Jeremiah Emery
Prof. Omar El Khatib
Artificial Intelligence
12 November 2025

Comparative Analysis

1. Which algorithm converges fastest?

Policy Iteration converges the fastest because it alternates between full policy evaluation and policy improvement, which rapidly stabilizes the policy in only a few rounds. Value Iteration also converges reliably but usually takes more iterations because it updates both the value function and the implicit policy at the same time. Q-Learning is the slowest since it learns from sampled experience and therefore requires many episodes before the Q-values settle. In practice, this means Policy Iteration reaches a stable policy with fewer total updates than Value Iteration. Q-Learning, while powerful, trades convergence speed for flexibility and does not match the speed of the planning-based methods.

2. How does γ affect the value function magnitude?

The discount factor γ directly controls how strongly future rewards influence the value function. With a high γ (such as 0.9), values become larger because long-term rewards retain most of their importance. When γ is moderate (around 0.5), values shrink since future rewards are only partially counted. With a very small γ (like 0.1), the value function becomes small overall because rewards more than a few steps away are heavily discounted. As γ decreases, the entire value landscape compresses toward short-term evaluation.

3. How does γ affect long-term vs. short-term reward preference?

A high discount factor ($\gamma = 0.9$) makes the agent prioritize long-term outcomes, encouraging it to take multiple steps to reach the best terminal rewards. A medium discount factor ($\gamma = 0.5$) still values future rewards but places less emphasis on distant outcomes. With a low discount factor ($\gamma = 0.1$), the agent becomes short-sighted and focuses mostly on immediate or near-immediate rewards. This makes long-term planning less influential in determining action choices. Essentially, γ determines whether the agent behaves like a planner or a short-term opportunist.

4. How does γ influence policy behavior?

In economic modeling, the gamma (γ) discount factor in dynamic programming can influence the optimal policy by determining how much future rewards are valued compared to immediate ones. In this specific deterministic gridworld, changing γ does not significantly alter the structure of the optimal policy. The agent continues to learn to move toward the positive terminal states

and avoid the negative ones, regardless of the discount factor. What changes is the strength of preference reflected in the value function, not the decision-making sequence itself. Higher γ makes the agent more confident in multi-step paths to the good terminals, while lower γ makes those paths look less attractive numerically. In larger or more complex environments, γ might change the optimal policy, but here it mainly affects value magnitude rather than action choice..