

**Brief comparison (3–5 sentences)**

RTDP quickly learns value estimates and produces consistent, relatively short paths to the goal — it converged to a stable policy within the training episodes and gives low variance execution (steps ≈ 9–16 typical). MCTS with the current settings is much more variable: some searches find short good paths but many produce long/wandering trajectories, increasing average steps and negative reward. The main reason is that MCTS here uses random rollouts and relatively modest search effort per decision; without a strong rollout policy or many rollouts, UCT exploration can waste time exploring suboptimal branches. In short: RTDP is efficient and stable on this small grid when you have a model; vanilla MCTS can work but needs more rollouts, a better rollout policy, or tuned c\_uct to be competitive.

**Summary statistics (from the printed run)**

* RTDP (20 eval episodes):
  + mean steps to goal ≈ **11.85**
  + mean total reward ≈ **-10.85**
* MCTS (20 searches / execution runs):
  + mean steps to goal ≈ **29.70**
  + mean total reward ≈ **-28.70**
  + steps standard deviation ≈ **20.57** (MCTS was much more variable)