Maze Shaping Report: Manhattan PBRS + Potential-Based Exploration (v1)

February 13, 2026

1 Experiment Setup

This run uses Manhattan PBRS and replaces the previous first-visit exploration bonus with a potential-based exploration term.

Reward used in training:

$$r_t = r_{\text{env}}(s_t, a_t, s_{t+1}) + \gamma\Phi_\kappa(s_{t+1}) - \Phi_\kappa(s_t) + \beta\left(\gamma\Psi(s_{t+1}) - \Psi(s_t)\right).$$

Definitions:

- Manhattan potential: $\Phi(s) = -(|r - r_g| + |c - c_g|)$, $\Phi_\kappa(s) = \kappa\Phi(s),\ \kappa \in \{0, 0.5, 1.0\}$
- Exploration potential: $\Psi(s) = -N(s)$, where $N(s)$ is in-episode visit count
- Exploration coefficient: $\beta = 0.05$
- Discount factor: $\gamma = 0.99$

Other settings: episodes = 500, runs = 12, $\alpha = 0.02$, $\epsilon = 0.10$, step reward = 0.0, goal reward = 1.0, max steps = 350.

2 Learning and Validation Curves

3 Final Summary (Episode 500)

Condition	Train mean steps	Validation success rate	Validation mean steps
`no_shaping`	299.42	0.1667	299.0
`phi_half`	350.00	0.0000	350.0
`phi_full`	350.00	0.0000	350.0

Table 1: Episode-500 metrics for potential-based exploration run.

Best validation success over training:

- `no_shaping`: 0.1667
- `phi_half`: 0.0000
- `phi_full`: 0.0000

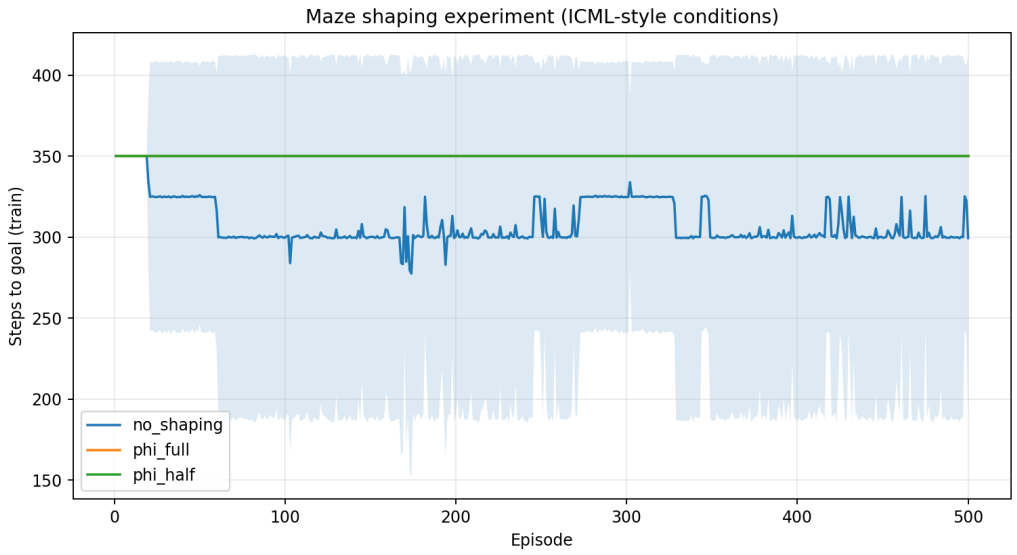

Figure 1: Training curve (mean steps to goal).

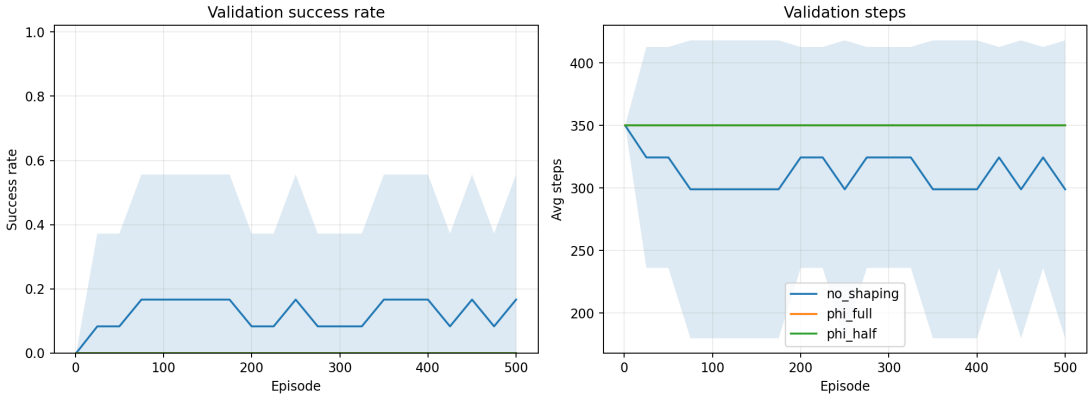

Figure 2: Validation success and validation average steps.

4 Interpretation

Under this configuration, adding potential-based exploration did not improve Manhattan PBRS conditions. The run is recorded as a negative result for this shaping design and coefficient scale.

5 Artifacts

- Output folder: `../../outputs/maze_shaping_icml_style_pbrs_manhattan_explore_potential_v1`
- CSVs: `learning_curve.csv`, `validation_progress.csv`
- Config dump: `run_summary.json`
- GIF rollouts: `gifs/policy_rollout_ep_*.gif`