

Figure 1: Additional results on Mujoco-P, where we allow GRU and LRU to use a larger truncation length than needed to solve these tasks. We also show results for GPT-2.

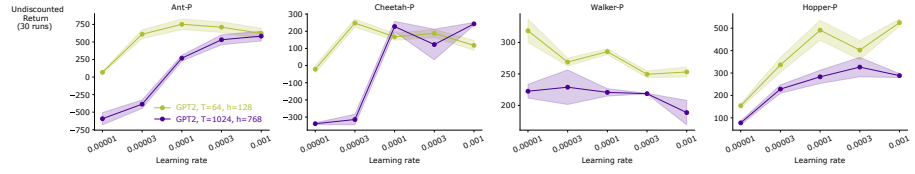


Figure 2: Learning rate sweep for GPT-2.