The estimate values from dynamic programming and MCTS with p=0.25, p=0.55 are all given separately. We can notice that when p = 0.25 they both perform good, however, when p comes to p=0.55, they differ a lot. The main reason is that when p=0.55, the optimal policy is easy: always bid one and since the chance of head is larger than the tail. It's highly likely to win even for a state with few capitals.

In this case, the **initial policy** for MCTS will be extremely important. If our initial policy is to bid 1 forever, we will get a close (almost the same) results as the DP outputs. However, if the initialization of the policy is min(i,100-i), MCTS is can barely correct the policy.