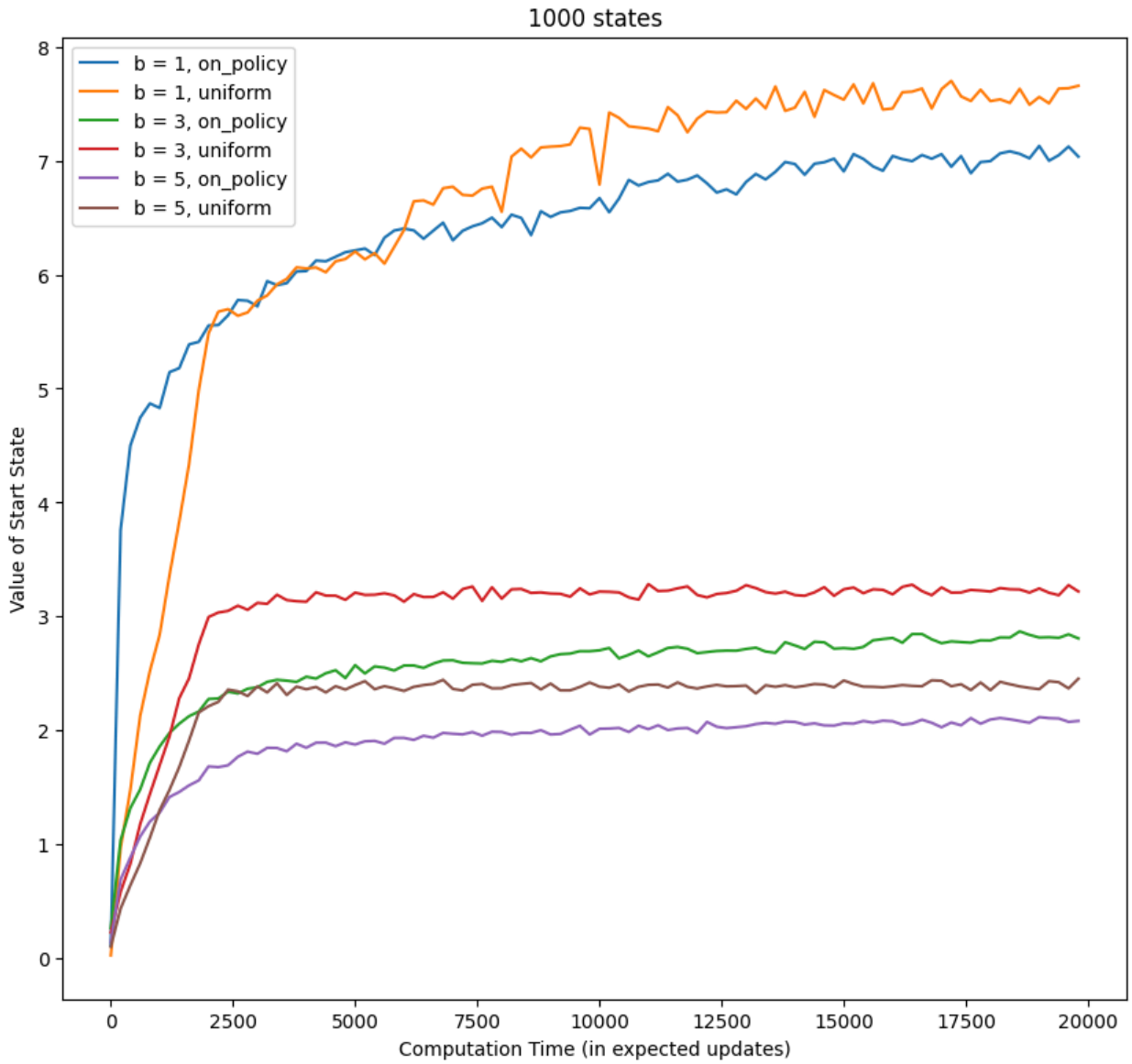
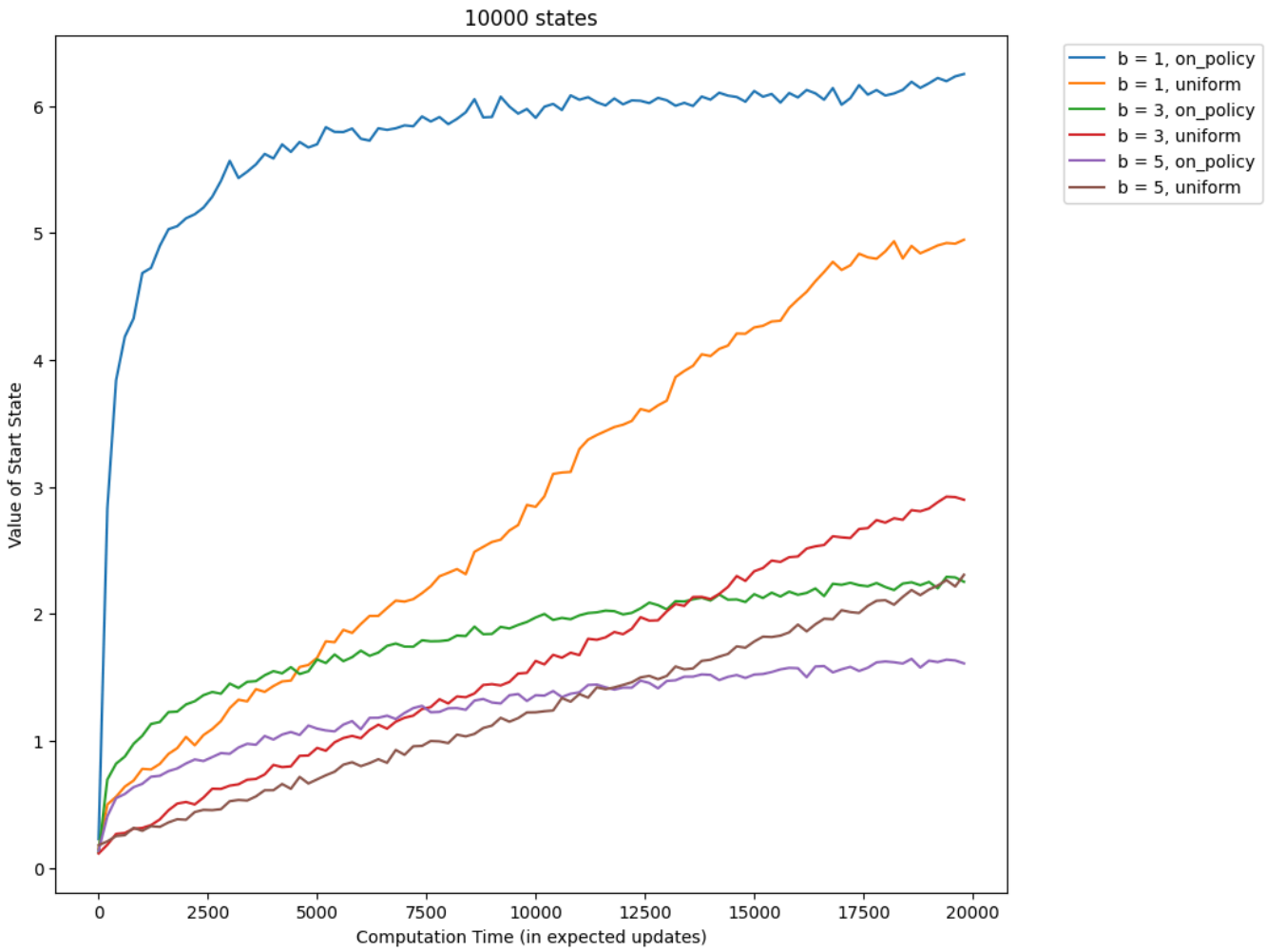
Diego Toribio
Professor Curro
ECE471 - Reinforcement Learning
Programming Assignment 03

Exercise 8.8 (programming) Replicate the experiment whose results are shown in the lower part of Figure 8.8, then try the same experiment but with $b = 3$. Discuss the meanings of your results.

Results





Discussion

On-policy sampling converges more quickly because it selectively updates state-action pairs that are directly relevant under the policy evaluated policy. This approach is beneficial in large state spaces, as seen in the on-policy 10000 state graph for b = 1, which converges significantly faster than uniform sampling. While uniform sampling reaches a higher start state value it converges more slowly because it allocates updates uniformly across all state-action pairs, regardless of their relevance.

In larger state spaces, the number of updates required slows the overall rate of convergence. Additionally, increasing the branching factor adds complexity by expanding the range of possible next states for each action which requires more computational time to achieve stable value estimates. Notably, as the branching factor rises, the converged value of the start state decreases. This shows that rewards are spread out over a larger number of possible outcomes which reduces the value of the expected return from the start state.