

Lecture 20: Decisions and Bandits

Instructor: Sergei V. Kalinin

Friday lecture: options



AREMS

The Appalachian Regional Microscopy Society

Institute of Advanced Materials and Manufacturing
2641 Osprey Vista Way
Knoxville, TN 37996

Annual Meeting in Knoxville, TN
October 12-13, 2023

[Learn More Here \(https://arems.org/fall-meeting-info\)](https://arems.org/fall-meeting-info)

A Celebration of the Life and Contributions to UT and Microscopy by Dr. David Joy

Oct. 12, 2023 (Thursday)

12:00	Registration Opens
1:00-3:30	Hands On Tutorials (Thermo-Fisher, Plasma Fib) Hands On Tutorials (Oxford Instruments, AFM)
3:30-5:30	Hands On Tutorials (Leica, Cryo-Ultramicrotome sample prep) Hands On Tutorials (Protochips, In-Situ Holders in TEM/STEM)
5:30-6:30	Social/Student Poster Session
6:15-7:30	Dinner
7:30-9:00	Banquet Speaker - Dr. Nicholas Ritchie (NIST)



AREMS web link

Oct. 13, 2022 (Friday)

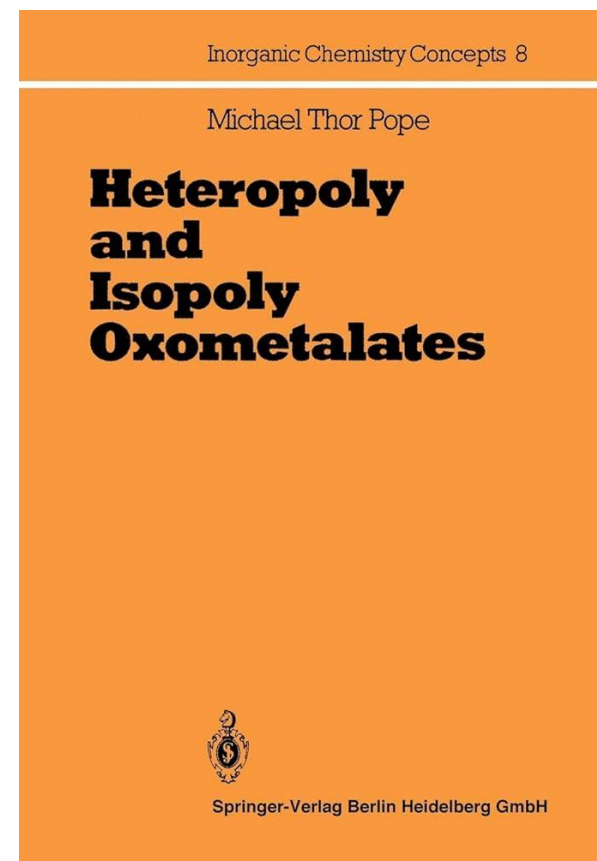
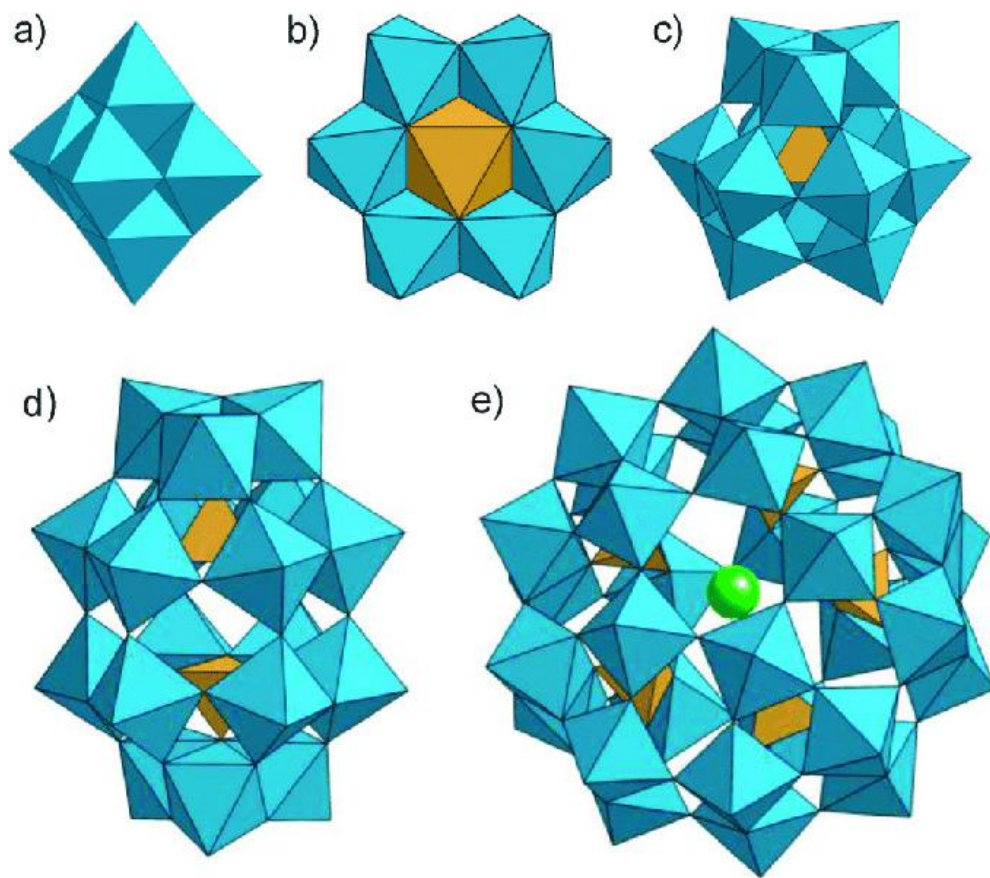
8:30-8:45	Welcome/Introductions to Officers & Attendees
8:45-9:15	Ed Principe: PanO Scientific CSSG
9:15-9:45	Dylan Wood: Protochips, In-Situ Holders in TEM/STEM
9:45-10:15	Alexis Williams: ORNL CryoTEM
10:15-10:45	Coffee Break
10:45-11:15	Rama Vasudevan: Machine Learning in Microscopy
11:15-11:45	Wonhee Ko: STM - AVS Invited Speaker
11:45-12:00	Business Meeting/Voting
12:00-12:30	Awards Ceremony
12:30-1:30	Lunch Break (on your own)
1:30-5:00	IAMM and ORNL lab tours

AREMS Awards

1. \$50 for the first 15 Student Posters
2. \$250 for Best Physical & Biological Posters
3. \$500 for McGill Award for Most Innovative Poster
4. \$500 for Russell Education Award

Meet your local fellow microscopists and vendors!

Synthesis of polyoxometallates

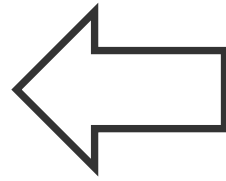


X. López, J. Carbó, C. Bo, J. Poblet, *Structure, properties and reactivity of polyoxometalates: A theoretical perspective*, Chemical Society reviews 71, 7537 (2012)

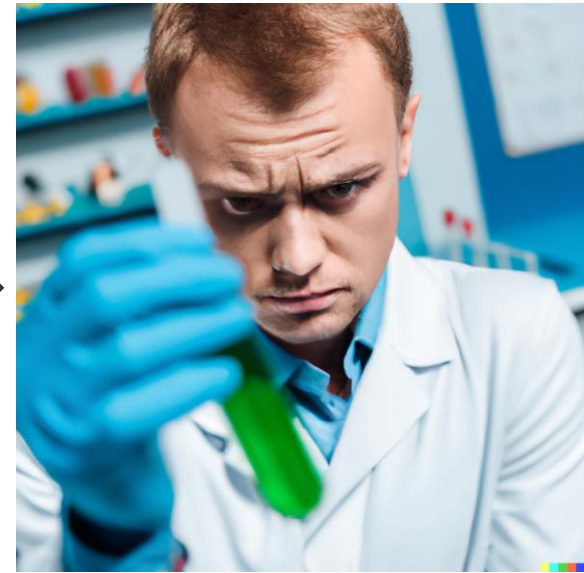
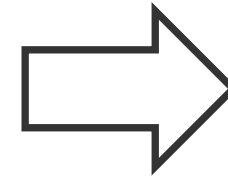
Synthesis of polyoxometallates



**Good
day**

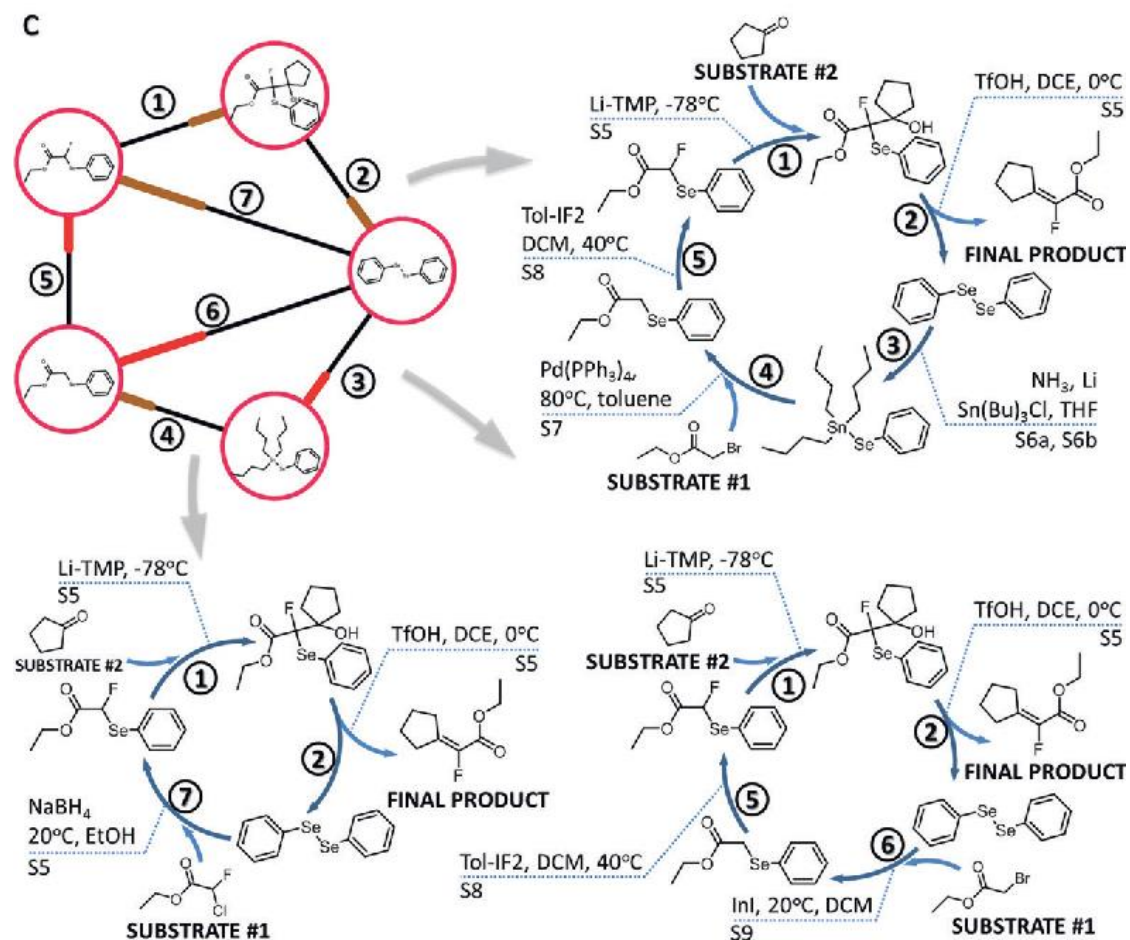


**Bad
day**



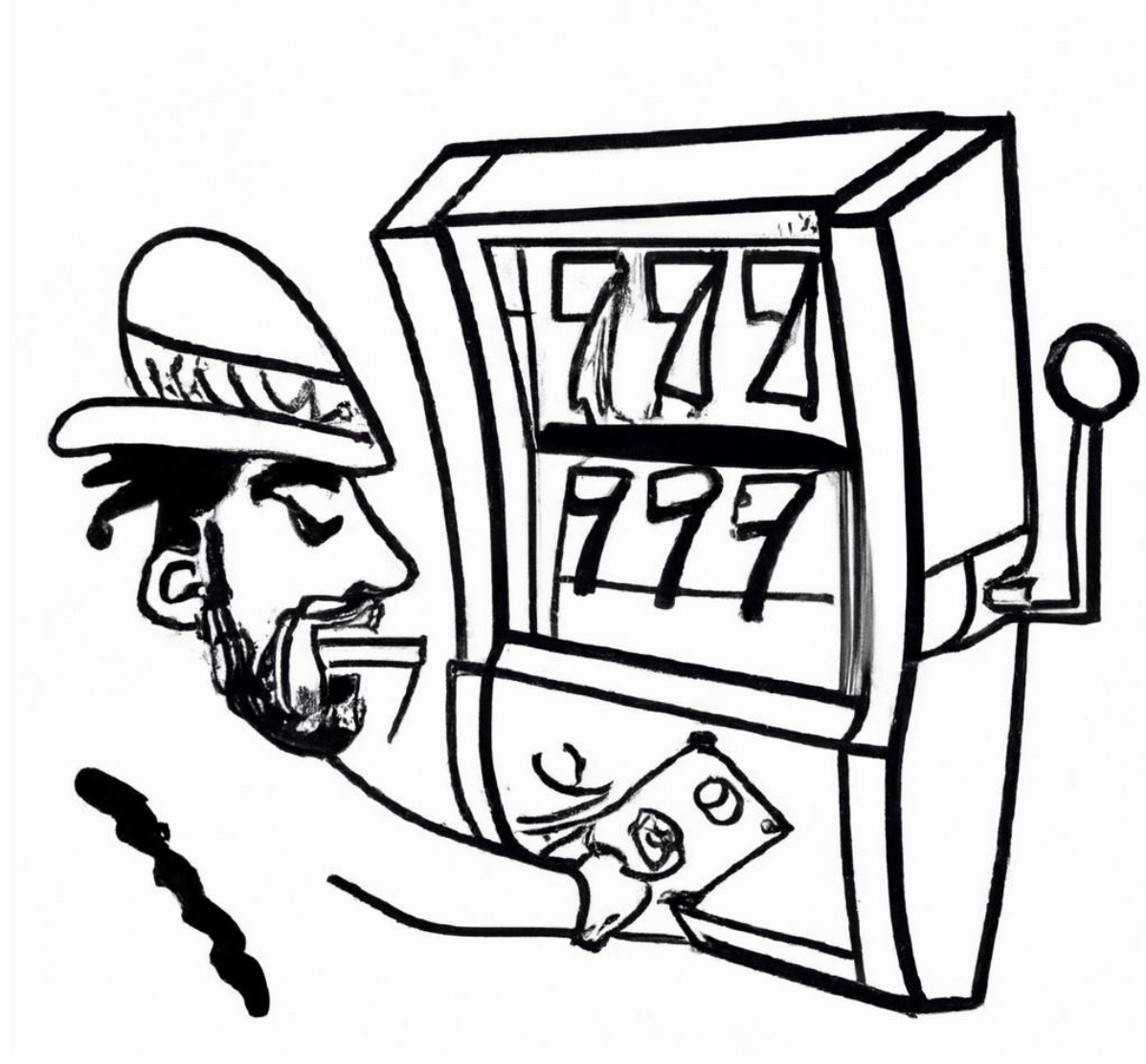
- Synthesis may or may not work!
- Depending on difficult to quantify factors
- It's a coin toss, a.k.a. Bernoulli distribution

Organic synthesis



M. Bajczyk, P. Dittwald, A. Wołoś, S. Szymkuć, B. Grzybowski, *Discovery and Enumeration of Organic-Chemical and Biomimetic Reaction Cycles within the Network of Chemistry*, DOI:10.1002/anie.201712052

Bandit problem



- Imagine that we have a number of slot machines with different probabilities of win...
- Or different web-sites to places ads on...
- Or groups of patients for specific medical protocol....
- Or team members to synthesize certain material...
- Or reaction pathways to choose

How do we optimize this problem and maximize our reward?

A/B Testing

- The most common exploration strategy is **A/B testing**, a method to determine which one of the two alternatives (of online products, pages, ads etc.) performs better.
- The users are randomly split into two groups to try different alternatives. At the end of the testing period, the results are compared to choose the best alternative, which is then used in production for the rest of the problem horizon.
- This approach can be applied for more than two alternatives - **A/B/n testing**.

Definitions: reward and value

$$Q_n \triangleq \frac{R_1 + R_2 + \dots + R_{n-1}}{n - 1}$$

- First, we denote the reward (i.e. 1 for click, 0 for no click) received after selecting the action a for the n^{th} time by R_i .
- Q_n estimates the expected value of the reward that this action yields, R , after $n-1$ observations.
- Q_n is also called the action value of a . Here, Q_n estimates of the action value after selecting this action $n - 1$ times.

Updating value

$$Q_{n+1} = \frac{R_1 + R_2 + \dots + R_n}{n} = Q_n + \frac{1}{n} \cdot (R_n - Q_n)$$

- Q_n is the estimate for the action value of a before we take it for the n^{th} time.
- When we observe the reward R_n , it gives us another signal for the action value.
- We adjust our current estimate, Q_n , in the direction of the **error** that we calculate based on the latest observed reward, R_n , with a **step size** $1/n$ and obtain a new estimate Q_{n+1}
- For convenience, $Q_0 = 0$ (But - Human heuristics!)

Updating value: generalization

$$Q_{n+1}(a) = Q_n(a) + \alpha(R_n(a) - Q_n(a))$$

- The rate at which we adjust our estimate will get smaller as we make more observations
- The step size must be smaller than 1 for the estimate to converge (and larger than 0 for a proper update).
- Using a fixed α will make the weights of the older observations to decrease exponentially as we take action a more and more.

Limitations of A/B testing

- **A/B/n testing is inefficient as it does not modify the experiment dynamically by learning from the observations.** It fails to benefit from the early observations in the test by writing off/promoting an alternative even if it is obviously under- or outperforming the others.
- **It is unable to correct a decision once it's made.** There is no way to correct the decision for the rest of the deployment horizon.
- **It is unable to adapt to the changes in a dynamic environment.** If the underlying reward distributions change over time, plain A/B/n testing has no way of detecting such changes after the selection is fixed.
- **The length of the test period is a hyperparameter to tune, affecting the efficiency of the test.** If this period is chosen to be shorter than needed, an incorrect alternative could be declared the best because of the noise in the observations. If the test period is chosen to be too long, too much money gets wasted in exploration.
- **A/B/n testing is simple.** Despite all these shortcomings, it is intuitive and easy to implement, therefore widely used in practice

ϵ -greedy policies

- Most of the time, greedily taking the action that is the best according to the rewards observed that far in the experiment (i.e. with $1-\epsilon$ probability)
- Once in a while (i.e. with ϵ probability) take a random action regardless of the action performances.
- Here ϵ is a number between 0 and 1, usually closer to zero (e.g. 0.1) to "exploit" in most decisions.
- Obviously, the number of alternatives has to be fairly small
- Parameter ϵ can change during the experiment

ϵ -greedy policies

