

Validation Reward

1.5
1.4
1.3
1.2
1.1
1.0

0

500

1000

1500

2000

2500

3000

Training Epoch

