

Upper Confidence Bound 101

Upper Confidence Bound

Introduction

- Reinforcement learning algorithm
- Problem: find a solution to a problem with incomplete information, uncertain rewards
- Takes „learnings“ into account to define future actions
- Solves the exploration/exploitation dilemma elegantly

Upper Confidence Bound

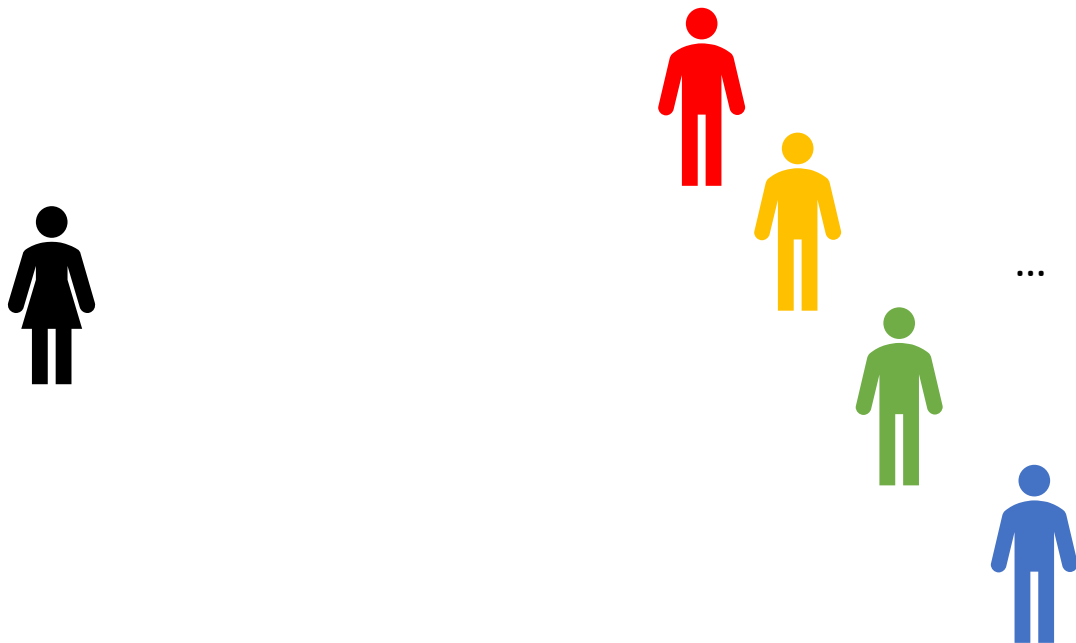
Exploration / Exploitation Dilemma

- Problem
 - incomplete information on a process
 - No simple solution
- Exploitation
 - Choose an action that you know
 - Getting a reward close to what I expect
- Exploration
 - Choose an action with an unsure outcome
 - Possibly learn something
- Best long-term solution might have short-term costs!

Upper Confidence Bound

Exploration / Exploitation Dilemma

- Exploration/exploitation dilemma found in many aspects of life
 - Example: Dating



Upper Confidence Bound

Multi-Armed-Bandit Problem

- Multi-Armed Bandit Problem



Bandit A



Bandit B



Bandit C



Bandit D

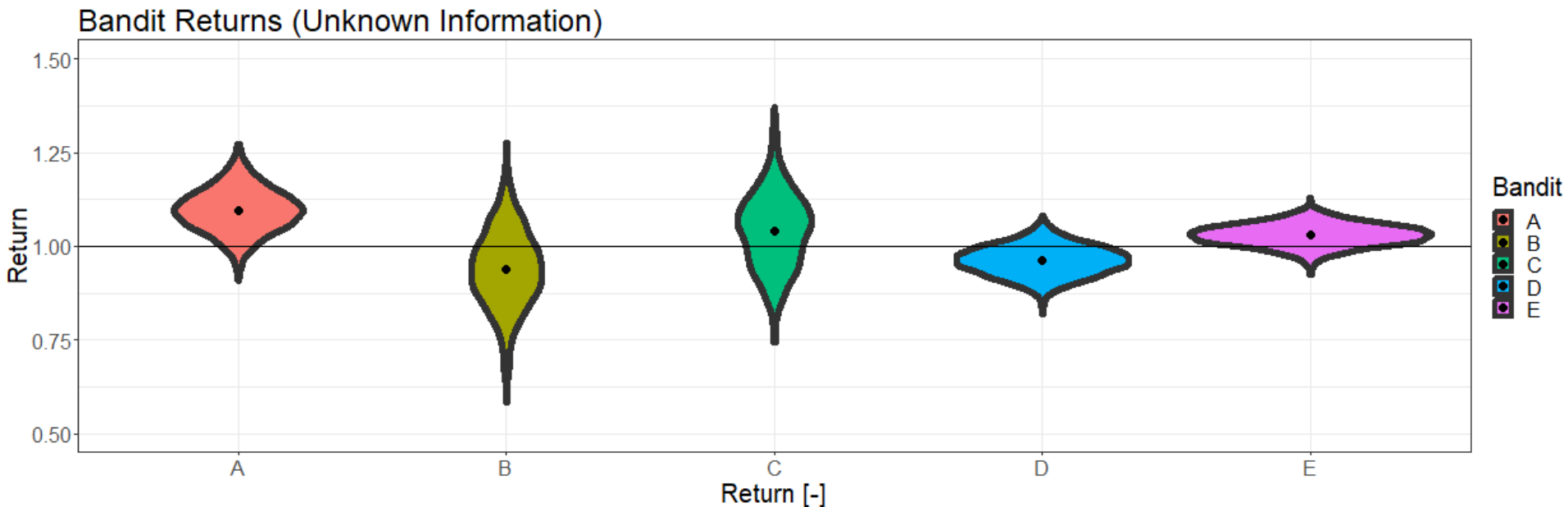


Bandit E

Upper Confidence Bound

Multi-Armed-Bandit Problem

- Multi-Armed Bandit Problem



Upper Confidence Bound

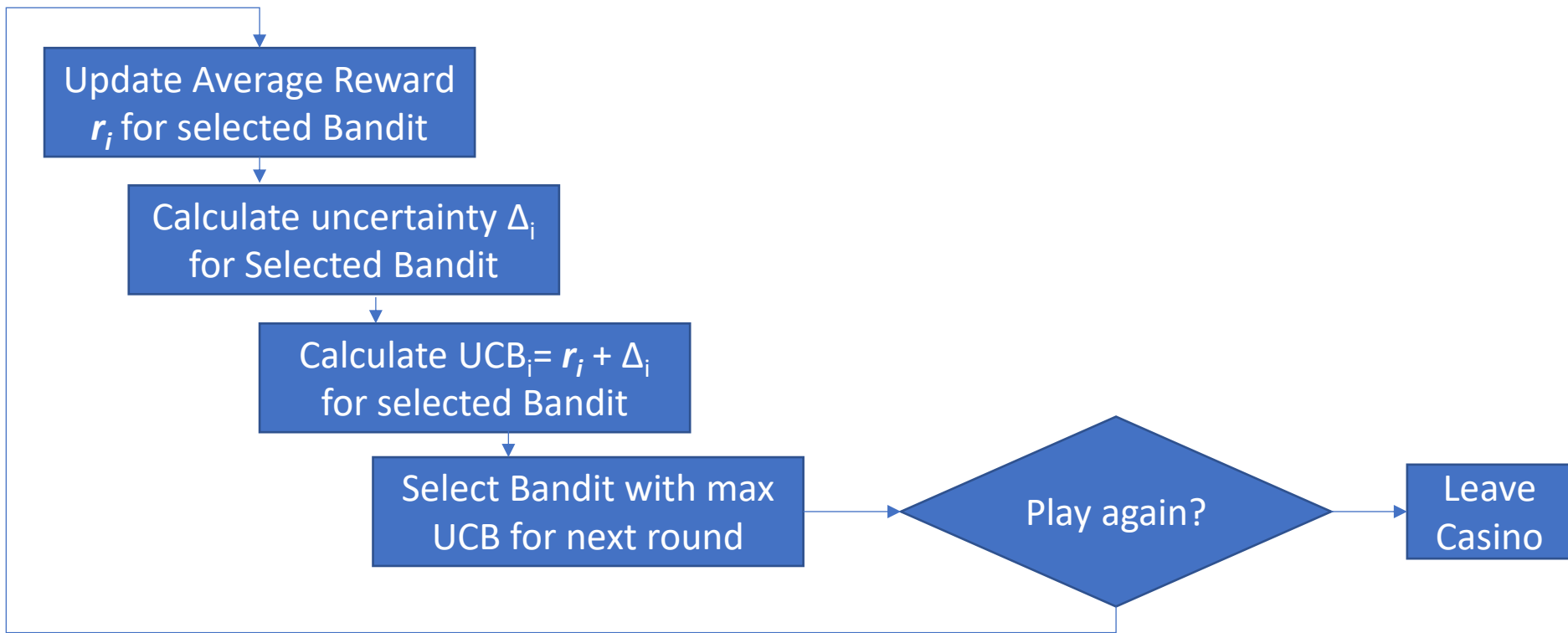
Multi-Armed-Bandit Problem

- Possible strategies
 - No exploration
 - Random exploration
 - Smart exploration

Upper Confidence Bound

Workflow

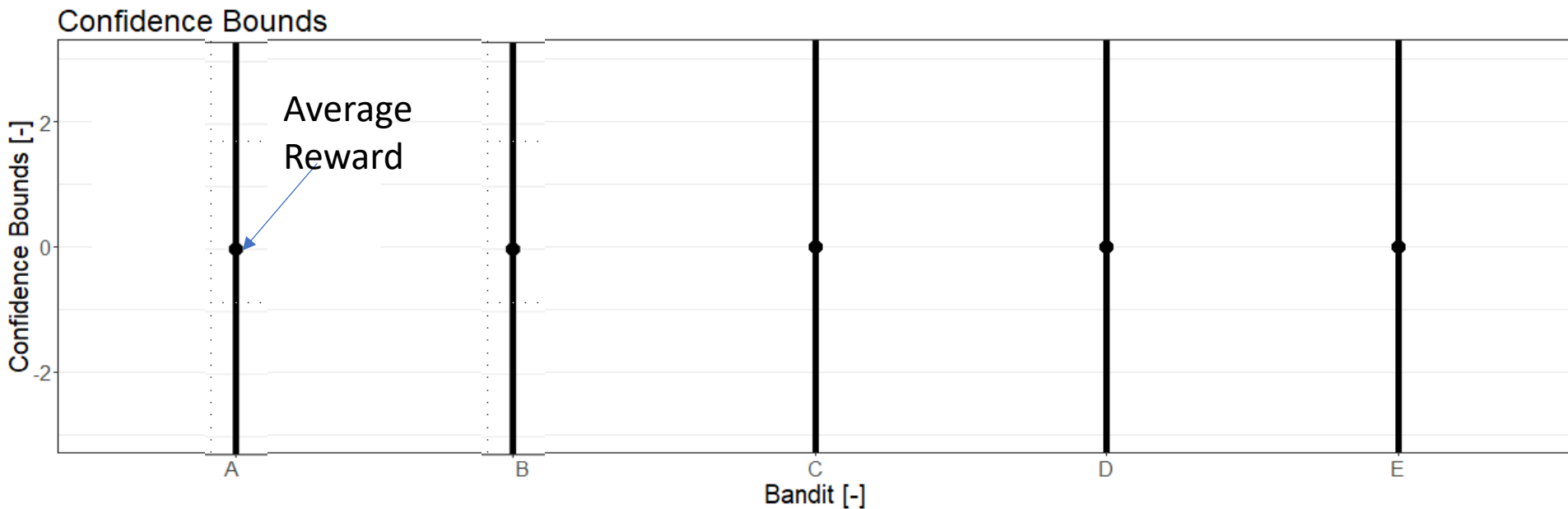
Workflow for Upper Confidence Bound



Upper Confidence Bound

Multi-Armed-Bandit Problem

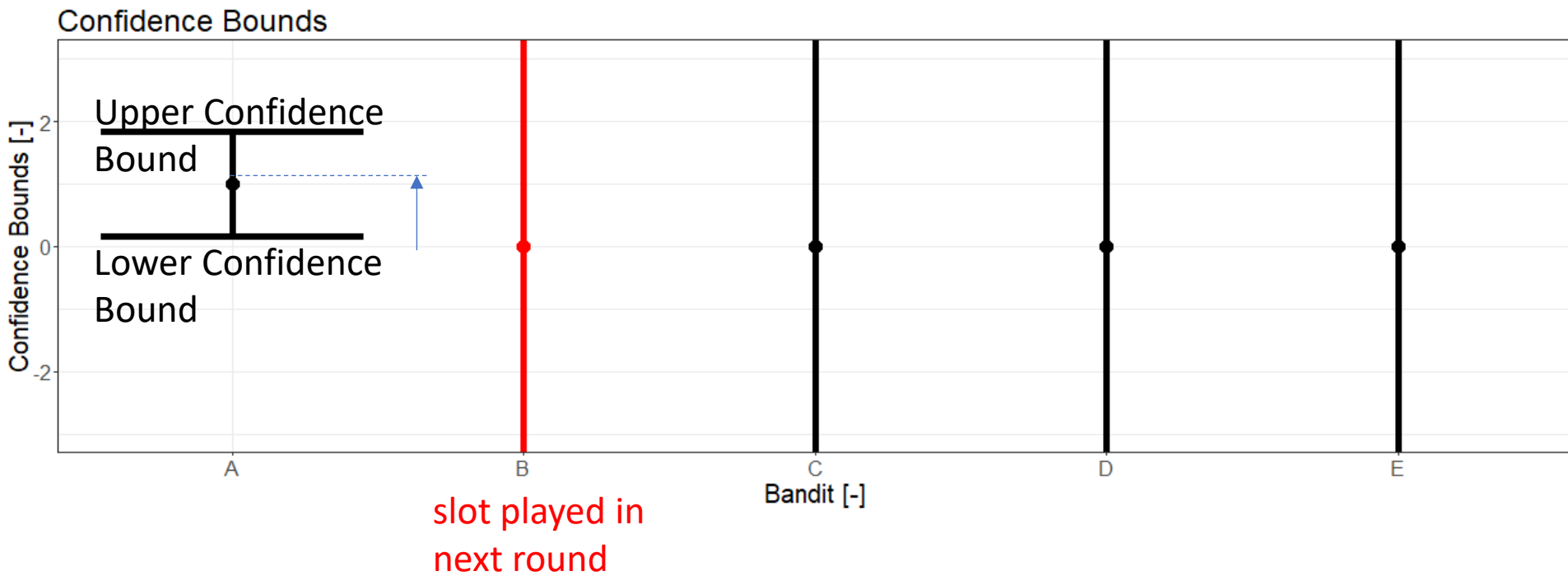
- Round: 0



Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 1

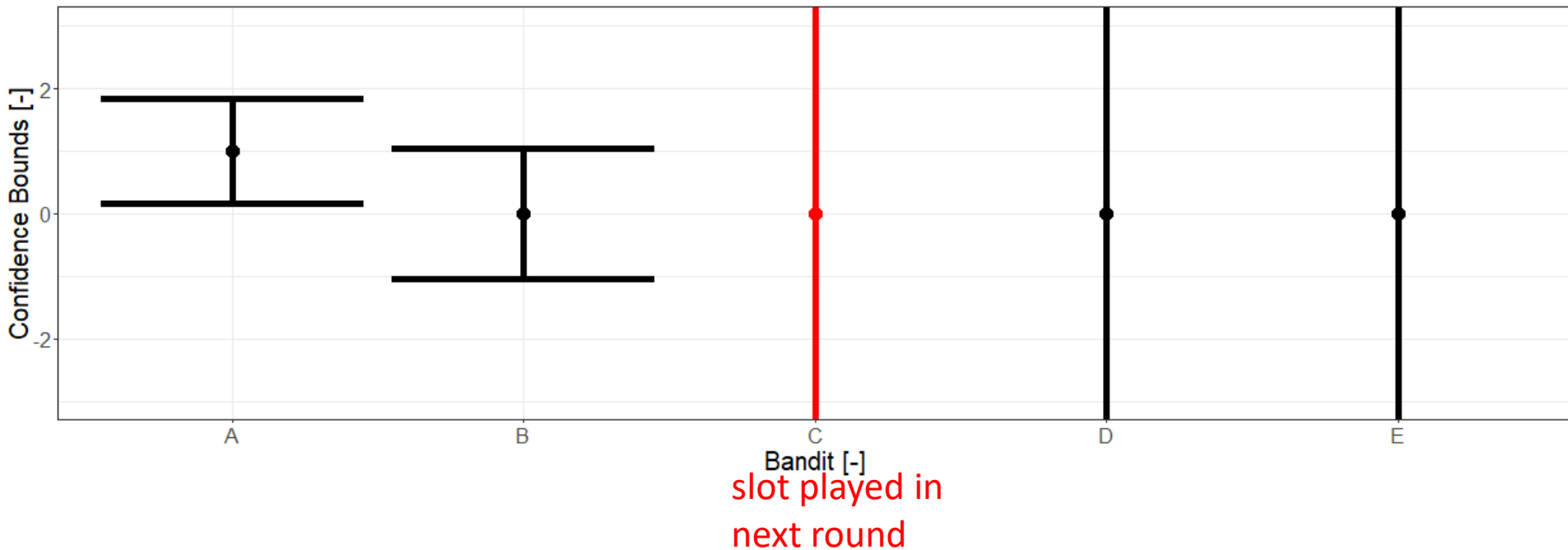


Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 2

Confidence Bounds

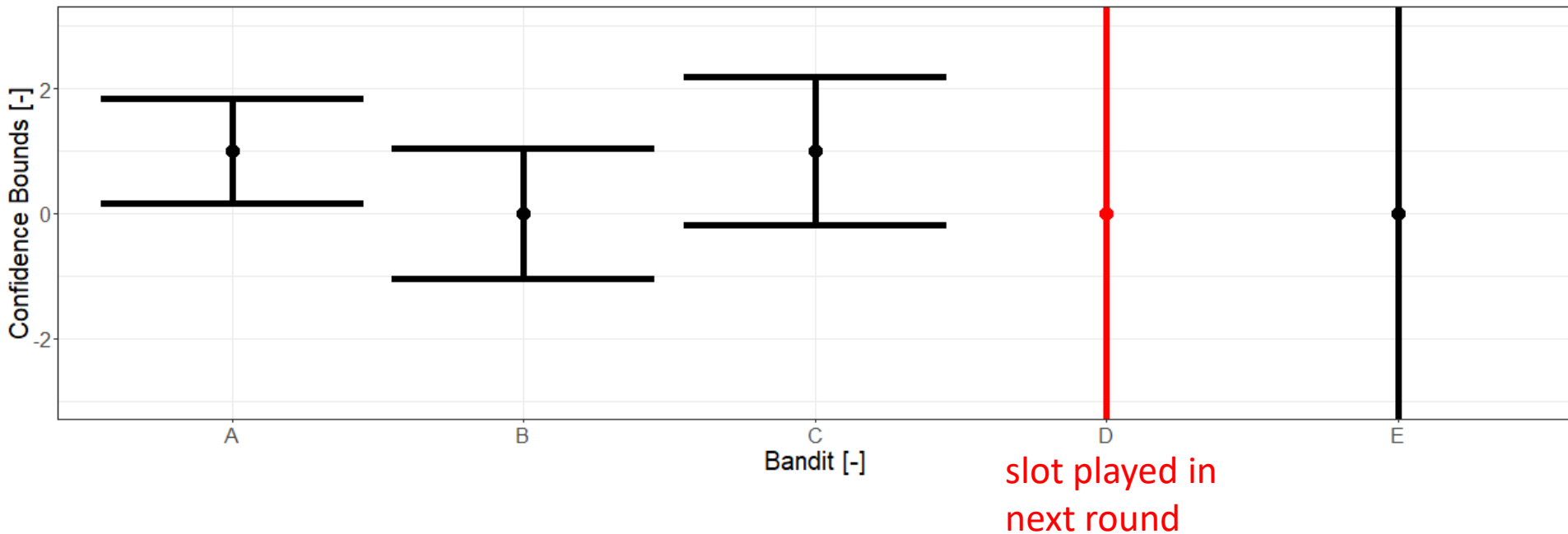


Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 3

Confidence Bounds

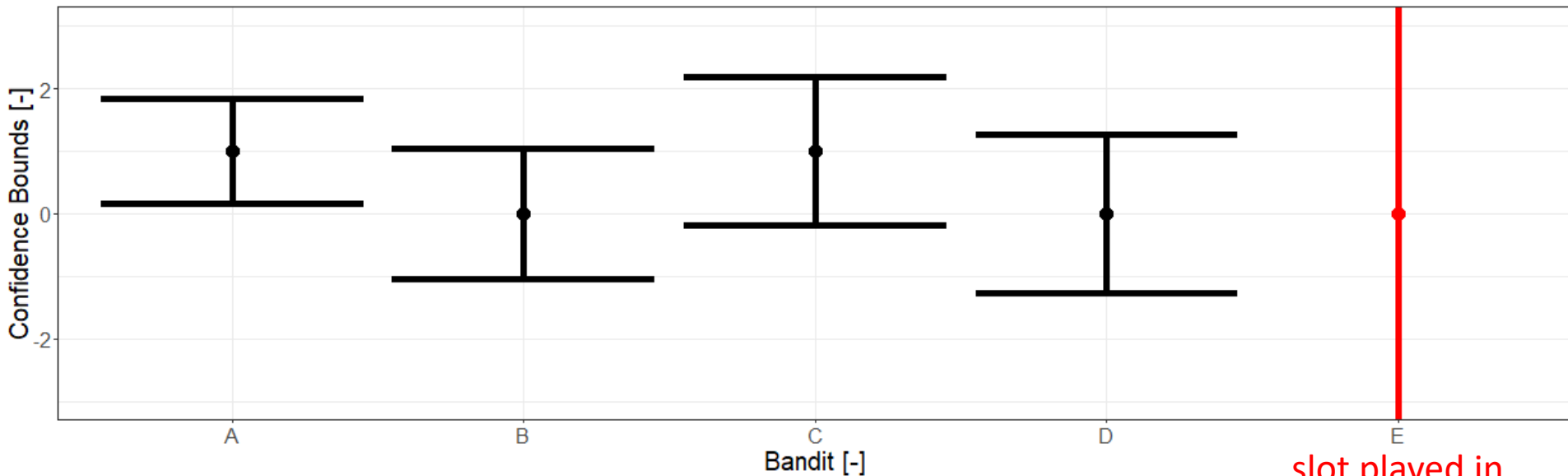


Upper Confidence Bound

Multi-Armed-Bandit Problem

■ Round: 4

Confidence Bounds



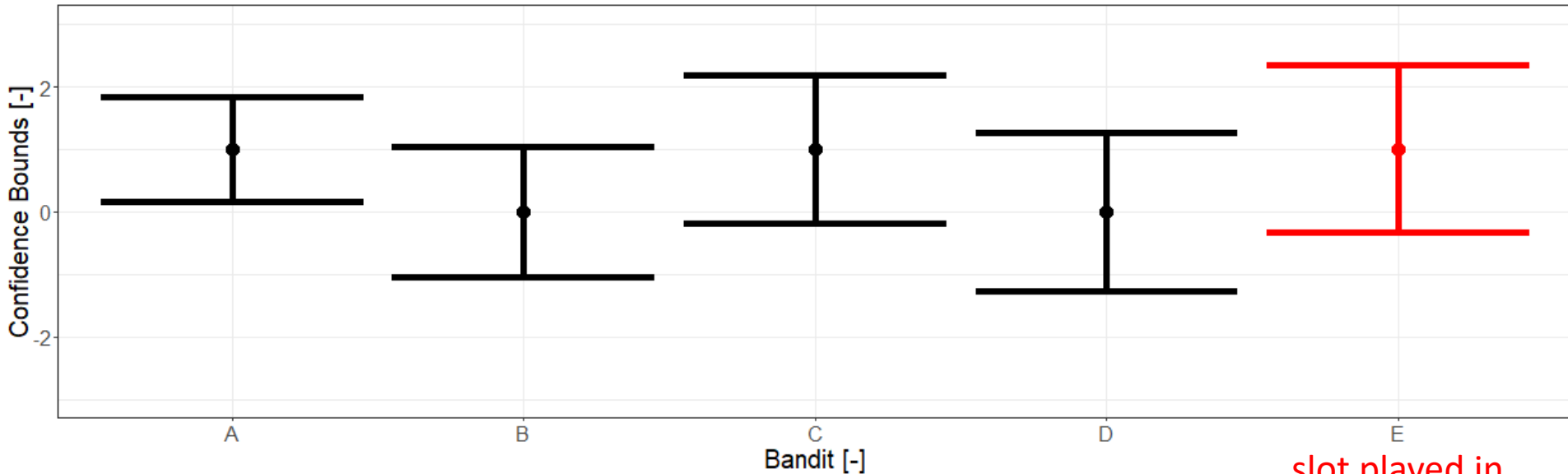
slot played in
next round

Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 5

Confidence Bounds



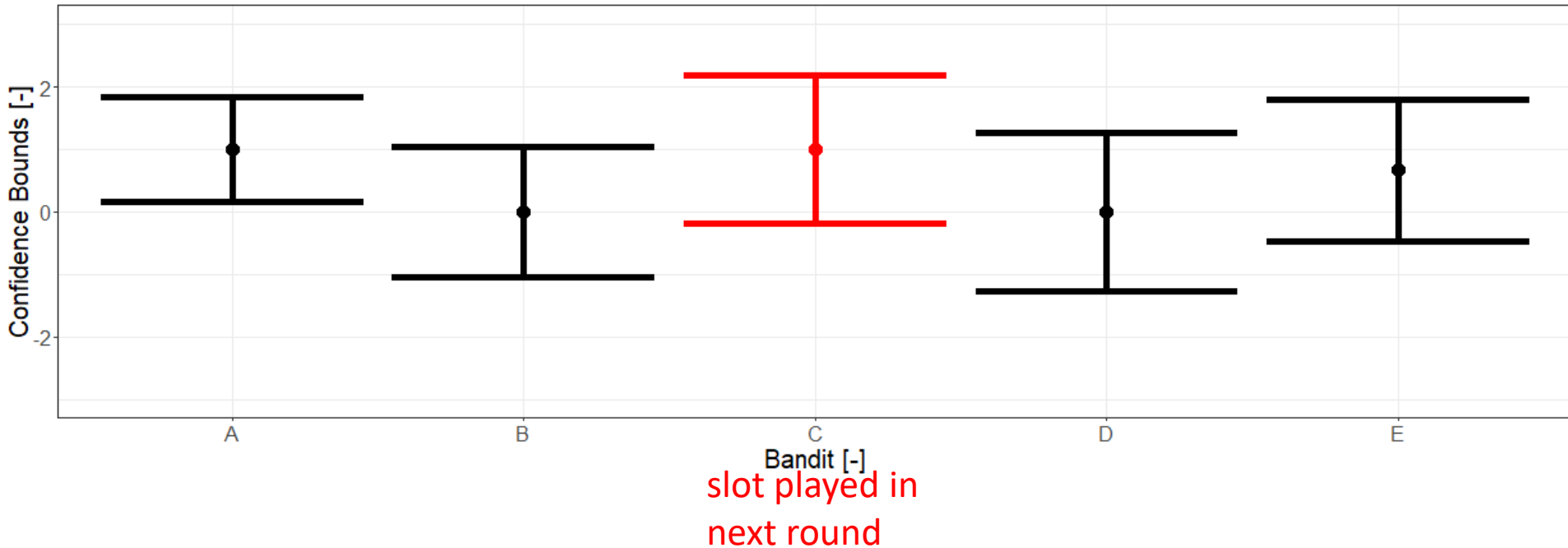
slot played in
next round

Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 6

Confidence Bounds

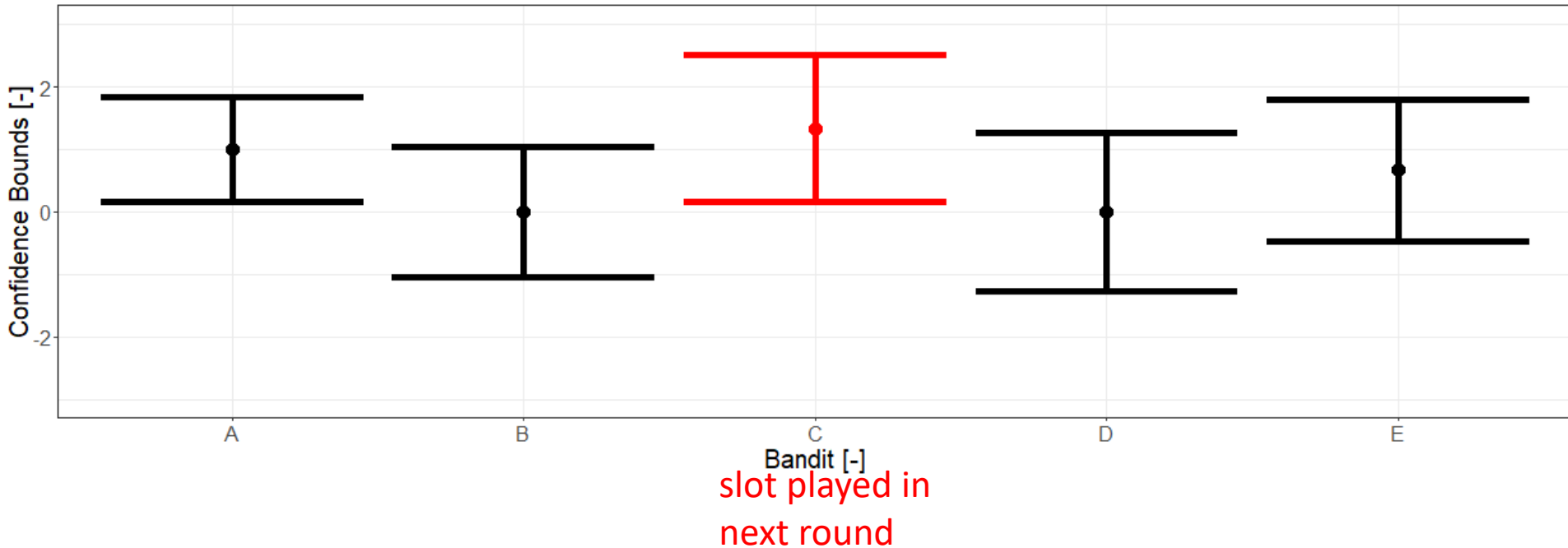


Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 7

Confidence Bounds

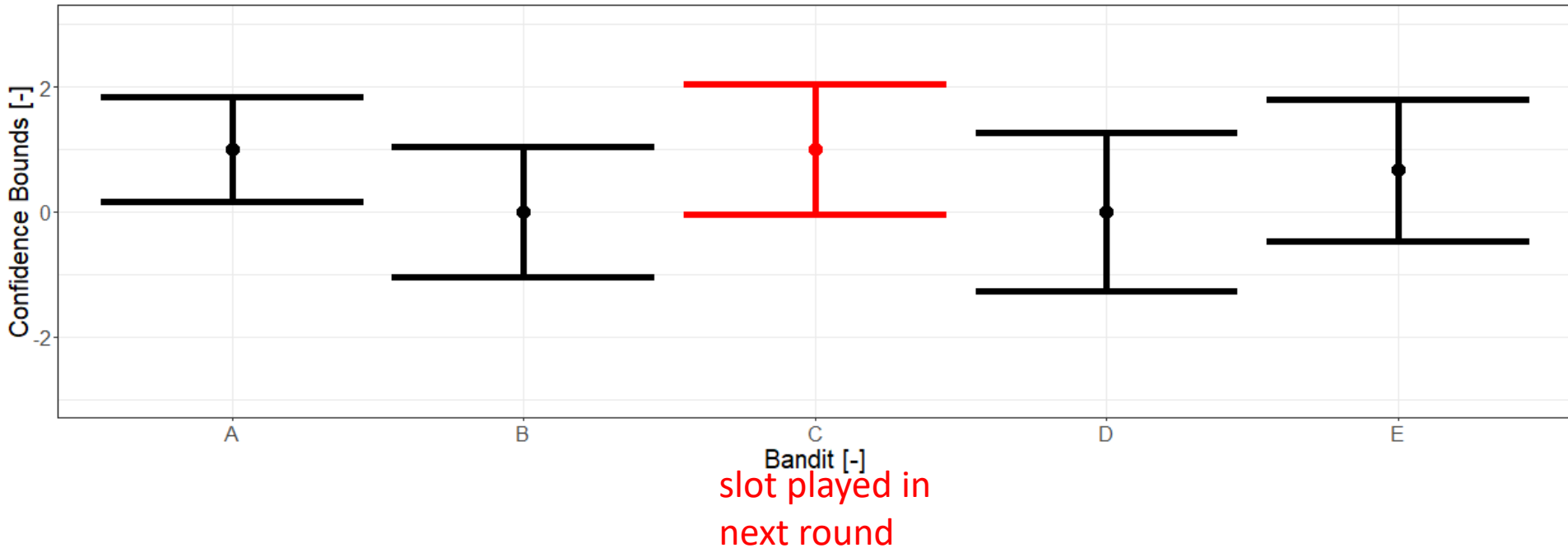


Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 8

Confidence Bounds

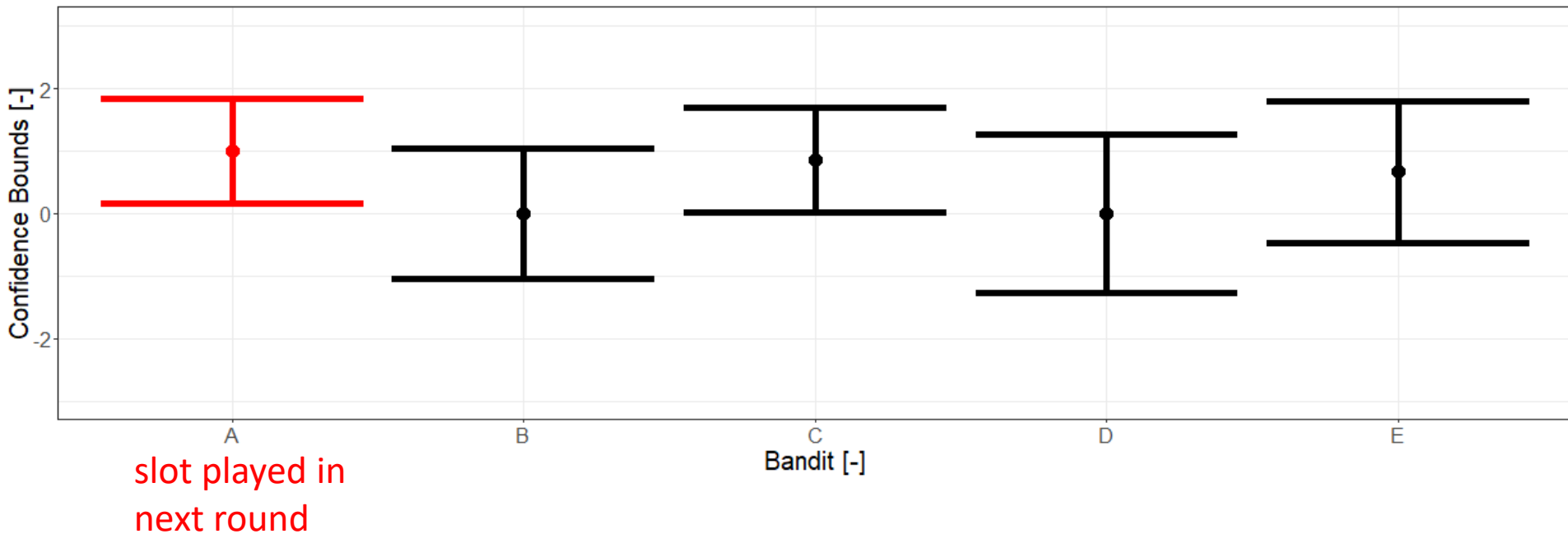


Upper Confidence Bound

Multi-Armed-Bandit Problem

▪ Round: 9

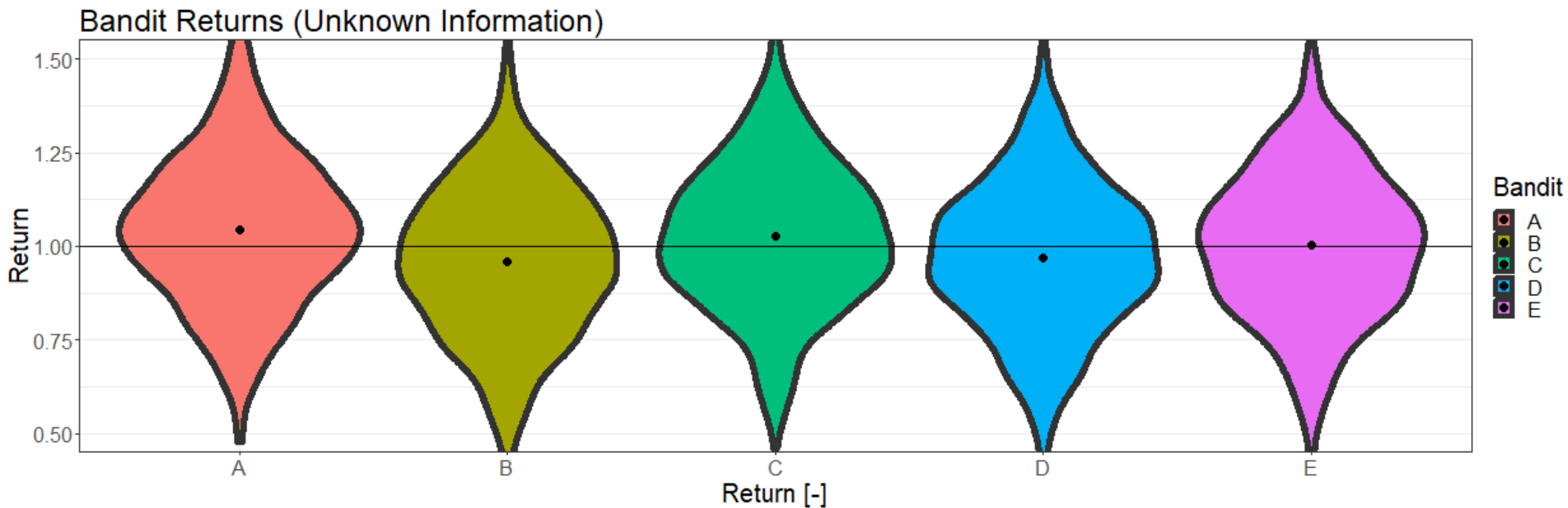
Confidence Bounds



Upper Confidence Bound

Multi-Armed-Bandit Problem

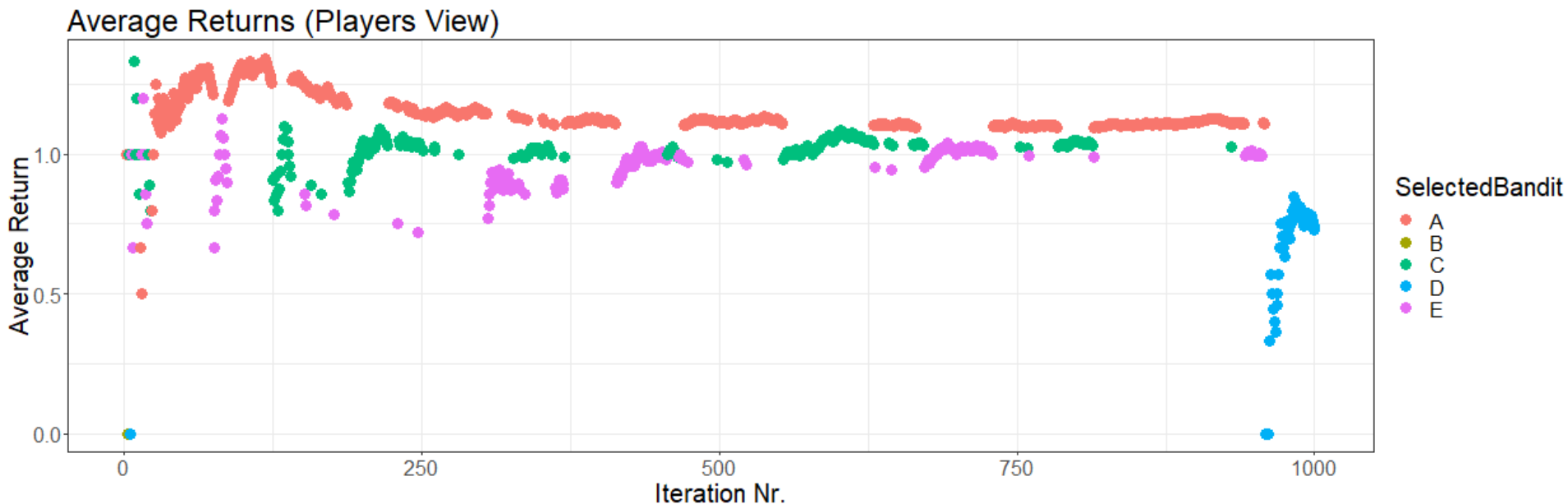
- Round: 1000



Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 1000



Upper Confidence Bound

Advantages / Disadvantages

+

- Adds some randomness
- Good balance of exploration and exploitation

-

- Bad action might be explored