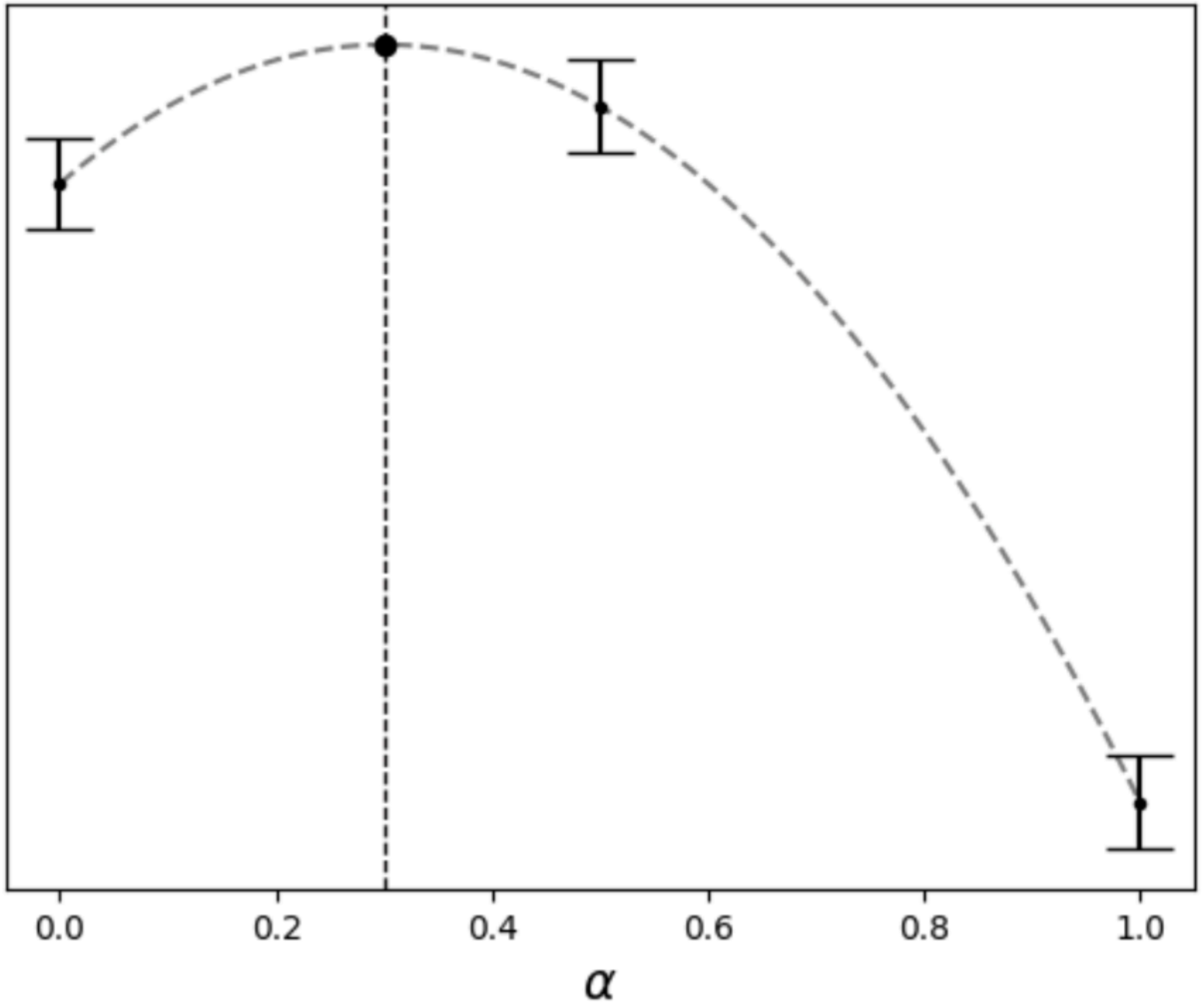
Module 12: Bayesian Optimization

DAV-6300-1: Experimental Optimization

David Sweet // 20241205

Review: Response Surface Methodology

- Surrogate: Model (regression)
  - Maps parameters, x, to measurements, y
- Analogy
  - $E[BM]$ is to observation y
  - as response function, $f(x)$, is to surrogate, $y(x)$

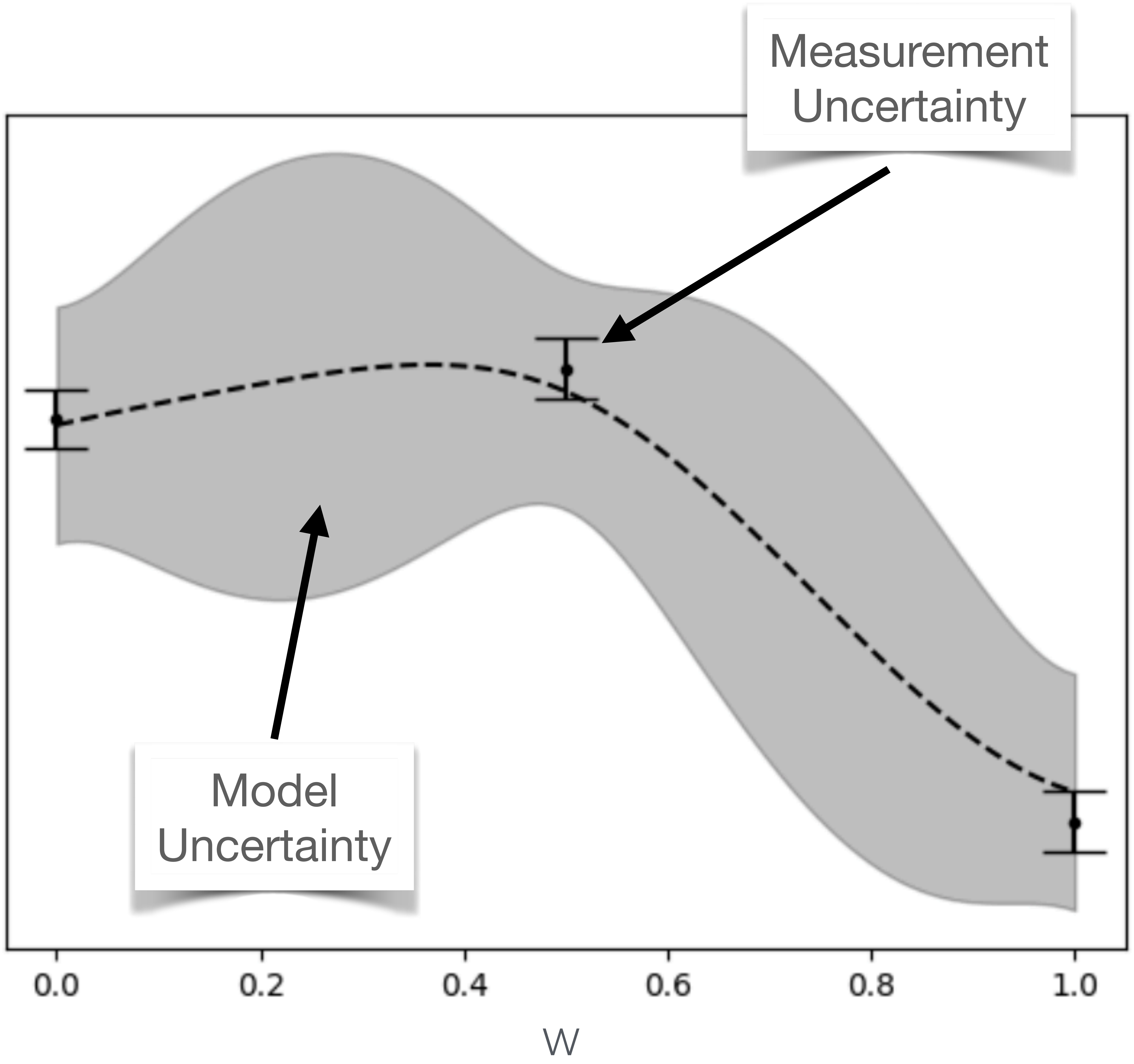


Review: Gaussian process regression

- $y(x) \sim \mathcal{N}(\mu(x), se^2(x))$
- Kernel function: $k(x, x') = e^{-d(x,x')^2/(2s^2)}$

$$\mu(x) = K_x^T (K_{xx} + se_0^2 I)^{-1} \mathbf{y}$$

$$se^2(x) = 1 - K_x^T (K_{xx} + se_0^2 I)^{-1} K_x$$



Review: Multi-armed bandits

- Model observations so far:
  - ε-greedy: μ_a
  - TS: $y_a \sim \mathcal{N}(\mu_a, se_a^2)$

$$\mu_a = \sum_i y_{a,i}/N$$

$$se_a = \left[\sqrt{\sum_i (y_{a,i} - \mu_a)^2/N}\right]/\sqrt{N}$$

- Select arm:
  - ε-greedy: 90% $\arg\max_a \mu_{a'}$ 10% random
  - TS:
    - Draw $m_a \sim \mathcal{N}(\mu_a, se_a^2)$
    - Arm $\arg\max_a m_a$

Bayesian optimization

- Combine surrogate with MAB arm-selection
- Use GPR as surrogate
- Like RSM+MAB+ML

TS: MAB vs. BO

- Model observations
  - MAB: $m_a \sim \mathcal{N}(\mu_a, se_a^2)$
  - BO: $y(x) \sim \mathcal{N}(\mu(x), se^2(x))$
- Select arm
  - MAB: $\arg\max_a m_a$
  - BO: $\arg\max_x y(x)$

- m_a is a sampled value from $\mathcal{N}(\mu_a, se_a^2)$
- $y(x)$ is a sampled *function* from $\mathcal{N}(\mu(x), se^2(x))$

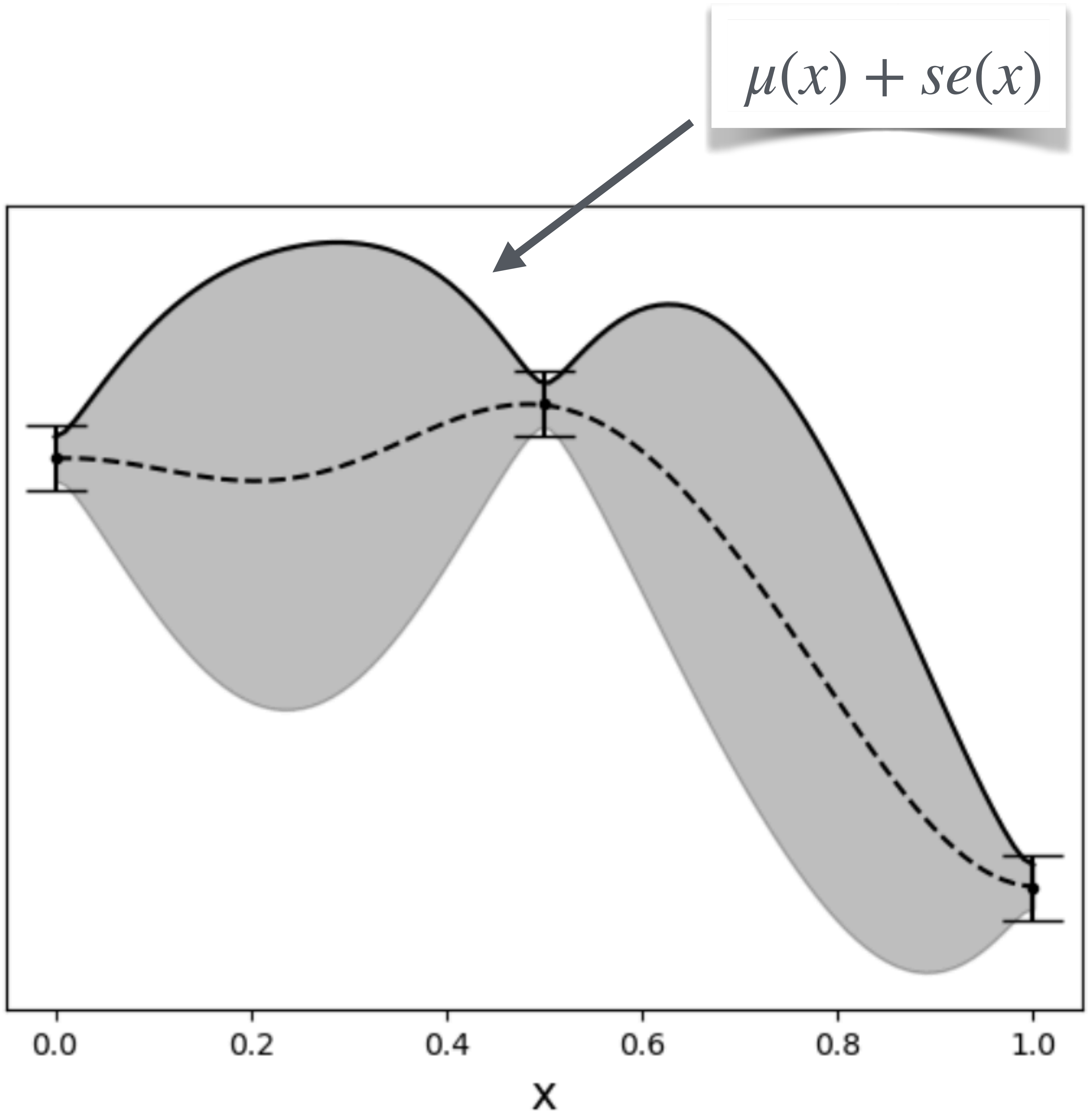
Acquisition functions

Hard to sample function, $y(x)$

- In practice, $\arg\max_x \mathcal{N}(\mu(x), se^2(x))$ is difficult
  - x is continuous, so infinite arms to check
  - Recall MAB's TS had finite arms (ex., $K = 3$)
- Instead, use heuristic: *acquisition function*
- Common ones: UCB and EI

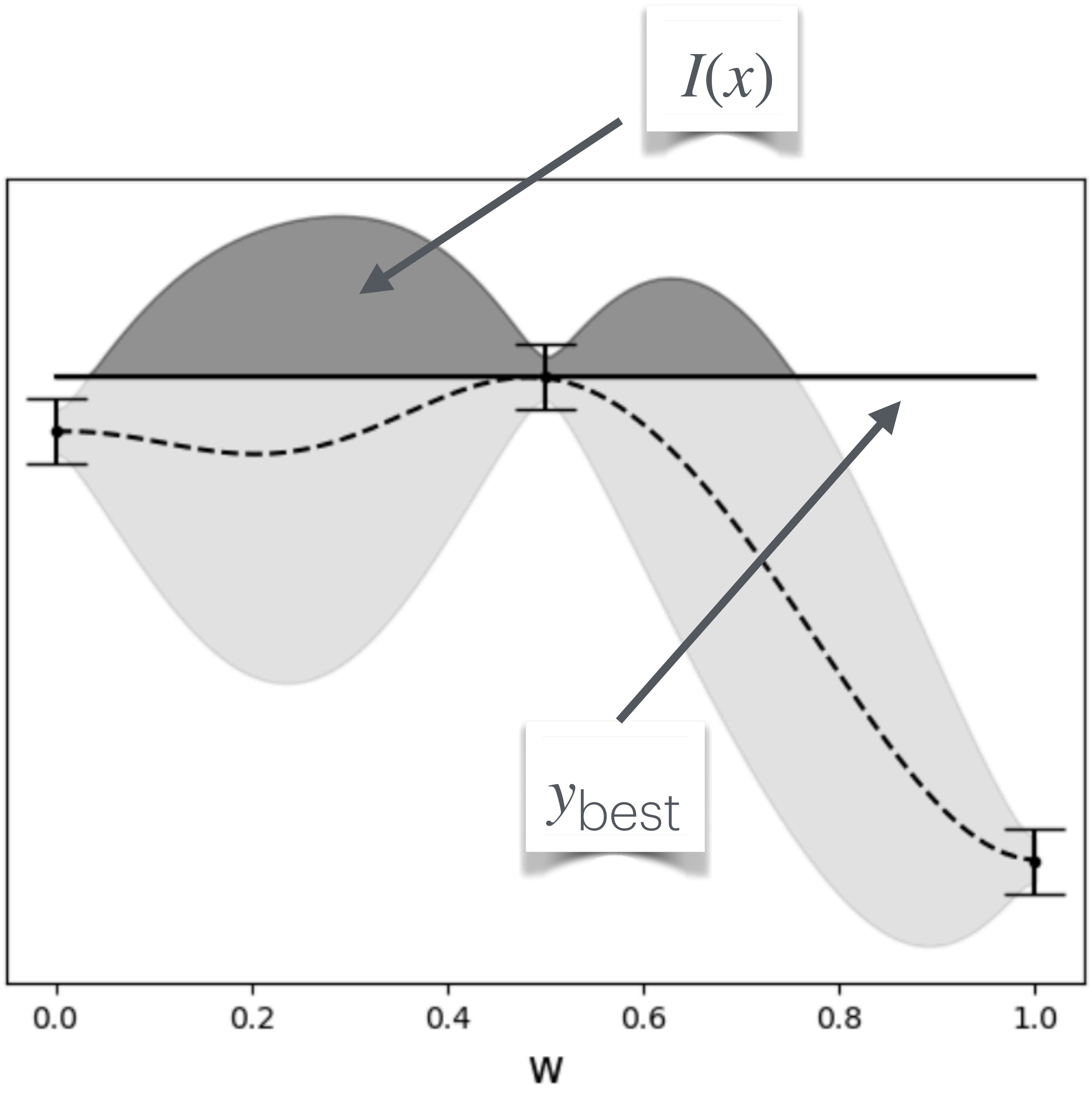
AF: Upper Confidence Bound

- Upper Confidence Bound (UCB):
  $\arg\max_{x} \mu(x) + se(x)$
- IOW: Maximize $\mu(x) + se(x)$ over x
- Exploration: $se(x)$
- Exploitation: $\mu(x)$
- Balancing: $+$



AF: Expected Improvement

- Expected Improvement (EI)
  - Improvement: $I(x) = max(0, y(x) - y_{\text{best}})$
  - Expected improvement:
    $E[I(x)] = \int I(x)\phi(y(x))dy(x)$



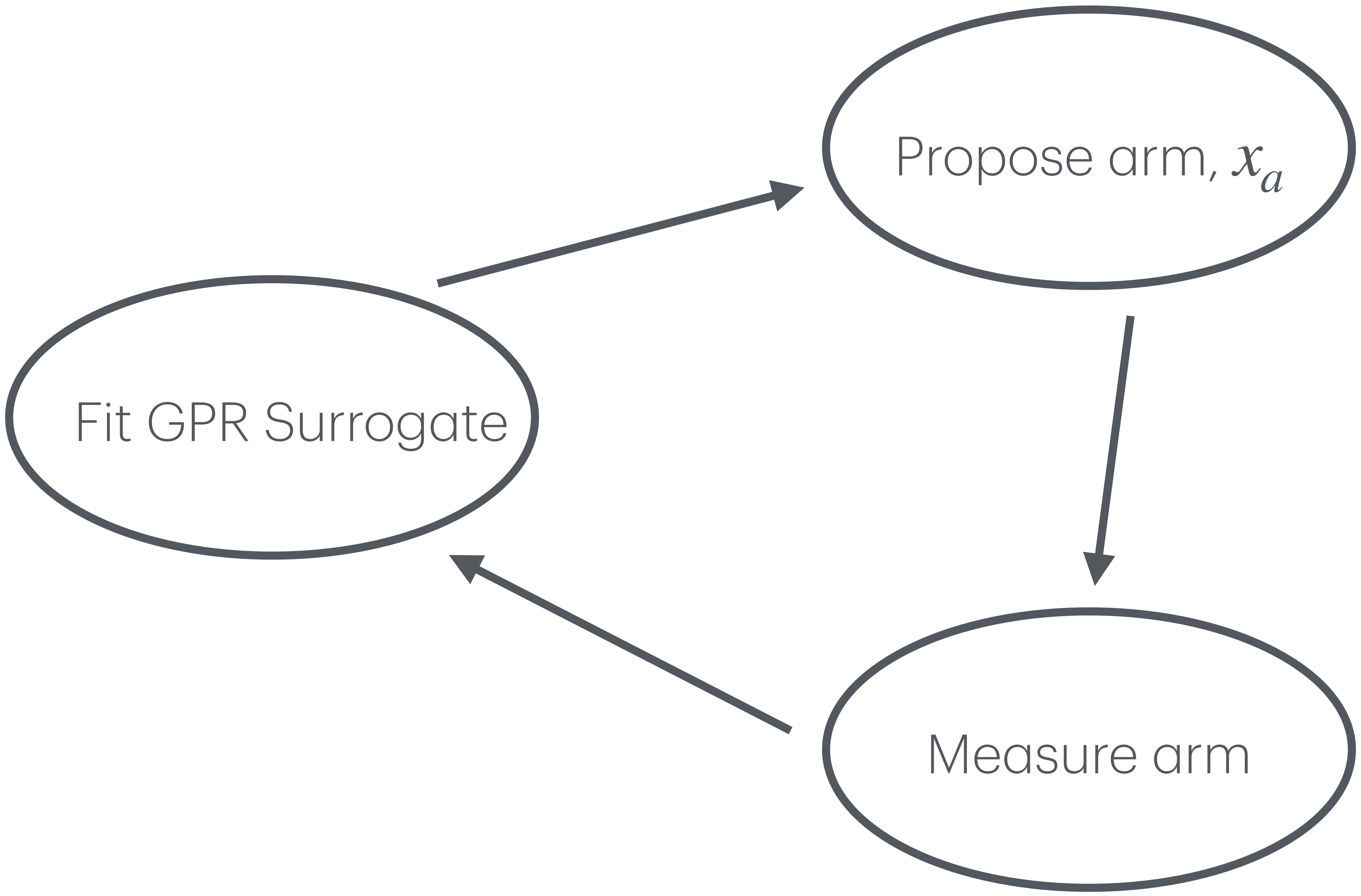
Proposing Arms

- Optimize the acquisition function
  - $x_a = \arg\max \text{acqf}(x)$
  - where $\text{acqf}(x)$ is, e.g., UCB or EI
  - x_a is the next arm to measure
- BO designs your next experiment
- aka, BO “proposes” or “suggests” arms

Proposing Arms

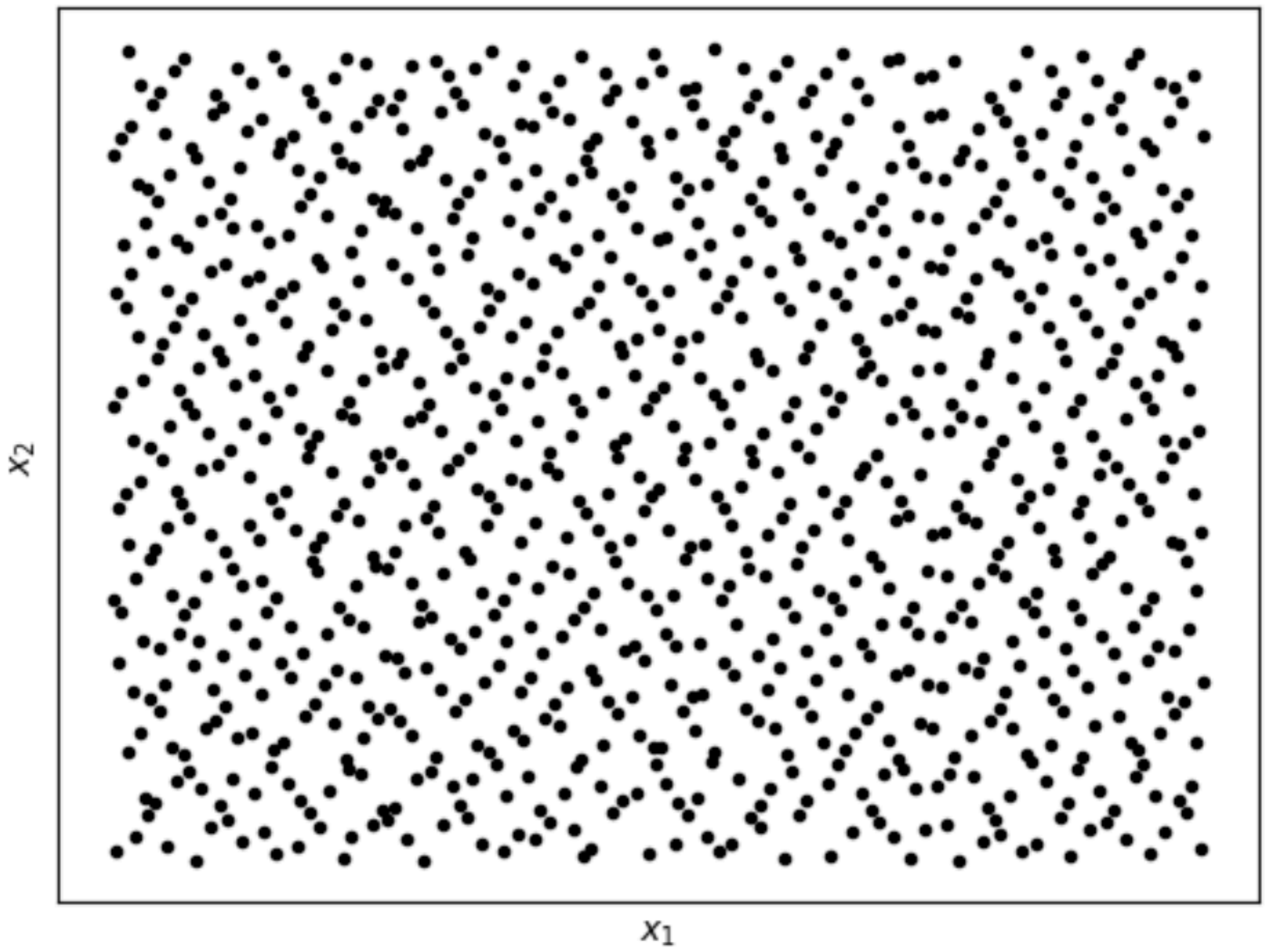
- This can be hard: $x_a = \arg\max \mathrm{acqf}(x)$
- Optimizers:
  - `scipy.optimize.minimize` — BFGS, multi-start, uses gradient
  - CMA-ES - gradient-free, repeated sampling
    - `pip install pycma`

Bayesian Optimization



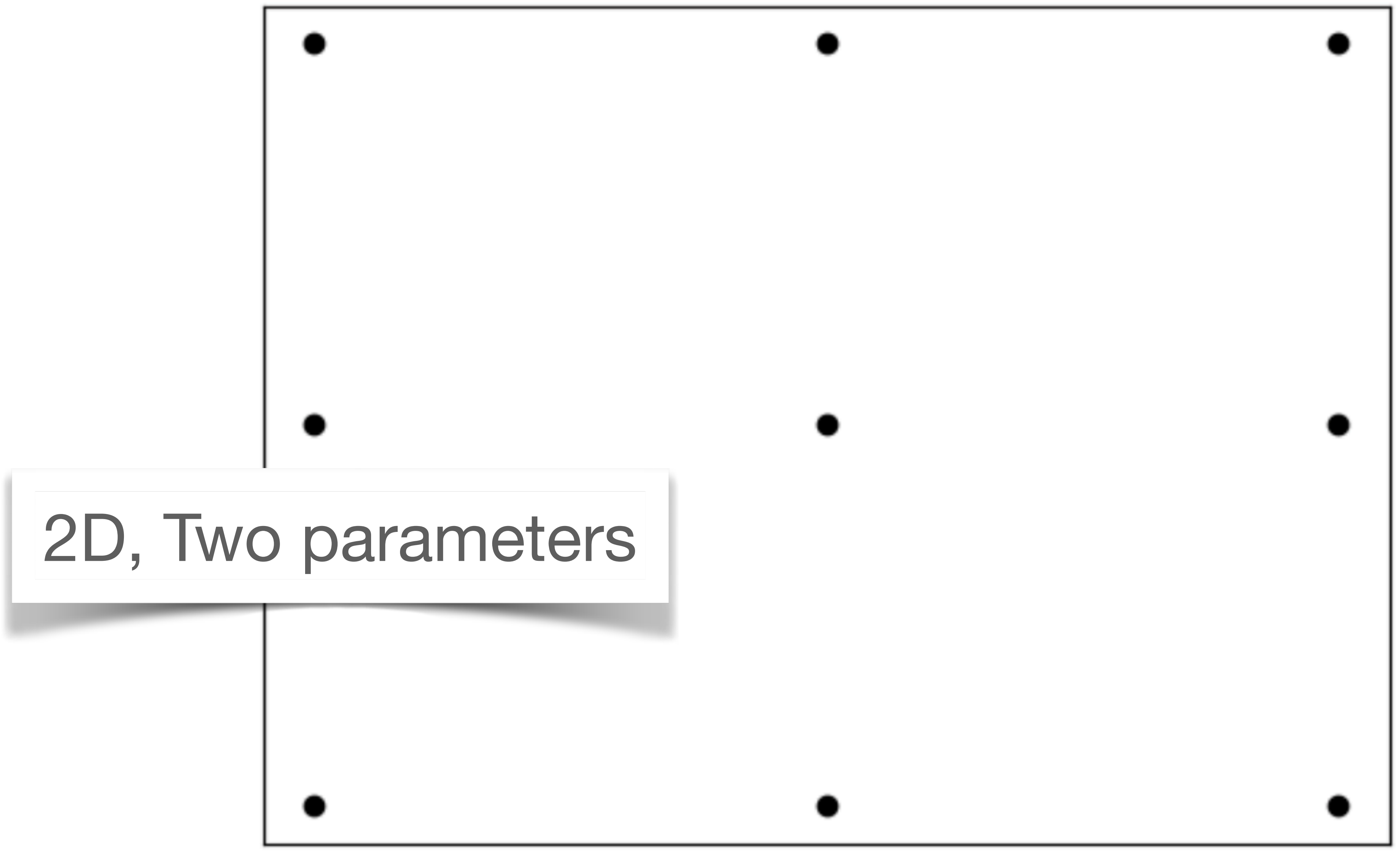
Thompson Sampling?

- Naive approach:
  - Randomly sample many x's
  - Many, MANY x's
- Two problems:
  - GP's slow to sample many x's
  - Curse of dimensionality



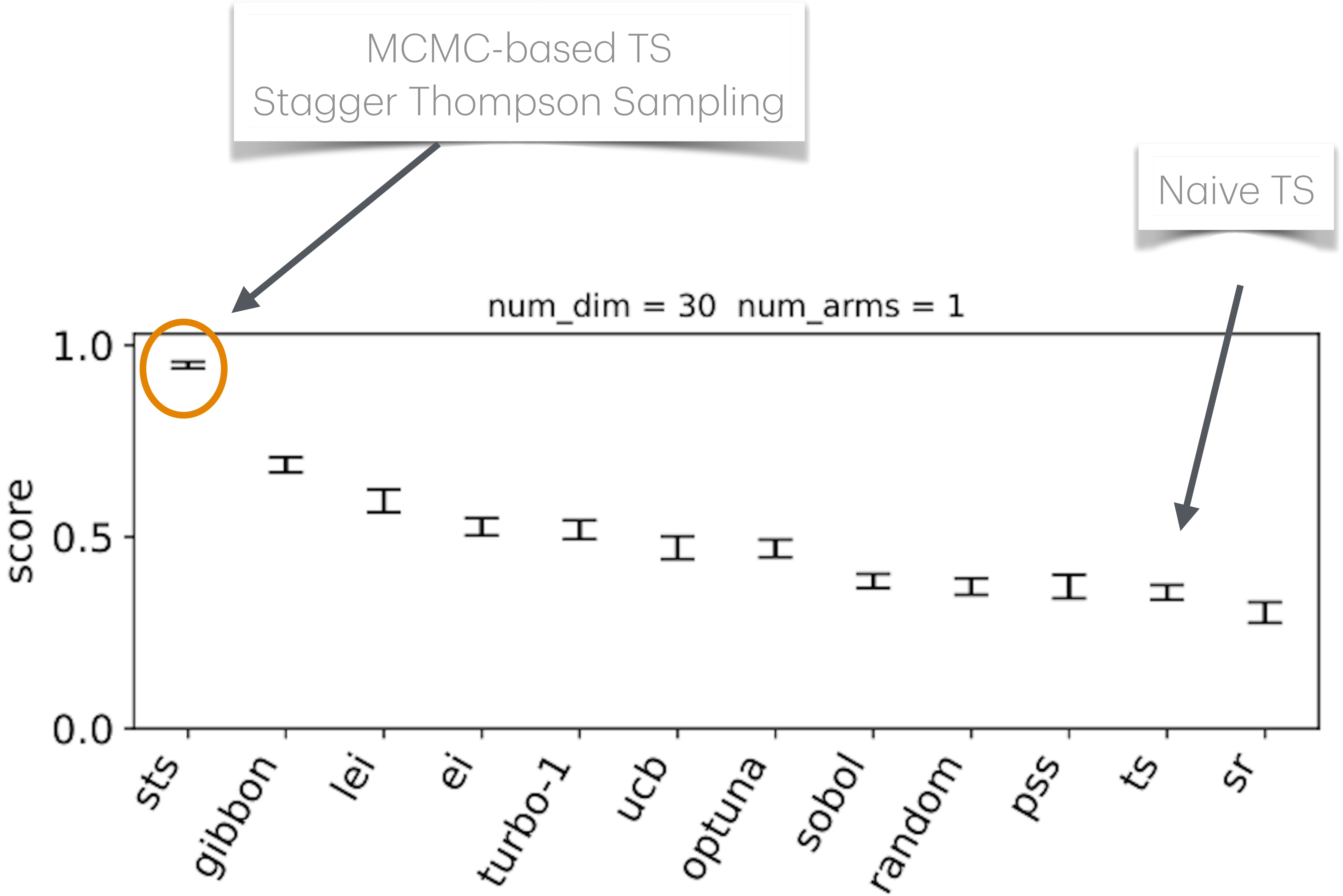
Aside: Curse of dimensionality

- Spread points out in parameter space
  - 1D: 3 points
  - 2D: $3 \times 3 = 3^2 = 9$ points
  - 3D: $3 \times 3 \times 3 = 3^3 = 27$ points
  - ... dD: $3^d =$ too many points



Thompson Sampling!

- Clever sampling can break the curse of dimensionality
  - Markov Chain Monte Carlo (MCMC) (advanced topic)
- MCMC makes Thompson sampling work great

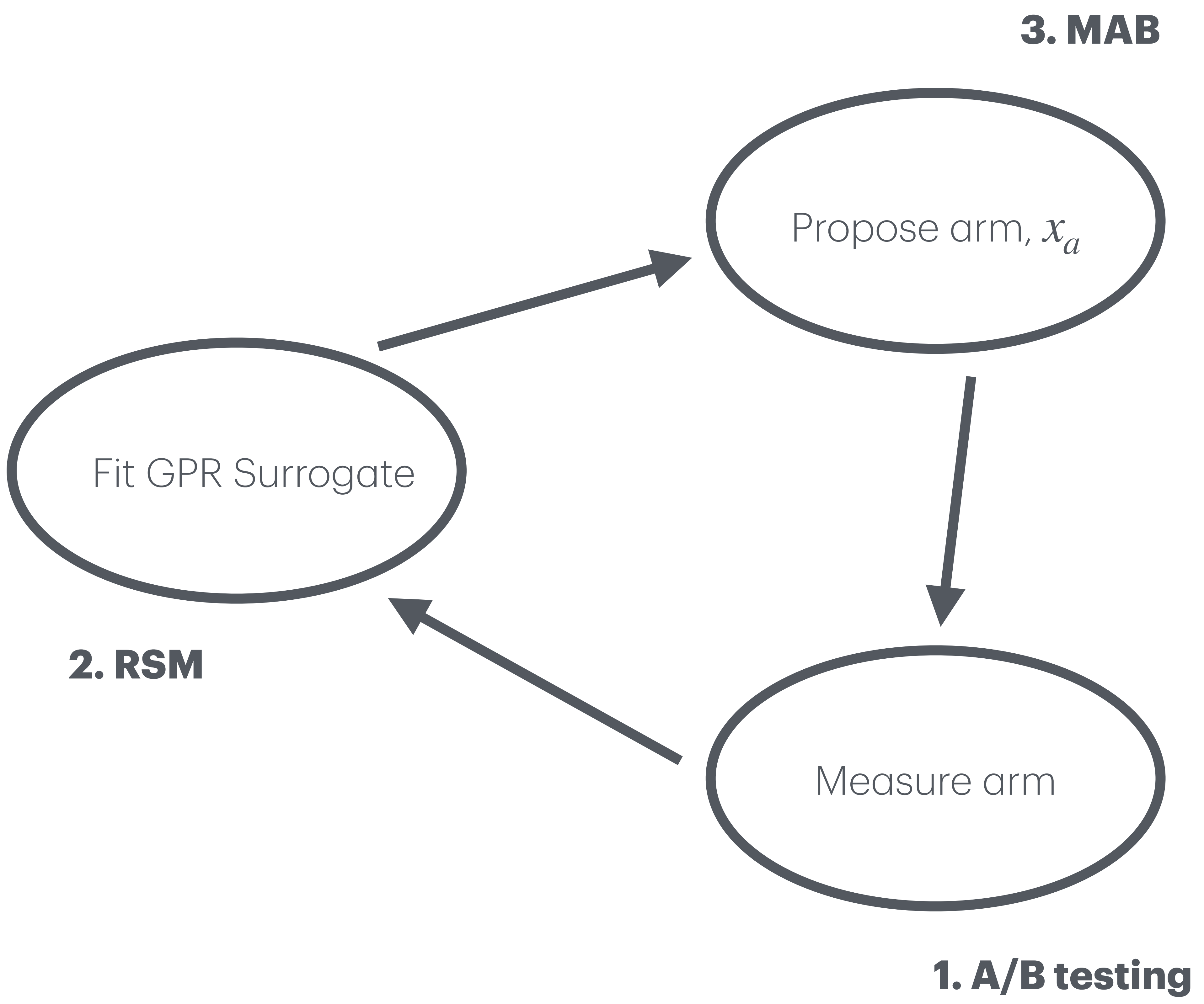


BO: Connections

1. A/B testing: Take a low-se, low-bias measurement

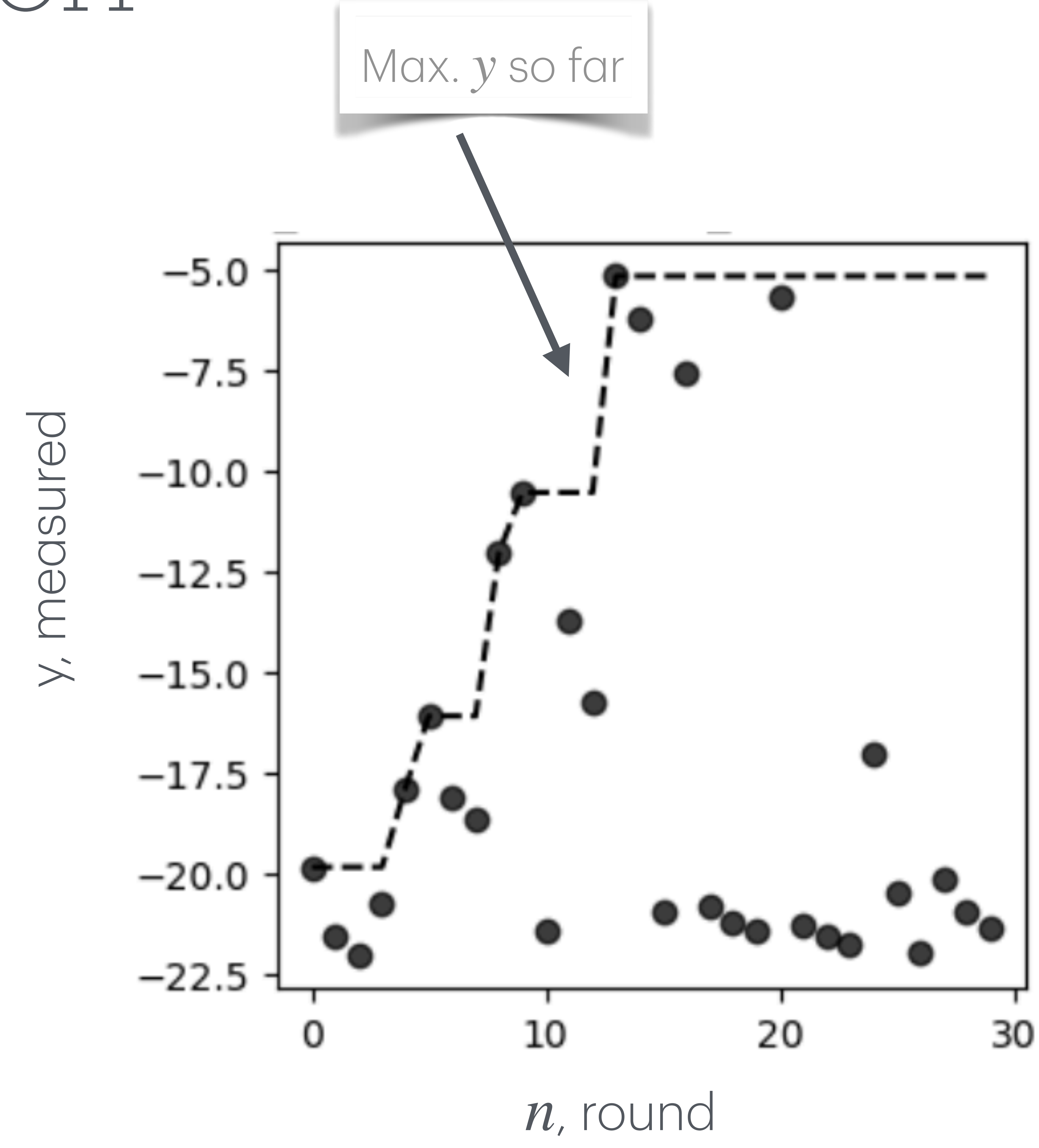
2. RSM: Build a surrogate

3. MAB: Balance exploration & exploitation in design



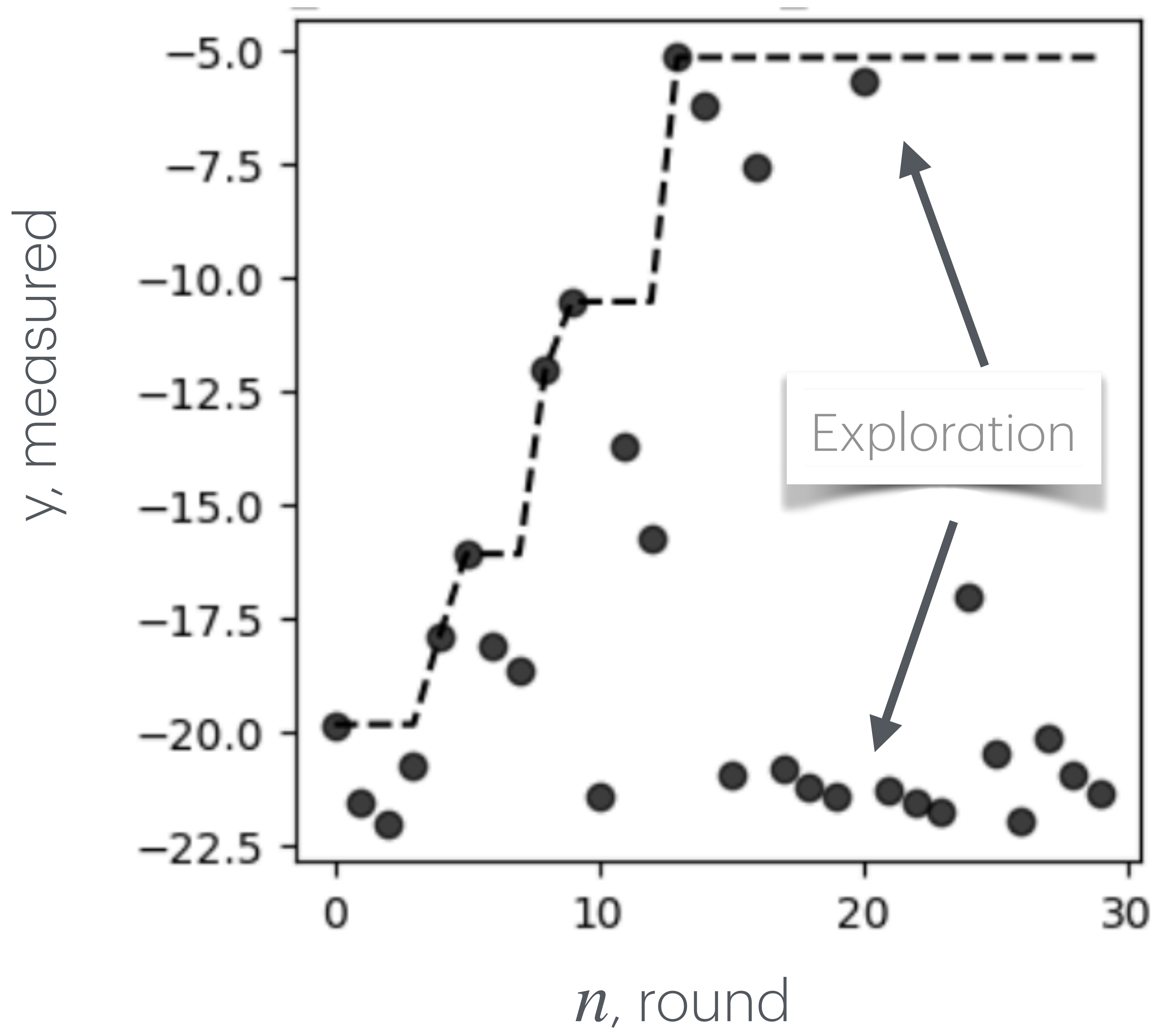
Bayesian Optimization

- BO may search for maximum
  - *simple regret*
  - aka, *instantaneous regret*
- Typical visualization:
  - Measured y vs. rounds, aka iterations
  - Track max y so far



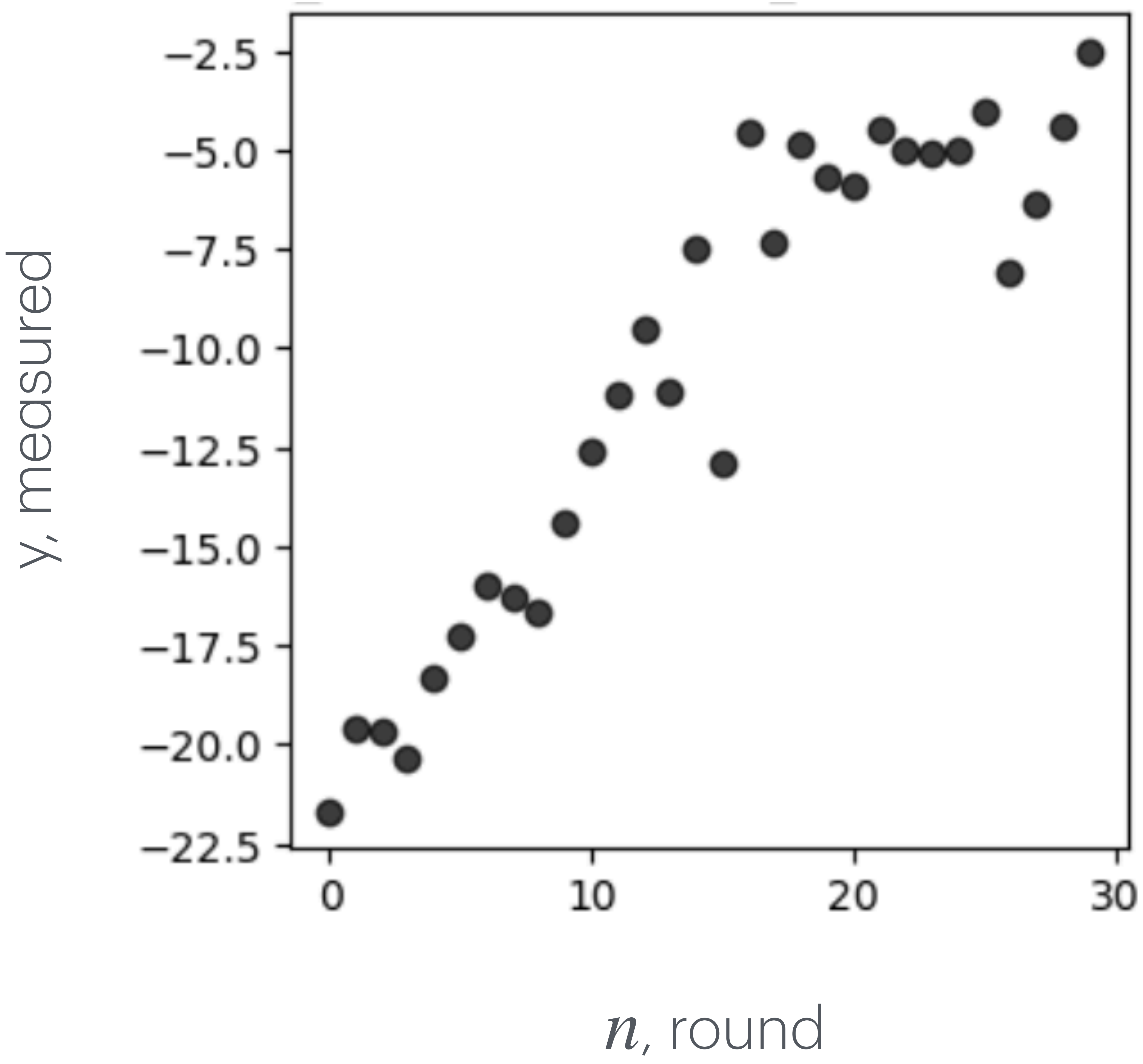
Bayesian Optimization

- Simple regret
  - $y_* - y_n$
  - y_* is (unknown) max
  - Lots of exploration, low y ok
  - Only care about max
- Black-box optimization (BBO)



Bayesian Optimization

- *Cumulative regret*
  - $\sum_{n}^{N} (y_* - y_n)$
  - y_* is (unknown) max
  - Care about every y_n; less "wild" exploration
  - Maximize BM while experimenting



BO Features & Advances

- Categorical, ordinal, and continuous variables
  - Also: images, molecules (probably any data object)
- Scales to hundreds of parameters
- Multiple metrics
- Parameter constraints
- Metric constraints

- Combine experiment with simulation
- Design experiments for multiple arms to be simultaneously measured
- Take measurements asynchronously
- Use all available measurements (whether proposed by BO or not)
- Account for context (which we can't control)

BO Applications

- Recommender systems
- Online advertising
- Quantitative trading
- Hyperparameter optimization
- Hardware configuration/design
- Free-electron laser tuning
- Design of new materials
- Drug discovery
- Gene design
- EV charging schedules
- Robotic cooking
- Self-driving cars