Variational Bayesian Optimal Experimental Design

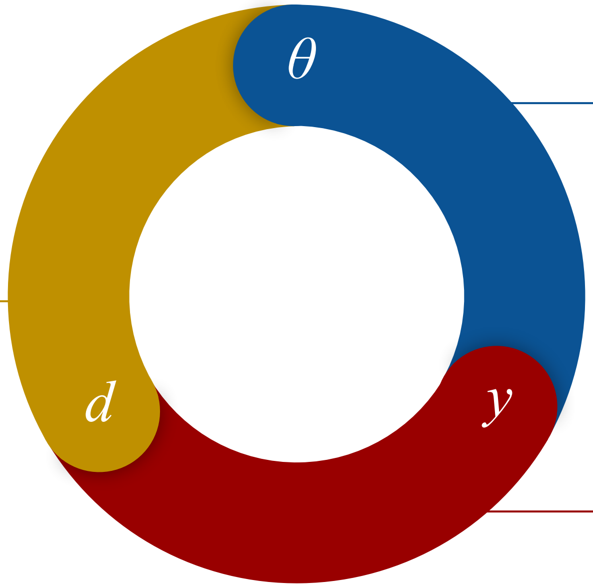
Adam Foster[†] **Martin Jankowiak[‡]** **Eli Bingham[‡]** **Paul Horsfall[‡]**
Yee Whye Teh[†] **Tom Rainforth[†]** **Noah D. Goodman[‡§]**

[†]Oxford Statistics, [‡]Uber AI, [§]Stanford

Spotlight, NeurIPS 2019

539,137
0.316
1.720
0.314
778,186
48,100
1.190
0.314
1.180
68,000
833,789
0.338
1.190
0.332
158,294
10,000
0.479
0.332
0.460
350,000
10,000
7.500

Adaptive experimentation



Inference

Data analyzed
Model fitted

Design

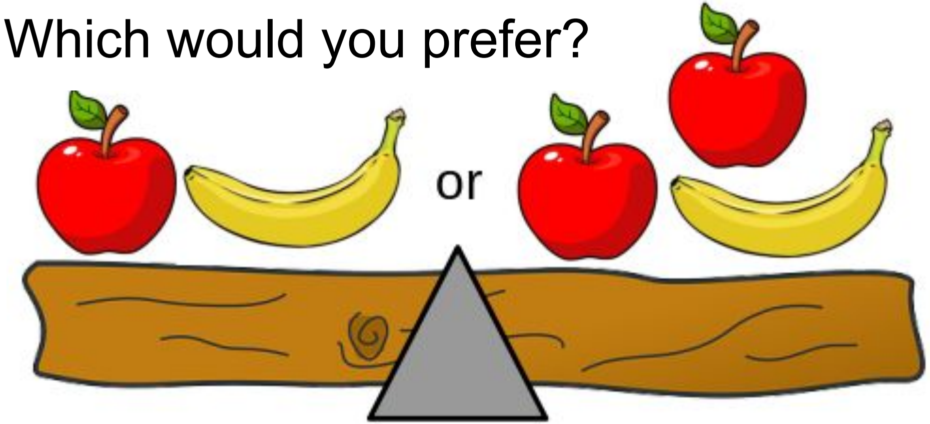
Experimental setup
Controlled by experimenter

Observation

Data generated
Response sampled

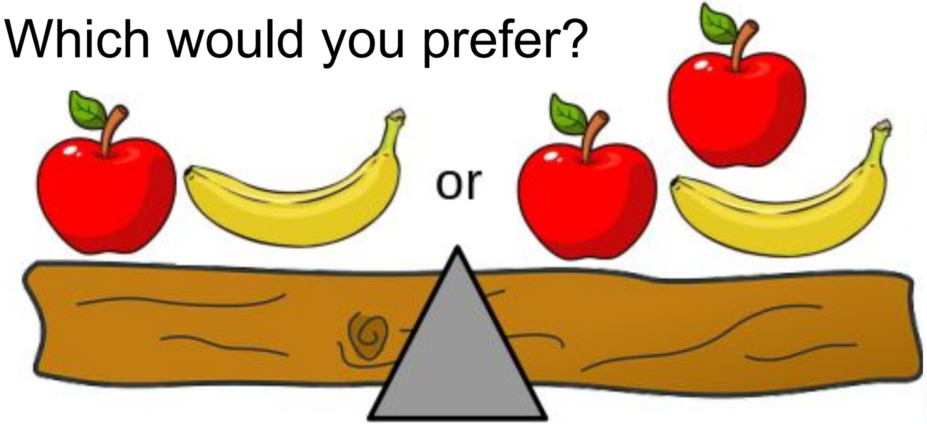
What makes a good experiment?

Which would you prefer?

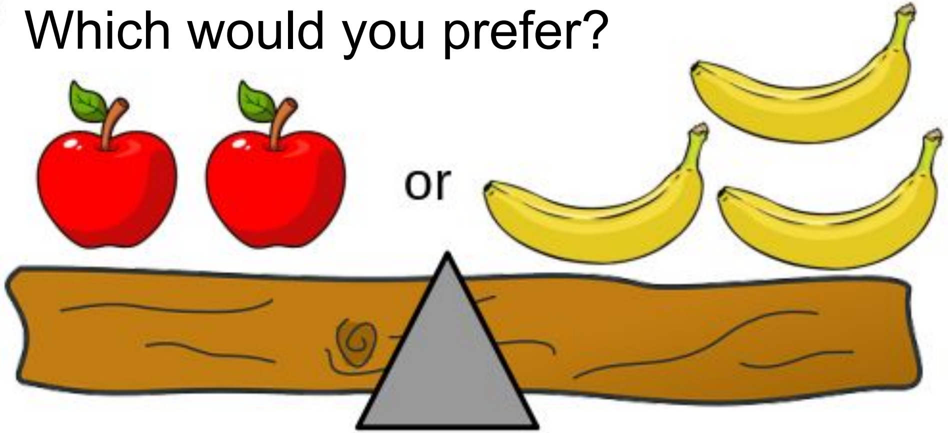


What makes a good experiment?

Which would you prefer?

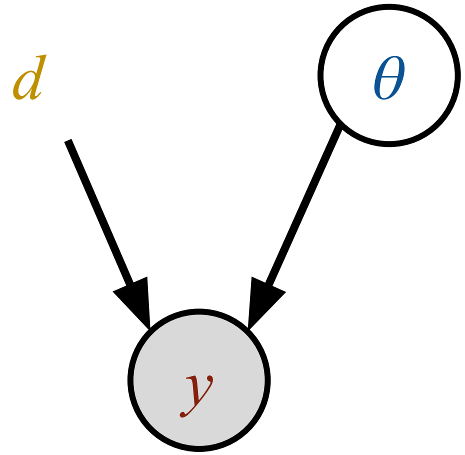


Which would you prefer?

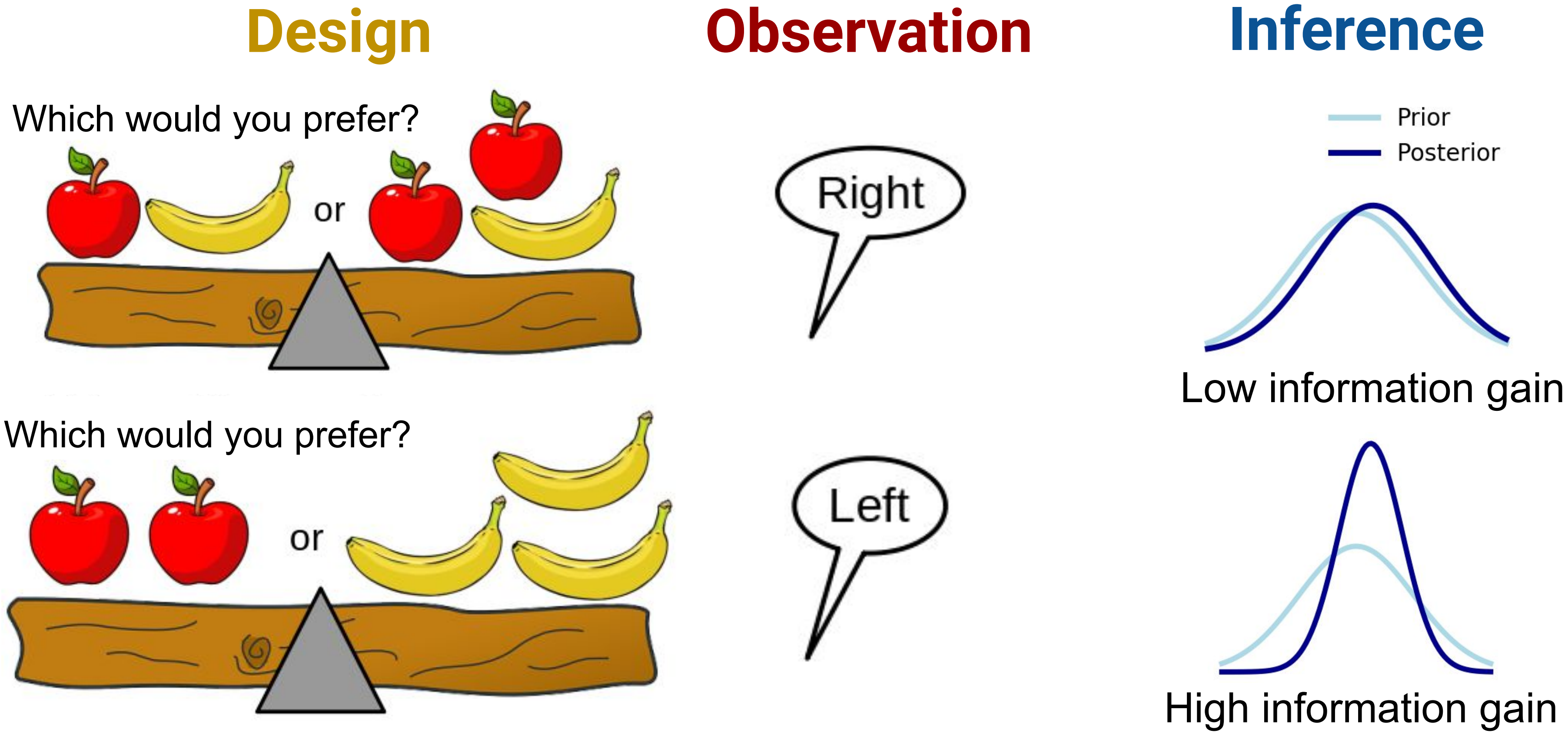


Bayesian experimental design

θ : **latent variable** of interest d : **design** y : **data**



$$\underbrace{p(\theta|y,d)}_{\text{posterior}} \propto \underbrace{p(\theta)}_{\text{prior}}\,\underbrace{p(y|\theta,d)}_{\text{likelihood}}$$



Design
Observation
Inference
Which would you prefer?
or
Right
Prior
Posterior
Low information gain
Which would you prefer?
or
Left
High information gain

Expected information gain (EIG)

Expected reduction in **entropy** from the **prior** to the **posterior**

prior entropy posterior entropy

$$\mathrm{EIG}(d) = \mathbb{E}_{p(y|d)}\{H[p(\theta)] - H[p(\theta|y,d)]\}$$

(Lindley, 1956)

Estimating the EIG is difficult!

$$\mathrm{EIG}(d) = \mathbb{E}_{p(y|d)}\{H[p(\theta)] - H[p(\theta|y,d)]\}$$

✓ simulate samples

✓ prior

✗ posterior

"Doubly intractable"

Our contribution: Variational estimators of the EIG

- Bound EIG to turn estimation into optimization
- This removes double intractability

$$\text{EIG}(d) \leq \mathbb{E}_{p(y,\theta|d)} \left\{ \log \frac{p(y|\theta, d)}{\boxed{q(y|d)}} \right\}$$

approximate marginal density

Variational estimator		Implicit?	Consistent?
Marginal	$\mathrm{EIG}(d) \leq \mathbb{E}_{p(\theta)p(y\|\theta,d)}\left[\log \frac{p(y\|\theta,d)}{q_m(y\|d,\phi)}\right]$	✗	✗
Posterior	$\mathrm{EIG}(d) \geq \mathbb{E}_{p(\theta)p(y\|\theta,d)}\left[\log \frac{q_p(\theta\|y,d,\phi)}{p(\theta)}\right]$	✓	✗
Variational NMC	$\mathrm{EIG}(d) \leq \mathbb{E}_{p(\theta_0)p(y\|\theta_0,d)q_v(\theta_{1:M}\|y,\phi)}\left[\log \frac{p(y\|\theta_0,d)}{\frac{1}{M}\sum_{m=1}^{M}\frac{p(\theta_m)p(y\|\theta_m,d)}{q_v(\theta_m\|y,d,\phi)}}\right]$	✗	✓
Marginal + likelihood	$\mathrm{EIG}(d) \approx \mathbb{E}_{p(\theta)p(y\|\theta,d)}\left[\log \frac{q_\ell(y\|\theta,d,\psi)}{q_m(y\|d,\phi)}\right]$	✓	✗

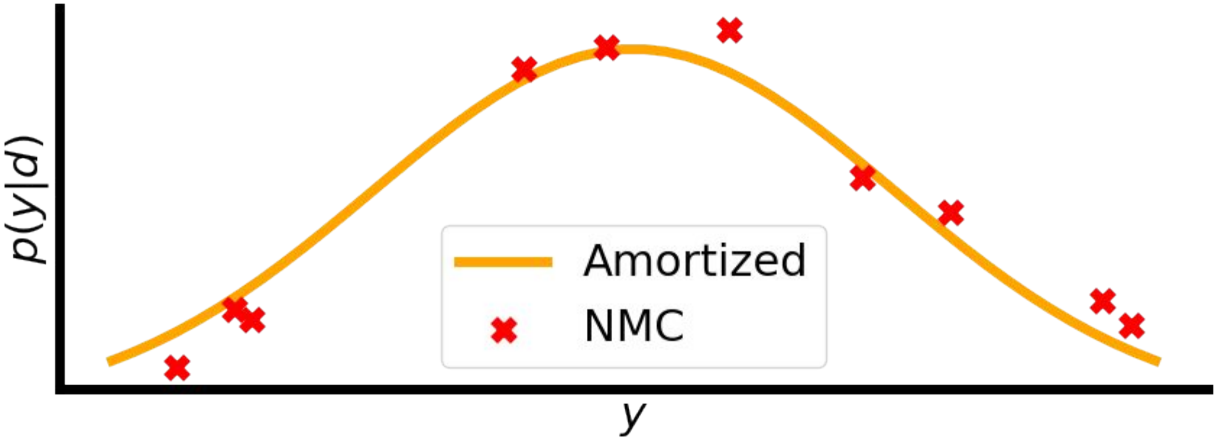
Much faster convergence rates!

Variational rate	$\mathcal{O}(T^{-1/2})$
Nested Monte Carlo rate	$\mathcal{O}(T^{-1/3})$

T = computational cost

Intuition: amortization

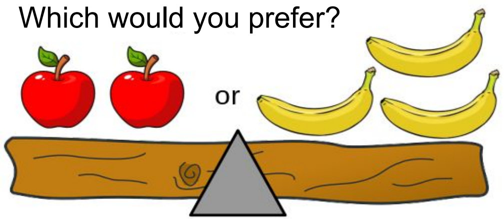
- Approximate the **functional form** rather than computing independent **point estimates**



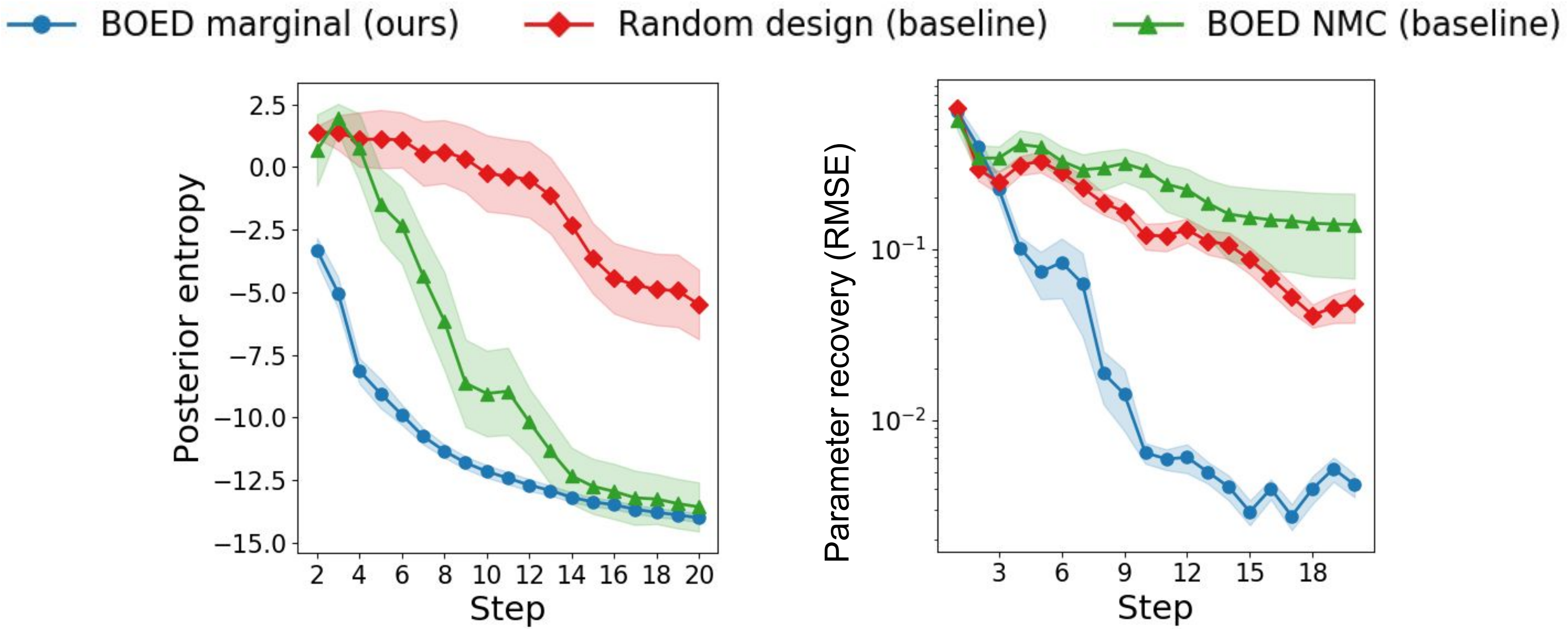
NMC = Nested Monte Carlo

Experiments: EIG estimation accuracy

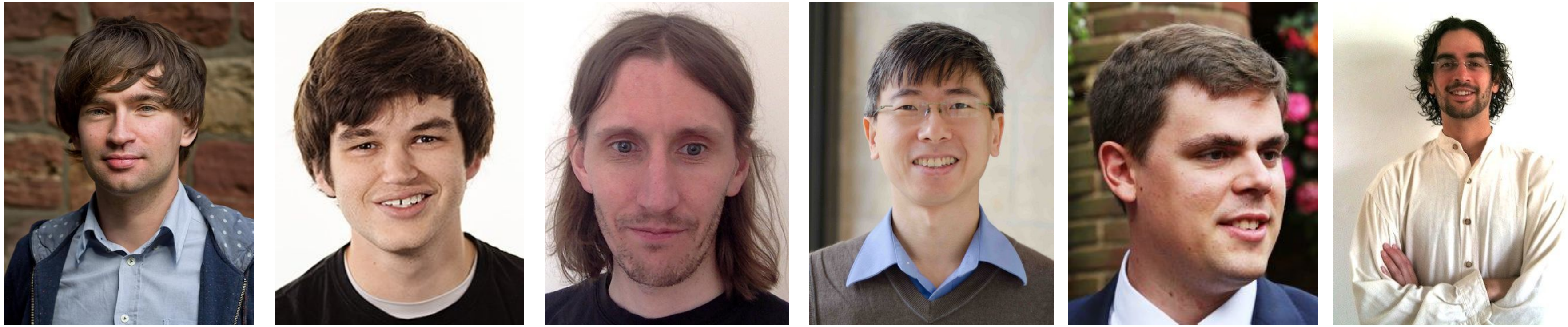
		Preference		Mixed effects		Extrapolation	
		$Bias^2$	Var	$Bias^2$	Var	$Bias^2$	Var
Ours	Posterior	4.26×10^{-2}	8.53×10^{-3}	2.21×10^{-3}	2.70×10^{-3}	1.24×10^{-4}	4.11×10^{-5}
	Marginal	$\mathbf{1.10\times10^{-3}}$	$\mathbf{1.99\times10^{-3}}$	n/a	n/a	n/a	n/a
	VNMC	4.17×10^{-3}	9.04×10^{-3}	n/a	n/a	n/a	n/a
	Marginal + likelihood	n/a	n/a	$\mathbf{3.05\times10^{-3}}$	$\mathbf{7.72\times10^{-5}}$	$\mathbf{6.90\times10^{-6}}$	$\mathbf{1.84\times10^{-5}}$
Baseline	NMC	7.60×10^{-2}	8.36×10^{-2}	n/a	n/a	n/a	n/a
	Laplace	8.42×10^{-2}	9.70×10^{-2}	n/a	n/a	n/a	n/a
	LFIRE	1.30×10^{-1}	1.41×10^{-2}	9.66×10^{-2}	7.69×10^{-2}	n/a	n/a
	DV	9.23×10^{-2}	8.07×10^{-3}	7.19×10^{-3}	6.76×10^{-4}	7.84×10^{-6}	4.11×10^{-5}



Experiments: End-to-end adaptive experimentation



Thank you



UNIVERSITY OF OXFORD

Uber AI

Stanford University

Implementation in Pyro docs.pyro.ai/en/stable/contrib.oed.html

Full paper papers.nips.cc/paper/9553-variational-bayesian-optimal-experimental-design.pdf