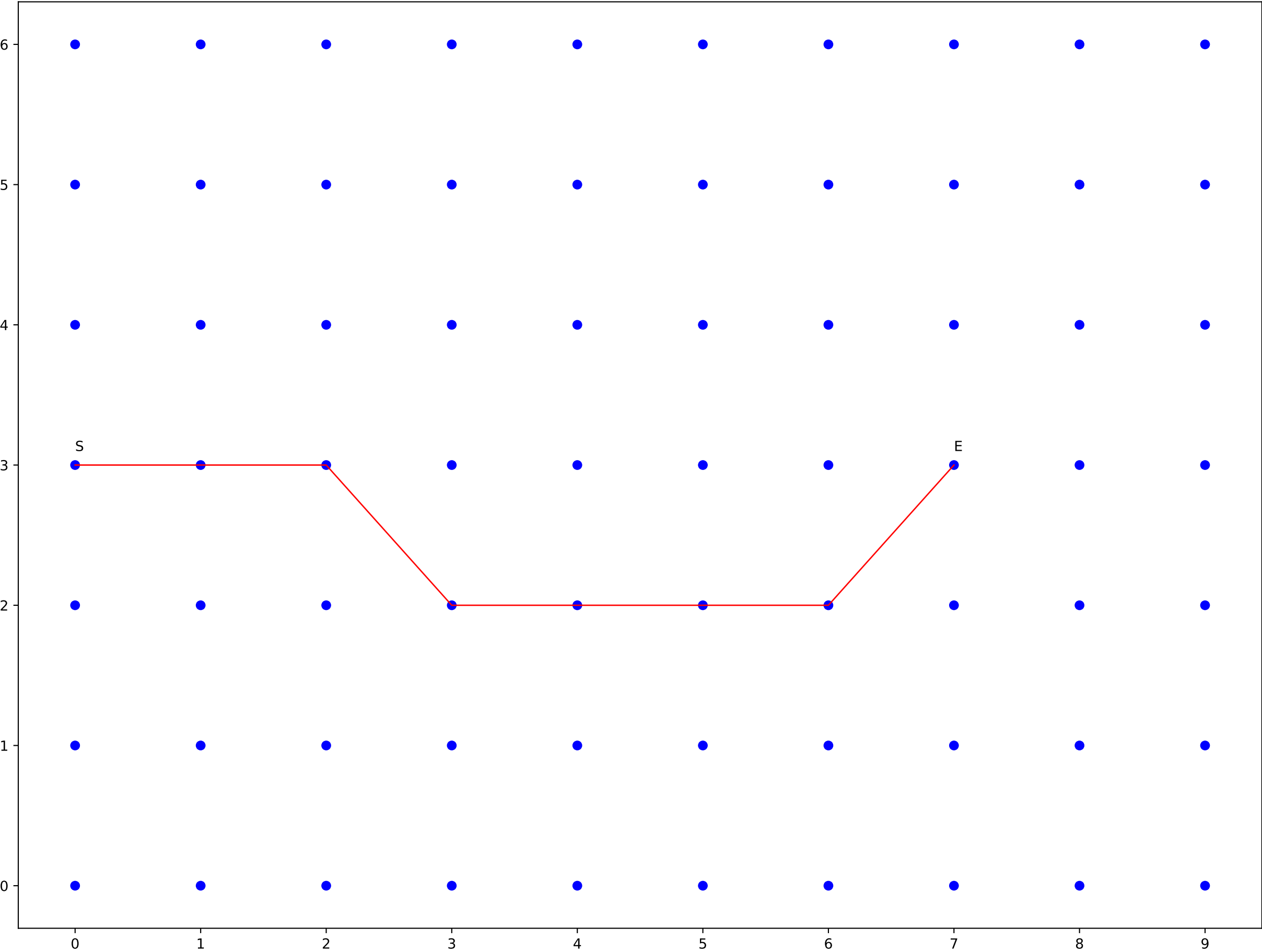


Deterministic optimal policy for Q-LEARNING after 100000 iterations.



WIND