Multi-armed Bandits

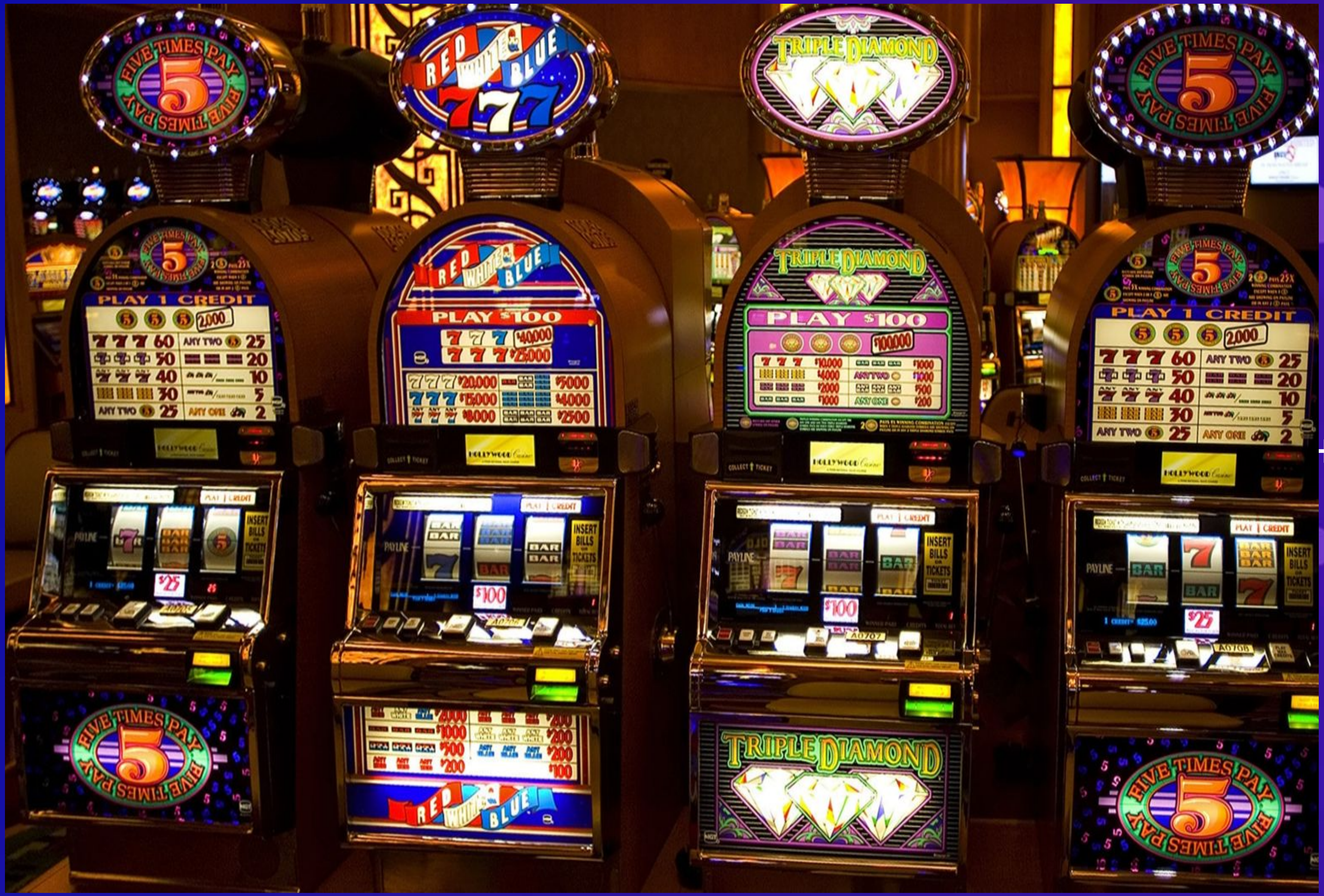
Tom Vodopivec

IADS Big Data and Analytics Summer School
2019-08-05

Bandits??

One-armed bandit

= slot machine



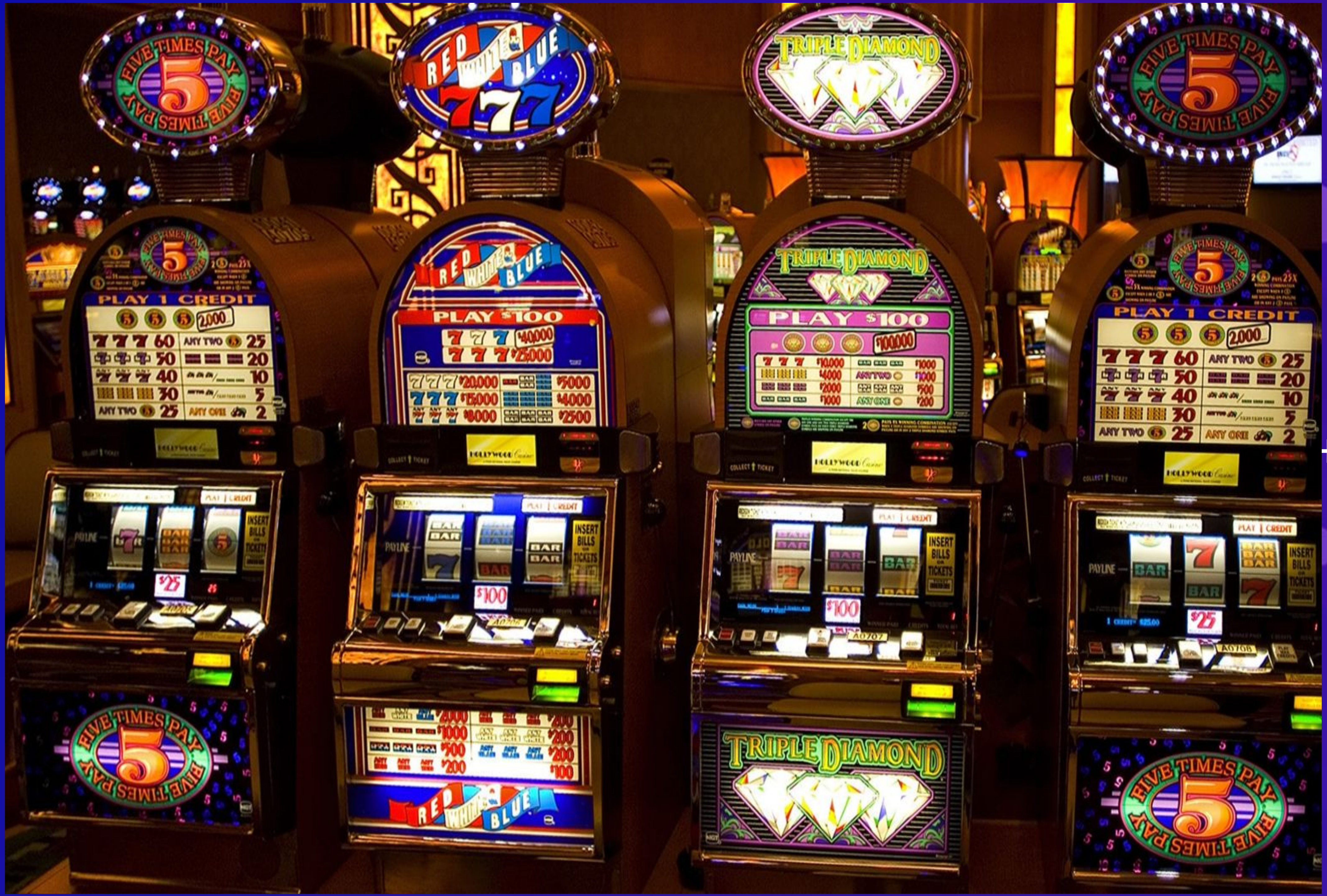
celtra

Bandits??

One-armed bandit

= slot machine

When you have more,
which one has better odds?



Problem definition

Single-state MDP

Maximize reward

Learn optimal policy

Exploration-exploitation trade-off

Types of Bandits

Stochastic

Adversarial (game theory and Nash equilibria)

Markovian (some underlying state space - use RL techniques)

+

Non-stationary

Budget-limited bandits

Contextual bandits (with covariates)

Sleeping bandits

Many-armed bandits

Solutions (policies)

Epsilon-greedy

Solutions (policies)

Epsilon-greedy

Upper Confidence Bounds (UCB)

- Optimism in the face of uncertainty
- Optimal in the limit

Solutions (policies)

Epsilon-greedy

Upper Confidence Bounds (UCB)

- Optimism in the face of uncertainty
- Optimal in the limit

Thompson Sampling

- Bayesian: draw samples from (Beta) distribution
- Number of pulls of an arm should match probability of the arm being optimal

Performance Metrics

Cumulative reward

Regret

Probability of choosing optimal arm

Use in Marketing

A/B testing with real-time optimization

Pros

- Faster identification of best variant - more cumulative reward
- Adapts to non-stationarity (no need to repeat A/B tests)
- No need to select best variant manually (additional step at end of A/B test)

Cons

- More difficult to understand and to implement
- More difficult to measure benefits

Reading Material

[Kuleshov and Precup, Algorithms for the Multi-armed Bandit Problem]

[Auer et al., Finite-time Analysis of the Multiarmed Bandit Problem]

[Vermorel and Mohri, Multi-Armed Bandit Algorithms and Empirical Evaluation]