 + \gamma^1 + \gamma^2 + \cdots + \gamma^{23} + \gamma^{24})}{25} = \frac{(1 + 0.95 + 0.95^2 + \cdots + 0.95^{23} + 0.95^{24})}{25}$$

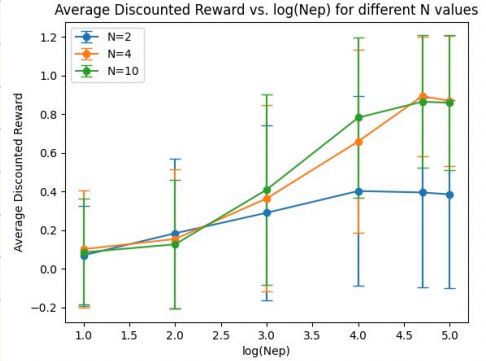
$$\approx 0.578$$

(b)

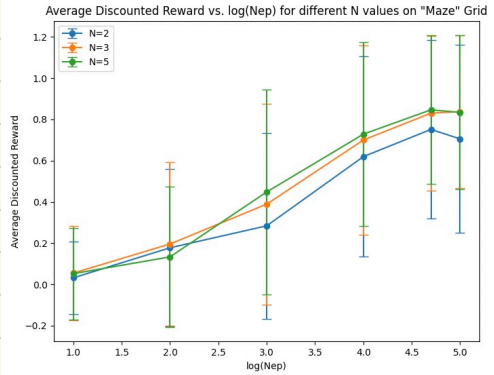



```
Receiver's Policy for message 0:
↓ → ▌ → →
↑ → ↑ → ↑
▌▌ ↓ ▌▌
↓ ← ↑ → ↑
← ← ▌ ↓ ←


Receiver's Policy for message 1:
↑ ↑ ▌ ↓ →
↑ ↑ ↑ ← ↑
▌▌ ↓ ▌▌
↓ ← ↑ ← ↑
↑ ← ▌ ← ↑


Receiver's Policy for message 2:
→ → ▌ → ↑
↓ → ↑ → ↑
▌▌ ↑ ▌▌
↓ → ↑ ← ↑
← ↑ ▌ ← ↑


Receiver's Policy for message 3:
→ ↑ ▌ → ↑
↓ ← ↑ → ↑
▌▌ ↑ ▌▌
↓ ← ↑ ← ↑
← ← ▌ ← ↑


Sender's Policy (preferred message for each prize location):
3 3 ▌ 1 1
3 3 0 1 1
▌▌ 0 ▌▌
2 2 0 0 0
2 2 ▌ 0 0
```

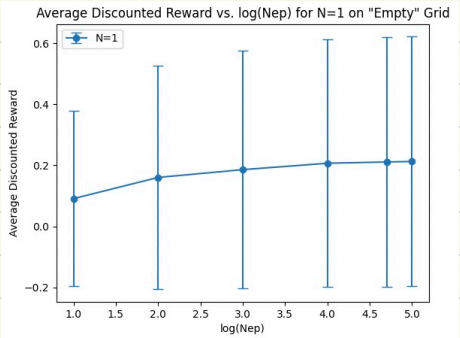
(c)



(d)


Average Discounted Reward vs. log(Nep) for different N values on "Maze" Grid
N=2
N=3
N=5
Average Discounted Reward
log(Nep)

(e)


Average Discounted Reward vs. log(Nep) for N=1 on "Empty" Grid
N=1
Average Discounted Reward
log(Nep)