maze learning: optimal pseudorewards during some episodes Q-learning use pseudorewards on first episode only on first 2 episodes only on first 5 episodes only on first 10 episodes only on first 20 episodes only steps episode