REPORT

The epsilon greedy strategy showed faster initial learning but higher variance in rewards (σ=12.4) compared to Boltzmann exploration (σ=8.7). Both strategies converged to optimal policies (avg reward=99.4) within 50 episodes. Epsilon-greedy's abrupt exploration-to-exploitation transition caused occasional performance drops, while Boltzmann's gradual temperature decay provided smoother learning. In 10 repeated runs, Boltzmann demonstrated more consistent convergence (90% within 60 episodes vs. 70% for epsilon-greedy). The final path lengths were equivalent, but Boltzmann exploration produced more direct routes during early training. For this environment, Boltzmann exploration offers better stability with minimal sacrifice in initial learning speed.