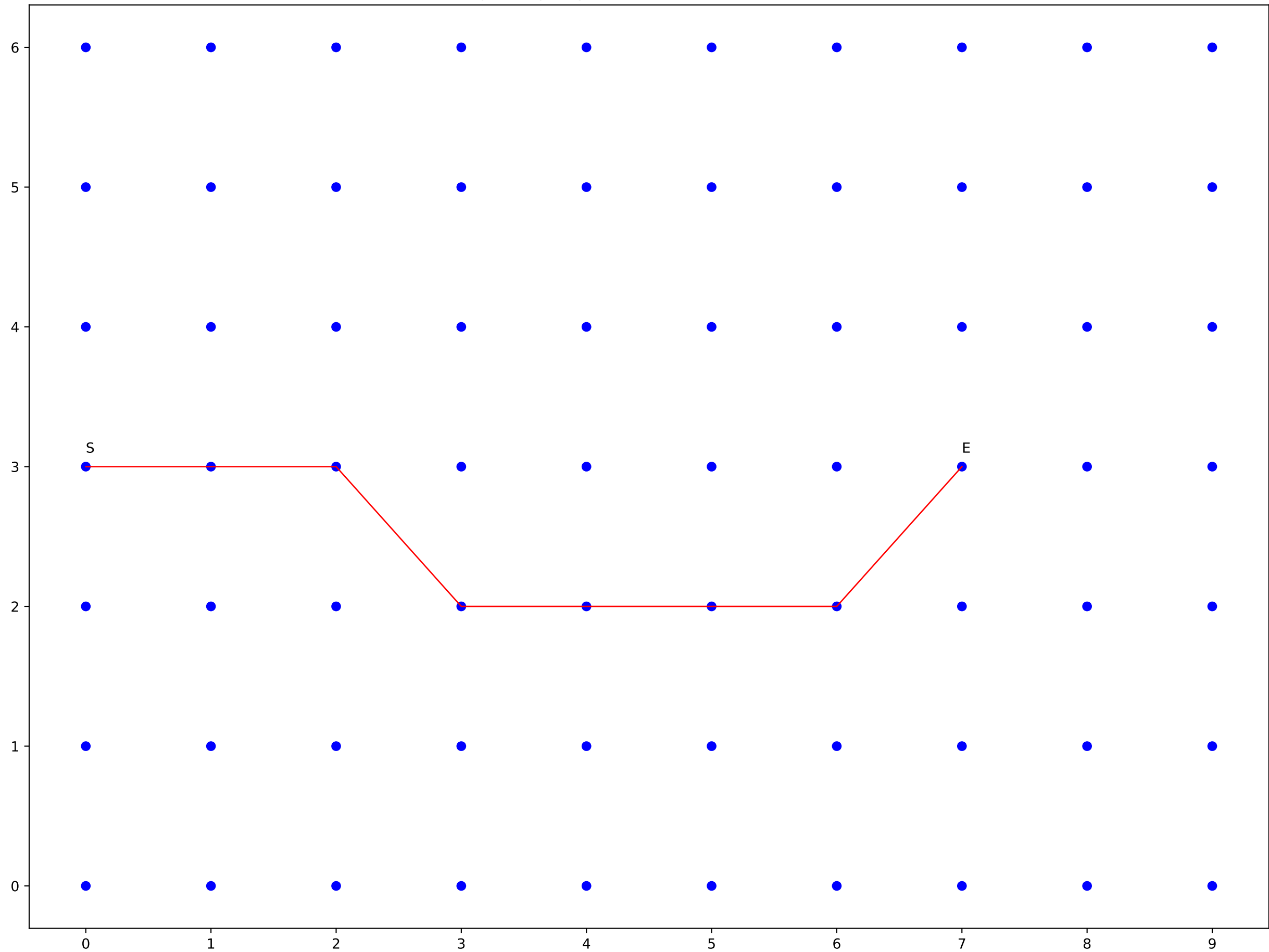


Deterministic optimal policy for 2 Q-LEARNING after 100000 iterations.



WIND

1

1

1

2

2

1