

# The Multi-Armed Bandit Problem

- We have  $d$  arms. For example, arms are ads that we display to users each time they connect to a web page.
- Each time a user connects to this web page, that makes a round.
- At each round  $n$ , we choose one ad to display to the user.
- At each round  $n$ , ad  $i$  gives reward  $r_i(n) \in \{0, 1\}$ :  $r_i(n) = 1$  if the user clicked on the ad  $i$ , 0 if the user didn't.
- Our goal is to maximize the total reward we get over many rounds.