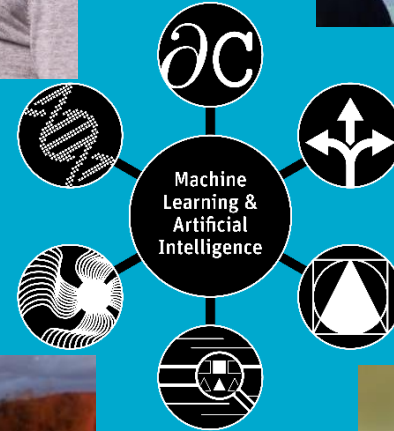




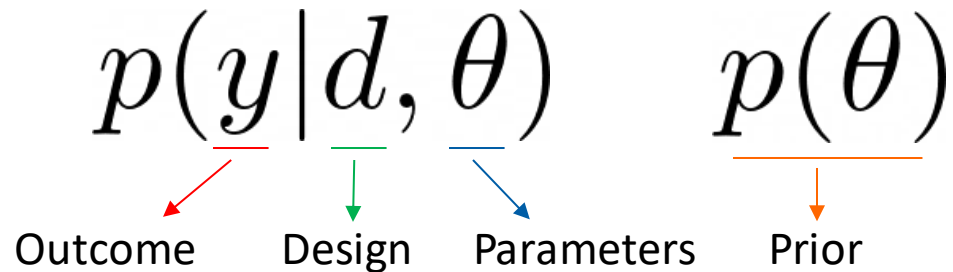
# Optimizing Sequential Experiment Design with Reinforcement Learning

**Tom Blau, Edwin V. Bonilla, Iadine Chades, Amir Dezfouli**



## What does it mean to design experiments?

- Given some prior model



## What does it mean to design experiments?

- Given some prior model

$$p(y|d, \theta) \quad p(\theta)$$

- Run experiment with design  $d$ .
- To maximise expected information gain:

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

Marginal likelihood

Prior

Posterior



# Motivation

## What's the problem?

- Hard to estimate EIG with Monte Carlo.
- Harder to optimise EIG w.r.t.  $d$ .
- Even harder to optimise a *sequence* of experiments.



# Sequential Experimental Design

## The DAD method

- Learn a design policy  $\pi$  [1].
  - Map history  $h_t = (d_{1:t}, y_{1:t})$  to next design.
  - I.E.  $d_{t+1} = \pi(h_t)$ .

[1] Foster et al. *Deep Adaptive Design: Amortizing Sequential Bayesian Experimental Design*, ICML 2021

## The DAD method

- Learn a design policy  $\pi$  [1].
  - Map history  $h_t = (d_{1:t}, y_{1:t})$  to next design.
  - I.E.  $d_{t+1} = \pi(h_t)$ .
- Can compute lower bound of EIG for  $\pi$ .

$$sPCE(\pi, L, T) = \mathbb{E}_{\underbrace{p(\theta_{0:L})}_{\text{Parameter samples}} \underbrace{p(h_T|\theta_0, \pi)}_{\text{History samples}}} \log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L+1} \sum_{l=0}^L p(h_T|\theta_l, \pi)}$$

## The DAD method

- Learn a design policy  $\pi$  [1].
  - Map history  $h_t = (d_{1:t}, y_{1:t})$  to next design.
  - I.E.  $d_{t+1} = \pi(h_t)$ .
- Can compute lower bound of EIG for  $\pi$ .

$$sPCE(\pi, L, T) = \mathbb{E}_{p(\theta_{0:L})p(h_T|\theta_0, \pi)} \log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L+1} \sum_{l=0}^L p(h_T|\theta_l, \pi)}$$

- Maximise by gradient ascent.



# Sequential Experimental Design

## Shortcomings of DAD

- Design space must be continuous.
- Likelihood model must be differentiable.
- How to do exploration?
- Proposed solution: use **Reinforcement Learning**



# Reinforcement Learning for SED

## Method Outline

- Formulate the problem as a Hidden Parameter Markov Decision Process [2]:  
 $\langle S, A, \Theta, T, R, \gamma, \rho_0, P_\Theta \rangle$
- Execute RL algorithm of choice.

[2] Doshi-Velez & Konidaris. *Hidden Parameter Markov Decision Processes: A Semiparametric Regression Approach for Discovering Latent Task Parametrizations*, IJCAI 2016

# Reinforcement Learning for SED

$$\langle \mathcal{S}, \mathcal{A}, \Theta, \mathcal{T}, R, \gamma, \rho_0, P_{\Theta} \rangle$$

State is  $S_t = (B_{\psi,t}, C_t, y_t)$

Learned History  
Representation

History  
Likelihood  
Vector

Current  
Experiment  
Outcome



# Reinforcement Learning for SED

$$\langle S, \mathbf{A}, \Theta, T, R, \gamma, \rho_0, P_{\Theta} \rangle$$

- The action is the next design.
- I.E.  $a_t = d_{t+1}$

# Reinforcement Learning for SED

$$\langle S, A, \Theta, T, R, \gamma, \rho_0, P_{\Theta} \rangle$$

- $\theta \in \Theta$  are the parameters of the likelihood model.
- $P_{\Theta}$  is the prior  $p(\theta)$ .
- Sample  $\theta_{0:L} \sim P_{\Theta}$  once at the start of an episode.



# Reinforcement Learning for SED

$$\langle S, A, \Theta, \mathbf{T}, R, \gamma, \rho_0, P_{\Theta} \rangle$$

- Transition dynamics are defined by the likelihood model.
  - Experiment outcome computed stochastically:  $y_t \sim p(y_t | d_t, \theta_0)$
  - $B_{\psi,t}$  and  $C_t$  are computed deterministically from  $B_{\psi,t-1}$  and  $C_{t-1}$

# Reinforcement Learning for SED

$\langle S, A, \Theta, T, \mathbf{R}, \gamma, \rho_0, P_{\Theta} \rangle$

- Reward is the immediate contribution to cumulative EIG

$$\begin{aligned} \mathcal{R}(s_{t-1}, a_{t-1}, s_t, \theta) &= \log p(y_t | \theta_0, d_t) \\ &\quad - \log(C_t \cdot \mathbf{1}) + \log(C_{t-1} \cdot \mathbf{1}) \end{aligned}$$

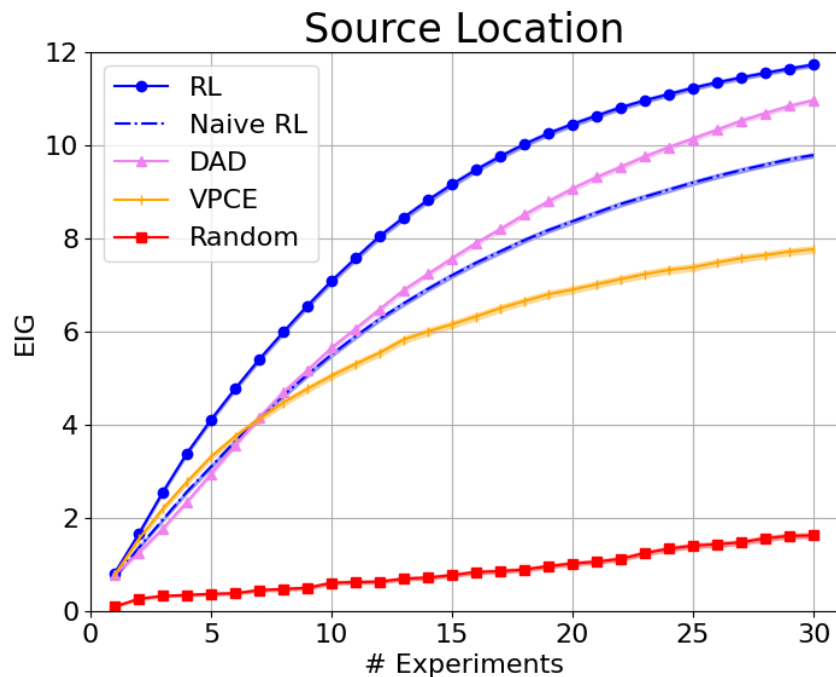


# Reinforcement Learning for SED

$$\langle S, A, \Theta, T, R, \gamma, \rho_0, P_\theta \rangle$$

- Initial state is  $(\mathbf{0}, \mathbf{1}, \emptyset)$ 
  - $B_{\psi,0}$  is an empty sum.
  - $C_0$  is an empty product.
  - $y_0$  is an empty set.
- If  $\gamma = 1$  there is an exact equivalence between MDP and SED problem.
  - Theorem 2 in the paper.

- Goal: find location of 2 signal sources in 2-d space.
- Designs: 2-d coordinate of noisy signal strength sample.
- Outcome: RL achieves higher RL than baselines.





Thank you for listening

Paper pre-print:

<https://arxiv.org/abs/2202.00821>

Code:

<https://github.com/csiro-mlai/RL-BOED>

Correspondence:

[Tom.blau@data61.csiro.au](mailto:Tom.blau@data61.csiro.au)