Sequential decision making in high-stakes environments

Eric B. Laber
SAS 2023

Duke

Acknowledgements

- Thank you to Anne Milley and Janice LeBeau
- Joint work with
  - Marie Davidian
  - Peter Norwood
  - I-SPY 2+ team
- Funding
  - NIH (R01 CA207689, R01 CA202779)
  - NSF (DMS 2136034)

Duke

Roadmap

- Canonical sequential decision problems
- Reinforcement learning and high-stakes problems
- Example of a high-stakes RL algorithm
- Concluding discussion

Duke

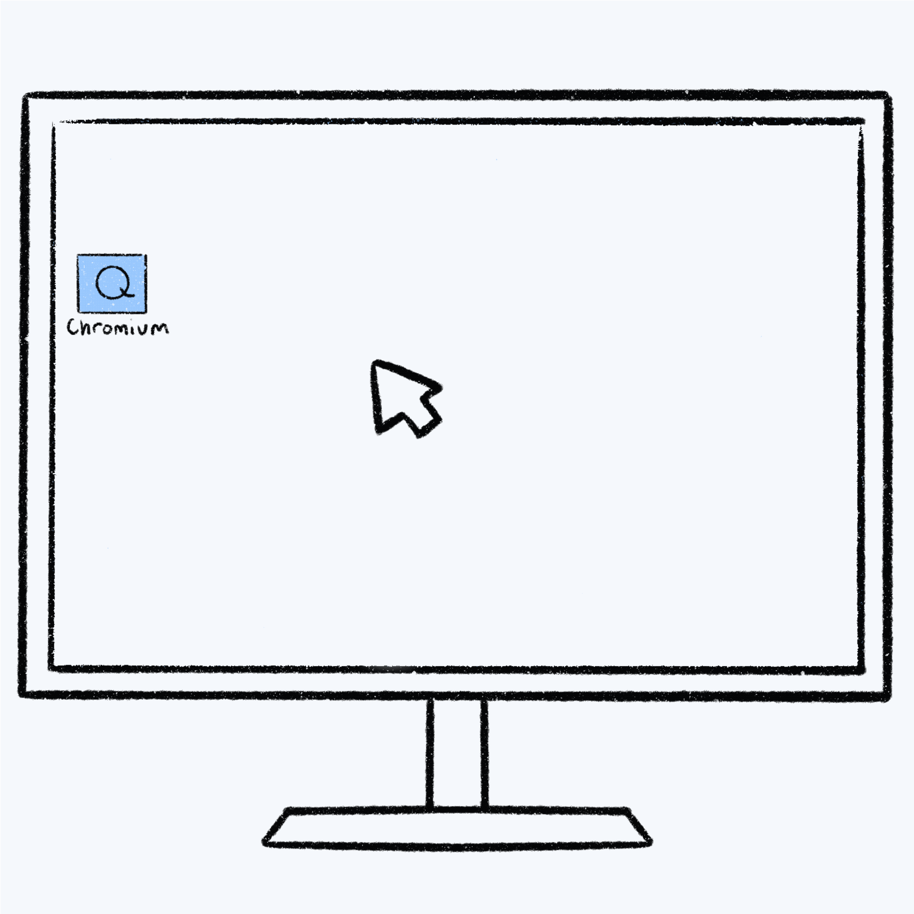
Roadmap

- **Canonical sequential decision problems**
- Reinforcement learning and high-stakes problems
- Example of a high-stakes RL algorithm
- Concluding discussion

Duke

A canonical decision problem

Recommender systems in e-commerce



Duke

Personalization in e-commerce

- Move towards personalized recommendations
  - Use customer attributes and history to drive recommendations
  - Search results
  - Ads and promotions
  - Streaming content
  - Etc.

Duke

Product category management

Duke

Assortment selection

Adaptive assortment selection

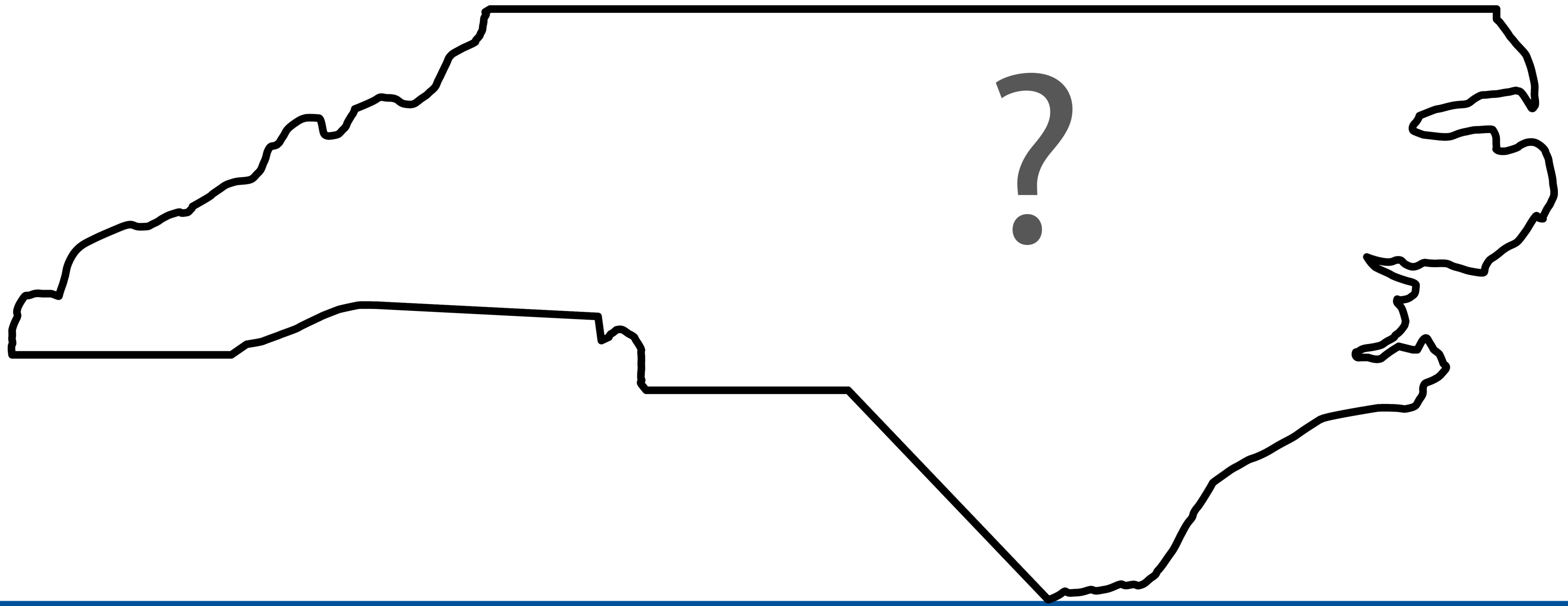
- [Select which items to put in stock at a store]

- Combinatorial decision problem
  - Select L items from catalog of size N → $O(N^L)$ choices
  - Even more complex as one considers more stores

- Goal: fresh and localized assortment

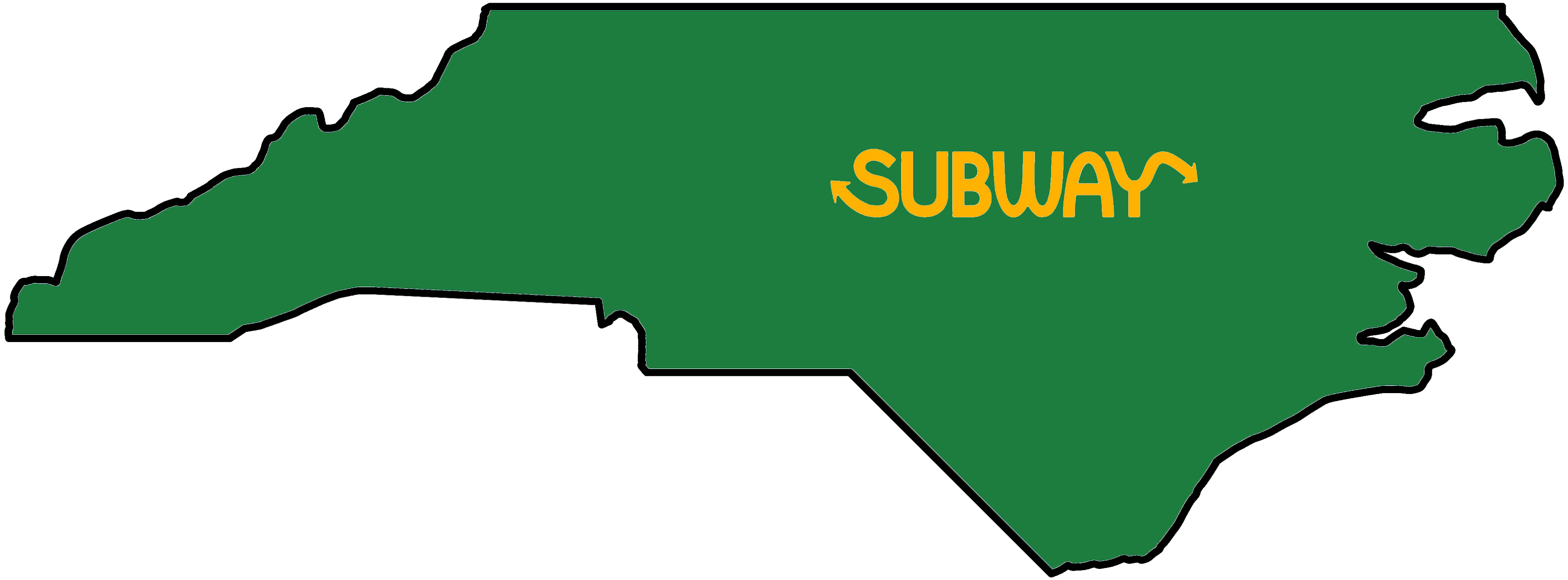
Duke

Which fast food restaurant has the most locations in North Carolina? [No googling ;)]



Duke

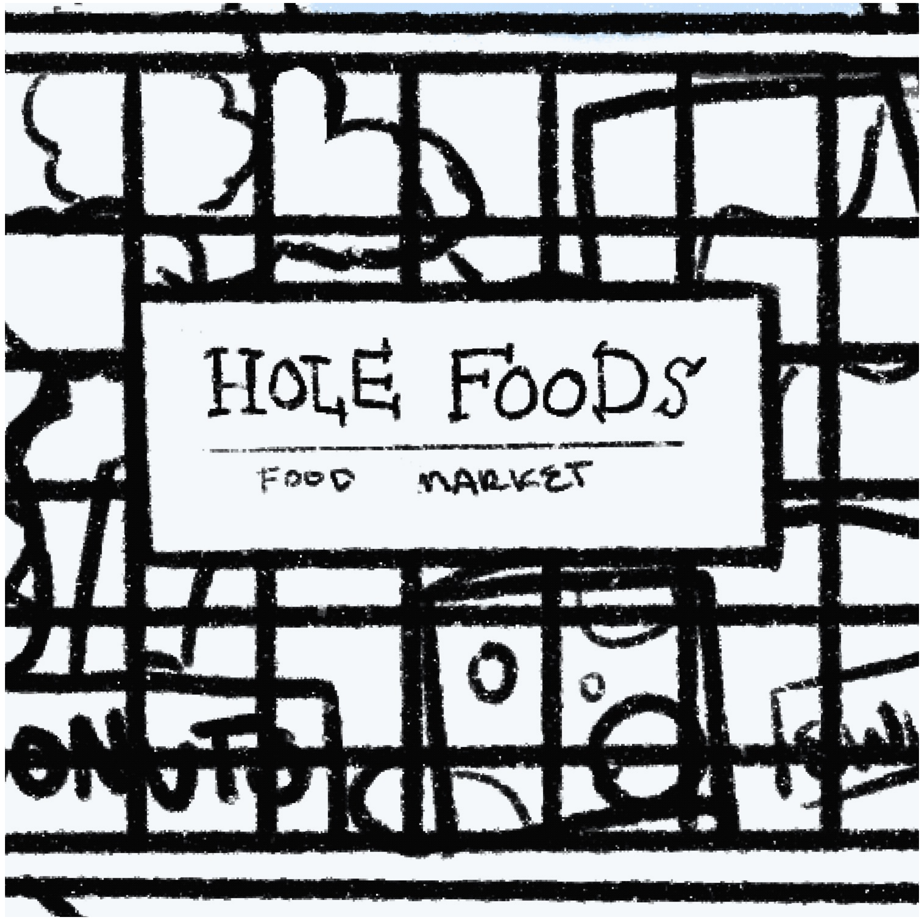
Which fast food restaurant has the most locations in North Carolina? [No googling ;)]



Which fast food restaurant has the most locations in USA?

Duke

Building out a network of stores



Duke

Facility location

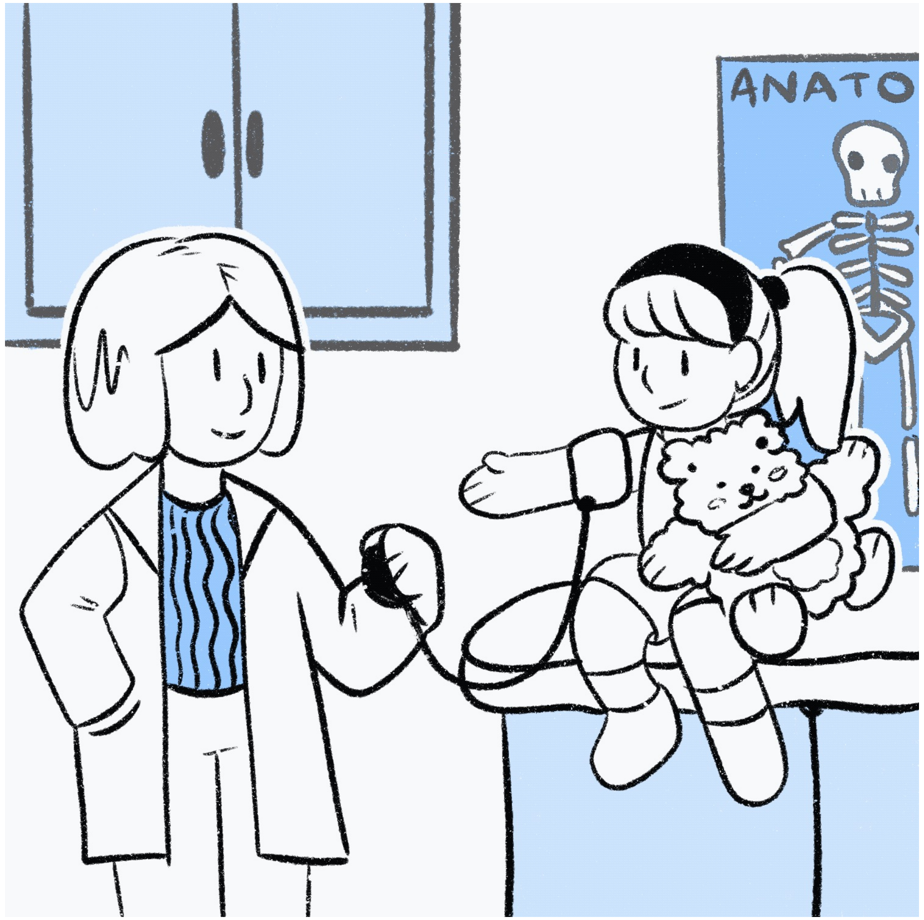
- Decide where and where to build stores/warehouses/hospitals/etc.

- Each decision carries high cost
  - Zero appetite for random exploration
  - Cannot easily undo a decision

- Requires coordination
  - Synergistic and cannibalization effects
  - Best location for single next store may not be optimal long-term
  - Current learnings inform future decisions

Duke

Medical decision making



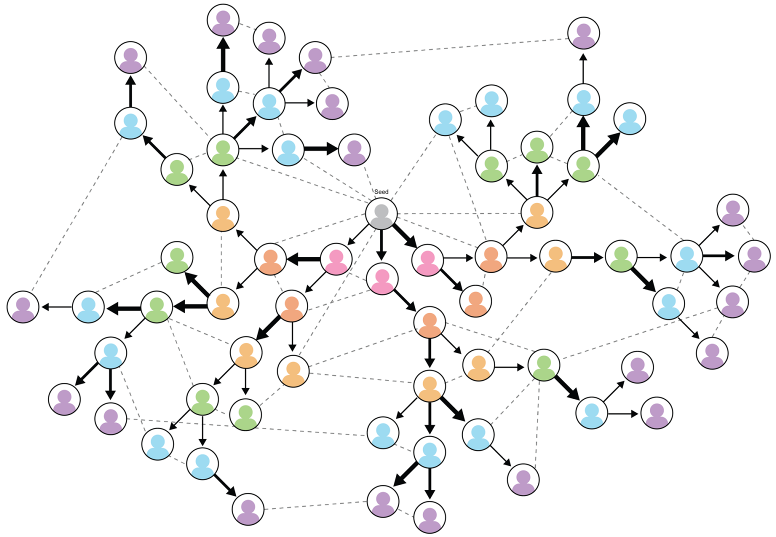
Duke

Personalization in healthcare

The I-SPY Trials

- Precision medicine
  - The right treatment for the right patient at the right time
  - Improve patient outcomes and reduce cost by giving treatment if and when needed

- Public health
  - Allocate resources if, when, and where needed
  - Adaptive network based sampling



Duke

Roadmap

- Canonical sequential decision problems
- Reinforcement learning and high-stakes problems
- Example of a high-stakes RL algorithm
- Concluding discussion

Duke

Roadmap

- Canonical sequential decision problems
- **Reinforcement learning and high-stakes problems**
- Example of a high-stakes RL algorithm
- Concluding discussion

Duke

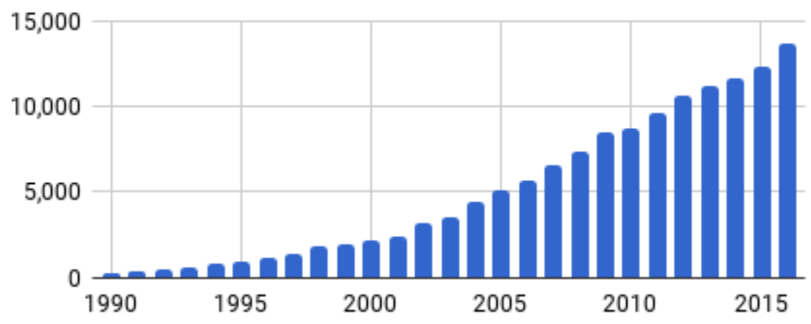
Reinforcement learning

Reinforcement learning (RL)

- Area of machine learning focused on optimal sequential decision making under uncertainty

- Massive and rapidly-expanding literature

Publications on RL (Henderson et al. 2017)



Duke

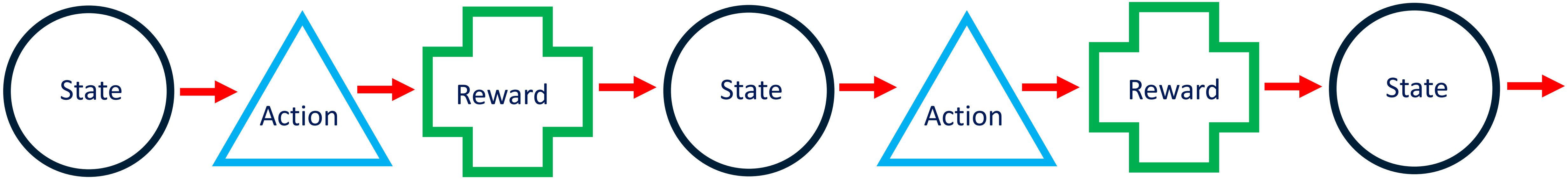
Reinforcement learning (RL) cont’d

- Application areas of RL [up to 2023 as per Bard]

Application area	Publication count
Robotics	43210
Games	31892
Control Theory	28102
Optimization	21321
Computer Vision	18201
Natural Language Processing	15120
Finance	12032
Healthcare	9821
Transportation	7610
Education	5392

Duke

Schematic for RL



Goal: select actions to maximize cumulative reward

Duke

RL background

- Formalize decision making as a policy

  State → Action

- Optimal policy maximizes cumulative utility, e.g., symptom reduction, disease-free survival, integrated quality of life, etc.

- Goal: learn optimal decision strategy as you go [i.e., online]
  - Balance generation of utility and information
  - I.e., earning v learning, exploration v exploitation, ethics v efficiency

Duke

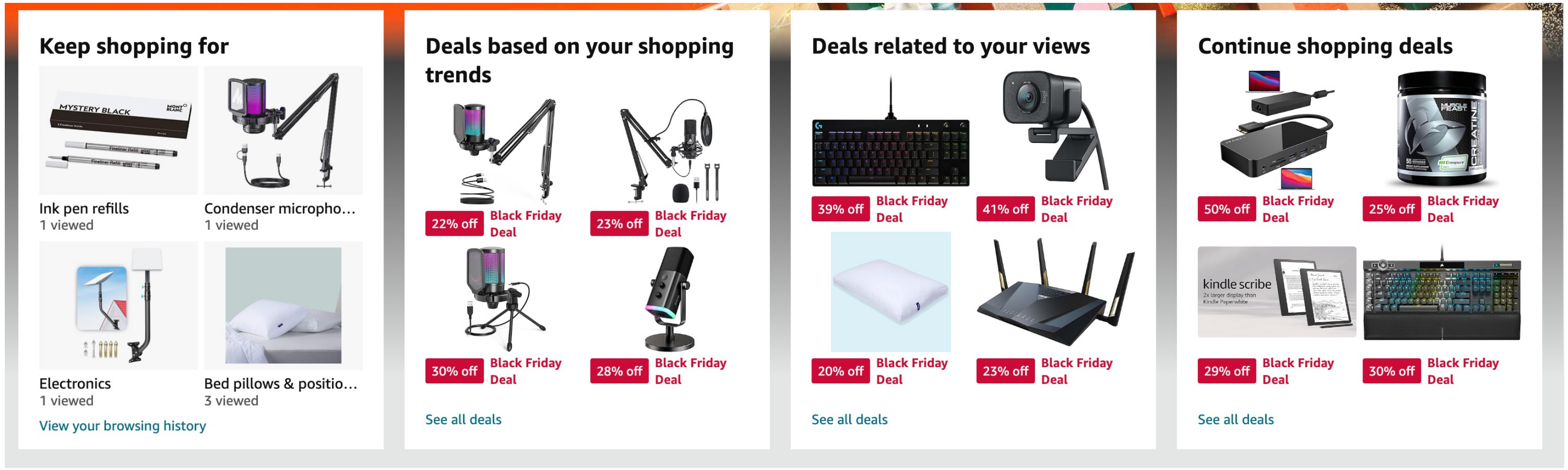
Ex. Thompson Sampling

- Widely used RL algorithm

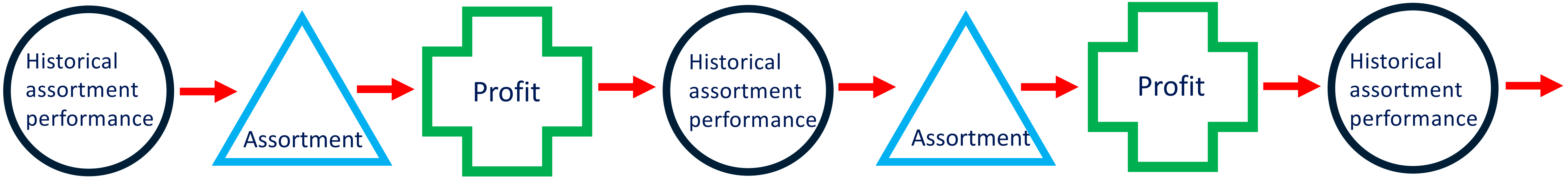
- Bayesian approach to uncertainty quantification
  - Posit class of models for system under study
  - At each time t
    - Draw a model from posterior
    - Select optimal decision assuming drawn model is correct
  - As information accumulates, posterior concentrates → balance experimentation and optimization

- Other algorithms inject exploration via randomization or *ad hoc* exploration bonus

Schematic for RL: recommender system

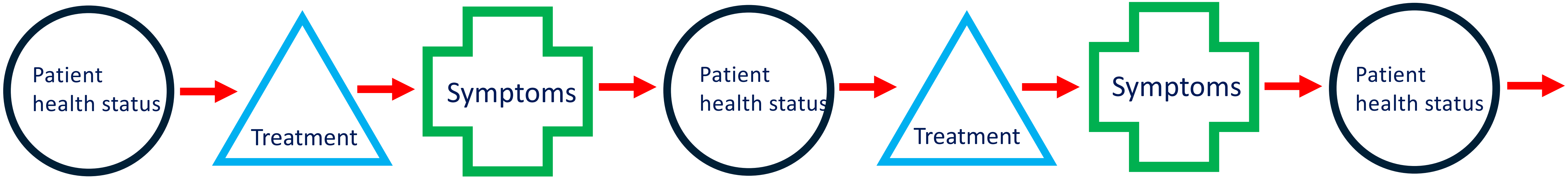


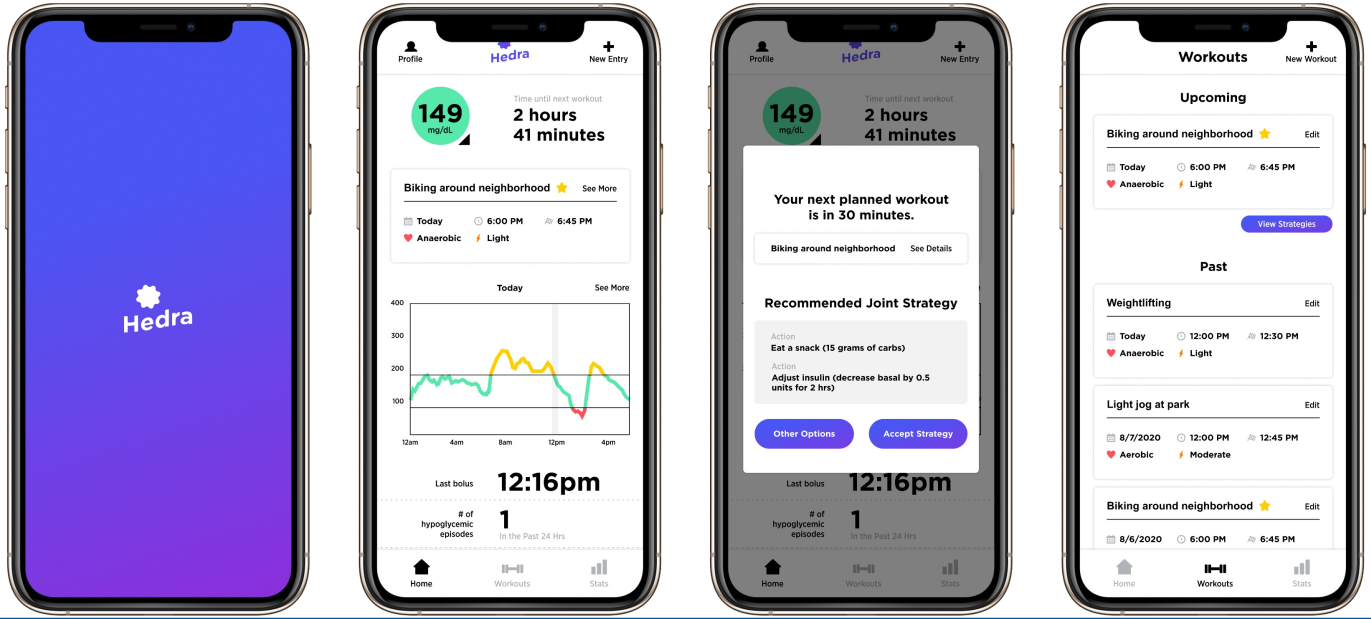
Schematic for RL: assortment selection



Duke

Schematic for RL: medical decision making



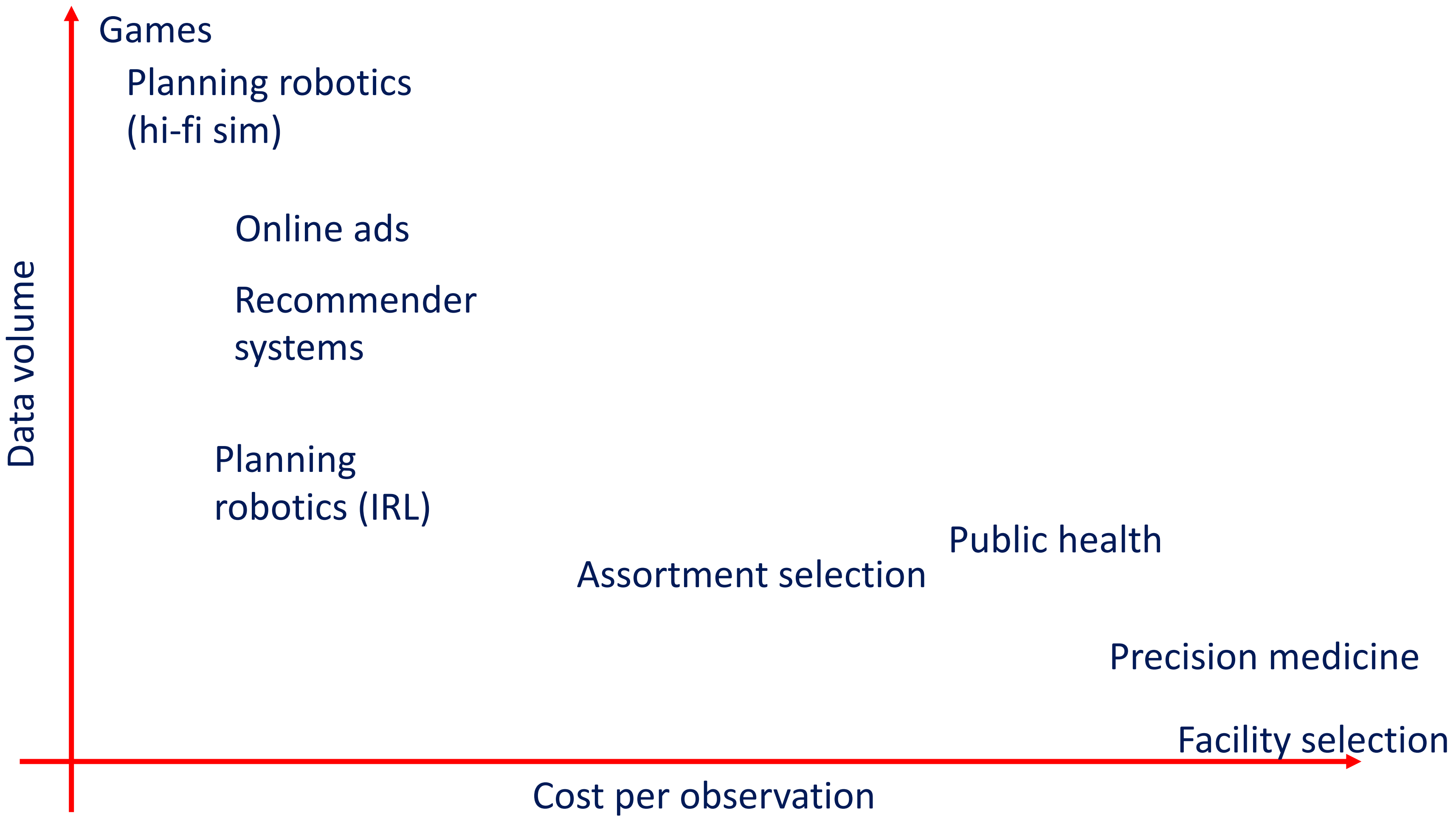


Duke

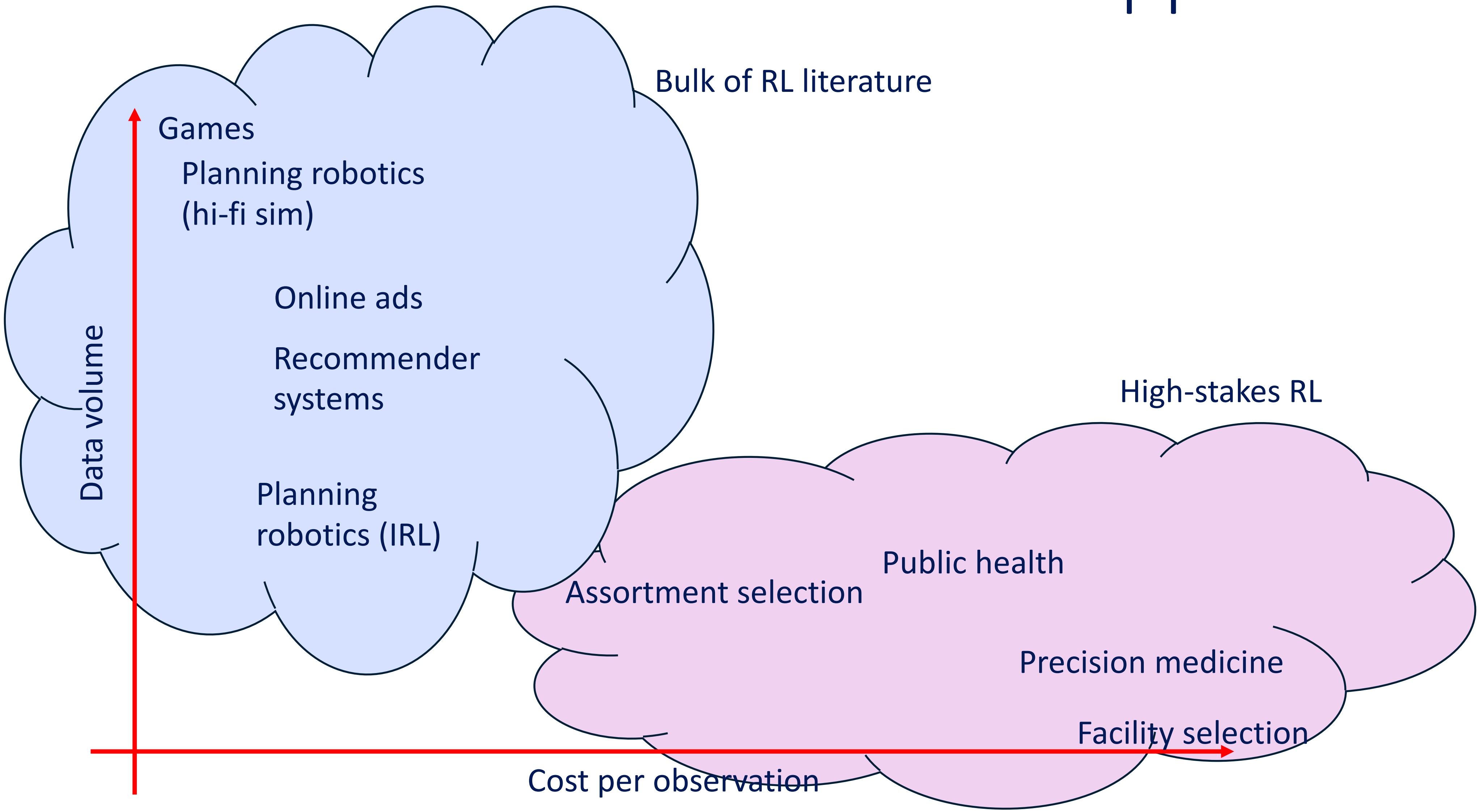
The danger of abstraction

- Nearly any repeated decision problem can be formulated as RL
- Bring existing literature to bear
  - Algorithms
  - Theory
  - Empirical benchmarks
- Heavily biased by focal applications

Cost and data volume across RL applications



Cost and data volume across RL applications



Duke

High-stakes RL

- High cost + low volume

- Exploration carries significant risk → efficiency and safety paramount
  - Every action must be justified in terms of short- and long-term benefits
  - Decisions typically on coarser time scale → large computation acceptable
  - Contrast: majority of RL algorithms focus on computational efficiency to accommodate high data throughput

- Statisticians have been thinking about these sorts of problems for a very long time [but with a slightly different objective]

Duke

Information and utility

- Every action generates information and utility

- Greedy selection: estimate utility gain for each action and pick maximizer
  - Best decision given current information [i.e., our best guess]
  - Can stagnate and fail to learn
  - Need not maximize long-term utility

- Sequential experimental design
  - Decision that yields greatest improvement in model
  - May incur high cost [e.g., poor in-trial outcomes]

Duke

Information and utility

- Every action generates information and utility

- Greedy selection: estimate utility gain for each action and pick maximizer
  - Best decision given current information [i.e., our best guess]
  - Can stagnate and fail to learn
  - Need not maximize long-term utility

- Sequential experimental design
  - Decision that yields greatest improvement in model
  - May incur high cost [e.g., poor in-trial outcomes]

Need to integrate principled experimental design into RL!

Duke

Roadmap

- Canonical sequential decision problems
- Reinforcement learning and high-stakes problems
- Example of a high-stakes RL algorithm
- Concluding discussion

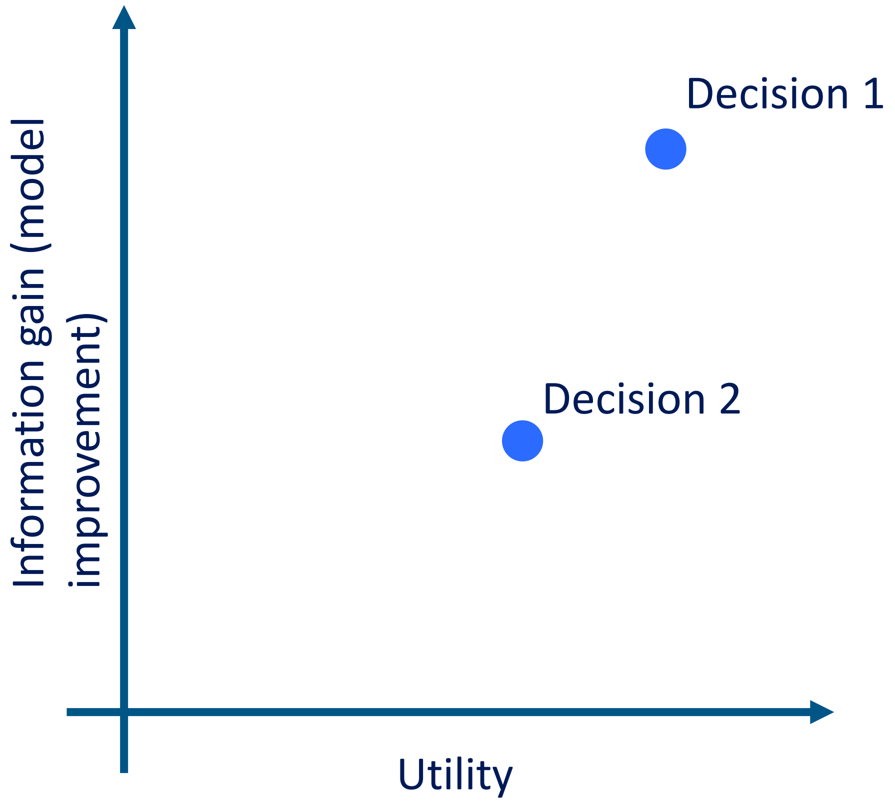
Duke

Roadmap

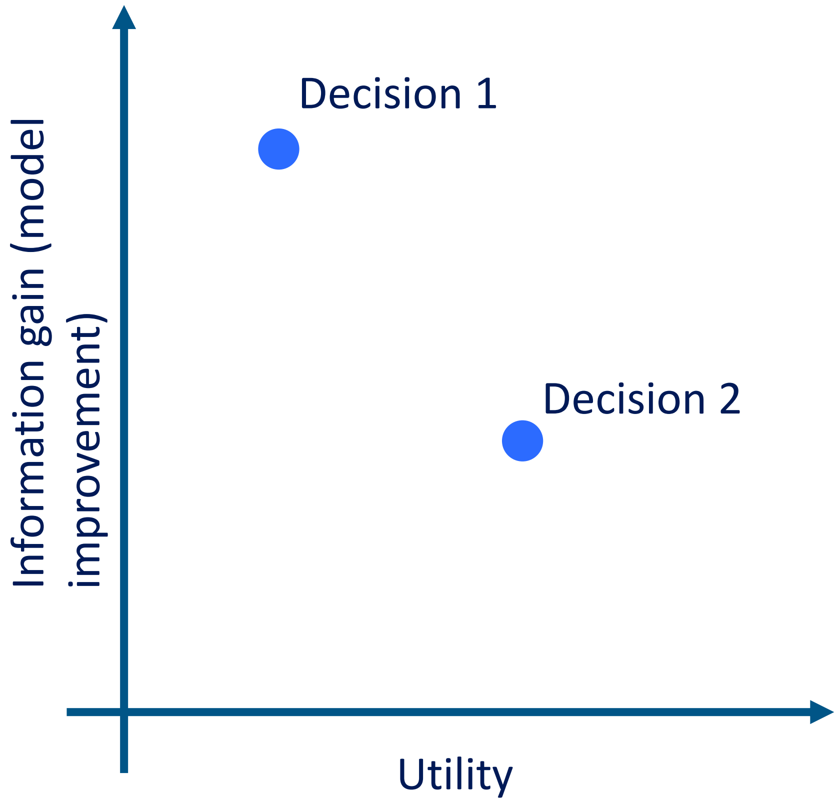
- Canonical sequential decision problems
- Reinforcement learning and high-stakes problems
- **Example of a high-stakes RL algorithm**
- Concluding discussion

Example: non-dominated selection

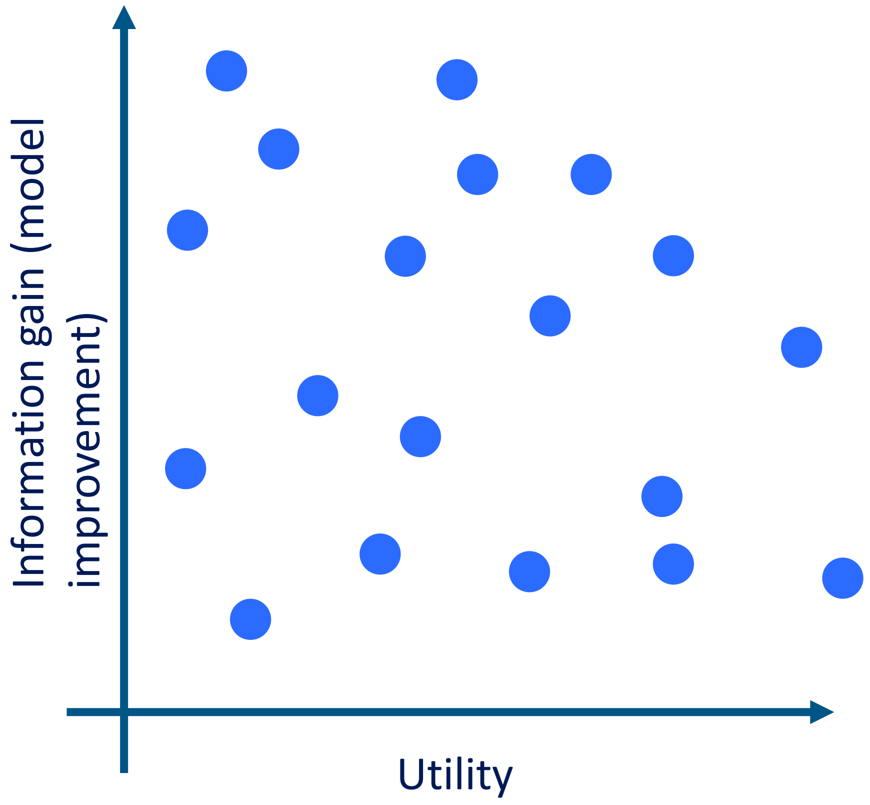
Which decision to select?



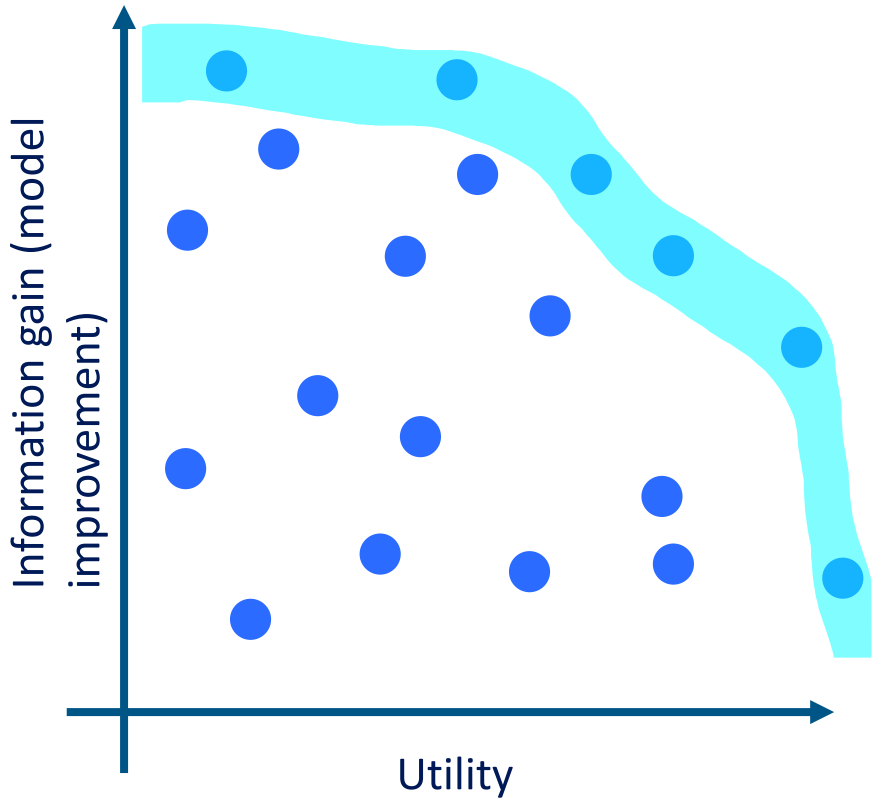
Which decision to select?



Which decision to select?



Which decision to select?



Duke

Non-dominated experiments

- An obvious conjecture

  One should never run an experiment if an alternative exists that generates more utility and more information

  - Ex., one should never prescribe a treatment that is worse for the patient being treated and that generates less information for the treatment of future patients
  - Ex., one should never recommend an ad to a customer if an alternative exists that will generate more revenue and more improvement in our forecast models

- Yet, many existing state-of-the-art online learning algorithms routinely select dominated interventions (actions)

Duke

Non-dominated selection example

- Small batched linear contextual bandit
  - Batches of size four
  - Mimic ongoing mHealth study at Duke
  - Binary treatments
  - Compare random selection, Thompson Sampling, ϵ-greedy ($\epsilon = 0.05$), and UCB
  - 1000 decision points

	Proportion of dominated selections	
Algorithm	Standard	Proposed [non-dom]
Random selection	0.82	0.52
Thompson Sampling	0.68	0.49
ϵ-greedy	0.63	0.59
UCB	0.62	0.50

Duke

Operationalizing non-dominated selection

- Posit model $\mathcal{M}_\theta$ for system under study indexed by $\theta \in \Theta$

- For every candidate action a compute

  $\mathcal{O}_\lambda(a) = \text{Expected Cumulative Utility}(a) \; + \; \lambda \, \text{Information Gain}\,(a; \theta)$

  action a is non-dominated if it maximizes $\mathcal{O}_\lambda(a)$ for some $\lambda > 0$

- Apply RL algorithm but restrict decisions to non-dominated actions

Duke

Advantages of non-dominated selection

- If RL algorithm consistent and rate optimal, so is non-dominated counterpart [Norwood et al.]
- In combinatorial problems, expected number of non-dominated points is log-order the size of the action space, e.g., $O(N^L)$ becomes $O(L \log N)$. [L. et al.]
- General framework that accommodates different measures of information gain (D-, A-, E-optimality, KL-divergence, etc.)

Duke

Discussion

Summary and future work

Summary

- RL increasingly used to inform decision making in high-cost low-volume settings [i.e., high-stakes settings]

- Exploration must be carefully considered
  - Incorporate principles from experimental design
  - Guardrails on performance
  - Limit or eliminate randomization

Future work

- Decision support tools for retail and medical applications
- Metrics for monitoring interim performance of RL
  - RL is designed to optimize long-term outcomes ←→ short-term performance may suffer, how do we reassure stakeholders?
- Other ideas? Let us know!

Thank you!

Please reach out if you have questions, suggestions, or want to team up!

eric.laber@duke.edu
laber-labs.com