

maze learning: optimal pseudorewards during some episodes, $\epsilon=0$

