We would like to thank all the reviewers for their thoughtful reviews and helpful suggestions. We were delighted to see that the paper was unanimously well received and were particularly happy to see that the reviewers agreed that the work has the potential to make a big impact. We are excited about variational BOED and follow-up work indicates that VBOED opens the door to further developments in machine learning, statistics and other fields. We turn now to specific comments and questions.

Reviewer 1 Thank you for your review.

1. More concrete examples in Section 2 is a great suggestion which we will implement in time for the camera ready, if accepted. To be specific, in the psychology trial example, the design d is the choice of question, θ represents the parameters of an underlying psychological model $p(y|\theta, d)$, and y is the participant's response.

Reviewer 2 Thank you for your review.

1. Thank you for pointing out our mistake with the reference for the variational marginal bound. We will be sure to correct this.
2. We are glad you brought the issue of high-level intuition for VNMC to our attention and we will make updates to be clearer here. To give some additional explanation, both the NMC and VNMC estimators take the form

$$\text{EIG}(d) \approx \frac{1}{N}\sum_{n=1}^{N} \log \frac{p(y_n|\theta_n, d)}{\hat{p}(y_n|d)} \text{ where } y_n, \theta_n \overset{\text{i.i.d.}}{\sim} p(\theta)p(y|\theta, d) \tag{1}$$

where, for NMC, $\hat{p}_{\text{NMC}}(y|d) = \frac{1}{M}\sum_{m=1}^{M} p(y|\theta_m, d)$ where $\theta_m \overset{\text{i.i.d.}}{\sim} p(\theta)$. Written in this way, we see that NMC is approximating $p(y|d)$ using samples from $p(\theta)$. We expect better approximations of $p(y|d)$ using samples from a proposal $q_v(\theta|y)$ that is close to the posterior $p(\theta|y, d)$, i.e.

$$\hat{p}_{\text{VNMC}}(y|d) = \frac{1}{M}\sum_{m=1}^{M} \frac{p(\theta)p(y|\theta_m, d)}{q_v(\theta|y)} \text{ where } \theta_m \overset{\text{i.i.d.}}{\sim} q_v(\theta|y) \tag{2}$$

which leads to the VNMC estimator. It is also important to establish the bounds of Lemma 1, because these allow a variational training method for q_v.
3. We agree that Poole, et al. (2019) is an important reference and will make sure it is discussed in the main text.
4. Table 1: thanks for picking this up. We agree a pointer to Section 5 would be helpful.
5. A1 and A2: we will be sure to indicate that these proofs were added for completeness and add references.
6. A3: thanks for picking up these typos!

Reviewer 3 Thank you for your review.

1. This is an interesting point where we could have been clearer. In the sequential setting, we assume that

$$p(y_{1:t}, \theta|d_{1:t}) = p(\theta)\prod_{\tau=1}^{t} p(y_\tau|\theta, d_\tau) \tag{3}$$

which says that experiments are conditionally independent given designs and θ. After conducting experiments $1, ..., t-1$ we have $p(y_t, \theta|y_{1:t-1}, d_{1:t-1}, d_t) = p(\theta|y_{1:t-1}, d_{1:t-1})p(y_t|\theta, d_t)$ and now select d_t conditional on $d_{1:t-1}, y_{1:t-1}$ using the new prior $p(\theta|d_{1:t-1}, y_{1:t-1})$. The entropy of this new prior distribution is a constant with respect to d_t which is why we can drop it on line 169. The new prior still makes its presence felt in the other term in $\mathcal{L}_{\text{post}}$, namely $\mathbb{E}_{p(\theta|d_{1:t-1}, y_{1:t-1})p(y|\theta, d_t)}[\log q_p(\theta|y, d_t)]$.
2. We agree that the lower bias of $\mu_{m+\ell}$ compared to μ_{post} may at first sight be unintuitive. Although $\mu_{m+\ell}$ uses two variational approximations compared to one for μ_{post}, the approximations are for variables which have different dimensionality. If y has a lower dimension than θ, it may make sense to use $\mu_{m+\ell}$ instead of μ_{post}. On the other hand, $\mu_{m+\ell}$ uses the same approximation as μ_{marg} plus an extra one. We would never recommend using $\mu_{m+\ell}$ in place of μ_{marg} (cf. line 186), but one may have to fall back on $\mu_{m+\ell}$ in an implicit likelihood setting.

   In our experiments, parameters were not shared between q_m and q_ℓ, although this is an interesting idea that could further reduce the bias. A limited discussion of this idea appears at the end of Section A.4 (it can actually lead to a new lower bound, but requires additional assumptions).

References

Poole, Ben, et al. "On variational bounds of mutual information." arXiv preprint arXiv:1905.06922 (2019).