FDR for a1false-a2false-b1true-c1false, reward function 1