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

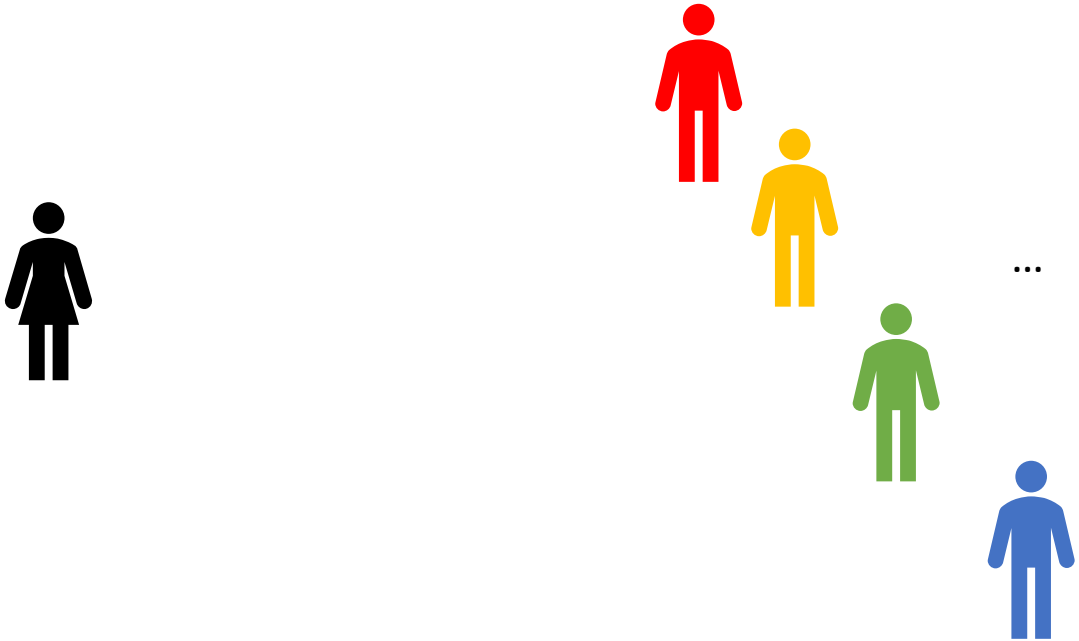
- Exploration
  - Choose an action with an unsure outcome
  - Possibly learn something

- 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

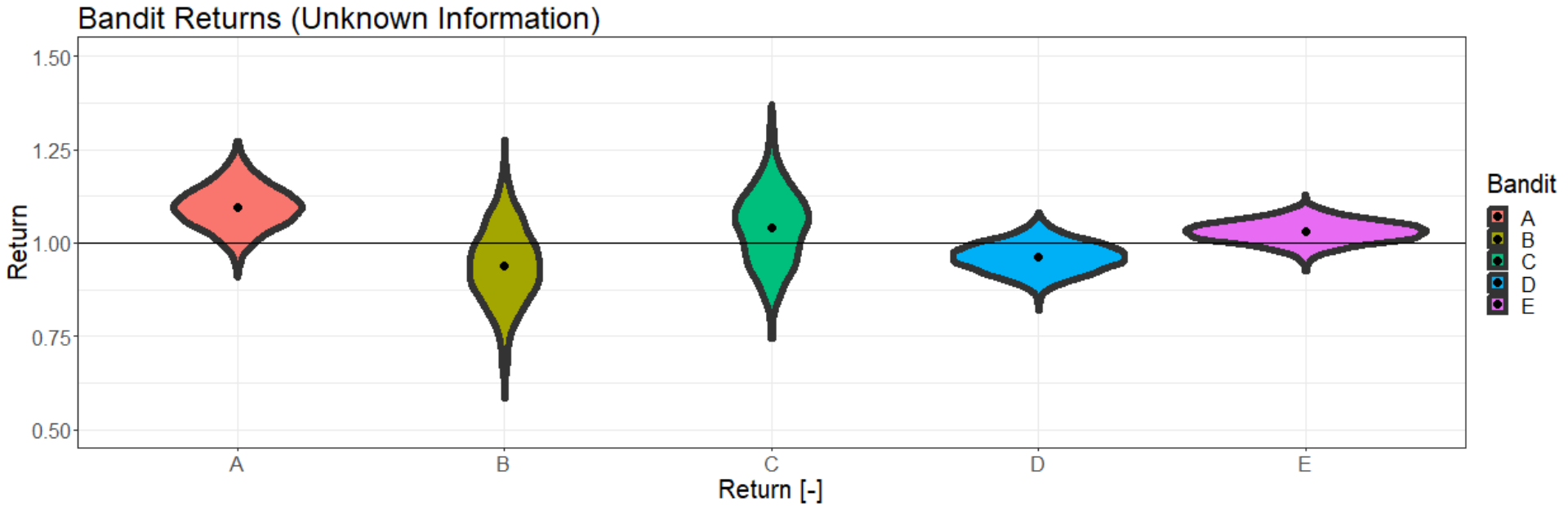
Bandit D

Bandit E

Upper Confidence Bound

Multi-Armed-Bandit Problem

- Multi-Armed Bandit Problem



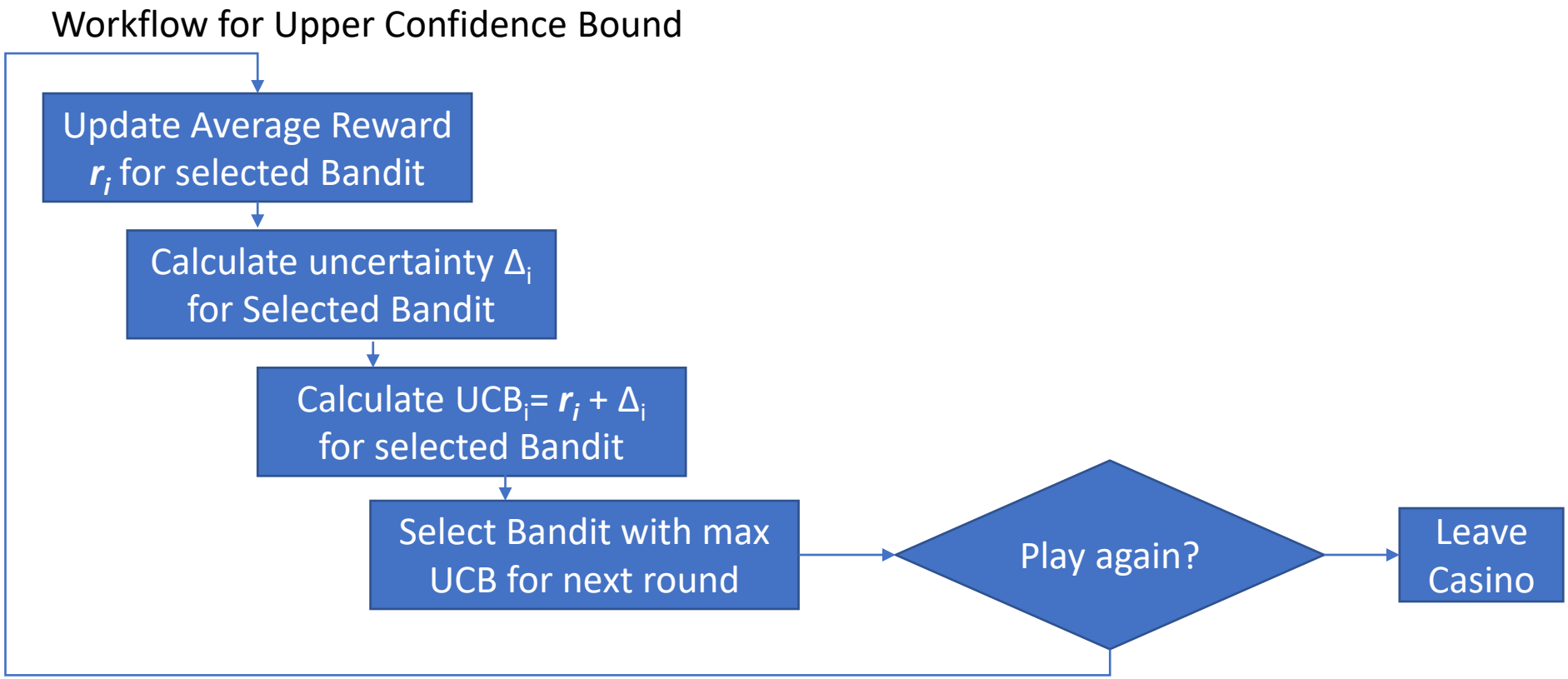
Upper Confidence Bound

Multi-Armed-Bandit Problem

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

Upper Confidence Bound

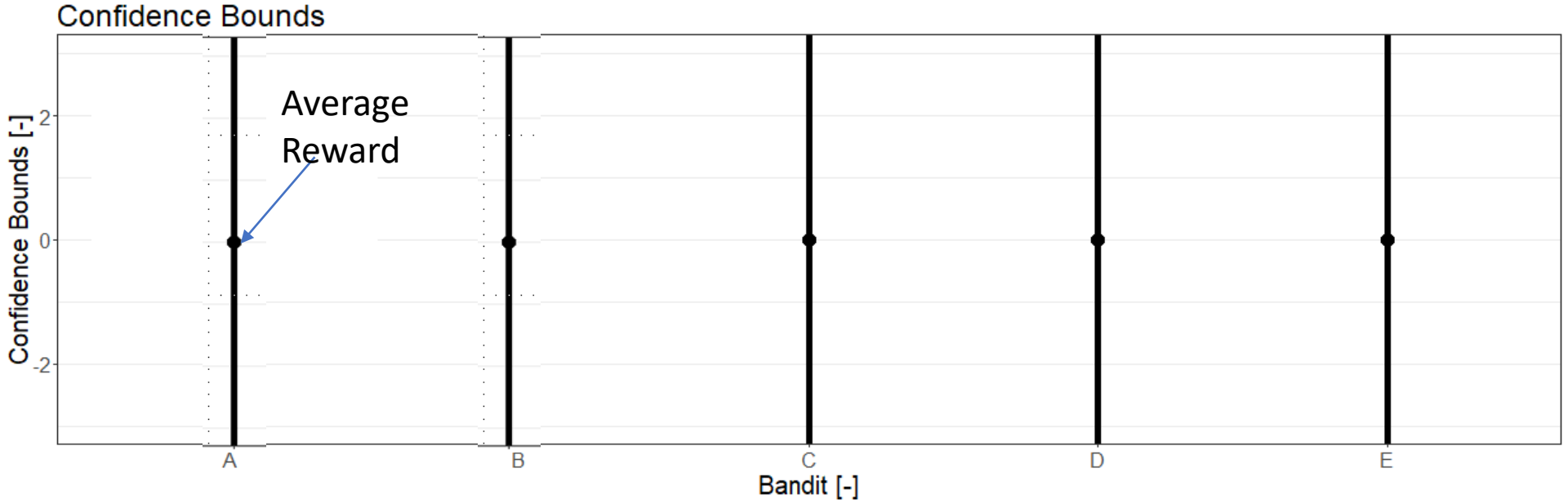
Workflow



Upper Confidence Bound

Multi-Armed-Bandit Problem

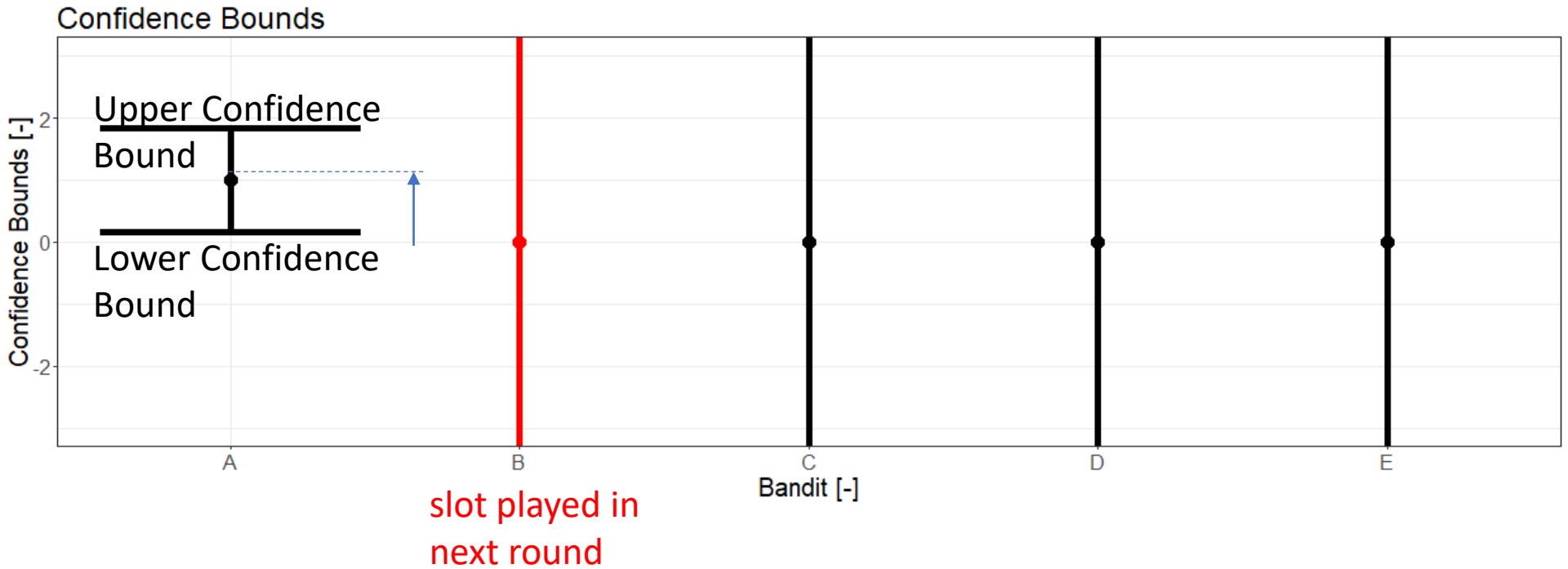
- Round: 0



Upper Confidence Bound

Multi-Armed-Bandit Problem

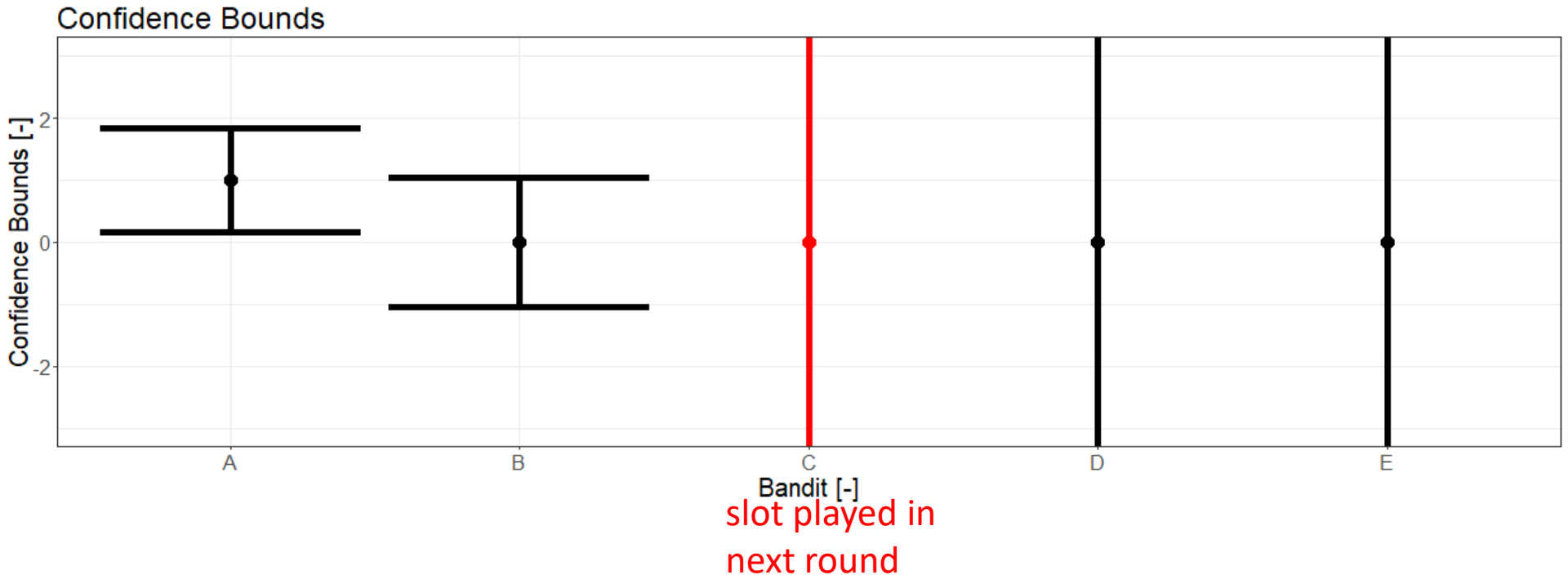
- Round: 1



Upper Confidence Bound

Multi-Armed-Bandit Problem

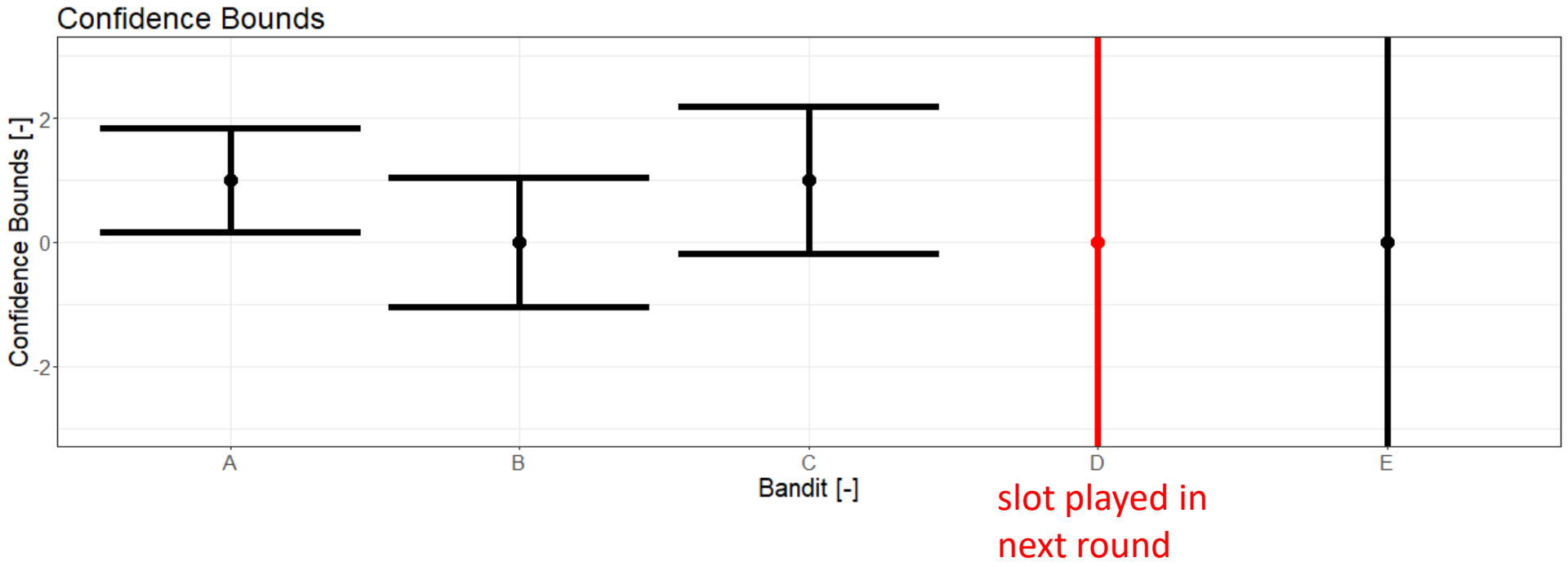
- Round: 2



Upper Confidence Bound

Multi-Armed-Bandit Problem

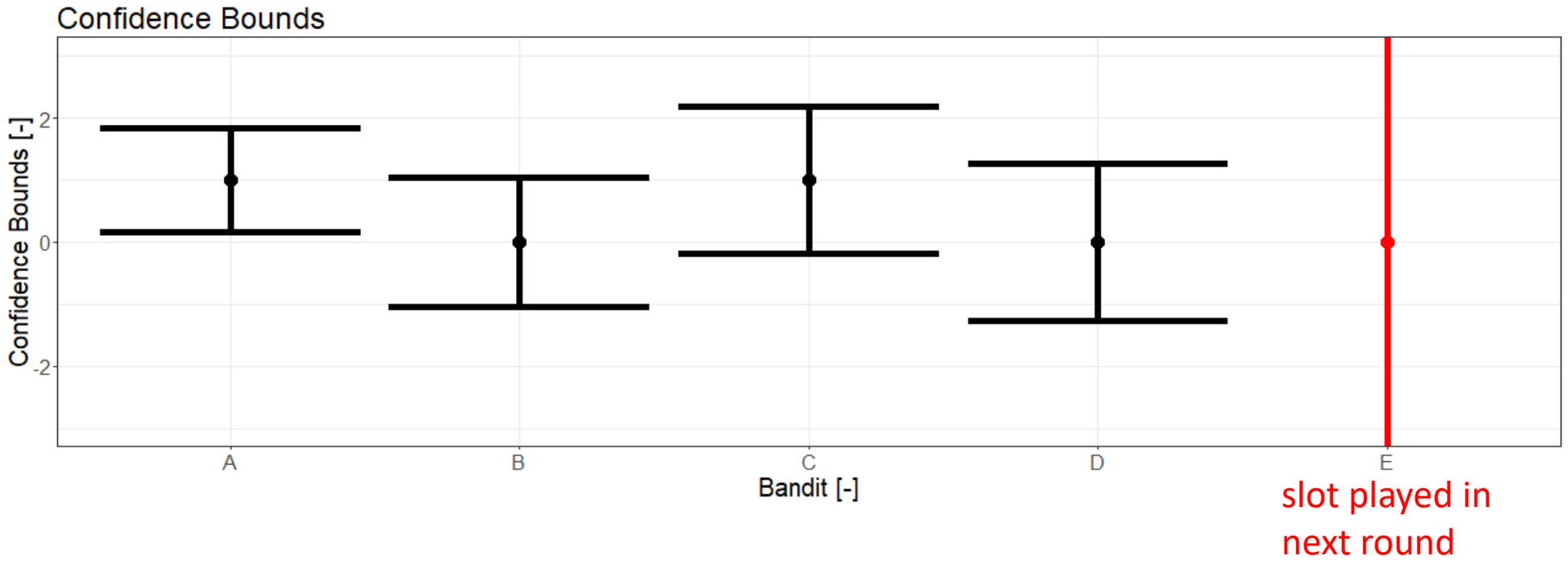
- Round: 3



Upper Confidence Bound

Multi-Armed-Bandit Problem

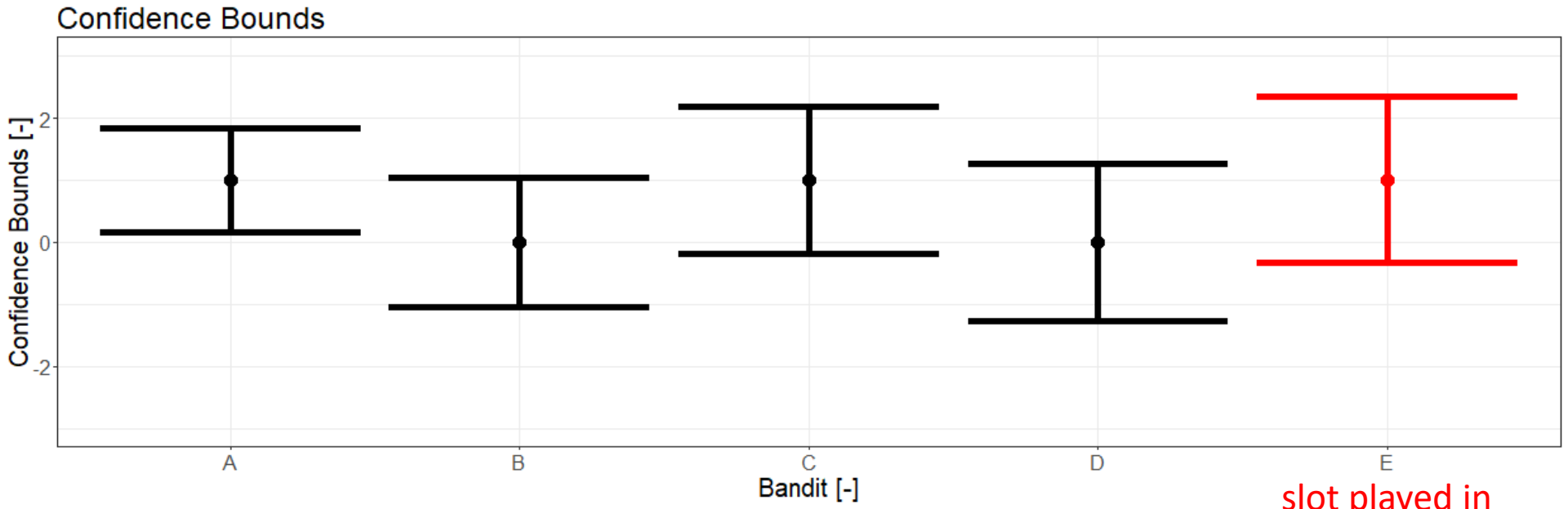
- Round: 4



Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 5

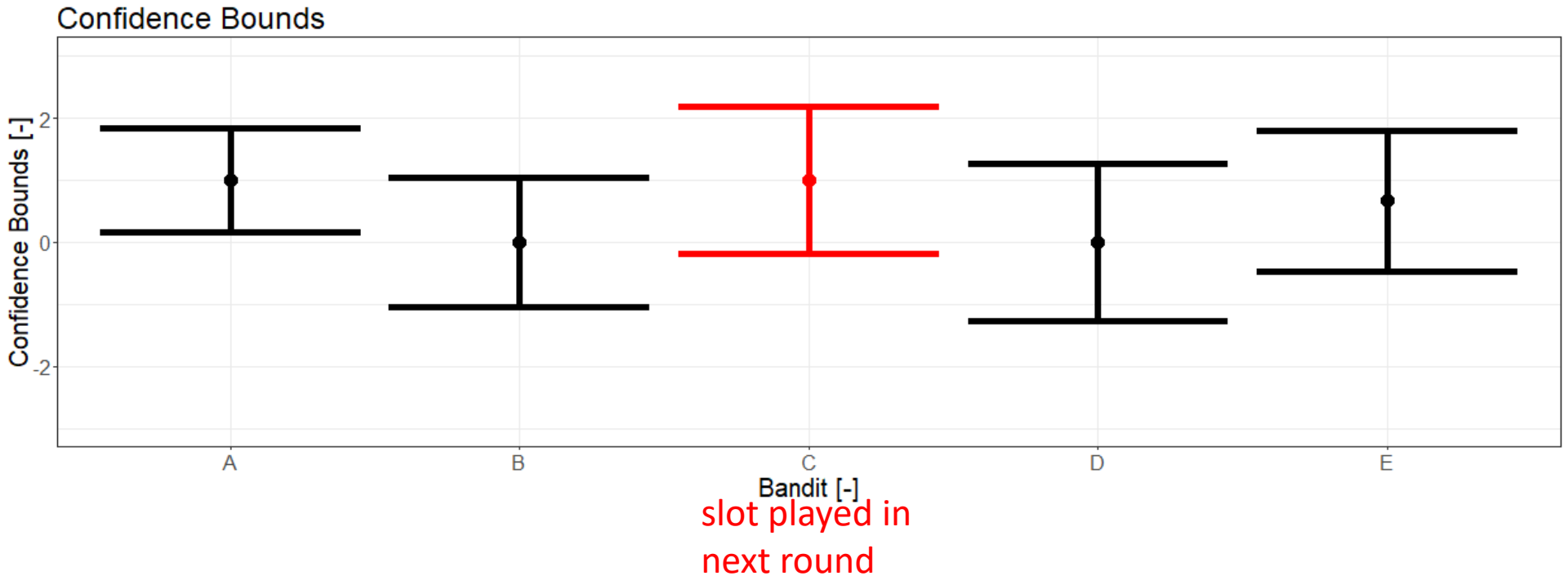


slot played in next round

Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 6

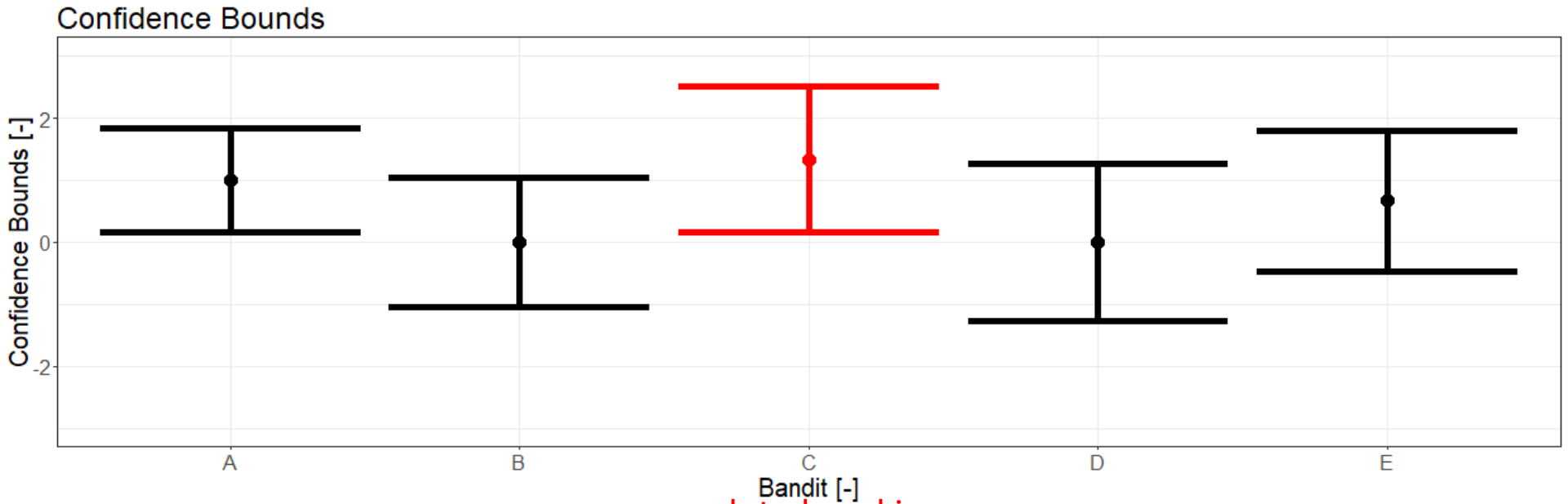


slot played in next round

Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 7

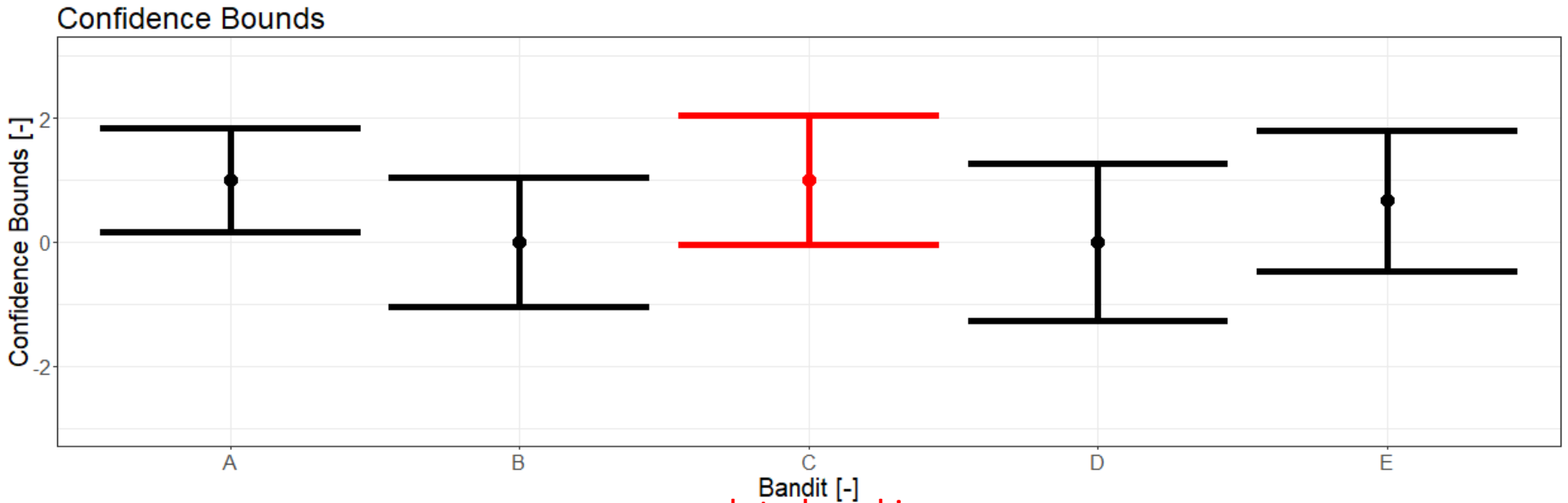


slot played in next round

Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 8

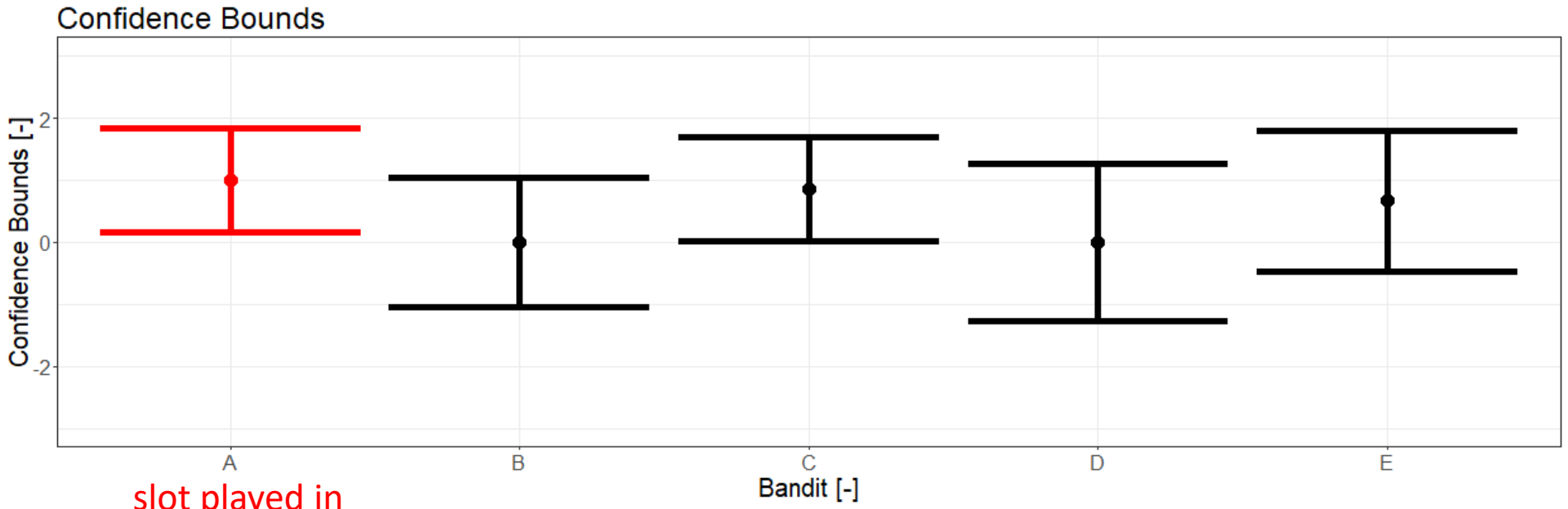


slot played in next round

Upper Confidence Bound

Multi-Armed-Bandit Problem

- Round: 9

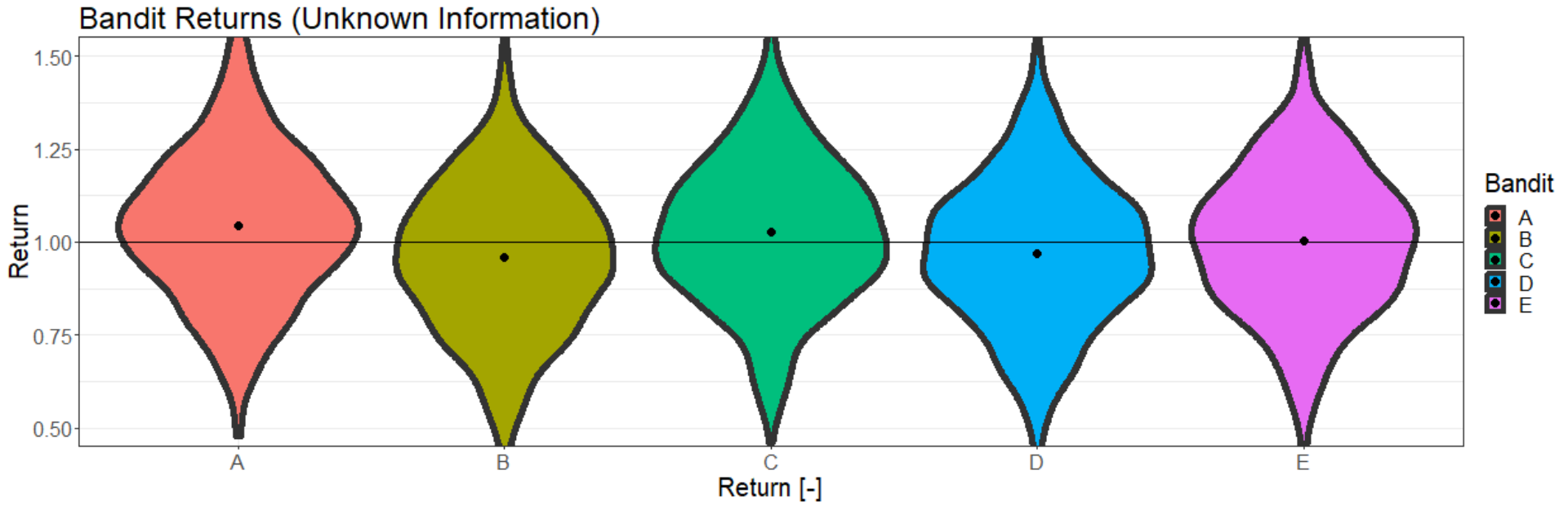


slot played in next round

Upper Confidence Bound

Multi-Armed-Bandit Problem

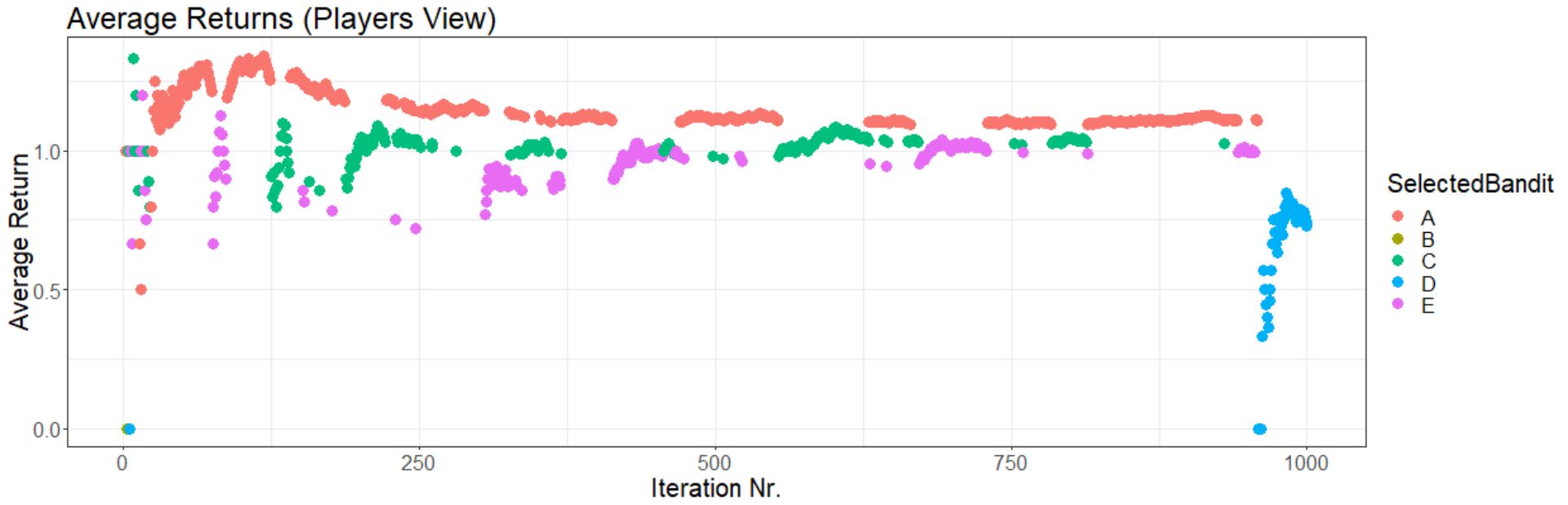
- Round: 1000



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