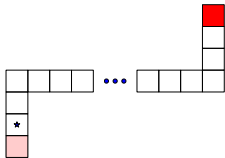
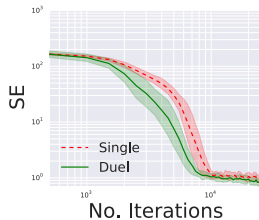


CORRIDOR ENVIRONMENT



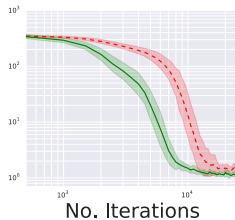
(a)

5 ACTIONS



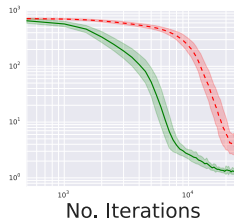
(b)

10 ACTIONS



(c)

20 ACTIONS



(d)

Figure 3. (a) The corridor environment. The star marks the starting state. The redness of a state signifies the reward the agent receives upon arrival. The game terminates upon reaching either reward state. The agent's actions are going up, down, left, right and no action. Plots (b), (c) and (d) shows squared error for policy evaluation with 5, 10, and 20 actions on a log-log scale. The dueling network (Duel) consistently outperforms a conventional single-stream network (Single), with the performance gap increasing with the number of actions.