Part 6 Reinforcement Learning

Table of Contents

[The Multi-Armed Bandit Problem 1](#__RefHeading___Toc123_3828085289)

[Lottery machines 1](#__RefHeading___Toc125_3828085289)

[Upper Confident Bound (UCB) 3](#__RefHeading___Toc127_3828085289)

[The Algorithm 3](#__RefHeading___Toc129_3828085289)

[Thompson Sampling Intuition 6](#__RefHeading___Toc137_3828085289)

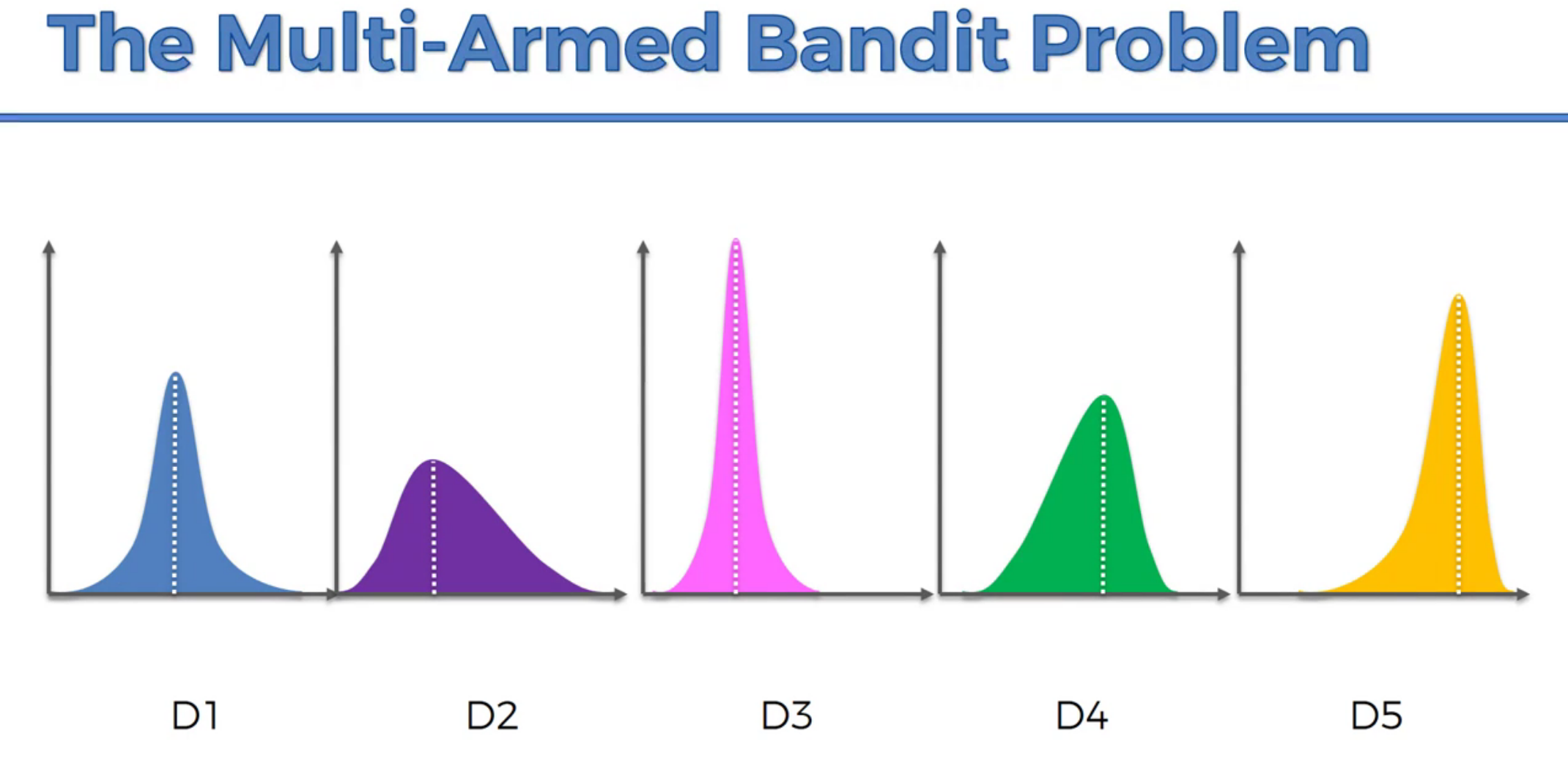
[UCB VS Thompson sampling algorithm 8](#__RefHeading___Toc108_1428834321)

# The Multi-Armed Bandit Problem

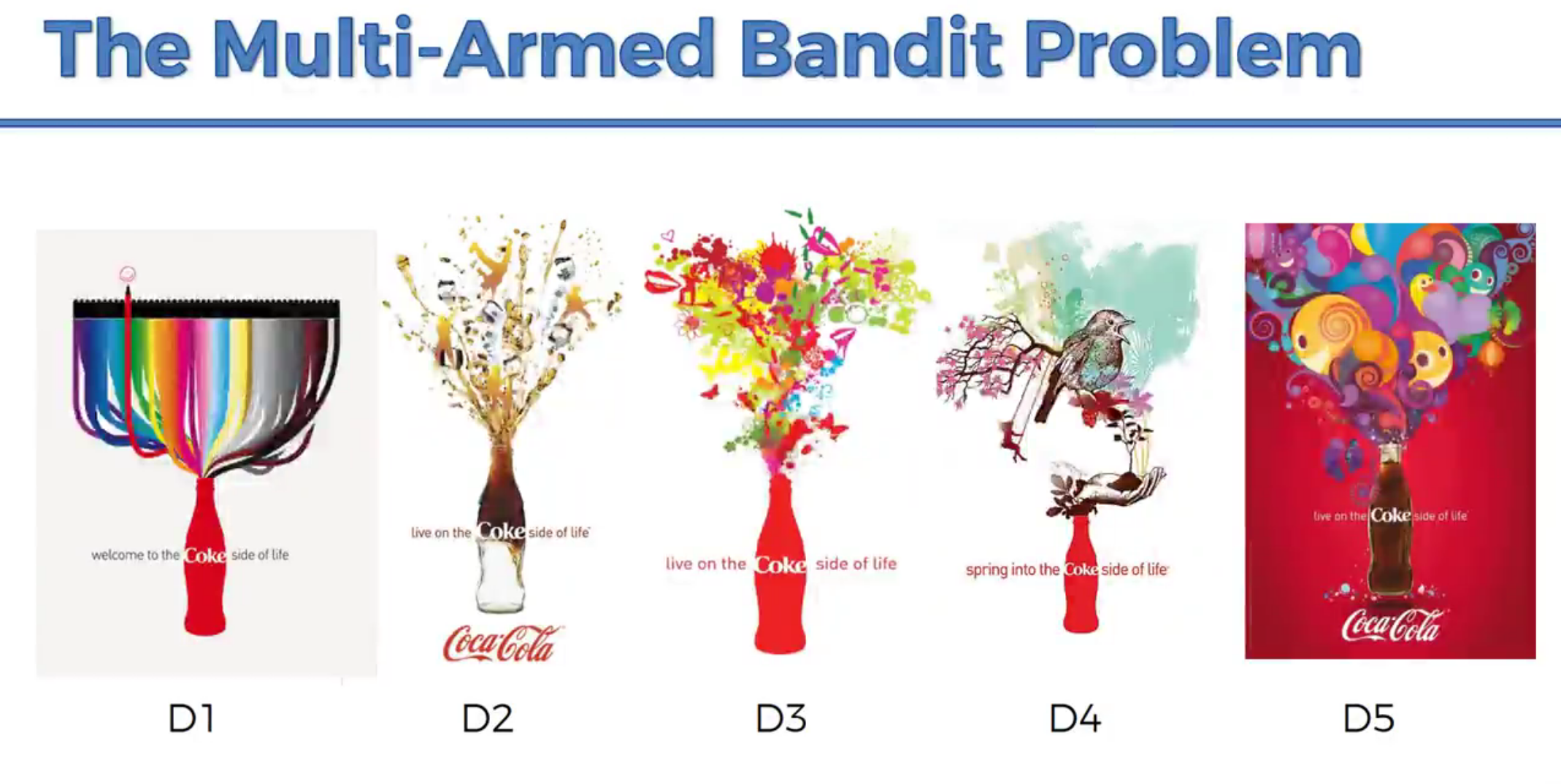
## Lottery machines



The machine on the right is the optimal machine fit with that problem



Sample of problems: This is Ad campaign for Coca-cola, which is the best Ad Campaign?



# Upper Confident Bound (UCB)

## The Algorithm

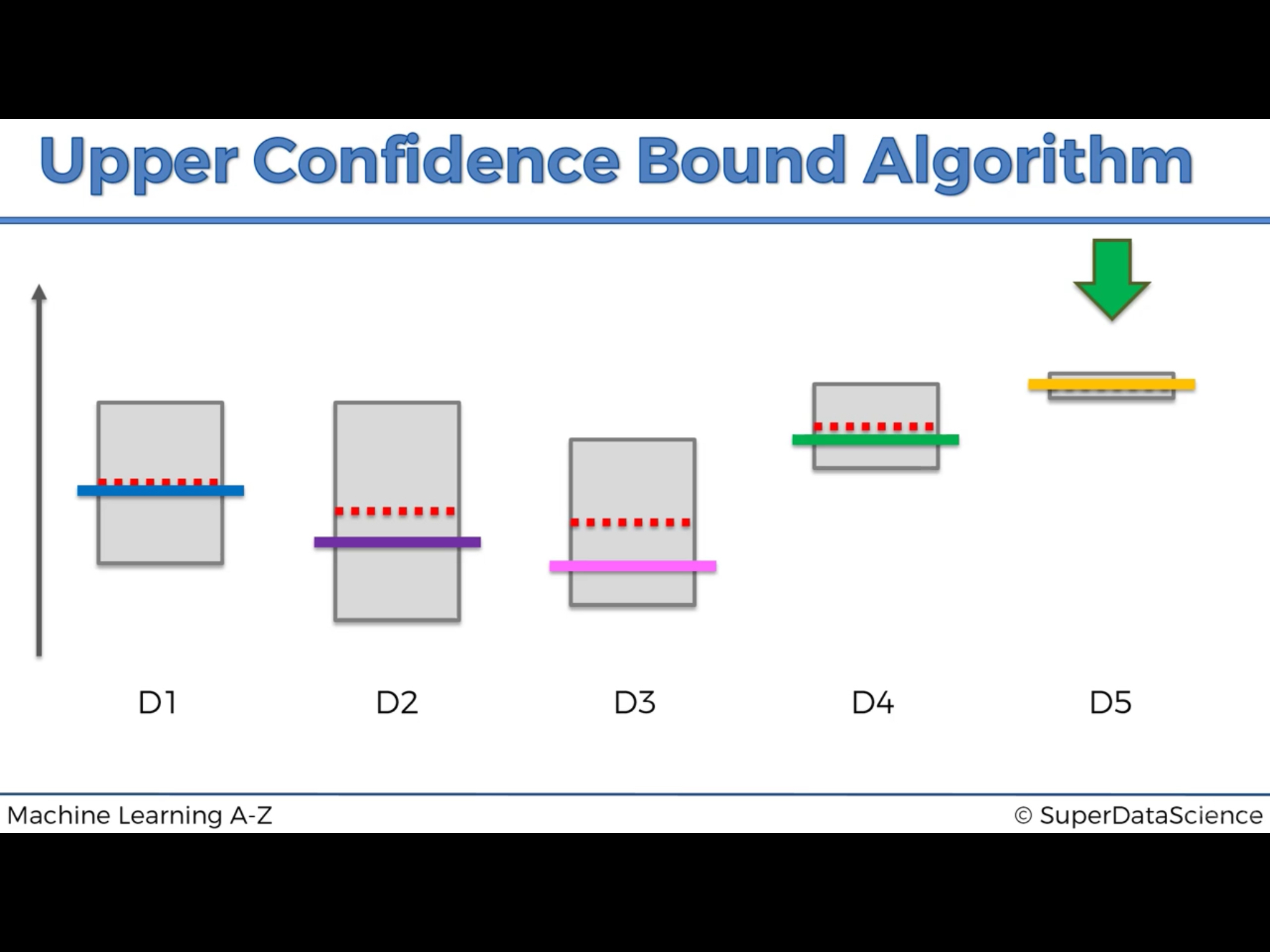
It’s more Robot and AI algorithm, that we can start with no data.

## 

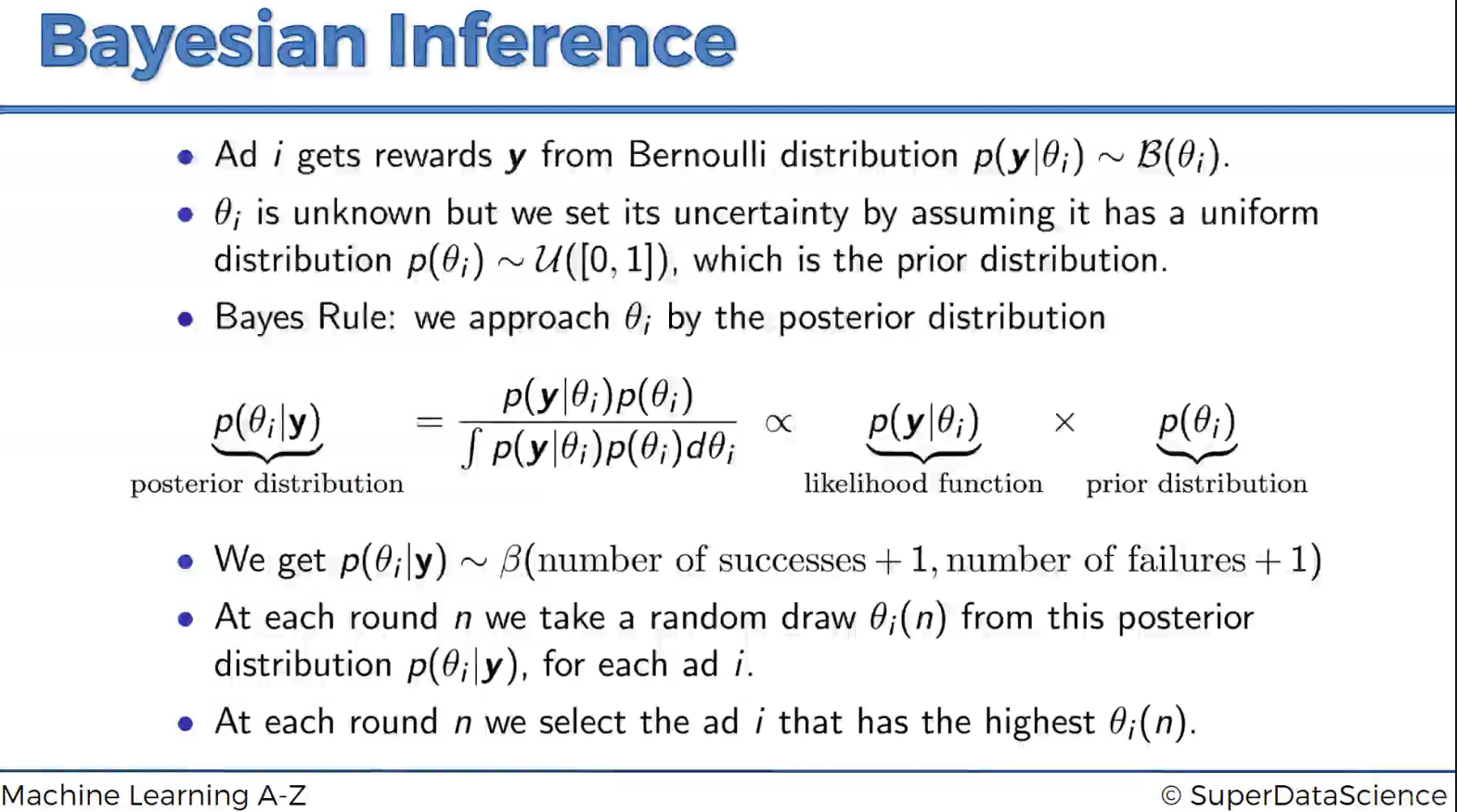
* 1. Ad Expected return levels

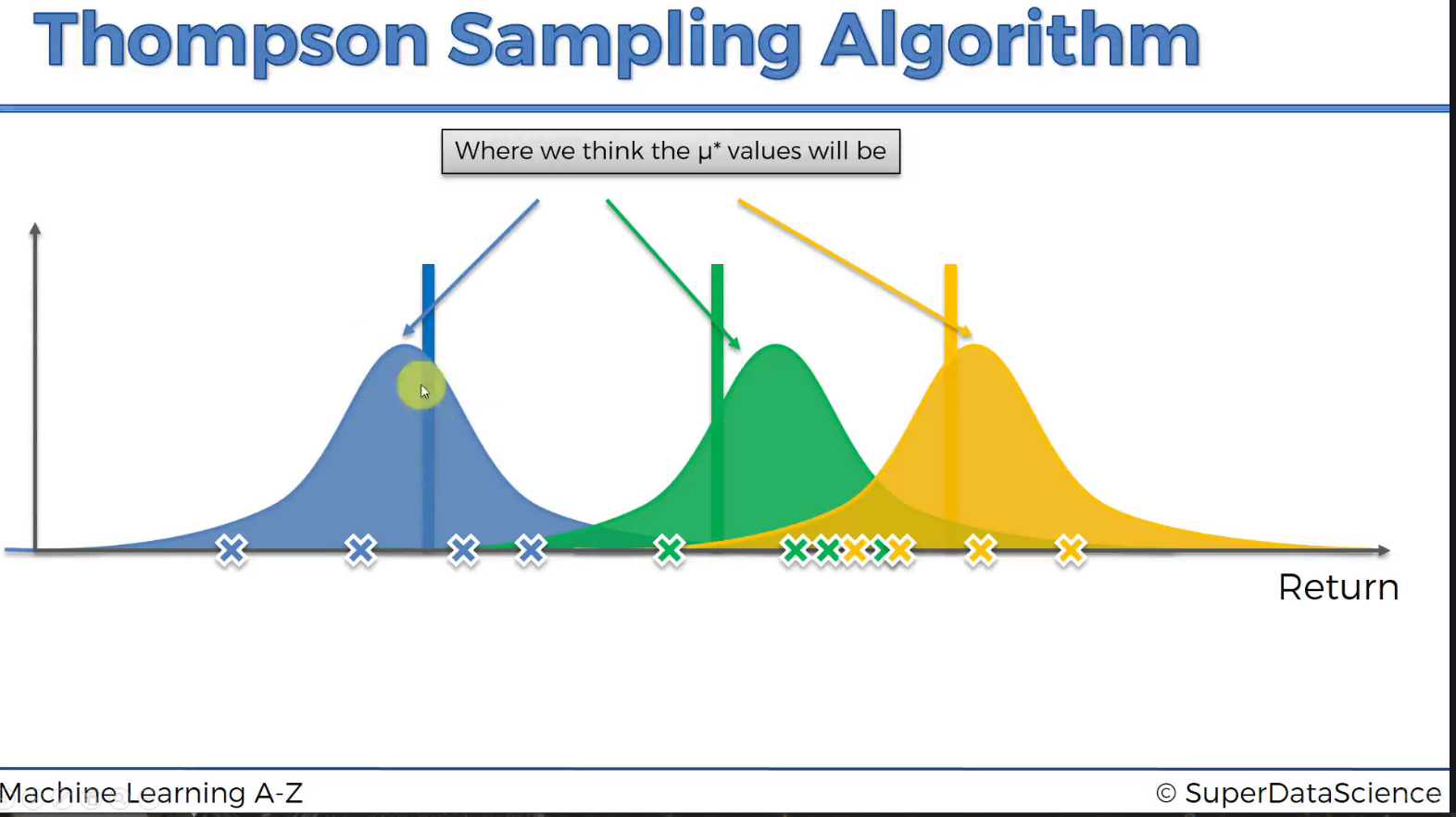
# 

* 1. Updating the confident level by time with the observations seen in the new data received

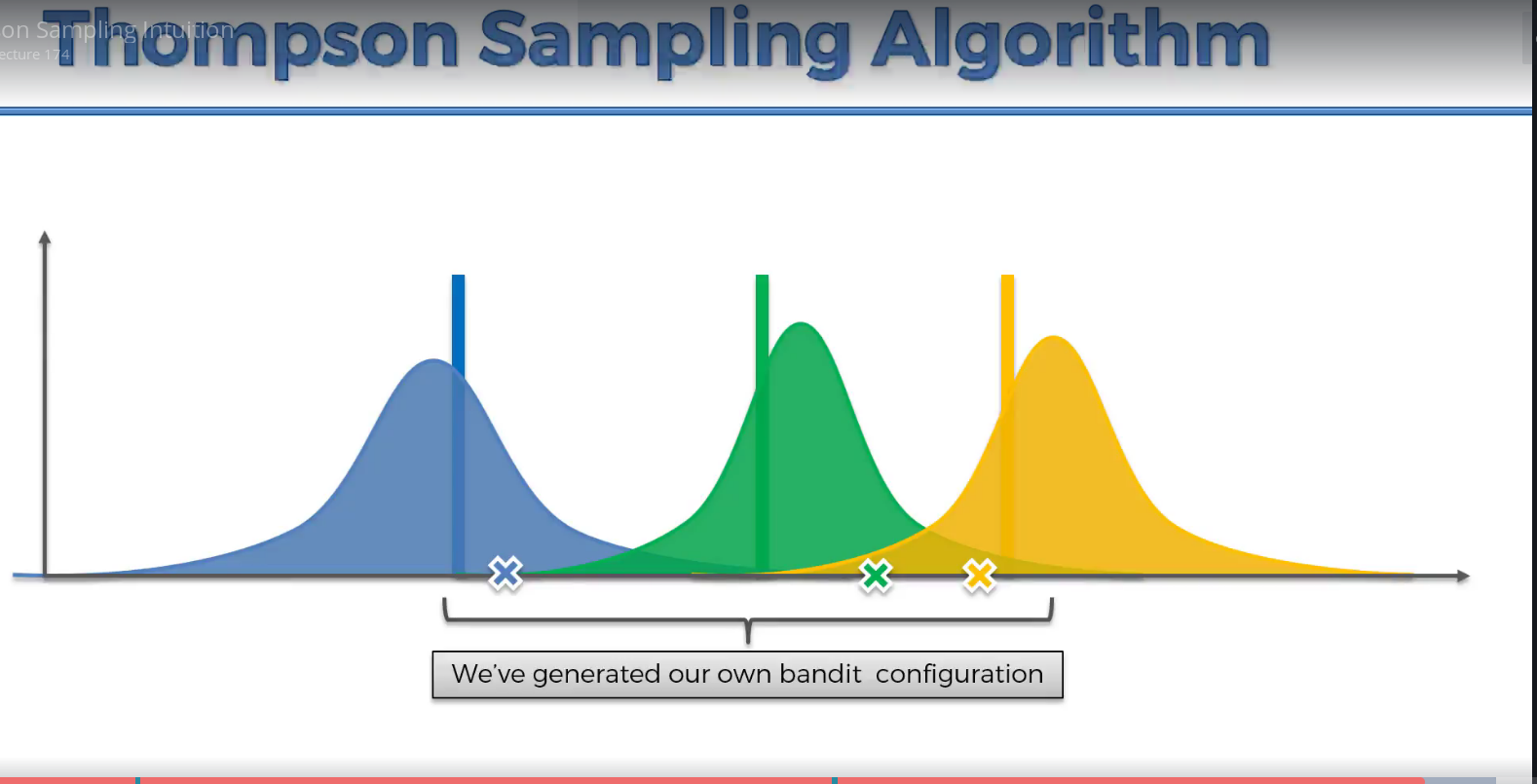
1. 

# Thompson Sampling Intuition

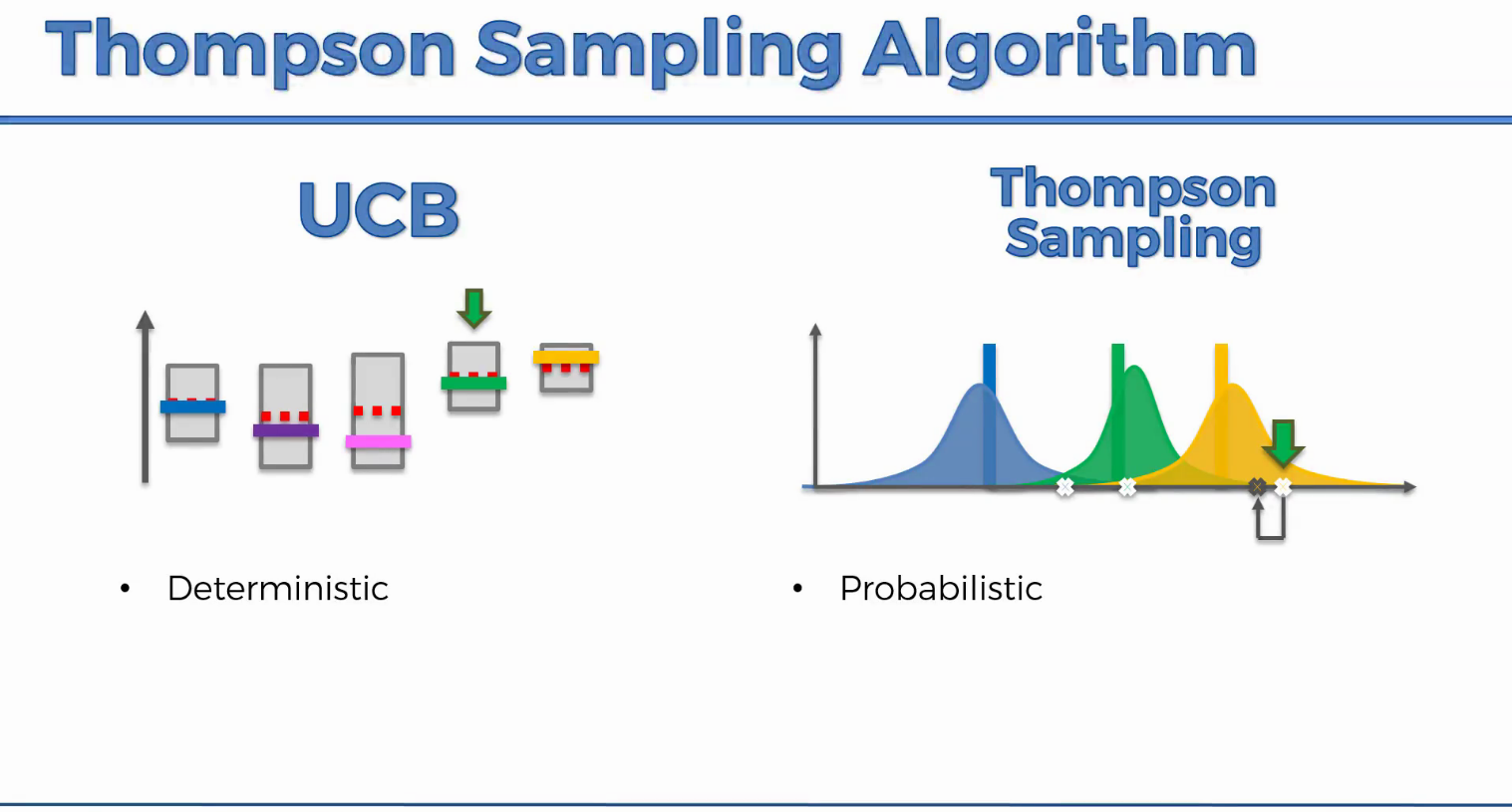




Adjust the perception after each round.



# UCB VS Thompson sampling algorithm



Thompson sampling algorithm is working better in real world.

