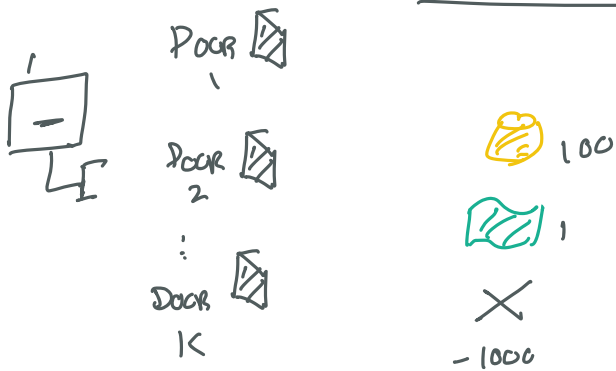


EXPLORE - EXPLOIT



STRATEGIES

① • FIRST N ROUNDS TRY ALL DOORS, UNIFORMLY

• SWITCHING TO PICKING DOOR WITH HIGHEST AVG REWARD

"EXPLORE - THEN - EXPLOIT"

③ THOMPSON SAMPLING

MAINTAIN A DISTRIBUTION

$P(\text{door})$ of how good door is

Sample door from this distrib

Update posterior

②

ϵ -prob EXPLORE "PICK A DOOR R. ANSWER"

$1-\epsilon$ prob EXPLOIT "GREEDY"

(ϵ -GREEDY)

④ UCB : Choose door with highest - upper confidence bound,