

Total Discounted Reward

50
40
30
20
10
0
-10

0.0

0.2

0.4

0.6

0.8

1.0

1e6

steps

— SM (Ours)
— td3 (Baseline)

