

Maze Shaping Report: PBRs Manhattan (v2)

February 13, 2026

1 Experiment Setup

This report documents the run with the following settings:

- Potential type: Manhattan distance PBRs
- Discount factor: $\gamma = 0.99$
- Base step reward: 0.0
- Goal reward: +1.0 on terminal transition
- Conditions: `no_shaping`, `phi_half`, `phi_full`
- Episodes: 500, Runs: 12
- Validation every 25 episodes, 30 greedy rollouts

Reward used for shaping:

$$R'_\kappa(s, a, s') = R(s, a, s') + \gamma \Phi_\kappa(s') - \Phi_\kappa(s), \quad \kappa \in \{0, 0.5, 1.0\}.$$

Here, $\Phi(s)$ is the potential value of state s . In this report we use Manhattan potential:

$$\Phi(s) = -(|r - r_g| + |c - c_g|),$$

where (r, c) is the agent state and (r_g, c_g) is the goal coordinate. Each condition scales this base potential as

$$\Phi_\kappa(s) = \kappa \Phi(s),$$

so $\kappa = 0$ is no shaping, $\kappa = 0.5$ is half-strength shaping, and $\kappa = 1.0$ is full-strength shaping.

2 Results

3 Final Episode Summary (Episode 500)

Condition	Validation success rate	Validation mean steps
<code>no_shaping</code>	0.0	350.0
<code>phi_half</code>	0.0	350.0
<code>phi_full</code>	0.0	350.0

Table 1: All conditions remained unsuccessful in this run budget.

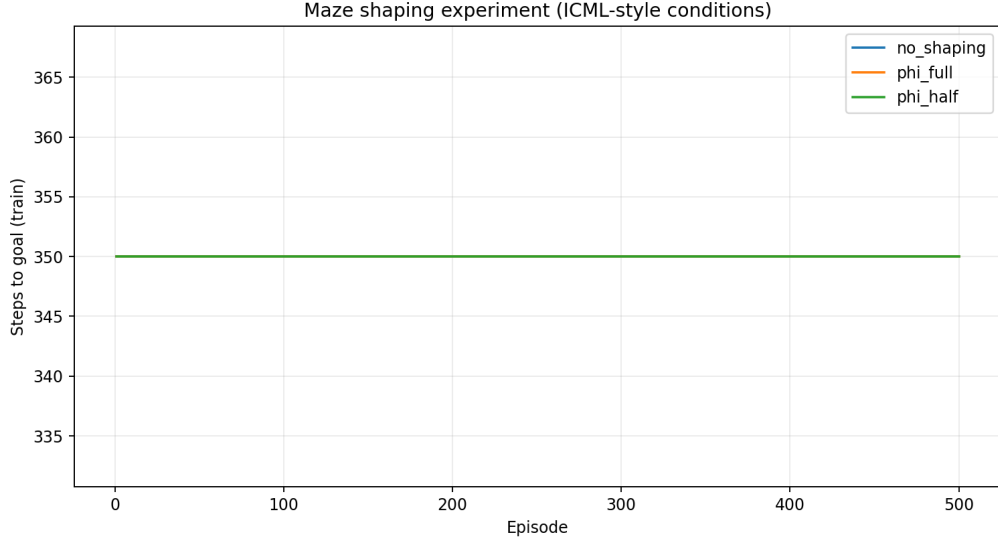


Figure 1: Training curve (steps to goal).

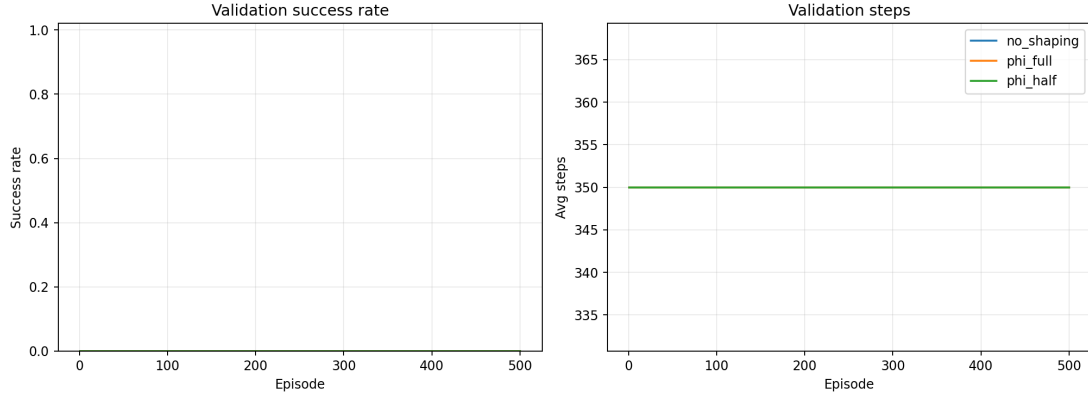


Figure 2: Validation progress (success rate and average steps).

4 Interpretation

Under this specific configuration, none of the three conditions reached successful greedy validation behavior within 500 episodes. The run is still useful as a record of a negative result for this reward design.

5 Artifacts

- Output folder: `../../outputs/maze_shaping_icml_style_pbrs_manhattan.v2`
- Learning CSV: `learning_curve.csv`
- Validation CSV: `validation_progress.csv`
- Run summary: `run_summary.json`
- GIFs: `gifs/policy_rollout_ep_*.gif`