



**Figure 2.5:** Average performance of the gradient bandit algorithm with and without a reward baseline on the 10-armed testbed when the  $q_*(a)$  are chosen to be near +4 rather than near zero.