FDR for a1false-a2false-b1true-c1true, reward function 1 0.9 0.8 0.7 0.6 bandit-eps bandit-opt 0.5 qlearning random bayes-high 0.4 bayes-low mlfq 0.3 tableau ann

0

50

100

150

200

experiments run

250

300

350

400