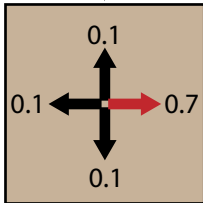
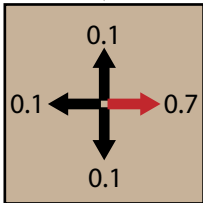
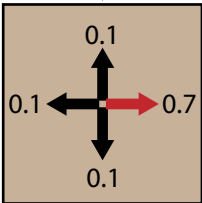


The objective is find
the optimal policy (π)
that maxime the reward



How quantify
the performance
of a policy (π)?