

Deterministic optimal policy for DYNA-Q+ B 0.05 after 6000 iterations.

