Appendix for Deep Adaptive Design: Amortizing Sequential Bayesian Experimental Design

Adam Foster * **Desi R. Ivanova** * **Ilyas Malik** **Tom Rainforth**

A. Proofs

We begin by showing that $g_L(\theta_{0:L}, h_T)$ from equation (9) is bounded by $\log(L+1)$ and that $f_L(\theta_{0:L}, h_T)$ from equation (10) can potentially be unbounded. For the former

$$g_L(\theta_{0:L}, h_T) = \log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell, \pi)} \tag{18}$$

$$= \log \frac{p(h_T|\theta_0, \pi)}{p(h_T|\theta_0, \pi) + \sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)} + \log(L+1) \tag{19}$$

$$\leq \log(1) + \log(L+1). \tag{20}$$

For the latter we have

$$f_L(\theta_{0:L}, h_T) = \log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)} \to +\infty \text{ as } \max_{1\leq\ell\leq L} p(h_T|\theta_\ell, \pi) \to 0 \text{ with } p(h_T|\theta_0, \pi) \text{ held constant.}$$

Next we present proofs for all Theorems in the main paper, with each restated for convenience.

Theorem 1. *The total expected information gain for policy π over a sequence of T experiments is*

$$\mathcal{I}_T(\pi) := \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\right] \tag{7}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log p(h_T|\theta, \pi) - \log p(h_T|\pi)\right] \tag{8}$$

where $p(h_T|\pi) = \mathbb{E}_{p(\theta)}[p(h_T|\theta, \pi)]$.

Proof. We begin by rewriting $I_{h_{t-1}}$ in terms of the information gain. This closely mimics the development that we presented in Section 2. By repeated appplication of Bayes Theorem we have

$$I_{h_{t-1}}(\xi_t) = \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi_t)}\left[\log \frac{p(y_t|\theta, \xi_t)}{p(y_t|h_{t-1}, \xi_t)}\right] \tag{21}$$

$$= \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi_t)}\left[\log \frac{p(\theta|h_{t-1})p(y_t|\theta, \xi_t)}{p(\theta|h_{t-1})p(y_t|h_{t-1}, \xi_t)}\right] \tag{22}$$

$$= \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi_t)}\left[\log \frac{p(\theta|h_{t-1}, \xi_t, y_t)}{p(\theta|h_{t-1})}\right] \tag{23}$$

$$= \mathbb{E}_{p(\theta|h_{t-1})}\left[-\log p(\theta|h_{t-1})\right] + \mathbb{E}_{p(y_t,\theta|\xi_t,h_{t-1})}\left[\log p(\theta|h_{t-1}, \xi_t, y_t)\right] \tag{24}$$

$$= \mathbb{E}_{p(\theta|h_{t-1})}\left[-\log p(\theta|h_{t-1})\right] + \mathbb{E}_{p(y_t|\xi_t,h_{t-1})p(\theta|h_{t-1},\xi_t,y_t)}\left[\log p(\theta|h_{t-1}, \xi_t, y_t)\right] \tag{25}$$

$$= \mathbb{E}_{p(y_t|\xi_t,h_{t-1})}\left[\, H[\, p(\theta|h_{t-1})\,] - H[\, p(\theta|h_{t-1}, \xi_t, y_t)\,]\,\right]. \tag{26}$$

Now noting that each $I_{h_{t-1}}(\xi_t)$ is completely determined by h_{t-1} and π (in particular noting that ξ_t is deterministic given

these, while θ is already marginalized out in each $I_{h_{t-1}}(\xi_t)$), we can write

$$\mathcal{I}_T(\pi) = \mathbb{E}_{p(h_T|\pi)}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\right] \tag{27}$$

$$= \sum_{t=1}^{T} \mathbb{E}_{p(h_{t-1}|\pi)}\left[I_{h_{t-1}}(\xi_t)\right] \tag{28}$$

and substituting in our earlier formulation for $I_{h_{t-1}}(\xi_t)$

$$= \sum_{t=1}^{T} \mathbb{E}_{p(h_{t-1}|\pi)}\left[\mathbb{E}_{p(y_t|\xi_t,h_{t-1})}\left[H[\,p(\theta|h_{t-1})\,] - H[\,p(\theta|h_{t-1},\xi_t,y_t)\,]\right]\right]. \tag{29}$$

We now observe that we can write $h_t = h_{t-1} \cup \{(\xi_t, y_t)\}$, which allows us to rewrite this as

$$= \sum_{t=1}^{T} \mathbb{E}_{p(h_t|\pi)}\left[H[\,p(\theta|h_{t-1})\,] - H[\,p(\theta|h_t)\,]\right] \tag{30}$$

$$= \sum_{t=1}^{T} \mathbb{E}_{p(h_T|\pi)}\left[H[\,p(\theta|h_{t-1})\,] - H[\,p(\theta|h_t)\,]\right] \tag{31}$$

$$= \mathbb{E}_{p(h_T|\pi)}\left[\sum_{t=1}^{T} H[\,p(\theta|h_{t-1})\,] - H[\,p(\theta|h_t)\,]\right] \tag{32}$$

$$= \mathbb{E}_{p(h_T|\pi)}\left[H[\,p(\theta)\,] - H[\,p(\theta|h_T)\,]\right], \tag{33}$$

where the last line follows from the fact that we have a telescopic sum. To complete the proof, we rearrange this as

$$= \mathbb{E}_{p(\theta,h_T|\pi)}\left[\log p(\theta|h_T) - \log p(\theta)\right] \tag{34}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log \frac{p(\theta)p(h_T|\theta,\pi)}{p(h_T|\pi)} - \log p(\theta)\right] \tag{35}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log p(h_T|\theta,\pi) - \log p(h_T|\pi)\right] \tag{36}$$

as required. □

Theorem 2 (Sequential PCE)**.** *For a design function π and a number of contrastive samples $L \geq 0$, let*

$$\mathcal{L}_T(\pi, L) = \mathbb{E}_{p(\theta_0,h_T|\pi)p(\theta_{1:L})}\left[g_L(\theta_{0:L}, h_T)\right] \tag{11}$$

where $g_L(\theta_{0:L}, h_T)$ is as per (9), and $\theta_0, h_T \sim p(\theta, h_T|\pi)$, and $\theta_{1:L} \sim p(\theta)$ independently. Given minor technical assumptions discussed in the proof, we have[4]

$$\mathcal{L}_T(\pi, L) \uparrow \mathcal{I}_T(\pi) \text{ as } L \to \infty \tag{12}$$

at a rate $\mathcal{O}\left(L^{-1}\right)$.

Proof. We first show that $\mathcal{L}_T(\pi, L)$ is a lower bound on $\mathcal{I}_T(\pi)$:

$$\mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) = \mathbb{E}_{p(\theta_0,h_T|\pi)}\left[\log \frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)}\right] - \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0,\pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)}\right] \tag{37}$$

$$= \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right] \tag{38}$$

$$= \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log\left(\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(\theta_\ell|h_T)}{p(\theta_\ell)}\right)\right] \tag{39}$$

[4] $x_L \uparrow x$ means that x_L is a monotonically increasing sequence in L with limit x.

now introducing the shorthand $p(\theta_{0:L}^{-\ell}) := p\left(\theta_{0:L\setminus\{\ell\}}\right) = \prod_{j=0,j\neq\ell}^{L} p(\theta_j)$,

$$= \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(\theta_\ell|h_T)p(\theta_{0:L}^{-\ell})}{p(\theta_{0:L})}\right]. \tag{40}$$

Now by the systemtry on term in side the log, we see that this expectation would be the same if it were instead taken over $p(\theta_i, h_T|\pi)p(\theta_{0:L}^{-i})$ for any $i \in \{0,\dots,L\}$ (with $i=0$ giving the original form). Furthermore, the result is unchanged if we take the expectation over the mixture distribution $\frac{1}{L+1}\sum_{i=0}^{L} p(\theta_i, h_T|\pi)p(\theta_{0:L}^{-i}) = p(h_T|\pi)\frac{1}{L+1}\sum_{i=0}^{L} p(\theta_i|h_T)p(\theta_{0:L}^{-i})$ and thus we have

$$= \mathbb{E}_{p(h_T|\pi)}\mathbb{E}_{\frac{1}{L+1}\sum_{i=0}^{L} p(\theta_i|h_T)p(\theta_{0:L}^{-i})}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(\theta_\ell|h_T)p(\theta_{0:L}^{-\ell})}{p(\theta_{0:L})}\right] \tag{41}$$

$$= \mathbb{E}_{p(h_T|\pi)}\left[\mathrm{KL}\left(\tilde{p}(\theta_{0:L}|h_T)||p(\theta_{0:L})\right)\right] \tag{42}$$

where $\tilde{p}(\theta_{0:L}|h_T) = \frac{1}{L+1}\sum_{\ell=0}^{L} p(\theta_\ell|h_T)p(\theta_{0:L}^{-\ell})$, which is indeed a distribution since

$$\int \tilde{p}(\theta_{0:L}|h_T)d\theta_{0:L} = \frac{1}{L+1}\sum_{\ell=0}^{L}\left(\int p(\theta_\ell|h_T)d\theta_\ell \cdot \int p(\theta_{0:L}^{-\ell})d\theta_{0:L}^{-\ell}\right) = 1. \tag{43}$$

Now by Gibbs' inequality the expected KL in (42) must be non-negative, establishing $\mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) \geq 0$ and thus $\mathcal{I}_T(\pi) \geq \mathcal{L}_T(\pi, L)$ as required.

We next show monotonicity in L, i.e. $\mathcal{L}_T(\pi, L_2) \geq \mathcal{L}_T(\pi, L_1)$ for $L_2 \geq L_1 \geq 0$, using similar argument as above

$$\mathcal{L}_T(\pi, L_2) - \mathcal{L}_T(\pi, L_1) = \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{\frac{1}{L_1+1}\sum_{i=0}^{L_1} p(h_T|\theta_i,\pi)}{\frac{1}{L_2+1}\sum_{j=0}^{L_2} p(h_T|\theta_j,\pi)}\right] \tag{44}$$

$$= \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{\frac{1}{L_1+1}\sum_{i=0}^{L_1}\left(p(\theta_i|h_T)/p(\theta_i)\right)}{\frac{1}{L_2+1}\sum_{j=0}^{L_2}\left(p(\theta_j|h_T)/p(\theta_j)\right)}\right] \tag{45}$$

$$= \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{\frac{1}{L_1+1}\sum_{i=0}^{L_1}\left(p(\theta_i|h_T)p(\theta_{0:L_1}^{-i})\right)/p(\theta_{0:L_1})}{\frac{1}{L_2+1}\sum_{j=0}^{L_2}\left(p(\theta_j|h_T)p(\theta_{0:L_2}^{-j})\right)/p(\theta_{0:L_2})}\right] \tag{46}$$

$$= \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{\frac{1}{L_1+1}\sum_{i=0}^{L_1} p(\theta_i|h_T)p(\theta_{0:L_2}^{-i})}{\frac{1}{L_2+1}\sum_{j=0}^{L_2} p(\theta_j|h_T)p(\theta_{0:L_2}^{-j})}\right] \tag{47}$$

$$= \mathbb{E}_{p(h_T|\pi)}\mathbb{E}_{\frac{1}{L+1}\sum_{\ell=0}^{L_1} p(\theta_\ell|h_T)p(\theta_{0:L_2}^{-\ell})}\left[\log \frac{\frac{1}{L_1+1}\sum_{i=0}^{L_1} p(\theta_i|h_T)p(\theta_{0:L_2}^{-i})}{\frac{1}{L_2+1}\sum_{j=0}^{L_2} p(\theta_j|h_T)p(\theta_{0:L_2}^{-j})}\right] \tag{48}$$

$$= \mathbb{E}_{p(h_T|\pi)}\left[\mathrm{KL}(\tilde{p}_1||\tilde{p}_2)\right] \geq 0 \tag{49}$$

where $\tilde{p}_1$ and $\tilde{p}_2$ are, respectively, the distributions in the numerator and denominator in (48). The result then again follows by Gibbs' inequality.

Next we show $\mathcal{L}_T(\pi, L) \to \mathcal{I}_T(\pi)$ as $L \to \infty$. First, note that the denominator in (11), $\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)$, is a consistent estimator of the marginal $p(h_T|\pi)$, since $\frac{1}{L+1}p(h_T|\theta_0,\pi) \to 0$, and by the Strong Law of Large Numbers

$$\frac{1}{L+1}\sum_{\ell=1}^{L} p(h_T|\theta_\ell,\pi) = \frac{L}{L+1}\cdot\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell,\pi) \xrightarrow{\text{a.s.}} \mathbb{E}_{p(\theta)}\left[p(h_T|\theta,\pi)\right] = p(h_T|\pi). \tag{50}$$

Now from (38) we also have that

$$\mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) = \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right] \tag{51}$$

and we have $\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)} \to 0$ almost surely as $L \to \infty$. The minor technical assumption, which is required to establish convergence is that there exist some $0 < \kappa_1, \kappa_2 < \infty$ such that[5]

$$\kappa_1 \leq \frac{p(h_T|\theta,\pi)}{p(h_T|\pi)} \leq \kappa_2 \quad \forall \theta, h_T. \tag{52}$$

using this assumption, the integrand of (51) is bounded, because

$$\begin{aligned}
\left|\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right| &= \left|\log\left(\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right)\right| && (53)\\
&\leq \max\left(\left|\log\left(\max_{\ell}\frac{p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right)\right|, \left|\log\left(\min_{\ell}\frac{p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right)\right|\right) && (54)\\
&\leq \max\left(|\log\kappa_2|, |\log\kappa_1|\right) && (55)\\
&< \infty. && (56)
\end{aligned}$$

Thus, the Bounded Convergence Theorem can be applied to conclude that $\mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) \to 0$ as $L \to \infty$.

Finally, for the rate of convergence we apply the inequality $\log x \leq x - 1$ to (38) to get

$$\begin{aligned}
\mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) &= \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right] && (57)\\
&\leq \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)} - 1\right] && (58)\\
&= \mathbb{E}_{p(\theta_0,h_T|\pi)}\left[\frac{\frac{1}{L+1}\left(p(h_T|\theta_0\pi) + \sum_{\ell=1}^{L}\mathbb{E}_{p(\theta_{1:L})}[p(h_T|\theta_\ell,\pi)]\right)}{p(h_T|\pi)} - 1\right] && (59)\\
&= \mathbb{E}_{p(\theta_0,h_T|\pi)}\left[\frac{\frac{1}{L+1}\left(p(h_T|\theta_0\pi) + Lp(h_T|\pi)\right)}{p(h_T|\pi)} - 1\right] && (60)\\
&= \frac{1}{L+1}\mathbb{E}_{p(\theta_0,h_T|\pi)}\left[\frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)} - 1\right] && (61)\\
&= \frac{C}{L+1}, && (62)
\end{aligned}$$

where we can conclude $C < \infty$ using (52). Combining this with the our previous result showing that $\mathcal{L}_T(\pi, L)$ is a lower bound on $\mathcal{I}_T(\pi)$, we have shown that

$$0 \leq \mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) \leq \frac{C}{L+1}. \tag{63}$$

This establishes the $\mathcal{O}(L^{-1})$ rate of convergence. □

Theorem 3 (Permutation invariance)**.** *Consider a permutation $\sigma \in S_k$ acting on a history h_k^1, yielding $h_k^2 = (\xi_{\sigma(1)}, y_{\sigma(1)}), ..., (\xi_{\sigma(k)}, y_{\sigma(k)})$. For all such σ, we have*

$$\mathbb{E}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\middle| h_k = h_k^1\right] = \mathbb{E}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\middle| h_k = h_k^2\right]$$

such that the EIG is unchanged under permutation. Further, the optimal policies starting in h_k^1 and h_k^2 are the same.

[5] In practice, we can actually weaken this assumption significantly if necessary by making κ_1 and κ_2 dependent on h_T and θ then assuming that the expectation $\mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}[\log|\kappa_i(\theta_j, h_T)|]$ is finite for $i \in \{1, 2\}$ and $j \in \{0, 1\}$. This then permits $\kappa_1(h_T, \theta) \to 0$ and $\kappa_2(h_T, \theta) \to \infty$ for certain h_T and θ, provided that these events are zero measure under both $p(\theta, h_T|\pi)$ and $p(\theta)p(h_T|\pi)$, thereby avoiding potential issues with tail behavior in the limits of extreme values for θ.

Technical note: In this statement, the first expectation is with respect to $p(h_T|\pi)$ for policy π and the second is with respect to $p(h_T|\pi')$, where for $t > k$ we set $\pi'(h_t) = \pi(\sigma^{-1}(h_t))$ where σ^{-1} acts on the first k labels by permutation and as the identity on other labels. This means we remove *explicit* variability under permutation caused by π, and show that no other source of variability can arise.

Proof. To begin, we set up some notation. Given the partial history $h_k = h_k^1$, we complete the experiment by sampling (ξ_t, y_t) for $t = k+1, ..., T$. We denote the resulting full history as h_T^1, and define h_T^2 similarly. Next, we use Theorem 1 to rewrite the conditional objective under consideration as

$$\mathbb{E}_{p(h_T^1|\pi)}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\middle| h_k = h_k^1\right] = \mathbb{E}_{p(\theta|h_k^1)\prod_{t=k+1}^{T} p(y_t|\theta,\xi_t)}\left[\log p(h_T^1|\theta,\pi) - \log p(h_T^1|\pi)\right] \tag{64}$$

$$= \mathbb{E}_{p(\theta|h_k^1)\prod_{t=k+1}^{T} p(y_t|\theta,\xi_t)}\left[\log p(\theta|h_T^1) - \log p(\theta)\right] \tag{65}$$

$$= \mathbb{E}_{p(\theta|h_k^1)p(h_T^1|h_k^1,\theta,\pi)}\left[\log p(\theta|h_T^1) - \log p(\theta)\right]. \tag{66}$$

A central point of the proof is that the posterior distribution $p(\theta|h_t)$ is invariant to the order of the history. Indeed, we have

$$p(\theta|h_t) \propto p(\theta)\prod_{s=1}^{t} p(y_s|\theta,\xi_s) \tag{67}$$

which shows that $p(\theta|h_k^1) = p(\theta|h_k^2)$. Given a continuation of the history $(\xi_{k+1}, y_{k+1}), ..., (\xi_T, y_T)$, if we use the same continuation starting from h_k^1 and h_k^2 to give h_T^1 and h_T^2 then we have $p(\theta|h_T^1) = p(\theta|h_T^2)$. However, we need to show that the continuations $(\xi_{k+1}, y_{k+1}), ..., (\xi_T, y_T)$ are equal in distribution.

We now show that the sampling distributions of $(\xi_{k+1}, y_{k+1}), ..., (\xi_T, y_T)$ are equal starting from h_k^1 and h_k^2. We have shown that $\theta \sim p(\theta|h_k^1)$ is unchanged in distribution if we instead sample $\theta \sim p(\theta|h_k^2)$. Further, we have

$$\xi_{k+1}^1 = \pi(h_k^1) \qquad \xi_{k+1}^2 = \pi'(h_k^2) \tag{68}$$

which, by the construction of π' implies $\xi_{k+1}^1 = \xi_{k+1}^2$. Together, these results imply that the observations y_{k+1}^1 and y_{k+1}^2 are equal in distribution. Proceeding inductively, since h_{k+1}^1 and h_{k+1}^2 are equal in distribution a similar argument shows that h_{k+2}^1 and h_{k+2}^2 have the same distribution. Continuing in this way, we have that h_T^1 and h_T^2 are equal in distribution. Together, these results imply that

$$\mathbb{E}_{p(\theta|h_k^1)p(h_T^1|h_k^1,\theta,\pi)}\left[\log p(\theta|h_T^1) - \log p(\theta)\right] = \mathbb{E}_{p(\theta|h_k^2)p(h_T^2|h_k^2,\theta,\pi')}\left[\log p(\theta|h_T^2) - \log p(\theta)\right] \tag{69}$$

which conclude the first part of the proof.

To establish the permutation invariance of the optimal policy π^*, we reason by induction starting with $k = T-1$, using a dynamic programming style argument. Given h_{T-1}, the total EIG is a function of $p(\theta|h_{T-1})$ and ξ_T. Since we do not need to account for future asymmetry in the policy, we immediately have that the optimal final design ξ_T only depends on $p(\theta|h_{T-1})$, which implies that is invariant to the order of the history.

We now assume that the optimal policy is permutation invariant starting from $k+2$. Using the previous result (69), we separete out the design ξ_{k+1} and substitute π^* for both π and π' (since it is permutation invariant for the steps after $k+1$ by inductive hypothesis) to give

$$\begin{aligned}&\mathbb{E}_{p(\theta|h_k^1)p(y_{k+1}|\theta,\xi_{k+1})\prod_{t=k+2}^{T} p(y_t|\theta,\pi^*(h_{t-1}))}\left[\log p(\theta|h_T^1) - \log p(\theta)\right]\\ &= \mathbb{E}_{p(\theta|h_k^2)p(y_{k+1}|\theta,\xi_{k+1})\prod_{t=k+2}^{T} p(y_t|\theta,\pi^*(h_{t-1}))}\left[\log p(\theta|h_T^2) - \log p(\theta)\right].\end{aligned} \tag{70}$$

To extend the optimal policy to $k+1$, we consider choosing ξ_{k+1} and then following π^* thereafter. As (70) shows us, the decision problem for ξ_{k+1} is the same starting from h_k^1 and h_k^2 because the posterior distributions $p(\theta|h_k^1)$ and $p(\theta|h_k^2)$ are equal, and the optimal policy after $k+1$ does not depend on history order. This implies that the optimal choice of ξ_{k+1} is the same for h_k^1 and h_k^2. This implies that the optimal policies starting in h_k^1 and h_k^2 are the same. This completes the proof. □

Theorem 4. *For a design function π and a number of contrastive samples $L \geq 1$, let*

$$\mathcal{U}_T(\pi, L) = \mathbb{E}\left[\log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)}\right] \tag{71}$$

where the expectation is over $\theta_0, h_T \sim p(\theta, h_T|\pi)$ and $\theta_{1:L} \sim p(\theta)$ independently. Then,

$$\mathcal{U}_T(\pi, L) \downarrow \mathcal{I}_T(\pi) \text{ as } L \to \infty \tag{72}$$

at a rate $\mathcal{O}(L^{-1})$.

Proof. We first show $\mathcal{U}_T(\pi, L)$ is an upper bound to $\mathcal{I}_T(\pi)$

$$\mathcal{U}_T(\pi, L) - \mathcal{I}_T(\pi) = \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)}\right] - \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log \frac{p(h_T|\theta_0, \pi)}{p(h_T|\pi)}\right] \tag{73}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log p(h_T|\pi) - \log\left(\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)\right)\right] \tag{74}$$

now using Jensen's inequality

$$\geq \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log p(h_T|\pi) - \log\left(\frac{1}{L}\sum_{\ell=1}^{L} \mathbb{E}_{p(\theta_\ell)}\left[p(h_T|\theta_\ell, \pi)\right]\right)\right] \tag{75}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log p(h_T|\pi) - \log\left(\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\pi)\right)\right] \tag{76}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log p(h_T|\pi) - \log p(h_T|\pi)\right] \tag{77}$$

$$= 0. \tag{78}$$

To show monotonicity in L, pick $L_2 \geq L_1 \geq 0$ and consider the difference

$$\delta := \mathcal{U}_T(\pi, L_1) - \mathcal{U}_T(\pi, L_2) = \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{\frac{1}{L_2}\sum_{j=1}^{L_2} p(h_T|\theta_j, \pi)}{\frac{1}{L_1}\sum_{i=1}^{L_1} p(h_T|\theta_i, \pi)}\right]. \tag{79}$$

Notice that we can write expression in the numerator $\frac{1}{L_2}\sum_{j=1}^{L_2} p(h_T|\theta_j, \pi) = \mathbb{E}_{J_1,\ldots,J_{L_1}}\left[\frac{1}{L_1}\sum_{k=1}^{L_1} p(h_T|\theta_{J_k}, \pi)\right]$, where the indices J_k have been uniformly drawn from $1, \ldots, L_2$. We have

$$\delta = \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \mathbb{E}_{J_1,\ldots,J_{L_1}}\left[\frac{1}{L_1}\sum_{k=1}^{L_1} p(h_T|\theta_{J_k}, \pi)\right] - \log \frac{1}{L_1}\sum_{i=1}^{L_1} p(h_T|\theta_i, \pi)\right] \tag{80}$$

now applying Jensen's Inequality

$$\geq \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\mathbb{E}_{J_1,\ldots,J_{L_1}}\left[\log \frac{1}{L_1}\sum_{k=1}^{L_1} p(h_T|\theta_{J_k}, \pi)\right] - \log \frac{1}{L_1}\sum_{i=1}^{L_1} p(h_T|\theta_i, \pi)\right] \tag{81}$$

then use the fact that any L_1-subset of $\theta_1, ..., \theta_{L_2}$ has the same distribution

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{1}{L_1}\sum_{i=1}^{L_1} p(h_T|\theta_i, \pi) - \log \frac{1}{L_1}\sum_{i=1}^{L_1} p(h_T|\theta_i, \pi)\right] = 0 \tag{82}$$

which establishes monotonicity.

Finally, convergence is shown analogously to Theorem 2. Again we adopt the assumption (52). The Strong Law of Large Numbers gives us almost sure convergence $\log\left(\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell,\pi)\right) \to \log p(h_T|\pi)$ as $L \to \infty$. Applying the Bounded Convergence Theorem, as in Theorem 2, we have

$$\lim_{L\to\infty}\left(\mathcal{U}_T(\pi,L) - \mathcal{I}_T(\pi,L)\right) = \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\lim_{L\to\infty}\log\frac{p(h_T|\pi)}{\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell,\pi)}\right] \tag{83}$$

$$= 0. \tag{84}$$

Finally, for the rate of convergence, we have

$$\mathcal{U}_T(\pi,L) - \mathcal{I}_T(\pi) = \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log\frac{p(h_T|\pi)}{\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell,\pi)}\right] \tag{85}$$

$$= \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[-\log\left(\frac{1}{L}\sum_{\ell=1}^{L}\frac{p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right)\right] \tag{86}$$

$$= \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[-\log\left(1+\frac{1}{L}\sum_{\ell=1}^{L}\left(\frac{p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}-1\right)\right)\right] \tag{87}$$

$$= \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\sum_{n=1}^{\infty}(-1)^n\frac{x^n}{n}\right] \tag{88}$$

where $x = \frac{1}{L}\sum_{\ell=1}^{L}\left(\frac{p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)} - 1\right)$ and we have applied the Taylor expansion for $\log(1+x)$. We have

$$\mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[x\right] = 0 \tag{89}$$

$$\mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[x^2\right] = \frac{1}{L}\mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\left(\frac{p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}-1\right)^2\right] \tag{90}$$

and higher order terms are $o(L^{-1})$ (Angelova, 2012; Nowozin, 2018). This shows that $\mathcal{U}_T(\pi,L) - \mathcal{I}_T(\pi) \to 0$ at a rate $\mathcal{O}(L^{-1})$. This concludes the proof. □

B. Additional bounds

In this section, we consider a more general lower bound on $\mathcal{I}_T(\pi)$ based on the ACE bound of Foster et al. (2020). We consider a parametrized proposal distribution $q(\theta; h_T)$ which can be used to approximate the posterior $p(\theta|h_T)$. One example of such a proposal would be an amortized variational approximation to the posterior that takes as input h_T and outputs a variational distribution over θ. It would be possible to share the representation $R(h_T)$ from (17) between the design network and the inference network. However, the following theorem is not limited to variational posteriors, and concerns any parametrized proposal distribution.

Theorem 5. *For a design function π, a number of contrastive samples $L \geq 1$, and a parametrized proposal $q(\theta; h_T)$, we have the sequential Adaptive Contrastive Estimation (sACE) lower bound*

$$\mathcal{I}_T(\pi) \geq \mathbb{E}_{p(\theta_0,h_T|\pi)q(\theta_{1:L};h_T)}\left[\log\frac{p(h_T|\theta_0,\pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(h_T|\theta_\ell,\pi)p(\theta_\ell)}{q(\theta_\ell;h_T)}}\right] \tag{91}$$

and the sequential Variational Nested Monte Carlo (sVNMC) upper bound

$$\mathcal{I}_T(\pi) \leq \mathbb{E}_{p(\theta_0,h_T|\pi)q(\theta_{1:L};h_T)}\left[\log\frac{p(h_T|\theta_0,\pi)}{\frac{1}{L}\sum_{\ell=1}^{L}\frac{p(h_T|\theta_\ell,\pi)p(\theta_\ell)}{q(\theta_\ell;h_T)}}\right]. \tag{92}$$

Proof. We begin by showing the sACE lower bound. The proof closely follows that of Theorem 2. We have the error term

$$\delta_{sACE} = \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log \frac{p(h_T|\theta_0, \pi)}{p(h_T|\pi)}\right] - \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{q(\theta_{1:L}; h_T)}\left[\log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(h_T|\theta_\ell, \pi)p(\theta_\ell)}{q(\theta_\ell; h_T)}}\right] \tag{93}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{q(\theta_{1:L}; h_T)}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(h_T|\theta_\ell, \pi)p(\theta_\ell)}{q(\theta_\ell; h_T)}}{p(h_T|\pi)}\right] \tag{94}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{q(\theta_{1:L}; h_T)}\left[\log \left(\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(\theta_\ell|h_T)}{q(\theta_\ell; h_T)}\right)\right] \tag{95}$$

now introducing the shorthand $q(\theta_{0:L}^{-\ell}; h_T) := q\left(\theta_{0:L\setminus\{\ell\}}; h_T\right) = \prod_{j=0, j\neq\ell}^{L} q(\theta_j; h_T)$,

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{q(\theta_{1:L}; h_T)}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(\theta_\ell|h_T) q(\theta_{0:L}^{-\ell}; h_T)}{q(\theta_{0:L}; h_T)}\right]. \tag{96}$$

Now by the systemtry on term in side the log, we see that this expectation would be the same if it were instead taken over $p(\theta_i, h_T|\pi)q(\theta_{0:L}^{-i}; h_T)$ for any $i \in \{0, \ldots, L\}$. It is also the same if we take the expectation over $\frac{1}{L+1}\sum_{i=0}^{L} p(\theta_i, h_T|\pi)q(\theta_{0:L}^{-i}; h_T) = p(h_T|\pi)\frac{1}{L+1}\sum_{i=0}^{L} p(\theta_i|h_T)q(\theta_{0:L}^{-i}; h_T)$ and thus we have

$$= \mathbb{E}_{p(h_T|\pi)}\mathbb{E}_{\frac{1}{L+1}\sum_{i=0}^{L} p(\theta_i|h_T)q(\theta_{0:L}^{-i}; h_T)}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(\theta_\ell|h_T) q(\theta_{0:L}^{-\ell}; h_T)}{q(\theta_{0:L}; h_T)}\right] \tag{97}$$

$$= \mathbb{E}_{p(h_T|\pi)}\left[\mathrm{KL}\left(\breve{q}(\theta_{0:L}; h_T)||q(\theta_{0:L}; h_T)\right)\right] \tag{98}$$

where $\breve{q}(\theta_{0:L}; h_T) = \frac{1}{L+1}\sum_{\ell=0}^{L} p(\theta_\ell|h_T)q(\theta_{0:L}^{-\ell}; h_T)$, which is indeed a distribution since

$$\int \breve{q}(\theta_{0:L}; h_T) d\theta_{0:L} = \frac{1}{L+1}\sum_{\ell=0}^{L}\left(\int p(\theta_\ell|h_T)d\theta_\ell \cdot \int q(\theta_{0:L}^{-\ell}; h_T)d\theta_{0:L}^{-\ell}\right) = 1. \tag{99}$$

Now by Gibb's inequality the expected KL in (98) must be non-negative, establishing the required lower bound.

Turning to the sVNMC bound, we use a proof that is close in spirit to Theorem 4. We have the error term

$$\delta_{sVNMC} = \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{q(\theta_{1:L}; h_T)}\left[\log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L}\sum_{\ell=1}^{L} \frac{p(h_T|\theta_\ell, \pi)p(\theta_\ell)}{q(\theta_\ell; h_T)}}\right] - \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log \frac{p(h_T|\theta_0, \pi)}{p(h_T|\pi)}\right] \tag{100}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{q(\theta_{1:L}; h_T)}\left[\log p(h_T|\pi) - \log\left(\frac{1}{L}\sum_{\ell=1}^{L} \frac{p(h_T|\theta_\ell, \pi)p(\theta_\ell)}{q(\theta_\ell; h_T)}\right)\right] \tag{101}$$

now using Jensen's inequality

$$\geq \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log p(h_T|\pi) - \log\left(\frac{1}{L}\sum_{\ell=1}^{L} \mathbb{E}_{q(\theta_\ell; h_T)}\left[\frac{p(h_T|\theta_\ell, \pi)p(\theta_\ell)}{q(\theta_\ell; h_T)}\right]\right)\right] \tag{102}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log p(h_T|\pi) - \log\left(\frac{1}{L}\sum_{\ell=1}^{L} \mathbb{E}_{p(\theta_\ell)}\left[p(h_T|\theta_\ell, \pi)\right]\right)\right] \tag{103}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log p(h_T|\pi) - \log\left(\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\pi)\right)\right] \tag{104}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log p(h_T|\pi) - \log p(h_T|\pi)\right] \tag{105}$$

$$= 0. \tag{106}$$

This establishes the upper bound. □

C. Gradient details

C.1. Score function gradient

Recall that our sPCE objective is

$$\mathcal{L}_T(\pi_\phi, L) = \mathbb{E}_{p(\theta_{0:L})p(h_T|\theta_0,\pi_\phi)}\left[g_L(\theta_{0:L}, h_T)\right] \tag{107}$$

$$= \mathbb{E}_{p(\theta_{0:L})p(h_T|\theta_0,\pi_\phi)}\left[\log \frac{p(h_T|\theta_0,\pi_\phi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\right] \tag{108}$$

$$= \mathbb{E}_{p(\theta_{0:L})p(h_T|\theta_0,\pi_\phi)}\left[\log \frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\right] + \log(L+1) \tag{109}$$

Differentiating this gives:

$$\frac{d\mathcal{L}_T}{d\phi} = \mathbb{E}_{p(\theta_{0:L})}\left[\int \frac{d}{d\phi}\left(p(h_T|\theta_0,\pi_\phi)\log\frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\right)dh_T\right] \tag{110}$$

$$= \mathbb{E}_{p(\theta_{0:L})}\left[\int \log\frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\frac{d}{d\phi}p(h_T|\theta_0,\pi_\phi) + p(h_T|\theta_0,\pi_\phi)\frac{d}{d\phi}\log\frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}dh_T\right] \tag{111}$$

$$= \mathbb{E}_{p(\theta_{0:L})}\Bigg[\int p(h_T|\theta_0,\pi_\phi)\log\frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\left(\frac{d}{d\phi}\log p(h_T|\theta_0,\pi_\phi)\right)dh_T \tag{112}$$

$$+ \int p(h_T|\theta_0,\pi_\phi)\left(\frac{d}{d\phi}\log p(h_T|\theta_0,\pi_\phi)\right)dh_T - \int p(h_T|\theta_0,\pi_\phi)\frac{d}{d\phi}\log\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)dh_T\Bigg] \tag{113}$$

$$= \mathbb{E}_{p(\theta_{0:L})}\mathbb{E}_{p(h_T|\theta_0,\pi_\phi)}\left[\log\frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\left(\frac{d}{d\phi}\log p(h_T|\theta_0,\pi_\phi)\right) - \frac{d}{d\phi}\log\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)\right]. \tag{114}$$

In line (112) we used the log-trick $\frac{d}{dx}f(x) = f(x)\left(\frac{d}{dx}\log f(x)\right)$ and again in line (114) (in the reverse direction), together with the fact $\int \frac{d}{d\phi}p(h_T|\theta_0,\pi_\phi)dh_T = \frac{d}{d\phi}\int p(h_T|\theta_0,\pi_\phi)dh_T = 0$.

C.2. Expanded reparametrized gradient

For completeness, we provided a fully expanded form of the gradient in (14), computed using the chain rule. In practice, derivatives of this form are calculated automatically in PyTorch (Paszke et al., 2019).

Initially, we set up some additional notation. Suppose ξ the design is of dimension D_1 and y the observation is of dimension D_2. Then $u = (\xi, y)$ is of dimension $D_1 + D_2$. For an arbitrary scalar quantity x, we have

$$\frac{\partial x}{\partial u} = \begin{pmatrix}\frac{\partial x}{\partial \xi^{(1)}} & \cdots & \frac{\partial x}{\partial \xi^{(D_1)}} & \frac{\partial x}{\partial y^{(1)}} & \cdots & \frac{\partial x}{\partial y^{(D_2)}}\end{pmatrix} \tag{115}$$

and

$$\frac{\partial u}{\partial x} = \begin{pmatrix}\frac{\partial \xi^{(1)}}{\partial x} & \cdots & \frac{\partial \xi^{(D_1)}}{\partial x} & \sum_{d=1}^{D_1}\frac{\partial y^{(1)}}{\partial \xi^{(d)}}\frac{\partial \xi^{(d)}}{\partial x} & \cdots & \sum_{d=1}^{D_1}\frac{\partial y^{(D_2)}}{\partial \xi^{(d)}}\frac{\partial \xi^{(d)}}{\partial x}\end{pmatrix}^\top. \tag{116}$$

This notation enables us to concisely and clearly deal with both scalar and vector quantities. In general, the derivatives $\partial a/\partial b$ and da/db represent a matrix of shape $(\dim a, \dim b)$ where one or both of a, b may have dimension 1. This notation is particularly attractive because the Chain Rule for partial derivatives can be concisely expressed as follows. Suppose $a = a(b_1(c), ..., b_n(c), c)$, then the total derivative is given by

$$\frac{da}{dc} = \frac{\partial a}{\partial c} + \sum_{i=1}^{n}\frac{\partial a}{\partial b_i}\frac{db_i}{dc} \tag{117}$$

where the normal rules of matrix multiplication apply. We now apply this in the context of the function $g(\theta_{0:L}, h_T)$ which was defined in Section 4.2.

We have $g = g(\theta_{0:L}, u_1, ..., u_T)$. The Chain Rule implies that

$$\frac{dg}{d\phi} = \sum_{t=1}^{T} \frac{\partial g}{\partial u_t} \frac{du_t}{d\phi}. \tag{118}$$

We also have, for $t = 1, ..., T$, that $u_t = u(\phi, h_{t-1}, \theta_0, \epsilon_t) = u(\phi, u_1, ..., u_{t-1}, \theta_0, \epsilon_t)$. This represents the dependence of ξ_t on h_{t-1} via π_ϕ, and the further dependence of y_t on θ_0 and ϵ_t. Expanding the derivatives again using the Chain Rule gives

$$\frac{dg}{d\phi} = \sum_{t=1}^{T} \frac{\partial g}{\partial u_t} \left(\frac{\partial u_t}{\partial \phi} + \sum_{s=1}^{t-1} \frac{\partial u_t}{\partial u_s} \frac{du_s}{d\phi} \right). \tag{119}$$

Again, we can expand the total derivative to give

$$\frac{dg}{d\phi} = \sum_{t=1}^{T} \frac{\partial g}{\partial u_t} \left(\frac{\partial u_t}{\partial \phi} + \sum_{s=1}^{t-1} \frac{\partial u_t}{\partial u_s} \left(\frac{\partial u_s}{\partial \phi} + \sum_{r=1}^{s-1} \frac{\partial u_s}{\partial u_r} \frac{du_r}{d\phi} \right) \right). \tag{120}$$

Rather than continuing in this manner, we observe that the current expansion (120) can be split up as follows

$$\frac{dg}{d\phi} = \sum_{t=1}^{T} \frac{\partial g}{\partial u_t} \frac{\partial u_t}{\partial \phi} + \sum_{1 \le s < t \le T} \frac{\partial g}{\partial u_t} \frac{\partial u_t}{\partial u_s} \frac{\partial u_s}{\partial \phi} + \sum_{1 \le r < s < t \le T} \frac{\partial g}{\partial u_t} \frac{\partial u_t}{\partial u_s} \frac{\partial u_s}{\partial u_r} \frac{du_r}{d\phi} \tag{121}$$

which shows that we have completely enumerated over all paths of length 1 and 2 through the computational graph, and the final term with a total derivative concerns paths of length 3 or more. This approach can be naturally extended to enumerate over all paths. To write this concisely, we introduce a new variable k which denotes the length of the path, and then a sum over all increasing sequences $1 \le t_1 < ... < t_k \le T$. This gives

$$\frac{dg}{d\phi} = \sum_{k=1}^{T} \left[\sum_{1 \le t_1 < ... < t_k \le T} \frac{\partial g}{\partial u_{t_k}} \frac{\partial u_{t_k}}{\partial u_{t_{k-1}}} \cdots \frac{\partial u_{t_2}}{\partial u_{t_1}} \frac{\partial u_{t_1}}{\partial \phi} \right]. \tag{122}$$

This can be written concisely as

$$\frac{dg}{d\phi} = \sum_{\substack{k \in \{1,...,T\} \\ 1 \le t_1 < ... < t_k \le T}} \frac{\partial g}{\partial u_{t_k}} \left(\prod_{j=1}^{k-1} \frac{\partial u_{t_{j+1}}}{\partial u_{t_j}} \right) \frac{\partial u_{t_1}}{\partial \phi} \tag{123}$$

where the product is interpretted in the order given in (122) for the matrix multiplication to operate correctly, and an empty product is equal to the identity.

D. Experiment details

Our experiments were implemented using PyTorch (Paszke et al., 2019) and Pyro (Bingham et al., 2018). An open-source implementation of DAD, including code for reproducing each experiment, is available at `https://github.com/ae-foster/dad`. Full details on running the code are given in the `README.md` file.

D.1. Location Finding

In this experiment we have K hidden objects or *sources* in $\mathbb{R}^d$, $d \in \{1, 2, 3\}$ and aim to learn their locations, $\theta = \{\theta_k\}_{k=1}^K$. The number of sources, K, is assumed to be known. Each of the sources emits a signal with intensity obeying the inverse-square law. In other words, if a source is located at θ_k and we perform a measurement at a point ξ, the signal strength will be proportional to $\frac{1}{\|\theta_k - \xi\|^2}$.

Since there are multiple sources, we consider the total intensity at location ξ, which is a superposition of the individual ones

$$\mu(\theta, \xi) = b + \sum_{k=1}^{K} \frac{\alpha_k}{m + \|\theta_k - \xi\|^2}, \tag{124}$$

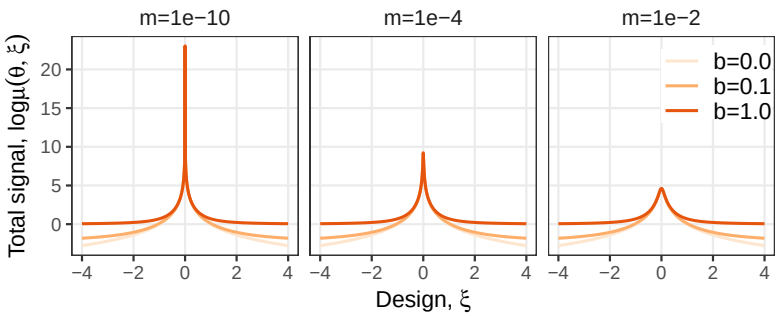

Figure 5. Log-total intensity

where α_k can be known constants or random variables, $b, m > 0$ are constants controlling background and maximum signal, respectively. Figure 5 shows the effect b and m have on log total signal strength.

We place a standard normal prior on each of the location parameters θ_k and we observe the log total intensity with some Gaussian noise. We therefore have the following prior and likelihood:

$$\theta_k \overset{\text{i.i.d.}}{\sim} N(0_d, I_d), \;\; \log y \mid \theta, \xi \sim N(\log \mu(\theta, \xi), \sigma). \tag{125}$$

The model hyperparameters used in our experiments can be found in the table below.

Parameter	Value
Number of sources, K	2
Base signal, b	10^{-1}
Max signal, m	10^{-4}
α_1, α_2	1
Signal noise, σ	0.5

We trained a DAD network to amortize experimental design for this problem, using the neural architecture outlined in Section 4.3. Both the encoder and the decoder are simple feed-forward neural networks with a single hidden layer; details in the following table. For the encoder

Layer	Description	Dimension	Activation
Input	ξ, y	3	-
H1	Fully connected	256	ReLU
Output	Fully connected	16	-

and for the emitter

Layer	Description	Dimension	Activation
Input	$R(h_t)$	16	-
H1	Fully connected	2	-
Output	ξ	2	-

Since the likelihood is reparametrizable, we use (14) to calculate approximate gradients. We optimized the network using Adam (Kingma & Ba, 2014) with exponential learning rate annealing with parameter γ. Full details are given in the following table.

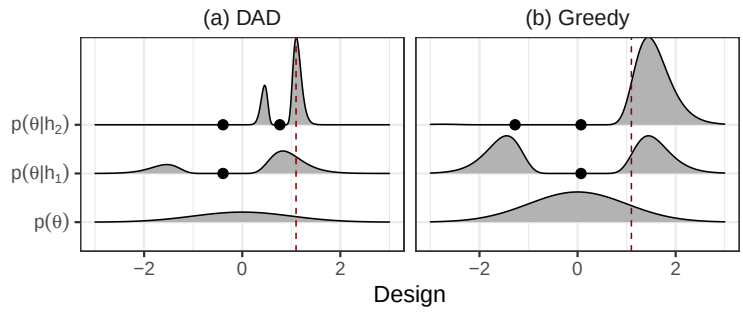

Figure 6. Posterior distributions of the location finding example with $K = 1$ source $\mathbb{R}$.

Parameter	Value
Inner samples, L	2000
Outer samples	2000
Initial learning rate	5×10^{-5}
Betas	(0.8, 0.998)
γ	0.98
Gradient steps	50000
Annealing frequency	1000

We used a greater number of inner and outer samples for a more accurate estimate of $\mathcal{I}_T(\pi)$ for evaluation when computing the presented values in Table 1 and in our Training Stability ablation, specifically $L = 5 \times 10^5$ inner samples, and 256 (variational) or 2048 (other methods) outer samples.

Deployment times Deployment speed tests were performed on a CPU-only machine witht the following specifications:

Memory	16 GB 2133 MHz LPDDR3
Processor	2.8 GHz Quad-Core Intel Core i7
Operating System	MacOS BigSur v.11.2.3

We took the mean and ± 1 s.e. over 10 realizations. Deployment times for all methods are given in the following table

Method	Deployment time (s)
Random	0.0026 ± 0.0001
Fixed	0.0018 ± 0.0001
DAD	0.0474 ± 0.0003
Variational	8963.2 ± 42.2

Discussion details In the discussion, we used a simpler form of the same model with $K = 1$ source and $\theta \in \mathbb{R}, \xi \in \mathbb{R}$. In this simplified setting, we can calculate the true optimal myopic (greedy) baseline using numerical integration. We evaluate equation (1) using line integrals as follows

$$I_t(\xi) = \int p(\theta|h_{t-1}) \mathbb{E}_{p(y|\theta)} \left[\log \frac{p(y|\theta)}{\int p(\theta'|h_{t-1}) p(y|\theta') d\theta'} \right] d\theta \tag{126}$$

$$= \int p(\theta|h_{t-1}) \mathbb{E}_{p(y|\theta)} \left[\log \int p(\theta'|h_{t-1}) p(y|\theta') d\theta' \right] d\theta + C \tag{127}$$

where $C = -H(p(y|\theta))$ is the entropy of a Gaussian, location independent and therefore constant with respect to ξ. We calculate (127) for a range of designs, $\xi \in \Xi_{\text{grid}}$, and select the optimal design $\xi^* = \arg\max_{\Xi_{\text{grid}}} I_t(\xi)$. The integrals themselves are also calculated using numerical integration on a grid, Θ_{grid}, and use sampling to calculate the inner expectation; further details can in the table below.

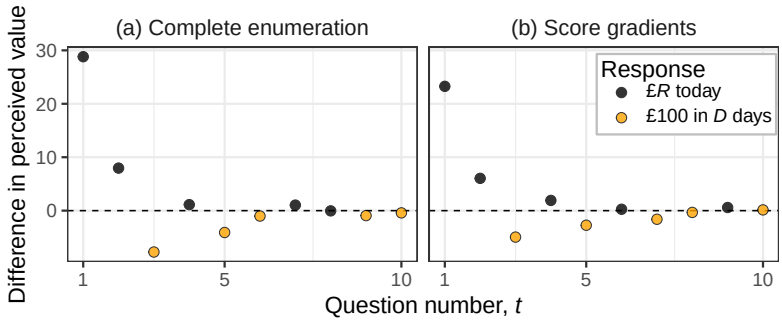

Figure 7. Comparison of two gradient methods for the hyperbolic temporal discounting model with $T = 10$ experiments.

Parameter	Value
Design grid, Ξ_{grid}	300 equally spaced from -3 to 3
θ grid, Θ_{grid}	600 equally spaced from -4 to 4
y samples for inner expectation	400

It is important to emphasize that even in this simple one-dimensional setting evaluating the myopic strategy is extremely costly and may require more sophisticated numerical integration techniques (e.g. quadrature) as posteriors become more peaked. Furthermore, as Figure 6 indicates, the resulting posteriors are complex and multi-modal even in 1D. This multi-modality may also be a reason why the variational method does not work well in this example.

D.2. Hyperbolic temporal discounting

We consider a hyperbolic temporal discounting model (Mazur, 1987; Vincent, 2016; Vincent & Rainforth, 2017) in which a participant's behaviour is characterized by the latent variables $\theta = (k, \alpha)$ with prior distributions

$$\log k \sim N(-4.25, 1.5) \qquad \alpha \sim \text{HalfNormal}(0, 2) \tag{128}$$

where the HalfNormal distribution is a Normal distribution truncated at 0. For given k, α, the value of the two propositions "£R today" and "£100 in D days" with design $\xi = (R, D)$ are given by

$$V_0 = R, \qquad V_1 = \frac{100}{1 + kD}. \tag{129}$$

The probability of the participant selecting the second option, V_1, rather than V_0 is then modelled as

$$p(y = 1|k, \alpha, R, D) = \epsilon + (1 - 2\epsilon)\Phi\left(\frac{V_1 - V_0}{\alpha}\right) \tag{130}$$

where Φ is the c.d.f. of the standard Normal distribution, i.e.

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}z^2\right) \tag{131}$$

and we fix $\epsilon = 0.01$. We considered the iterated version of this experiment, modelling $T = 20$ experiments with each sampled setting for the latents k, α.

We began by training a DAD network to amortize experimental design for this problem. The design parameters R, D have the constraints $D > 0$ and $0 < R < 100$. We represented R, D in an unconstrained space ξ_d, ξ_r and transformed them using the maps

$$D = \exp(\xi_d) \qquad R = 100 \text{ sigmoid}(\xi_r) \tag{132}$$

We used the neural architecture outlined in Section 4.3. For the encoder E_{ϕ_1} we used the following network with two hidden layers

Layer	Description	Dimension	Activation
Design input	ξ_d, ξ_r	2	-
H1	Fully connected	256	Softplus
H2	Fully connected	256	Softplus
H3	Fully connected	16	-
H3'	Fully connected	16	-
Output	$y \odot H3 + (1-y) \odot H3'$	16	-

The emitter network F_{ϕ_2} similarly used two hidden layers as follows

Layer	Description	Dimension	Activation
Input	$R(h_t)$	16	-
H1	Fully connected	256	Softplus
H2	Fully connected	256	Softplus
Output	ξ_d, ξ_r	2	-

Since the number of experiments we perform is relatively large ($T = 20$), we constructed a score function gradient estimator of (16) (see also § C.1 for details) and optimized this network with Adam (Kingma & Ba, 2014).We used exponential learning rate annealing with parameter γ. Full details are given in the following table.

Parameter	Value
Inner samples, L	500
Outer samples	500
Initial learning rate	10^{-4}
Betas	(0.9, 0.999)
γ	0.96
Gradient steps	100000
Annealing frequency	1000

For the fixed baseline, we used the same optimization settings, except we set the initial learning rate to 10^{-1}. We trained the DAD and fixed methods on a machine with 8 Intel(R) Xeon(R) CPU E5-2637 v4 @ 3.50GHz CPUs, one GeForce GTX 1080 Ti GPU, 126 GiB memory running Fedora 32. Note this is *not* the machine used to conudct speed tests. For the Badapted baseline of Vincent & Rainforth (2017), we used the public code provided at `https://github.com/drbenvincent/badapted`. We used 50 PMC steps with 100 particles. For the baselines of Frye et al. (2016) and Kirby (2009), we used the public code provided at `https://github.com/drbenvincent/darc-experiments-matlab/tree/master/darc-experiments`, which we reimplemented in Python. These methods do not involve a pre-training step, except that we did not include time to compute the first design ξ_1 within the speed test, as this can be computed before the start of the experiment.

To implement the deployment speed tests fairly, we ran each method on a lightweight CPU-only machine, which more closely mimics the computer architecture that we might expect to deploy methods such as DAD on. The specifications of the machine we used are described below

Memory	7.7GiB
Processor	Intel Core M-5Y10c CPU @ 0.80GHz $\times$ 4
Operating System	Ubuntu 16.04 LTS

The values in Table 2 show the mean and standard error of the times observed from 10 independent runs on a idle system. To make the final evaluation for each method in Table 3, we computed the sPCE and sNMC bounds using $L = 5000$ inner samples and 10000 outer samples of the outer expectation. We present the mean and standard error from the outer expectation over 10000 rollouts.

D.2.1. Ablation: total enumeration

We compare the two methods for estimating gradients for the case of discrete observations: total enumeration of histories (Equation 15) and score function gradient estimator (Equation 16). To this end we train DAD networks to perform $T = 10$ experiments, which gives rise to a total of $2^{10} = 1024$ possible histories.

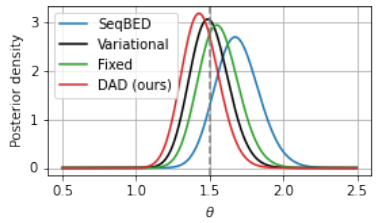

Figure 8. Comparison of posteriors obtained from a single rollout of the Death Process, used to compute the information gains quoted in Section 6.3. The dashed line indicates the true value $\theta = 1.5$ used to simulate responses.

Find that the two methods perform the same, both quantitatively and qualitatively. Table 5 reports the estimated upper and lower bounds on the mutual information objective, indicating statistically equal performance of the two methods (mean estimates are within 2 standard errors of each other). Figure 7 demonstrates the qualitative similarity in the designs learnt by the two networks.

	Lower bound, $\mathcal{L}_{10}$	Upper bound, $\mathcal{U}_{10}$
Complete enumeration	4.068 ± 0.0124	4.090 ± 0.0126
Score function gradient	4.037 ± 0.0126	4.058 ± 0.0128

Table 5. Final lower and upper bounds on the total information $\mathcal{I}_{10}(\pi)$ for the Hyperbolic Temporal Discounting experiment with $T = 10$ experiments and different gradient estimation schemes (see § 4.2 and § C.1 for details). The bounds are finite sample estimates of $\mathcal{L}_{10}(\pi, L)$ and $\mathcal{U}_{10}(\pi, L)$ with $L = 5000$. The errors indicate ± 1 s.e. over the sampled histories.

D.3. Death process

For the Death Process model (Cook et al., 2008), we use the settings that were described by Kleinegesse et al. (2020). Specifically, we use a truncated Normal prior for the infection rate

$$\theta \sim \text{TruncatedNormal}(\mu = 1, \sigma = 1, \min = 0, \max = \infty). \tag{133}$$

The likelihood is then given by

$$\eta = 1 - \exp(-\xi\theta) \qquad y|\theta, \xi \sim \text{Binomial}(N, \eta) \tag{134}$$

where we set $N = 50$. We consider a sequential version of this experiment as in Kleinegesse et al. (2020), with $T = 4$ and in which an independent stochastic process is observed at each step, meaning there are no constraints relating $\xi_1, ..., \xi_4$ other than the natural constraint $\xi_t > 0$.

We began by training a DAD network to perform experimental design for this problem. We used the neural architecture outlined in Section 4.3. For the encoder E_{ϕ_1} we used the following network with two hidden layers

Layer	Description	Dimension	Activation
Input	ξ, y	2	-
H1	Fully connected	128	Softplus
H2	Fully connected	128	Softplus
Output	Fully connected	16	-

The emitter network F_{ϕ_2} similarly used two hidden layers as follows

Layer	Description	Dimension	Activation
Input	$R(h_t)$	16	-
H1	Fully connected	128	Softplus
H2	Fully connected	128	Softplus
Output	ξ	1	Softplus

Although the number of experiments we perform is relatively small ($T = 4$), we could not use complete enumeration due to the prohibitively large size of the outcome space ($|\mathcal{Y}| = 51$). Hence, we constructed a score function gradient estimator of (16) (see also § C.1 for details) and optimized the DAD network with Adam (Kingma & Ba, 2014).We used exponential learning rate annealing with parameter γ. Full details are given in the following table.

Parameter	Value
Inner samples, L	500
Outer samples	500
Initial learning rate	0.001
Betas	$(0.9, 0.999)$
γ	0.96
Gradient steps	100000
Annealing frequency	1000

For the fixed baseline, we used the same optimization settings, except we set the initial learning rate to 10^{-1} and we set $\gamma = 0.85$. We trained the DAD and fixed methods using the same machine as used for training in Section D.2. For the variational baseline, we used a truncated Normal variational family to approximate the posterior at each step. We used SGD with momentum to optimize the design at each step, and to optimize the variational approximation to the posterior at each step. We used exponential learning rate annealing with paramter γ. The settings used were

Parameter	Value
Design inner samples	250
Design outer samples	250
Design initial learning rate	10^{-2}
Design γ	0.9
Design gradient steps	5000
Inference initial learning rate	10^{-3}
Inference γ	0.2
Inference gradient steps	5000
Momentum	0.1
Annealing frequency	1000

For the SeqBED baseline, we used the code publicly available at `https://github.com/stevenkleinegesse/seqbed`. The speed tests except for SeqBED were implemented as in Section D.2. For SeqBED and the variational method, we did not include the time to compute the first design as deployment time, as this can be computed before the start of the experiment. Due to its long-running nature, we implemented the speed test for SeqBED using a more powerful machine with 40 Intel(R) Xeon(R) CPU E5-2680 v2 @ 2.80GHz processors and 189GiB memory. Therefore, the timing value for SeqBED given in Table 4 represents a significant *under-estimate* of the expected computational time required to deploy this method. However, we note that SeqBED can be applied to a broader class of implicit likelihood models.

For evaluation of $\mathcal{I}_4(\pi)$ in the Death Process, it is possible to compute the information gain $H[p(\theta)] - H[p(\theta|h_T)]$ to high accuracy using numerical integration. We then took the expectation of the information gain over rollouts, see Table 4 for the exact number of rollouts used. This gives us an estimate

$$\mathcal{I}_4(\pi) = \mathbb{E}_{p(h_T|\pi)}\left[H[p(\theta)] - H[p(\theta|h_T)]\right] \tag{135}$$

which is shown to be a valid form for the total EIG in Section A.

For a comparison with SeqBED which is too slow to use this evaluation, we instead performed one rollout of each of our methods using a fixed value $\theta = 1.5$. This is close in spririt to the evaluation used in Kleinegesse et al. (2020). Figure 8 shows the posterior distributions obtained from this rollout. The information gains were then computed using the aforementioned numerical integration and are quoted in Section 6.3. We observe that, visually, the posterior distributions are similar, and cluster near to the true value of θ.

References

Amzal, B., Bois, F. Y., Parent, E., and Robert, C. P. Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical association*, 101(474):773–785, 2006.

Angelova, J. A. On moments of sample mean and variance. *Int. J. Pure Appl. Math*, 79(1):67–85, 2012.

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., and Siskind, J. M. Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, 18, 2018.

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P., Horsfall, P., and Goodman, N. D. Pyro: Deep universal probabilistic programming. *Journal of Machine Learning Research*, 2018.

Bloem-Reddy, B. and Teh, Y. W. Probabilistic symmetry and invariant neural networks. *arXiv preprint arXiv:1901.06082*, 2019.

Cappé, O., Guillin, A., Marin, J.-M., and Robert, C. P. Population monte carlo. *Journal of Computational and Graphical Statistics*, 13(4):907–929, 2004.

Chaloner, K. and Verdinelli, I. Bayesian experimental design: A review. *Statistical Science*, pp. 273–304, 1995.

Cook, A. R., Gibson, G. J., and Gilligan, C. A. Optimal observation times in experimental epidemic processes. *Biometrics*, 64(3):860–868, 2008.

Cremer, C., Li, X., and Duvenaud, D. Inference suboptimality in variational autoencoders. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1078–1086. PMLR, 2018.

Critchfield, T. S. and Kollins, S. H. Temporal discounting: Basic research and the analysis of socially important behavior. *Journal of applied behavior analysis*, 34(1):101–122, 2001.

Del Moral, P., Doucet, A., and Jasra, A. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

Drovandi, C. C., McGree, J. M., and Pettitt, A. N. A sequential monte carlo algorithm to incorporate model uncertainty in bayesian sequential design. *Journal of Computational and Graphical Statistics*, 23(1):3–24, 2014.

Dushenko, S., Ambal, K., and McMichael, R. D. Sequential bayesian experiment design for optically detected magnetic resonance of nitrogen-vacancy centers. *Physical Review Applied*, 14(5):054036, 2020.

Edwards, H. and Storkey, A. Towards a neural statistician. *arXiv preprint arXiv:1606.02185*, 2016.

Evans, J. R. and Mathur, A. The value of online surveys. *Internet research*, 2005.

Foster, A., Jankowiak, M., Bingham, E., Horsfall, P., Teh, Y. W., Rainforth, T., and Goodman, N. Variational Bayesian Optimal Experimental Design. In *Advances in Neural Information Processing Systems 32*, pp. 14036–14047. Curran Associates, Inc., 2019.

Foster, A., Jankowiak, M., O'Meara, M., Teh, Y. W., and Rainforth, T. A unified stochastic gradient approach to designing bayesian-optimal experiments. volume 108 of *Proceedings of Machine Learning Research*, pp. 2959–2969, Online, 26–28 Aug 2020. PMLR.

Frye, C. C., Galizio, A., Friedel, J. E., DeHart, W. B., and Odum, A. L. Measuring delay discounting in humans using an adjusting amount task. *JoVE (Journal of Visualized Experiments)*, (107):e53584, 2016.

Garnelo, M., Rosenbaum, D., Maddison, C. J., Ramalho, T., Saxton, D., Shanahan, M., Teh, Y. W., Rezende, D. J., and Eslami, S. Conditional neural processes. *arXiv preprint arXiv:1807.01613*, 2018a.

Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Eslami, S., and Teh, Y. W. Neural processes. *arXiv preprint arXiv:1807.01622*, 2018b.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

Goda, T., Hironaka, T., and Kitade, W. Unbiased mlmc stochastic gradient-based optimization of bayesian experimental designs. *arXiv preprint arXiv:2005.08414*, 2020.

González, J., Osborne, M., and Lawrence, N. Glasses: Relieving the myopia of bayesian optimisation. In *Artificial Intelligence and Statistics*, pp. 790–799. PMLR, 2016.

Green, L. and Myerson, J. A discounting framework for choice with delayed and probabilistic rewards. *Psychological bulletin*, 130(5):769, 2004.

Hainy, M., Drovandi, C. C., and McGree, J. M. Likelihood-free extensions for bayesian sequentially designed experiments. In Kunert, J., Müller, C. H., and Atkinson, A. C. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, pp. 153–161. Springer International Publishing, 2016.

Huan, X. and Marzouk, Y. Gradient-based stochastic optimization methods in bayesian experimental design. *International Journal for Uncertainty Quantification*, 4(6), 2014.

Huan, X. and Marzouk, Y. M. Sequential bayesian optimal experimental design via approximate dynamic programming. *arXiv preprint arXiv:1604.08320*, 2016.

Jiang, S., Chai, H., Gonzalez, J., and Garnett, R. Binoculars for efficient, nonmyopic sequential experimental design. In *International Conference on Machine Learning*, pp. 4794–4803. PMLR, 2020.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Kingma, D. P. and Welling, M. Auto-encoding variational Bayes. In *ICLR*, 2014.

Kirby, K. N. One-year temporal stability of delay-discount rates. *Psychonomic bulletin & review*, 16(3):457–462, 2009.

Kleinegesse, S. and Gutmann, M. Bayesian experimental design for implicit models by mutual information neural estimation. In *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pp. 5316–5326. PMLR, 2020. URL http://proceedings.mlr.press/v119/kleinegesse20a.html.

Kleinegesse, S., Drovandi, C., and Gutmann, M. U. Sequential bayesian experimental design for implicit models via mutual information. *arXiv preprint arXiv:2003.09379*, 2020.

Kruschke, J. Doing bayesian data analysis: A tutorial with r, jags, and stan. 2014.

Lewi, J., Butera, R., and Paninski, L. Sequential optimal design of neurophysiology experiments. *Neural Computation*, 21 (3):619–687, 2009.

Lindley, D. V. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pp. 986–1005, 1956.

Long, Q., Scavino, M., Tempone, R., and Wang, S. Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Computer Methods in Applied Mechanics and Engineering*, 259:24–39, 2013.

Lyu, J., Wang, S., Balius, T. E., Singh, I., Levit, A., Moroz, Y. S., OMeara, M. J., Che, T., Algaa, E., Tolmachova, K., et al. Ultra-large library docking for discovering new chemotypes. *Nature*, 566(7743):224, 2019.

Mazur, J. E. An adjusting procedure for studying delayed reinforcement. *Commons, ML.; Mazur, JE.; Nevin, JA*, pp. 55–73, 1987.

Mohamed, S., Rosca, M., Figurnov, M., and Mnih, A. Monte carlo gradient estimation in machine learning. *Journal of Machine Learning Research*, 21(132):1–62, 2020.

Müller, P. Simulation based optimal design. *Handbook of Statistics*, 25:509–518, 2005.

Myung, J. I., Cavagnaro, D. R., and Pitt, M. A. A tutorial on adaptive design optimization. *Journal of mathematical psychology*, 57(3-4):53–67, 2013.

Nowozin, S. Debiasing evidence approximations: On importance-weighted autoencoders and jackknife variational inference. In *International Conference on Learning Representations*, 2018.

Pasek, J. and Krosnick, J. A. Optimizing survey questionnaire design in political science. In *The Oxford handbook of American elections and political behavior*. 2010.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.

Poole, B., Ozair, S., van den Oord, A., Alemi, A., and Tucker, G. On variational bounds of mutual information. In *International Conference on Machine Learning*, pp. 5171–5180, 2019.

Rainforth, T. *Automating Inference, Learning, and Design using Probabilistic Programming*. PhD thesis, University of Oxford, 2017.

Rainforth, T., Cornish, R., Yang, H., Warrington, A., and Wood, F. On nesting monte carlo estimators. In *International Conference on Machine Learning*, pp. 4267–4276. PMLR, 2018.

Robbins, H. and Monro, S. A stochastic approximation method. *The annals of mathematical statistics*, pp. 400–407, 1951.

Ryan, E. G., Drovandi, C. C., McGree, J. M., and Pettitt, A. N. A review of modern computational algorithms for bayesian optimal design. *International Statistical Review*, 84(1):128–154, 2016.

Sekar, R., Rybkin, O., Daniilidis, K., Abbeel, P., Hafner, D., and Pathak, D. Planning to explore via self-supervised world models. In *International Conference on Machine Learning*, pp. 8583–8592. PMLR, 2020.

Senarathne, S., Drovandi, C., and McGree, J. A laplace-based algorithm for bayesian adaptive design. *Statistics and Computing*, 30(5):1183–1208, 2020.

Sheng, X. and Hu, Y. H. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *IEEE Transactions on Signal Processing*, 2005. ISSN 1053587X. doi: 10.1109/TSP.2004.838930.

Stuhlmüller, A., Taylor, J., and Goodman, N. Learning stochastic inverses. In *Advances in neural information processing systems*, pp. 3048–3056, 2013.

Tucker, G., Mnih, A., Maddison, C. J., Lawson, D., and Sohl-Dickstein, J. Rebar: Low-variance, unbiased gradient estimates for discrete latent variable models. *arXiv preprint arXiv:1703.07370*, 2017.

van den Oord, A., Li, Y., and Vinyals, O. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Vanlier, J., Tiemann, C. A., Hilbers, P. A., and van Riel, N. A. A Bayesian approach to targeted experiment design. *Bioinformatics*, 28(8):1136–1142, 2012.

Vincent, B. T. Hierarchical bayesian estimation and hypothesis testing for delay discounting tasks. *Behavior research methods*, 48(4):1608–1620, 2016.

Vincent, B. T. and Rainforth, T. The DARC toolbox: automated, flexible, and efficient delayed and risky choice experiments using bayesian adaptive design. 2017.

Williams, R. J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 1992. ISSN 0885-6125. doi: 10.1007/bf00992696.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R., and Smola, A. Deep sets. *arXiv preprint arXiv:1703.06114*, 2017.

Zheng, S., Pacheco, J., and Fisher, J. A robust approach to sequential information theoretic planning. In *International Conference on Machine Learning*, pp. 5941–5949, 2018.

Zheng, S., Hayden, D., Pacheco, J., and Fisher III, J. W. Sequential bayesian experimental design with variable cost structure. *Advances in Neural Information Processing Systems*, 33, 2020.