

---

# qEUBO: A Decision-Theoretic Acquisition Function for Preferential Bayesian Optimization

---

Raul Astudillo  
Caltech

Zhiyuan Jerry Lin  
Meta

Eytan Bakshy  
Meta

Peter I. Frazier  
Cornell University

## Abstract

Preferential Bayesian optimization (PBO) is a framework for optimizing a decision maker’s latent utility function using preference feedback. This work introduces the expected utility of the best option (qEUBO) as a novel acquisition function for PBO. When the decision maker’s responses are noise-free, we show that qEUBO is one-step Bayes optimal and thus equivalent to the popular knowledge gradient acquisition function. We also show that qEUBO enjoys an additive constant approximation guarantee to the one-step Bayes-optimal policy when the decision maker’s responses are corrupted by noise. We provide an extensive evaluation of qEUBO and demonstrate that it outperforms the state-of-the-art acquisition functions for PBO across many settings. Finally, we show that, under sufficient regularity conditions, qEUBO’s Bayesian simple regret converges to zero at a rate  $o(1/n)$  as the number of queries,  $n$ , goes to infinity. In contrast, we show that simple regret under qEI, a popular acquisition function for standard BO often used for PBO, can fail to converge to zero. Enjoying superior performance, simple computation, and a grounded decision-theoretic justification, qEUBO is a promising acquisition function for PBO.

## 1 INTRODUCTION

Bayesian optimization (BO) is a framework for global optimization of objective functions with expensive or time-consuming evaluations (Shahriari et al., 2015). BO algorithms have been successful in broad range of applications, such as sensor set selection (Garnett et al., 2010), hyperparameter tuning of machine learning algorithms (Snoek

et al., 2012), chemical design (Griffiths and Hernández-Lobato, 2020), and culture media optimization for cellular agriculture (Cosenza et al., 2022). In many problems, it is not possible to observe (potentially noisy) objective values directly. Instead, a decision-maker (DM) provides preference feedback, often in the form of pairwise comparisons between options shown. This arises in applications such as animation design (Brochu et al., 2010), where a DM is shown two different images and chooses the one with better characteristics (e.g. realism or resemblance to a target image); and exoskeleton gait design (Tucker et al., 2020), where a DM assisted by an exoskeleton walks for a short period of time using two different gait configurations and indicates the one that resulted in more comfortable walking. Preferential Bayesian optimization (PBO) (Brochu et al., 2010; González et al., 2017), a subframework within BO, has emerged as a powerful tool for tackling such problems.

As in standard BO, a PBO algorithm consists of two main components: a probabilistic surrogate model of the DM’s latent utility function; and an acquisition function (AF), computed from the probabilistic surrogate model, whose value at a given set of  $q$  alternatives quantifies the benefit of DM feedback about their preferred alternative in the set. Several AFs for PBO have been proposed (Brochu et al., 2010; González et al., 2017; Benavoli et al., 2021; Siivola et al., 2021; Nguyen et al., 2021). However, most are derived from heuristic arguments and lack a proper decision-theoretic or information-theoretic justification. For example, Brochu et al. (2010) selects the point that maximizes the posterior mean of the model over points in previous queries as the first alternative, and the point that maximizes the expected improvement with respect to the posterior mean value of the first point as the second alternative. Other works simply adopt AFs from the standard BO literature (Siivola et al., 2021), ignoring the fact that preference feedback is observed rather than direct utility values.

To address the shortcomings of existing approaches, we study the *expected utility of the best option* (qEUBO), which generalizes the EUBO AF proposed by Lin et al. (2022) for a different problem setting, as a novel AF for PBO with a proper decision-theoretic justification.

**Contributions** Our contributions are as follows:

- We propose qEUBO, an AF for PBO. qEUBO has a sound decision-theoretic interpretation, is simple to compute, and exhibits strong empirical performance.
- We show that qEUBO outperforms the state-of-the-art AFs for PBO in several synthetic and realistic test problems. Moreover, we show that qEUBO’s closest competitor performs well in early iterations because it is *similar* to qEUBO but its performance degrades as the number of queries grows.
- We show that, under sufficient regularity conditions, qEUBO’s Bayesian simple regret converges to zero at a rate  $o(1/n)$  as the number of queries,  $n$ , goes to infinity. Moreover, we show there exist problem instances where qEI, a popular acquisition from the standard BO setting that is often used in the PBO setting, has Bayesian simple regret bounded *below* by a strictly positive constant.
- We demonstrate significant benefit of asking queries with more than two alternatives. This contrasts with previous work by Siivola et al. (2021), which concluded that  $q > 2$  only provides limited performance improvement.

## 2 RELATED WORK

Several AFs for PBO have been proposed in the literature. Most of them are designed via heuristic arguments (Brochu et al., 2010) or simply reused from the standard BO setting (Siivola et al., 2021). For example, Brochu et al. (2010) selects the point that maximizes the posterior mean of the model over points in previous queries as the first alternative, and the point that maximizes the expected improvement with respect to the posterior mean value of the first point as the second alternative. Nielsen et al. (2014) proposes to use the point preferred by the user in the previous query as the first alternative, and the point that maximizes the expected improvement with respect to this point as the second alternative. For  $q = 2$ , qEUBO recovers this AF if we force the first alternative to be equal to the point preferred by the user in the previous query and optimize only over the second alternative. González et al. (2017) proposes a pure exploration sampling policy along with two AFs based on the expected improvement and Thompson sampling AFs that aim to maximize the soft-Copeland’s score. However, the computation of this score requires integration over the optimization domain, thus making these algorithms intractable even for problems of moderate dimension. Siivola et al. (2021) proposes using *batch* versions of the expected improvement and Thompson sampling AFs from standard BO for selecting the points in each query. Since utility values are not observed directly, the batch expected improvement

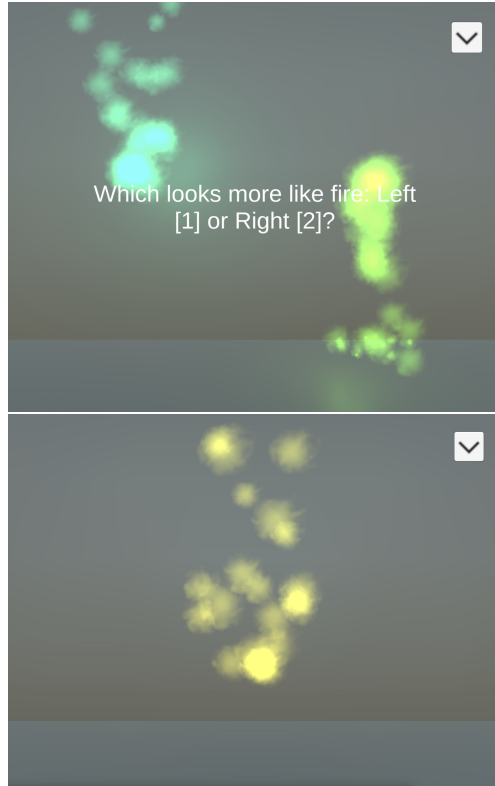


Figure 1: Fire particle rendering problem from Section 5, in which a human user is asked which of two animations looks more like fire (top). Final rendering results based on fitting a support vector machine model to 100 comparisons between random particle effects and then optimizing the predicted latent decision function over animation parameters (bottom).

is adapted by defining the improvement with respect to the maximum posterior mean value over points in previous queries, along the lines of the approach followed by Brochu et al. (2010). Batch Thompson sampling is defined as in the standard BO setting: each point in the batch is selected as the point that maximizes an independently drawn sample from the utility’s posterior distribution. Nguyen et al. (2021) proposes the multinomial predictive entropy search (MPES) AF for top- $k$  ranking BO, a slightly more general framework where the DM selects her most  $k$  preferred alternatives in each query. MPES selects the query that maximizes the information gain on the utility function’s maximizer through observing the DM’s feedback. It can be seen as a principled adaptation of the predictive entropy search (PES) AF for standard BO (Hernández-Lobato et al., 2014). Like with PES, the computation of MPES requires approximating an intractable multi-dimensional integral with respect to the posterior distribution on the utility function’s maximizer. This is computationally challenging, especially in problems of moderate dimension, and inaccurate approximation can lead to a degraded performance. To our knowledge, the



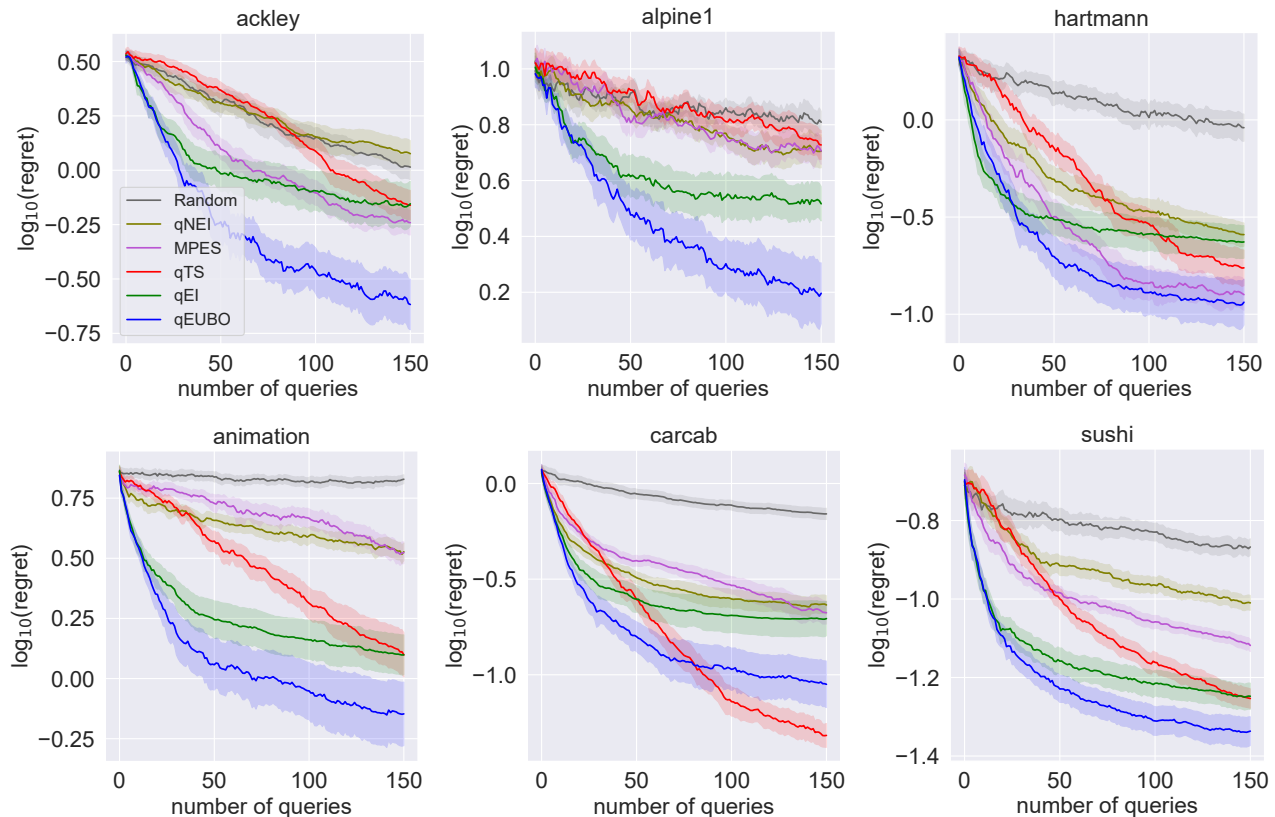


Figure 2:  $\log_{10}(\text{optimum value} - \text{utility value at the maximizer of the posterior mean})$  using moderate logistic noise and  $q = 2$  comparisons per DM query. All algorithms are shown up to 150 queries. qEUBO outperforms other algorithms on all but one problem.

AFs proposed by Siivola et al. (2021) and Nguyen et al. (2021) are the only existing ones allowing for queries with more than two alternatives. Finally, Benavoli et al. (2021) proposes an AF where the first alternative is the point chosen by the user in the previous query, and the second one is obtained by maximizing a linear combination between the logarithm of the probability of improvement and the information gain; the weight of this linear combination is a hyperparameter of the algorithm. It also proposes two other AFs based on Thompson sampling and GP-UCB (Srinivas et al., 2012).

The above AFs (except MPES) were derived via heuristic arguments. In contrast, qEUBO is derived following a principled decision-theoretic analysis modeling the fact that, in PBO, observations are comparisons instead of direct utility values. qEUBO’s approach is consistent with the rigorous decision-theoretic or information-theoretic analysis used to derive principled AFs in standard BO. Moreover, unlike MPES, qEUBO is easy to compute and comes with a convergence guarantee. Finally, qEUBO outperforms MPES significantly in our empirical evaluation.

qEUBO, restricted to the case  $q = 2$ , was first discussed by Lin et al. (2022) in the context of preference exploration for

multi-attribute BO. In this context, the DM does not express preferences directly over alternatives but over attributes of these alternatives, which are assumed to be time-consuming to evaluate. As a consequence, qEUBO is not used directly to find alternatives to show to the DM. Instead, it is combined with a probabilistic surrogate model of the mapping from alternatives to attributes to find hypothetical attribute vectors over which the DM expresses preferences. Our work places qEUBO in the context of PBO and extends its definition to queries with  $q > 2$  alternatives. We also generalize the analysis of Lin et al. (2022) by showing that maximizing qEUBO recovers a one-step optimal solution when responses are noise-free for  $q > 2$ . Finally, we provide a novel convergence analysis for qEUBO.

The connection between qEUBO and the one-step Bayes optimal policy relates qEUBO to the knowledge gradient class of sampling policies for sequential data collection, which are, by definition, one-step Bayes optimal (Frazier et al., 2008). Knowledge gradient AFs have been widely used in standard BO (Wu and Frazier, 2016; Wu et al., 2017; Cakmak et al., 2020) and are known for their superior performance over simpler AFs such as expected improvement or Thompson sampling, especially when observations are

noisy or the objective function is highly multi-modal (Wu and Frazier, 2016; Balandat et al., 2020) or when these simpler AFs are used in settings where they lack a meaningful interpretation. At the same time, they are typically very challenging to maximize (Balandat et al., 2020), often resulting in high computation times and degraded performance in problems of moderate dimension. The former is particularly problematic in the PBO setting where queries are often required to be generated in real time. Since qEUBO is significantly simpler to compute, we effectively overcome the computational burden commonly faced by knowledge gradient AFs. While this equivalence does not hold anymore when responses are noisy, we show that qEUBO still enjoys an additive constant approximation guarantee to the one-step Bayes optimal policy.

Our work is also related to the literature on dueling-bandits (Yue et al., 2012; Bengs et al., 2021). Like in our problem setting, the DM is assumed to express preference feedback over sets (typically pairs) of alternatives. However, most of these approaches assume a finite number of alternatives and often also independence across pairs of alternatives. The double Thomson sampling strategy for dueling bandits proposed by Wu and Liu (2016) is analogous to the Thompson sampling AF for PBO proposed by Siivola et al. (2021).

Finally, our work is also related to a broader stream of research on computational preference elicitation (Braziunas, 2006). This work focuses on problems with a finite set of alternatives, where it suffices to estimate a ranking of the alternatives, or on estimating the underlying DM’s utility function within a parametric class of functions. In particular, our work is closely related to Viappiani and Boutilier (2010), which derived an analogous result relating optimal recommendation sets with one-step Bayes optimal query sets. However, Viappiani and Boutilier (2010) proposes using optimal recommendation sets to query the DM, which differ from those queries selected by qEUBO when responses are noisy. As a consequence, our analysis under noisy responses also focuses on relating qEUBO to the one-step Bayes optimal policy rather than the policy that selects optimal recommendation sets.

### 3 PROBLEM SETTING

We denote the space of *alternatives* or *options* by  $\mathbb{X}$ . Succinctly, our goal is to find the best possible alternative in  $\mathbb{X}$  according to the DM’s underlying preferences. These preferences are encoded via a latent utility function,  $f : \mathbb{X} \rightarrow \mathbb{R}$ . We model  $f$  through a general Bayesian prior distribution. In our experiments we use a Gaussian process (GP) prior, but our derivation and analysis of qEUBO does not make this assumption and is applicable to more general priors.

At every interaction with the DM, an algorithm selects a *query*,  $X = (x_1, \dots, x_q) \in \mathbb{X}^q$ . The DM then expresses her most preferred alternative among these  $q$  points. This re-

sponse is denoted by  $r(X) \in \{1, \dots, q\}$ , where  $r(X) = i$  if  $x_i$  is the alternative chosen by the DM. The DM’s responses may be not be always consistent with the underlying utility function. We model this via a parametric likelihood function  $L(\cdot; \lambda) : \mathbb{R}^q \rightarrow \mathbb{R}^q$  such that

$$\mathbf{P}(r(X) = i \mid f(X)) = L_i(f(X); \lambda),$$

where  $L_i(f(X); \lambda)$  is the  $i$ -th component of  $L(f(X); \lambda)$  and  $\lambda$  is estimated along with other parameters of the model. Our numerical experiments and Theorem 2 assume a logistic likelihood function of the form

$$L_i(f(X); \lambda) = \frac{\exp(f(x_i)/\lambda)}{\sum_{j=1}^q \exp(f(x_j)/\lambda)},$$

for  $i = 1, \dots, q$ , where  $\lambda \geq 0$  is the *noise level* parameter. For  $\lambda = 0$ , the above expression is defined as its right-hand limit as  $\lambda$  converges to 0. It can be easily shown that  $\lambda = 0$  recovers a noise-free response likelihood. Theorem 3 allows for a broader class of likelihood functions. Details are provided in Section B.

Let  $\mathcal{D}^{(n)} = \{(X_m, r(X_m))\}_{m=1}^n$  denote the data collected after  $n$  queries and  $\mathbf{E}_n$  denote the conditional expectation given  $\mathcal{D}^{(n)}$ . Following the decision-theory literature, if we decide to stop at time  $N$ , we will recommend the point that maximizes the DM’s expected utility given the data collected so far; i.e., an element of  $\arg\max_{x \in \mathbb{X}} \mathbf{E}_N[f(x)]$ . Thus, we wish to select the queries  $X_1, \dots, X_N$  so that the expected utility received by the DM under our recommendation,  $\max_{x \in \mathbb{X}} \mathbf{E}_N[f(x)]$ , is as large as possible.

## 4 qEUBO

### 4.1 qEUBO and the One-Step Bayes Optimal Policy

To motivate our AF, we begin by discussing the one-step Bayes optimal policy, i.e., the policy that chooses at every iteration the query that would be optimal if it were the last one. To this end, we define for an arbitrary query  $X \in \mathbb{X}^q$ ,

$$V_n(X) = \mathbf{E}_n \left[ \max_{x \in \mathbb{X}} \mathbf{E}_{n+1}[f(x)] \mid X_{n+1} = X \right].$$

This is the expected utility received by the DM if one last additional query  $X_{n+1} = X$  is performed. The one-step Bayes optimal policy chooses at every iteration the query that maximizes  $V_n$ .

Since  $\max_{x \in \mathbb{X}} \mathbf{E}_n[f(x)]$  does not depend on  $X_{n+1}$ , maximizing  $V_n$  is equivalent to maximizing

$$\mathbf{E}_n \left[ \max_{x \in \mathbb{X}} \mathbf{E}_{n+1}[f(x)] - \max_{x \in \mathbb{X}} \mathbf{E}_n[f(x)] \mid X_{n+1} = X \right].$$

The above expression is analogous to the knowledge gradient AF from standard BO (Frazier et al., 2009; Scott et al.,

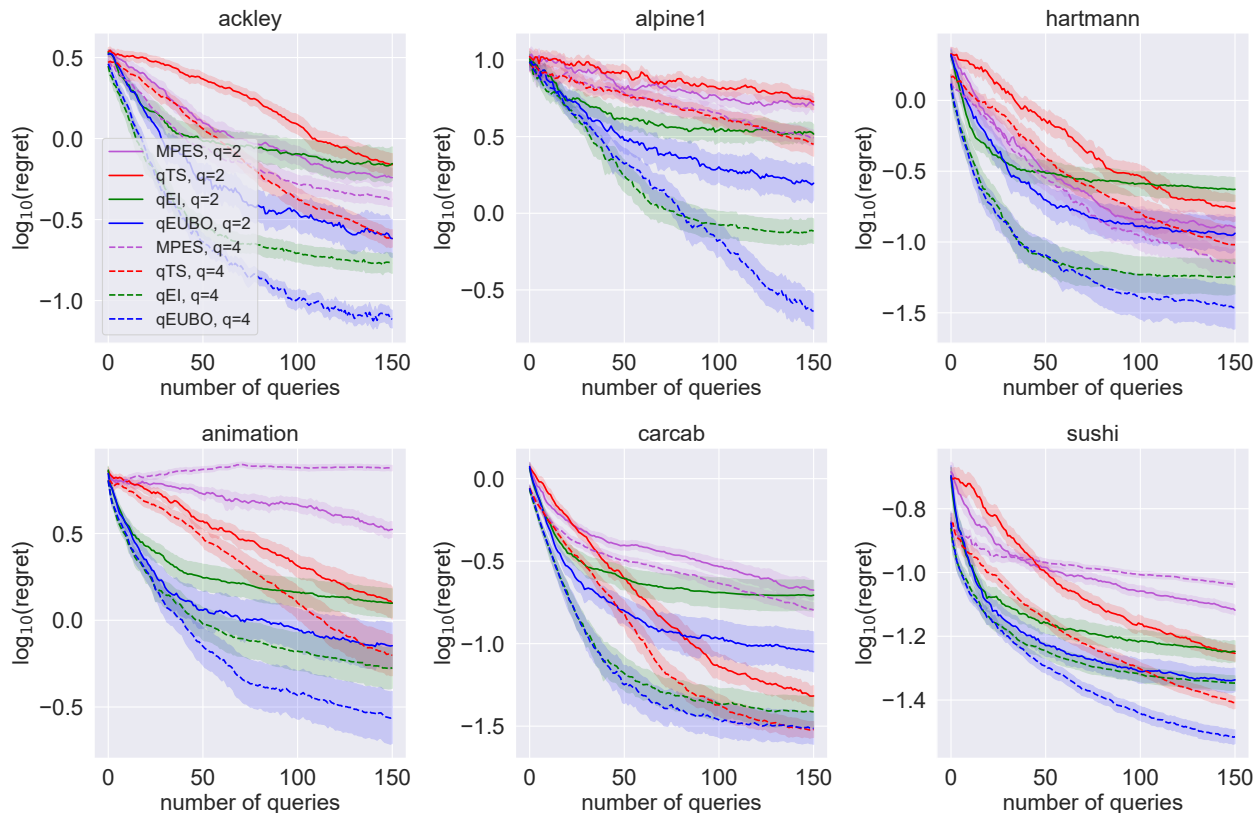


Figure 3:  $\log_{10}(\text{optimum value} - \text{utility value at the maximizer of the posterior mean})$  for  $q \in \{2, 4\}$ . Algorithms are shown up to 150 queries. qEUBO outperforms all other algorithms on all problems. Including more alternatives per query ( $q = 4$ ) allows regret to decline more quickly.

2011; Wu and Frazier, 2016). As mentioned earlier, knowledge gradient AFs often outperform simpler AFs. However they are also very challenging to maximize due to their nested expectation-maximization structure.

Our main result shows that, when the DM responses are noise-free, maximizing  $V_n$  is equivalent to maximizing a simpler AF. We define the *expected utility of the best option* (qEUBO) AF by

$$\text{qEUBO}_n(X) = \mathbf{E}_n [\max\{f(x_1), \dots, f(x_q)\}].$$

Under this definition, the following result holds.

**Theorem 1.** *Suppose the DM’s responses are noise-free. Then,*

$$\operatorname{argmax}_{X \in \mathbb{X}^q} \text{qEUBO}_n(X) \subseteq \operatorname{argmax}_{X \in \mathbb{X}^q} V_n(X).$$

Thus, to find a maximizer of  $V_n$ , it suffices to maximize  $\text{qEUBO}_n$ . This is a significantly simpler task as it does not require solving a nested stochastic optimization problem. When the posterior over  $f$  is Gaussian or approximated via a Gaussian distribution (e.g., via a Laplace approximation),  $\text{qEUBO}_n$  can be efficiently maximized via sample

average approximation (Balandat et al., 2020). This is the approach we pursue in our experiments. Moreover, if  $q = 2$ ,  $\text{qEUBO}_n$  has a closed form expression in terms of the posterior mean and covariance functions (Lin et al., 2022).

When the DM’s responses are noisy, maximizing  $\text{qEUBO}_n$  is no longer equivalent to maximizing  $V_n$ . However, the result below shows that if noise in the DM’s responses follows a logistic likelihood, maximizing  $\text{qEUBO}_n$  still recovers a high-quality query. Formally, we show the following.

**Theorem 2.** *Suppose that the DM’s responses follow the logistic likelihood function with parameter  $\lambda$  defined above. Denote  $V_n$  as  $V_n^\lambda$  to make its dependence on  $\lambda$  explicit. If  $X^* \in \operatorname{argmax}_{X \in \mathbb{X}^q} \text{qEUBO}_n(X)$ , then*

$$V_n^\lambda(X^*) \geq \max_{X \in \mathbb{X}^q} V_n^0(X) - \lambda C,$$

where  $C = L_W((q-1)/e)$ , and  $L_W$  is the Lambert W function (Corless et al., 1996).

The above two results extend those shown by Lin et al. (2022) to the logistic likelihood and  $q > 2$ . Their proofs can be found in Section A.

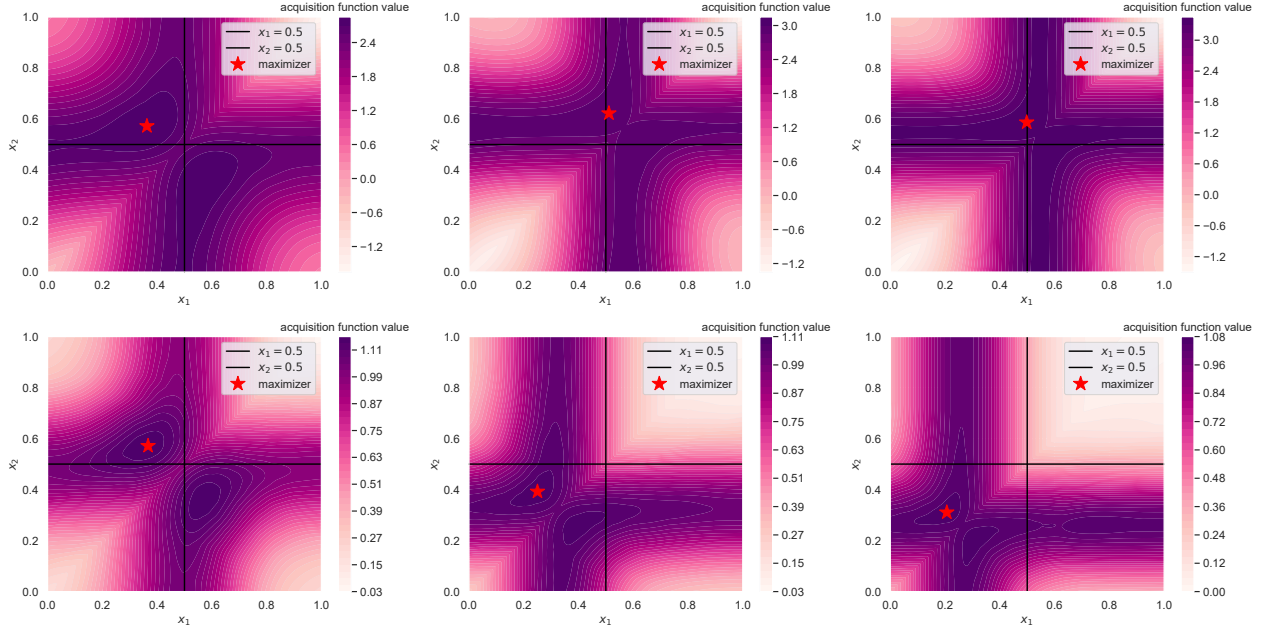


Figure 4: Comparison of qEUBO (top row) against qEI (bottom row) on a 1-dimensional problem (a quadratic function with a single maximum at  $x = 0.5$ ) with  $q = 2$ . The first column shows the AF over the two alternatives to be included in the query after training a preferential GP on 5 randomly generated queries. The second column shows the AF after 5 more queries, generated by the given AF. The third column shows the AF after 5 more queries generated via that row’s AF.

### 4.2 qEUBO and qEI

The batch expected improvement AF, commonly known as qEI, was developed in the context of parallel BO (Ginsbourger et al., 2008; Wang et al., 2016), where it enjoys a meaningful decision-theoretic interpretation. It was adapted to the PBO setting by Siivola et al. (2021). While qEI lacks a meaningful interpretation in the PBO setting, it often has good performance. Here, we show that qEUBO is related to qEI. This connection sheds light on qEI’s strong empirical performance as an AF for PBO. However, we also show qEI has significant drawbacks that cause it to have poor performance in some practical scenarios.

Observe that  $\text{qEUBO}_n(X) = \mathbf{E}_n[F(X)]$  where  $F : \mathbb{X}^q \rightarrow \mathbb{R}$  is defined by  $F(X) = \max_{i=1, \dots, q} \{f(x_1), \dots, f(x_q)\}$ . Moreover, if  $I_n$  is any value that does not depend on  $X$ , then maximizing  $\text{qEUBO}_n$  produces the same query as maximizing  $\mathbf{E}_n[F(X) - I_n]$ . Now observe that, if  $I_n = \max \{\mathbf{E}_n[f(x)] : x \in \cup_{m=1}^n X_m\}$ , then  $\text{qEI}_n(X) = \mathbf{E}_n\{[F(X) - I_n]^+\}$  recovers the qEI AF proposed by Siivola et al. (2021). From these expressions we observe that, if  $F(X)$  is typically larger than  $I_n$  so that  $F(X) - I_n = \{F(X) - I_n\}^+$ , then optimizing  $\text{qEUBO}_n(X)$  should produce an optimal  $X$  similar to the one obtained by optimizing  $\text{qEI}_n(X)$ . This is often the case in early iterations, when  $I_n$  is small. However, as  $I_n$  becomes larger, we expect this to occur less frequently, making qEI and qEUBO produce more different queries. Moreover, when they diverge, qEI can perform quite poorly, as we will see later.

When the incumbent alternative (the one whose posterior mean achieves  $I_n$ ) has low variance, as typically results from comparing a good alternative against many other alternatives, then qEI will become increasingly reluctant to include it or points near it into the next query. In standard BO, this reluctance is appropriate because re-measuring the incumbent will not generate an improvement. But, in PBO, there is great value in comparing an incumbent alternative to another alternative that might be better — this is a primary way that we evaluate new alternatives.

This is also consistent with experimental results discussed later in Section 5 and shown in Figures 2, 3 and 5: qEUBO and qEI tend to perform similarly early on when we have asked the DM few queries; later, qEI tends to stall, while qEUBO continues to reduce its simple regret. This intuition is also codified in Theorem 4 in the next section, which shows an example in which qEI fails to be consistent.

To illustrate this further, Figure 4 compares qEUBO and qEI on a simple 1-dimensional example problem (a quadratic function with a single maximum at  $x = 0.5$ ). For each AF, we first trained a preferential GP model using 5 randomly chosen comparisons (left column of Figure 4), then generated 5 more (middle column of Figure 4), and additionally 5 more comparisons (right column of Figure 4) using qEUBO (top rows) and qEI (bottom rows) respectively. After the first 5 randomly generated comparisons, the posteriors are the same, the contours of the two AFs are similar because  $I_n$  is small, and the two methods make similar queries. After 5

more queries, generated using each AF, qEUBO has already learned that 0.5 is a good solution and is comparing this with other alternatives. In contrast, qEI is choosing not to compare with 0.5. This pattern continues after 5 additional queries.

### 4.3 Convergence Analysis of qEUBO and qEI

We end this section by discussing the convergence properties of qEUBO and qEI. We show that, under sufficient regularity conditions, qEUBO’s Bayesian simple regret converges to zero at a rate  $o(1/n)$ . We also show that there are problem instances where qEI has Bayesian simple regret bounded below by a positive constant; in particular, qEI is not asymptotically consistent.

Our analysis assumes that  $\mathbb{X}$  is finite,  $q = 2$ , and other technical conditions described in Section B. These conditions hold in a broad range of settings. For example, they hold under the logistic likelihood function discussed above if the prior distribution on  $f$  is such that

$$\delta \leq |f(x) - f(y)| \leq \Delta$$

almost surely whenever  $x \neq y$  for some  $\Delta \geq \delta > 0$ . They also hold for general non-degenerate GP prior distributions if the likelihood function satisfies

$$L(f(X); \lambda) = a$$

for some fixed  $a > 1/2$  whenever  $f(x_1) \neq f(x_2)$ .

Under these conditions, we show the following results. Formal statements and proofs can be found in Section B.

**Theorem 3.** *Assume the sequence of queries is chosen by maximizing qEUBO and the assumptions described in Section B hold. Then,  $\mathbf{E}[f(x^*) - f(\hat{x}_n^*)] = o(1/n)$ , where  $x^* = \operatorname{argmax}_{x \in \mathbb{X}} f(x)$  and  $\hat{x}_n^* \in \operatorname{argmax}_{x \in \mathbb{X}} \mathbf{E}_n[f(x)]$ .*

**Theorem 4.** *There exists a problem instance (i.e.,  $\mathbb{X}$  and Bayesian prior distribution over  $f$ ) satisfying the assumptions described in Section B such that if the sequence of queries is chosen by maximizing qEI, then  $\mathbf{E}[f(x^*) - f(\hat{x}_n^*)] \geq R$  for all  $n$ , for a constant  $R > 0$ .*

The problem instance in which qEI fails to be consistent has the characteristics previously described in Section 4.2 — the incumbent, i.e., the alternative with the best posterior mean, also has known value. As a result, qEI is unwilling to include it in the queries asked. This makes qEI unable to learn about the value of other alternatives — it can learn about the relative value of other alternatives with *each other*, but not about their value relative to the incumbent.

## 5 EXPERIMENTS

We compare qEUBO with various state-of-the-art AFs for PBO from the literature. We consider MPES from Nguyen

et al. (2021), which, as described, is arguably the only existing PBO AF with a proper justification. We also consider qEI and batch Thompson sampling (qTS) from Siivola et al. (2021), which were both shown to have excellent empirical performance. We also consider qNEI, a version of qEI that accounts for the uncertainty in latent function values through Monte Carlo integration over fantasized values Balandat et al. (2020). qEUBO, qEI, and qNEI are optimized via sample average approximation with multiple restarts (Balandat et al., 2020). qTS uses approximate sample paths obtained via 1000 random Fourier features (Rahimi and Recht, 2007). For reference, we also include the performance of random search (Random), which selects queries uniformly at random over the space of alternatives. All algorithms use a Gaussian process prior with a constant mean function and RBF covariance function to model  $f$ . We approximate the posterior distribution over  $f$  via the variational inducing point approach introduced by Hensman et al. (2015). Our approach is equivalent to the one pursued by Nguyen et al. (2021) if we take the set of inducing points equal to the set of all points in the queries asked so far. Our set of inducing points includes these points in addition to a small set of quasi-random Sobol points, which improves performance slightly.

We report results across three synthetic test functions and three test functions built from real-world data. In contrast with most existing papers from the literature, which limit themselves to low-dimensional problems, we focus on more challenging problems of moderate dimension ( $> 3$ ). Synthetic functions include 6-dimensional Ackley, 7-dimensional Alpine1, and 6-dimensional Hartmann. Realistic problems include a 7-dimensional car cab design problem (Carcab) (Lin et al., 2022), a 4-dimensional problem involving real-world human preferences over 100 sushi items (Sushi) (Siivola et al., 2021), and a novel 5-dimensional animation optimization problem (Animation). Noise is added to simulate inconsistency in the DM’s responses.

To create our novel animation optimization problem, we use real human comparison data from a real-world particle effect rendering animation based on the publicly available demo in the AEPsych package (Owen et al., 2021). In this setting, a human user is asked to compare two rendered animations of particles side by side and to determine which one looks more like fire (Figure 1, top). The particle animation is parameterized by 5 parameters. We collected 100 such pairwise comparisons from human users with random particle animation parameters. We then confirmed that by fitting a support vector machine model on this data and optimizing, we are able to obtain a realistic fire-like particle effect. A screenshot of the resulting animation shown in the bottom of Figure 1. We then use this fitted model as the ground-truth test function to perform simulation.

In all problems, a first stage of interaction with the DM is performed using  $4d$  queries chosen uniformly at random



Problem/Acquisition function	qNEI	MPES	qTS	qEI	qEUBO
Ackley	8.1	24.8	6.5	6.3	12.4
Alpine1	11.4	16.4	6.4	8.9	11.3
Hartmann	8.7	15.3	7.1	6.0	8.3
Animation	9.4	13.5	8.6	7.4	8.2
Carcab	7.2	12.7	6.9	7.1	7.2
Sushi	8.9	23.9	9.5	5.9	7.5

Table 1: Average runtimes in seconds across all test problems. MPES is consistently the slowest algorithm, followed by qNEI. MPES is slow because it requires approximating an intractable integral involving the posterior distribution on the utility function’s maximizer. qTS and qEI are the fastest algorithms, followed closely by qEUBO.

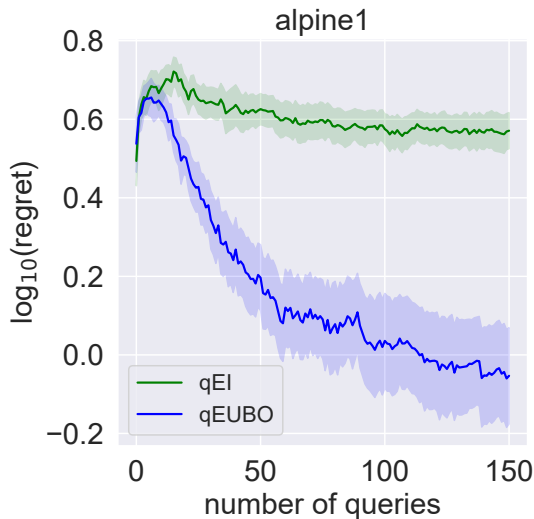


Figure 5: Comparison between qEUBO and qEI on the 7-dimensional Alpine1 function seeded with many comparisons between a good solution and other randomly chosen ones. This setting is similar to Theorem 4. When we have a reasonably good status quo solution whose value is known with high precision, qEI is unable to significantly reduce its simple regret while qEUBO steadily learns.

over  $\mathbb{X}^d$ , where  $d$  is the input dimension of the problem. After this initial stage, each algorithm was used to select 150 additional queries sequentially. Figures 2 and 3 show the mean of the log simple regret, plus and minus 1.96 times the standard deviation divided by the square root of the number of replications, as a function of the number of queries. Here, simple regret is defined as the maximum objective value minus the objective value at the maximizer of the posterior mean. We average over 100 replications for the Animation and Sushi problems and 50 replications for the other problems. Figure 2 shows results for  $q = 2$  for all algorithms. Figure 3 shows results for both  $q = 2$  and  $q = 4$  for MPES, qTS, qEI, and qEUBO; we only focus on the best-performing algorithms to reduce visual clutter. All problems use moderate levels of Gumbel noise,

consistent with the use of a logistic likelihood. qEUBO outperforms all other AFs in all problems except Carcab for  $q = 2$ , followed by qEI and then by qTS. In Section C we also include results for  $q = 4$ ,  $q = 6$ , and varying levels of noise. In these results, qEUBO continues to consistently outperform competitor methods.

Figure 3 shows that including more alternatives in each query ( $q = 4$  vs.  $q = 2$ ) allows qEUBO to achieve a given simple regret using fewer queries. Other AFs also benefit from including more alternatives in each query, but qEUBO seems to benefit the most. This contrasts with Siivola et al. (2021), which found only a marginal benefit of using larger values of  $q$ . At the same time, our results are consistent with those from Mikkola et al. (2020), which also observed significant benefits from using queries with *larger information content*. Our work provides complementary evidence because each query in Mikkola et al. (2020) is equivalent to an infinite number of pairwise comparisons, while our queries use only  $q - 1$  comparisons. Results in Section C suggest that there is also a benefit in going from  $q = 4$  to  $q = 6$  for all algorithms considered there, including qEUBO, but that this benefit is smaller and less consistent than that of going from  $q = 2$  to  $q = 4$ .

Table 1 shows the AF optimization walltime per iteration for each AF and each test problem, averaged over all the iterations. qEI is competitive in terms of its computational requirements, often outperforming all the other AFs. qEUBO is fast enough to support interactive learning applications, such as those for psychophysics experimentation (Owen et al., 2021) and animation (Brochu et al., 2010), despite the challenging dimensionality of the experiments presented here. To better support interactive applications, one can begin optimizing qEUBO to generate the next query while the user is considering the current query. This can be done by initiating qEUBO for all possible user responses to the current query.

Figure 5 compares qEI and qEUBO on an example problem similar to the one analyzed in Theorem 4 in which qEI fails to be consistent. The objective function is the 7-dimensional Alpine1 function. The initial data set contains



several queries constituted by pairs where the first point is a known high-utility point close to the optimum, and the second point is drawn uniformly at random over the domain. After these comparisons, the value of this point has relatively low variance and has a posterior mean relatively high compared to the posterior mean elsewhere. This mimics a setting common in practice where we have an in-use status quo solution that is reasonably good, has well-understood performance because it is currently in use, and on which we would like to improve. In this setting, the incumbent value  $I_n$  has a reasonably high value relative to the posterior mean elsewhere and the variance of the latent utility near the incumbent solution is small. As a result, qEI does not include the incumbent solution or nearby values in DM queries, hampering its ability to learn. Consequently, qEI's simple regret stalls while qEUBO, on the other hand, makes steady progress as the number of queries grows.

The code used to conduct our empirical evaluation can be found at <https://github.com/facebookresearch/qEUBO>.

## 6 CONCLUSION

This work introduces the expected utility of the best option (qEUBO) acquisition function for preferential Bayesian optimization. qEUBO is simple to compute, has a sound decision-theoretic interpretation, and exhibits a strong empirical performance across a broad range of problems. We also draw a connection between qEUBO and its closest competitor, qEI, showing that qEI tends to perform well in early iterations because it is similar to qEUBO but its performance degrades as the number of queries grows or when the variance around the optimum is very small. Furthermore, we show that qEUBO's Bayesian simple regret converges to zero at a rate  $o(1/n)$  as the number of queries,  $n$ , goes to infinity. In contrast, we show that simple regret under qEI can fail to converge to zero. Finally, we demonstrate the substantial benefit of performing queries with more than two alternatives, in contrast with previous work, which found only a marginal benefit. Future directions include studying qEUBO's performance under other probabilistic models and extending qEUBO to more structured problem settings such as contextual preferential Bayesian optimization.

### Acknowledgements

We thank Stryker Buffington and Michael Shvartsman for their help setting up the animation example. We also thank the anonymous reviewers for their helpful comments. PF was supported by AFOSR FA9550-19-1-0283 and FA9550-20-1-0351-20-1-0351.

### References

- Balandat, M., Karrer, B., Jiang, D., Daulton, S., Letham, B., Wilson, A. G., and Bakshy, E. (2020). Botorch: A framework for efficient Monte-Carlo Bayesian optimization. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 21524–21538. Curran Associates, Inc.
- Benavoli, A., Azzimonti, D., and Piga, D. (2021). Preferential Bayesian optimisation with skew Gaussian processes. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 1842–1850.
- Bengs, V., Busa-Fekete, R., El Mesaoudi-Paul, A., and Hüllermeier, E. (2021). Preference-based online learning with dueling bandits: A survey. *J. Mach. Learn. Res.*, 22:7–1.
- Brazian, D. (2006). Computational approaches to preference elicitation. Technical report, University of Toronto, Department of Computer Science.
- Brochu, E., Brochu, T., and De Freitas, N. (2010). A Bayesian interactive optimization approach to procedural animation design. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 103–112.
- Cakmak, S., Astudillo, R., Frazier, P., and Zhou, E. (2020). Bayesian optimization of risk measures. *Advances in Neural Information Processing Systems*, 33:20130–20141.
- Corless, R. M., Gonnet, G. H., Hare, D. E., Jeffrey, D. J., and Knuth, D. E. (1996). On the LambertW function. *Advances in Computational mathematics*, 5(1):329–359.
- Cosenza, Z., Astudillo, R., Frazier, P. I., Baar, K., and Block, D. E. (2022). Multi-information source Bayesian optimization of culture media for cellular agriculture. *Biotechnology and bioengineering*, 119(9):2447–2458.
- Cover, T. M. (1999). *Elements of information theory*. John Wiley & Sons.
- Frazier, P., Powell, W., and Dayanik, S. (2009). The knowledge-gradient policy for correlated normal beliefs. *INFORMS journal on Computing*, 21(4):599–613.
- Frazier, P. I., Powell, W. B., and Dayanik, S. (2008). A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439.
- Garnett, R., Osborne, M. A., and Roberts, S. J. (2010). Bayesian optimization for sensor set selection. In *Proceedings of the 9th ACM/IEEE international conference on information processing in sensor networks*, pages 209–219.

- Ginsbourger, D., Le Riche, R., and Carraro, L. (2008). A multi-points criterion for deterministic parallel global optimization based on gaussian processes.
- González, J., Dai, Z., Damianou, A., and Lawrence, N. D. (2017). Preferential Bayesian optimization. In *International Conference on Machine Learning*, pages 1282–1291. PMLR.
- Griffiths, R. R. and Hernández-Lobato, J. M. (2020). Constrained Bayesian optimization for automatic chemical design using variational autoencoders. *Chemical Science*, 11(2):577–586.
- Hensman, J., Matthews, A., and Ghahramani, Z. (2015). Scalable variational Gaussian process classification. In *Artificial Intelligence and Statistics*, pages 351–360. PMLR.
- Hernández-Lobato, J. M., Hoffman, M. W., and Ghahramani, Z. (2014). Predictive entropy search for efficient global optimization of black-box functions. *Advances in Neural Information Processing Systems*, 27.
- Lin, Z. J., Astudillo, R., Frazier, P., and Bakshy, E. (2022). Preference exploration for efficient Bayesian optimization with multiple outcomes. In Camps-Valls, G., Ruiz, F. J. R., and Valera, I., editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 4235–4258. PMLR.
- Mikkola, P., Todorović, M., Järvi, J., Rinke, P., and Kaski, S. (2020). Projective preferential Bayesian optimization. In *International Conference on Machine Learning*, pages 6884–6892. PMLR.
- Nguyen, Q. P., Tay, S., Low, B. K. H., and Jaillet, P. (2021). Top-k ranking Bayesian optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9135–9143.
- Nielsen, J. B. B., Nielsen, J., and Larsen, J. (2014). Perception-based personalization of hearing aids using gaussian processes and active learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(1):162–173.
- Owen, L., Browder, J., Letham, B., Stocck, G., Tymms, C., and Shvartsman, M. (2021). Adaptive nonparametric psychophysics.
- Rahimi, A. and Recht, B. (2007). Random features for large-scale kernel machines. *Advances in Neural Information Processing Systems*, 20.
- Scott, W., Frazier, P., and Powell, W. (2011). The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression. *SIAM Journal on Optimization*, 21(3):996–1026.
- Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175.
- Siivola, E., Dhaka, A. K., Andersen, M. R., González, J., Moreno, P. G., and Vehtari, A. (2021). Preferential batch Bayesian optimization. In *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing*, pages 1–6. IEEE.
- Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, pages 2951–2959, Red Hook, New York. Curran Associates, Inc.
- Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. W. (2012). Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE transactions on information theory*, 58(5):3250–3265.
- Tucker, M., Novoseller, E., Kann, C., Sui, Y., Yue, Y., Burdick, J. W., and Ames, A. D. (2020). Preference-based learning for exoskeleton gait optimization. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2351–2357. IEEE.
- Viappiani, P. and Boutilier, C. (2010). Optimal Bayesian recommendation sets and myopically optimal choice query sets. *Advances in Neural Information Processing Systems*, 23.
- Wang, J., Clark, S. C., Liu, E., and Frazier, P. I. (2016). Parallel Bayesian global optimization of expensive functions. *arXiv preprint arXiv:1602.05149*.
- Wu, H. and Liu, X. (2016). Double thompson sampling for dueling bandits. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.
- Wu, J. and Frazier, P. (2016). The parallel knowledge gradient method for batch Bayesian optimization. *Advances in Neural Information Processing Systems*, 29.
- Wu, J., Poloczek, M., Wilson, A. G., and Frazier, P. (2017). Bayesian optimization with gradients. *Advances in Neural Information Processing Systems*, 30.
- Yue, Y., Broder, J., Kleinberg, R., and Joachims, T. (2012). The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556.

## A PROOFS OF THEOREMS 1 AND 2

### A.1 Proof of Theorem 1

**Theorem A.1** (Theorem 1). *Suppose the DM's responses are noise-free. Then,  $\operatorname{argmax}_{X \in \mathbb{X}^q} \text{qEUBO}_n(X) \subseteq \operatorname{argmax}_{X \in \mathbb{X}^q} V_n(X)$ .*

*Proof.* For any given  $X \in \mathbb{X}^q$  and each  $i \in \{1, \dots, q\}$  let  $x^+(X, i) \in \operatorname{argmax}_{x \in \mathbb{X}} \mathbf{E}_n[f(x) \mid (X, i)]$  and define  $X^+(X) = (x^+(X, 1), \dots, x^+(X, q))$ . We claim that

$$V_n(X) \leq \text{qEUBO}_n(X^+(X)). \quad (1)$$

To see this, note that

$$\begin{aligned} V_n(X) &= \sum_{i=1}^q \mathbf{P}_n(r(X) = i) \mathbf{E}_n[f(x^+(X, i)) \mid (X, i)] \\ &\leq \sum_{i=1}^q \mathbf{P}_n(r(X) = i) \mathbf{E}_n \left[ \max_{i=1, \dots, q} f(x^+(X, i)) \mid (X, i) \right] \\ &= \mathbf{E}_n \left[ \max_{i=1, \dots, q} f(x^+(X, i)) \right] \\ &= \text{qEUBO}_n(X^+(X)), \end{aligned}$$

as claimed.

On the other hand, for any given  $X \in \mathbb{X}^q$  we have

$$\mathbf{E}_n[f(x_{r(X)}) \mid (X, r(X))] \leq \max_{x \in \mathbb{X}} \mathbf{E}_n[f(x) \mid (X, r(X))].$$

Since  $f(x_{r(X)}) = \max_{i=1, \dots, q} f(x_i)$ , taking expectations over  $r(X)$  on both sides we obtain

$$\text{qEUBO}_n(X) \leq V_n(X). \quad (2)$$

Now, building on the arguments above, let  $X^* \in \operatorname{argmax}_{X \in \mathbb{X}^q} \text{qEUBO}_n(X)$  and suppose for the sake of contradiction that  $X^* \notin \operatorname{argmax}_{X \in \mathbb{X}^q} V_n(X)$ . Then, there exists  $\tilde{X} \in \mathbb{X}^q$  such that  $V_n(\tilde{X}) > V_n(X^*)$ . By the arguments above we have

$$\begin{aligned} \text{qEUBO}_n(X^+(\tilde{X})) &\geq V_n(\tilde{X}) \\ &> V_n(X^*) \\ &\geq \text{qEUBO}_n(X^*) \\ &\geq \text{qEUBO}_n(X^+(\tilde{X})). \end{aligned}$$

The first inequality follows from (1). The second inequality is due to our supposition for contradiction. The third inequality is due to (2). Finally, the fourth inequality holds since  $X^* \in \operatorname{argmax}_{X \in \mathbb{X}^q} \text{qEUBO}_n(X)$ .

This contradiction concludes the proof.  $\square$

### A.2 Proof of Theorem 2

Before proving Theorem 2, we introduce notation and prove several lemmas.

Throughout this section we assume that

$$\mathbf{P}(r(X) = i \mid f(X)) = \frac{\exp(f(x_i)/\lambda)}{\sum_{j=1}^q \exp(f(x_j)/\lambda)},$$

for  $i = 1, \dots, q$ . We also define the functions

$$U_n^\lambda(X) = \mathbf{E}_n[f(x_{r(X)})],$$

and

$$V_n^\lambda(X) = \mathbf{E}_n \left[ \max_{x \in \mathbb{X}} \mathbf{E}_n[f(x) \mid (X, r(X))] \right].$$

We note that  $V_n^\lambda$  makes the dependence of  $V_n$  on  $\lambda$  explicit. On the other hand,  $U_n^\lambda$  generalizes the definition of qEUBO $_n$ , which is obtained as a special case for  $\lambda = 0$ .

Our analysis is similar to the one pursued by Viappiani and Boutilier (2010) to relate optimal recommendation sets and optimal query sets. In particular, we leverage the following result, which can be deduced from the proof of Theorem 3 in the supplement of Viappiani and Boutilier (2010).

**Lemma A.1.** *For any  $s_1, \dots, s_q \in \mathbb{R}$ ,*

$$\sum_{i=1}^q \frac{\exp(s_i/\lambda)}{\sum_{j=1}^q \exp(s_j/\lambda)} s_i \geq \max_{i=1, \dots, q} s_i - \lambda C,$$

where  $C = L_W((q-1)/e)$ , where  $L_W$  is the Lambert  $W$  function (Corless et al., 1996).

*Proof.* We may assume without loss of generality that  $\max_{i=1, \dots, q} s_i = s_q$ . Let  $t_i = (s_q - s_i)/\lambda$  for  $i = 1, \dots, q-1$ . After some algebra, we see that the inequality we want to show is equivalent to

$$\sum_{i=1}^{q-1} \frac{t_i \exp(-t_i)}{1 + \sum_{j=1}^{q-1} \exp(-t_j)} \leq C.$$

Thus, it suffices to show that the function  $\eta : [0, \infty)^{q-1} \rightarrow \mathbb{R}$  given by

$$\eta(t_1, \dots, t_{q-1}) = \sum_{i=1}^{q-1} \frac{t_i \exp(-t_i)}{1 + \sum_{j=1}^{q-1} \exp(-t_j)}$$

is bounded above by  $C$ . We refer the reader to the supplement of Viappiani and Boutilier (2010) for a proof. □

**Lemma A.2.**  $U_n^\lambda(X) \geq \text{qEUBO}_n(X) - \lambda C$  for all  $X \in \mathbb{X}^q$ .

*Proof.* Note that

$$\mathbf{E}[f(x_{r(X)}) \mid f(X)] = \sum_{i=1}^q \frac{\exp(f(x_i)/\lambda)}{\sum_{j=1}^q \exp(f(x_j)/\lambda)} f(x_i).$$

Thus, Lemma A.1 implies that

$$\mathbf{E}[f(x_{r(X)}) \mid f(X)] \geq \max_{i=1, \dots, q} f(x_i) - \lambda C.$$

Taking expectations over both sides of the inequality yields the desired result. □

**Lemma A.3.**  $V_n^\lambda(X) \geq U_n^\lambda(X)$  for all  $X \in \mathbb{X}^q$ .

*Proof.* Observe that

$$\begin{aligned} V_n^\lambda(X) &= \mathbf{E}_n \left[ \max_{x \in \mathbb{X}} \mathbf{E}[f(x) \mid (X, r(X))] \right] \\ &\geq \mathbf{E}_n \left[ \mathbf{E}[f(x_{r(X)}) \mid (X, r(X))] \right] \\ &= \mathbf{E}_n[f(x_{r(X)})] \\ &= U_n^\lambda(X), \end{aligned}$$

where the penultimate equality follows by the law of iterated expectation. □

**Theorem A.2** (Theorem 2). *If  $X^* \in \arg\max_{X \in \mathbb{X}^q} \text{qEUBO}_n(X)$ , then  $V_n^\lambda(X^*) \geq \max_{X \in \mathbb{X}^q} V_n^0(X) - \lambda C$ .*

*Proof.* Let  $X^{**} \in \operatorname{argmax}_{X \in \mathbb{X}^q} V_n^0(X)$ . We have the following chain of inequalities:

$$\begin{aligned} V_n^\lambda(X^*) &\geq U_n^\lambda(X^*) \\ &\geq U_n^0(X^*) - \lambda C \\ &= \text{qEUBO}_n(X^*) - \lambda C \\ &\geq \text{qEUBO}_n(X^+(X^{**})) - \lambda C \\ &\geq V_n^0(X^{**}) - \lambda C \\ &= \max_{X \in \mathbb{X}^q} V_n^0(X) - \lambda C. \end{aligned}$$

The first inequality follows from Lemma A.3. The second inequality follows from Lemma A.2. The third line (first equality) follows from the definition of  $U_n^0$ . The fourth line (third inequality) follows from the definition of  $X^*$ . The fifth line (fourth inequality) can be obtained as in the proof of Theorem 1. Finally, the last line (second equality) follows from the definition of  $X^{**}$ .  $\square$

## B PROOFS OF THEOREMS 3 AND 4

In this section, we prove Theorems 3 and 4. In contrast with other sections, here we use super indices with parentheses to denote the iteration number,  $n$ . We denote the query presented to the used by  $X^{(n)}$ , and the corresponding user's response by  $r(X^{(n)})$ . We shall sometimes denote this response more compactly by  $r^{(n)}$ . We let  $\mathcal{D}^{(n)} = \{(X^{(m)}, r^{(m)})\}_{m=1}^n$  denote the data collected up to time  $n$ . Similarly, we denote the conditional expectation given  $\mathcal{D}^{(n)}$  by  $\mathbf{E}^{(n)}$ .

Recall that Theorems 3 and 4 assume that  $\mathbb{X}$  is finite and  $q = 2$ . We make these assumptions throughout this section without stating them explicitly. The theorems also assume that  $\mathbf{E}[\max_{x \in \mathbb{X}} |f(x)|] < \infty$ , which guarantees that all expectations involved in our analysis are finite.

The statement of Theorems 3 and 4 rely on the following assumptions (called Assumptions 1-3 respectively):

1.  $\mathbf{P}(f(x) = f(y)) = 0$  for any  $x, y \in \mathbb{X}$  with  $x \neq y$ .
2. There exists  $a > 1/2$  such that  $\mathbf{P}(r(X) \in \operatorname{argmax}_{i=1, \dots, 2} f(x_i) \mid f(X)) \geq a$  for any  $X = (x_1, x_2) \in \mathbb{X}^2$  with  $x_1 \neq x_2$  almost surely under the prior on  $f$ .
3. There exist  $\Delta \geq \delta > 0$  such that for any  $\mathcal{D}^{(n)}$  and any  $x, y \in \mathbb{X}$  (potentially depending on  $\mathcal{D}^{(n)}$ ),

$$\delta \mathbf{P}^{(n)}(f(x) > f(y)) \leq \mathbf{E}^{(n)}[\{f(x) - f(y)\}^+] \leq \Delta \mathbf{P}^{(n)}(f(x) > f(y))$$

almost surely under the prior on  $f$ .

Before we begin with the proofs of Theorems 3 and 4, we show that Assumptions 1-3 hold in two general settings. This is summarized in the following two lemmas.

First, we show that if the difference in utilities between items is bounded below (away from 0) and above almost surely under the prior, then Assumptions 1-3 hold for a broad class of likelihoods for the query response.

**Lemma B.1.** *Suppose that there exist  $\Delta \geq \delta > 0$  such that  $\delta \leq |f(x) - f(y)| \leq \Delta$  almost surely whenever  $x \neq y$ . Additionally, suppose that the likelihood function  $L_i(f(X); \lambda) = \mathbf{P}(r(X) = i \mid f(X))$  is bounded below by  $a > 1/2$  when  $\delta \leq |f(x) - f(y)|$ . Then Assumptions 1-3 hold.*

*Proof.* Since  $\delta \leq |f(x) - f(y)|$  almost surely whenever  $x \neq y$ , Assumption 1 holds trivially. The assumed lower bound on the likelihood implies Assumption 2.

Now we show that Assumption 3 holds. Observe that

$$\begin{aligned} \{f(x) - f(y)\}^+ &= (f(x) - f(y))1\{f(x) > f(y)\} \\ &\geq \delta 1\{f(x) > f(y)\} \end{aligned}$$

almost surely under the prior (and thus under the posterior at time  $n$ ) on  $f$ . Thus,

$$\begin{aligned} \mathbf{E}^{(n)} [\{f(x) - f(y)\}^+] &\geq \mathbf{E}^{(n)} [\delta 1(f(x) > f(y))] \\ &= \delta \mathbf{P}^{(n)}(f(x) > f(y)). \end{aligned}$$

We can show similarly that

$$\mathbf{E}^{(n)} [\{f(x) - f(y)\}^+] < \Delta \mathbf{P}^{(n)}(f(x) > f(y)).$$

□

This lemma's lower bound on the likelihood is satisfied for the logistic likelihood, the probit likelihood, or for any other likelihood in which  $\mathbf{P}(r(X) = 1 \mid f(X))$  is a strictly increasing function of  $f(x_1) - f(x_2)$  and equal to  $1/2$  when  $f(x_1) = f(x_2)$ .

Our next result considers the situation where the difference in item utilities is not bounded. In this situation, it shows that Assumptions 1-3 hold provided the DM's response likelihood satisfies a stronger condition than that assumed by Lemma B.1: the DM's responses are correct with constant probability greater than  $1/2$  whenever the items in the query have non-identical utility value. This result applies to general non-degenerate GP prior distributions; i.e., for GP prior distributions with a positive definite covariance function.

**Lemma B.2.** *Suppose that Assumption 1 holds and there exists a  $> 1/2$  such that  $\mathbf{P}(r(X) = \operatorname{argmax}_{i=1,\dots,2} f(x_i) \mid f(X)) = a$  whenever  $f(x_1) \neq f(x_2)$ . Then, Assumption 3 holds.*

*Proof.* We first make a fundamental observation. For a permutation  $\pi : \{1, \dots, |\mathbb{X}|\} \rightarrow \{1, \dots, |\mathbb{X}|\}$ , let  $A_\pi$  denote the event  $\{f(x_{\pi(1)}) < \dots < f(x_{\pi(|\mathbb{X}|)})\}$ . We note that the distribution of  $f$  given  $\mathcal{D}^{(n)}$  and  $A_\pi$  coincides with the distribution of  $f$  given  $A_\pi$ . This observation may be surprising at first glance, but it has a very intuitive interpretation. Since the DM's responses depend exclusively on the relative order of the utility values, they only provide ordinal information about the utility values. Thus, if the relative order of the utilities of all points is known, we learn nothing from the DM's responses; i.e., the conditional distribution given  $A_\pi$  remains unchanged if we observe  $\mathcal{D}^{(n)}$ . This property relies heavily on the choice of the likelihood and does not hold in general for other likelihoods such as the logistic likelihood. We will make use of this observation in the proof of this lemma.

We first prove the existence of  $\delta$ . Let  $\Pi$  be the set of all permutations over  $\{1, \dots, |\mathbb{X}|\}$  and let  $\pi$  be an arbitrary element of  $\Pi$ . Define

$$\gamma(x, y, \pi) = \mathbf{E}[\{f(x) - f(y)\}^+ \mid A_\pi].$$

We note that  $\mathbf{E}[\{f(x) - f(y)\}^+ \mid A_\pi] > 0$  if and only if  $f(x) > f(y)$  under  $A_\pi$ .

Let

$$\delta = \min_{x, y \in \mathbb{X}, \pi \in \Pi} \{\gamma(x, y, \pi) : \gamma(x, y, \pi) > 0\},$$

We have

$$\begin{aligned} \mathbf{E}^{(n)}[\{f(x) - f(y)\}^+] &= \sum_{\pi \in \Pi} \mathbf{P}^{(n)}(A_\pi) \mathbf{E}^{(n)}[\{f(x) - f(y)\}^+ \mid A_\pi] \\ &= \sum_{\pi \in \Pi} \mathbf{P}^{(n)}(A_\pi) \mathbf{E}[\{f(x) - f(y)\}^+ \mid A_\pi] \\ &\geq \sum_{\pi \in \Pi} \mathbf{P}^{(n)}(A_\pi) 1\{f(x) > f(y) \mid A_\pi\} \delta \\ &= \delta \mathbf{P}^{(n)}(f(x) > f(y)), \end{aligned}$$

where the second equation follows from the observation earlier that  $f$  and the DM's responses are conditionally independent given  $A_\pi$  and the inequality follows from the observation above that  $\mathbf{E}[\{f(x) - f(y)\}^+ \mid A_\pi] > 0$  if and only if  $f(x) > f(y)$  under  $A_\pi$ .

The existence of  $\Delta$  can be proved similarly by taking

$$\Delta = \max_{x, y \in \mathbb{X}, \pi \in \Pi} \{\delta(x, y, \pi) : \delta(x, y, \pi) > 0\}.$$

□



### B.1 Proof of Theorem 3

The proof of Theorem 3 is achieved via a series of lemmas. We assume that

$$X^{(n+1)} = (x_1^{(n)}, x_2^{(n)}) \in \operatorname{argmax}_{X \in \mathbb{X}^2} \text{qEUBO}^{(n)}(X)$$

for all  $n$  throughout the proofs of these lemmas. We define  $x^* = \operatorname{argmax}_{x \in \mathbb{X}} f(x)$ , and  $x_*^{(n)} = \operatorname{argmax}_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)]$ .

**Lemma B.3.** *Let  $p_{**}^{(n)} = \mathbf{P}^{(n)}(x_*^{(n)} \neq x^*)$ . Then,*

$$\max_{X \in \mathbb{X}^2} \text{qEUBO}^{(n)}(X) \geq \max_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)] + \frac{\delta}{|\mathbb{X}| - 1} p_{**}^{(n)}.$$

*Proof.* Observe that

$$\begin{aligned} \max_{X \in \mathbb{X}^2} \text{qEUBO}^{(n)}(X) &\geq \max_{x \in \mathbb{X}} \text{qEUBO}^{(n)}(x, x_*^{(n)}) \\ &= \max_{x \in \mathbb{X}} \mathbf{E}^{(n)}[\max\{f(x), f(x_*^{(n)})\}] \\ &= \max_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x_*^{(n)}) + \{f(x) - f(x_*^{(n)})\}^+] \\ &= \mathbf{E}^{(n)}[f(x_*^{(n)})] + \max_{x \in \mathbb{X}} \mathbf{E}^{(n)}[\{f(x) - f(x_*^{(n)})\}^+] \\ &\geq \max_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)] + \max_{x \in \mathbb{X}} \delta \mathbf{P}^{(n)}(f(x) > f(x_*^{(n)})) \\ &\geq \max_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)] + \frac{\delta}{|\mathbb{X}| - 1} \sum_{x \in \mathbb{X}} \mathbf{P}^{(n)}(f(x) > f(x_*^{(n)})) \\ &\geq \max_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)] + \frac{\delta}{|\mathbb{X}| - 1} \mathbf{P}^{(n)}(\cup_{x \in \mathbb{X}} \{f(x) > f(x_*^{(n)})\}) \\ &= \max_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)] + \frac{\delta}{|\mathbb{X}| - 1} \mathbf{P}^{(n)}(x_*^{(n)} \neq x^*), \end{aligned}$$

where the second inequality follows from Assumption 2 and the last inequality follows from the union bound.  $\square$

**Lemma B.4.** *Let  $p_1^{(n)} = \mathbf{P}^{(n)}(f(x_1^{(n+1)}) > f(x_2^{(n+1)}))$  and  $p_2^{(n)} = \mathbf{P}^{(n)}(f(x_2^{(n+1)}) > f(x_1^{(n+1)}))$ . Then,*

$$p_1^{(n)}, p_2^{(n)} \geq C p_{**}^{(n)},$$

where  $C = \frac{\delta}{(|\mathbb{X}| - 1)\Delta}$ .

*Proof.* We have

$$\begin{aligned} \text{qEUBO}^{(n)}(X^{(n+1)}) &= \mathbf{E}^{(n)}[f(x_1^{(n+1)})] + \mathbf{E}^{(n)}[\{f(x_2^{(n+1)}) - f(x_1^{(n+1)})\}^+] \\ &\leq \mathbf{E}^{(n)}[f(x_1^{(n+1)})] + \Delta \mathbf{P}^{(n)}(f(x_2^{(n+1)}) > f(x_1^{(n+1)})) \\ &\leq \max_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)] + \Delta \mathbf{P}^{(n)}(f(x_2^{(n+1)}) > f(x_1^{(n+1)})), \end{aligned}$$

where the first inequality is again due to Assumption 2. Combining this with Lemma B.2., we get

$$\max_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)] + \Delta \mathbf{P}^{(n)}(f(x_2^{(n+1)}) > f(x_1^{(n+1)})) \geq \max_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)] + \frac{\delta}{|\mathbb{X}| - 1} p_{**}^{(n)};$$

i.e.,

$$\mathbf{P}^{(n)}(f(x_2^{(n+1)}) > f(x_1^{(n+1)})) \geq \frac{\delta}{(|\mathbb{X}| - 1)\Delta} p_{**}^{(n)}.$$

Finally, it follows by symmetry that

$$\mathbf{P}^{(n)}(f(x_1^{(n+1)}) > f(x_2^{(n+1)})) \geq \frac{\delta}{(|\mathbb{X}| - 1)\Delta} p_{**}^{(n)},$$

which finishes the proof.  $\square$

**Lemma B.5.** Let  $X$  denote a generic discrete random vector and  $Y$  denote a Bernoulli random variable correlated with  $X$ . Let  $p_X$  and  $p_Y$  denote the marginal distributions of  $X$  and  $Y$ , respectively, and  $p_{X|Y}$  denote the conditional distribution of  $X$  given  $Y$ . Let  $\mathbf{H}(X)$  and  $\mathbf{H}(X|Y = y)$  denote the entropy of  $X$  and the conditional entropy of  $X$  given  $Y = y$ ; i.e.,

$$\mathbf{H}(X) = - \sum_x p_X(x) \log(p_X(x)),$$

and

$$\mathbf{H}(X|Y = y) = - \sum_x p_{X|Y}(x|y) \log(p_{X|Y}(x|y)).$$

Define  $q(x) = \mathbf{P}(Y = 0|X = x)$ . Then,

$$- \sum_{y=0}^1 p_Y(y) \mathbf{H}(X|Y = y) = \mathbf{H}(X) - \left[ h \left( \sum_x p_X(x) q(x) \right) - \sum_x p_X(x) h(q(x)) \right],$$

where  $h(q) = -q \log q - (1 - q) \log(1 - q)$  is the binary entropy function.

*Proof.* Observe that  $-\sum_y p_Y(y) \mathbf{H}(X|Y = y)$  is the conditional entropy of  $X$  given  $Y$  (see definition 2.10 in Cover (1999)). By basic information theory results (see equations 2.43 and 2.44 in Cover (1999)) we know that

$$\mathbf{H}(X) - \mathbf{H}(X|Y) = \mathbf{H}(Y) - \mathbf{H}(Y|X)$$

Rearranging terms, we obtain

$$\mathbf{H}(X|Y) = \mathbf{H}(X) - (\mathbf{H}(Y) - \mathbf{H}(Y|X))$$

Consider the term  $\mathbf{H}(Y) - \mathbf{H}(Y|X)$ . We have

$$\begin{aligned} \mathbf{P}(Y = 0) &= \sum_x p_X(x) \mathbf{P}(Y = 0|X = x) \\ &= \sum_x p_X(x) q(x). \end{aligned}$$

Moreover, since  $Y$  is a Bernoulli random variable,  $\mathbf{H}(Y) = h(\mathbf{P}(Y = 0))$ .

The second term is

$$\begin{aligned} \mathbf{H}(Y|X) &= \sum_x p_X(x) \mathbf{H}(Y|X = x) \\ &= \sum_x p_X(x) h(q(x)). \end{aligned}$$

Hence,

$$\mathbf{H}(Y) - \mathbf{H}(Y|X) = h(\mathbf{P}(Y = 0)) - \sum_x p_X(x) h(q(x)),$$

which concludes the proof. □

**Lemma B.6.** Enumerate the elements of  $\mathbb{X}$  as  $x_1, \dots, x_{|\mathbb{X}|}$  and let  $\pi_f$  be the random permutation satisfying  $f(x_{\pi_f(1)}) < \dots < f(x_{\pi_f(|\mathbb{X}|)})$ . Let  $p_{\pi_f}^{(n)}$  denote the posterior on  $\pi$  given  $\mathcal{D}^{(n)}$ . Then,

$$\mathbf{E}^{(n)}[\mathbf{H}(p_{\pi_f}^{(n+1)})] \leq \mathbf{H}(p_{\pi_f}^{(n)}) - \varphi(p_1^{(n)})$$

where the expectation on the left-hand side is over  $r^{(n+1)}$ , and  $\varphi(u) = h(au + (1 - a)(1 - u)) - h(a)$ .

*Proof.* Observe that  $\mathbf{E}^{(n)}[\mathbf{H}(p^{(n+1)})] = \mathbf{H}(p^{(n)} \mid r^{(n+1)})$ . Thus, from Lemma B.5 it follows that

$$\mathbf{E}^{(n)}[\mathbf{H}(p^{(n+1)})] = \mathbf{H}(p^{(n)}) - \left[ h \left( \sum_{\pi} p_{\pi_f}^{(n)}(\pi) q(\pi) \right) - \sum_{\pi} p_{\pi_f}^{(n)}(\pi) h(q(\pi)) \right],$$

where  $q(\pi) = \mathbf{P}^{(n)}(r^{(n+1)} = 1 \mid \pi)$ . Thus, it suffices to show that

$$h \left( \sum_{\pi} p_{\pi_f}^{(n)}(\pi) q(\pi) \right) - \sum_{\pi} p_{\pi_f}^{(n)}(\pi) h(q(\pi)) \geq \varphi(p_1^{(n)}).$$

Let  $\Pi_1$  and  $\Pi_2$  be the set of permutations such that  $f(x_1^{(n)}) < f(x_2^{(n)})$  and  $f(x_1^{(n)}) > f(x_2^{(n)})$ , respectively. Define

$$q_i = \left( \sum_{\pi \in \Pi_i} p_{\pi_f}^{(n)}(\pi) q(\pi) \right) / p_i^{(n)}$$

for  $i = 1, 2$ . After some algebra we see that

$$h \left( \sum_{\pi} p_{\pi_f}^{(n)}(\pi) q(\pi) \right) - \sum_{\pi} p_{\pi_f}^{(n)}(\pi) h(q(\pi)) = h(p_1^{(n)} q_1 + p_2^{(n)} q_2) - p_1^{(n)} h(q_1) - p_2^{(n)} h(q_2) + p_1^{(n)} \psi_1 + p_2^{(n)} \psi_2,$$

where

$$\psi_i = h(q_i) \sum_{\pi \in \Pi_i} p_{\pi_f}^{(n)}(\pi) h(q(\pi)) / p_i^{(n)}$$

for  $i = 1, 2$ . Moreover, since  $h$  is concave,  $\psi_i \geq 0$  by Jensen's inequality. Thus,

$$h \left( \sum_{\pi} p_{\pi_f}^{(n)}(\pi) q(\pi) \right) - \sum_{\pi} p_{\pi_f}^{(n)}(\pi) h(q(\pi)) \geq h(p_1^{(n)} q_1 + p_2^{(n)} q_2) - p_1^{(n)} h(q_1) - p_2^{(n)} h(q_2)$$

Recall that  $\mathbf{P}(r(X) = \operatorname{argmax}_{i=1,2} f(x_i) \mid f(X)) \geq a$  whenever  $x_1 \neq x_2$  almost surely by Assumption 3. Also recall that  $X^{(n+1)} = (x_1^{(n+1)}, x_2^{(n+1)}) \in \operatorname{argmax}_{X \in \mathbb{X}^2} \text{qEUBO}_n(X)$ . It is not hard to see that we can always choose  $x_1^{(n+1)}$  and  $x_2^{(n+1)}$  such that  $x_1^{(n+1)} \neq x_2^{(n+1)}$ . It follows from this that  $q(\pi) \geq a$  for  $\pi \in \Pi_1$  and  $q(\pi) \leq 1 - a$  for  $\pi \in \Pi_2$ . From the definition of  $q_i$  for  $i = 1, 2$ , this in turn implies that  $q_1 \geq a$  and  $q_2 \leq 1 - a$ .

Taking the derivative of  $h(p_1^{(n)} q_1 + p_2^{(n)} q_2) - p_1^{(n)} h(q_1) - p_2^{(n)} h(q_2)$  with respect to  $q_1$  and recalling that the derivative of  $h$  is decreasing since  $h$  is concave, we can see that  $h(p_1^{(n)} q_1 + p_2^{(n)} q_2) - p_1^{(n)} h(q_1) - p_2^{(n)} h(q_2)$  is minimal when  $q_1 = a$  under the constraint  $q_1 \geq a$ . Similarly, we can see that  $h(p_1^{(n)} q_1 + p_2^{(n)} q_2) - p_1^{(n)} h(q_1) - p_2^{(n)} h(q_2)$  is minimal when  $q_2 = 1 - a$  under the constraint  $q_2 \leq 1 - a$ . Hence,

$$h \left( \sum_{\pi} p_{\pi_f}^{(n)}(\pi) q(\pi) \right) - \sum_{\pi} p_{\pi_f}^{(n)}(\pi) h(q(\pi)) \geq h(p_1^{(n)} a + p_2^{(n)} (1 - a)) - p_1^{(n)} h(a) - p_2^{(n)} h(1 - a) \\ \varphi(p_1^{(n)}),$$

where the last equation holds because  $p_2^{(n)} = 1 - p_1^{(n)}$  and  $h(a) = h(1 - a)$ . This concludes the proof.  $\square$

**Lemma B.7.**  $\varphi(u) \geq 2(h(1/2) - h(a))u$  for  $u \in [0, 1/2]$ .

*Proof.* Note that  $\varphi$  is concave in  $[0, 1]$ . Applying Jensen's inequality we obtain

$$\varphi((1 - 2u)0 + (2u)(1/2)) \geq (1 - 2u)\varphi(0) + 2u\varphi(1/2);$$

i.e.,

$$\varphi(u) \geq 2(h(1/2) - h(a))u.$$

$\square$

**Lemma B.8.** Let  $R^{(n)} = 2(h(1/2) - h(a))p_{**}^{(n)}C$ . Then,

$$\mathbf{E}^{(n)}[\mathbf{H}(p^{(n+1)})] \leq \mathbf{H}(p^{(n)}) - R^{(n)}.$$

*Proof.* From Lemma B.4 we know that  $p_1^{(n)}, p_2^{(n)} \geq p_{**}^{(n)}C$ . Since  $p_1^{(n)} + p_2^{(n)} = 1$ , it follows that

$$0 \leq p_{**}^{(n)}C \leq \min\{p_1^{(n)}, p_2^{(n)}\} \leq 1/2.$$

Now observe that the function  $\varphi$  is increasing in  $[0, 1/2]$  and symmetric around  $1/2$ . Thus,

$$\varphi(p_{**}^{(n)}C) \leq \varphi(\min\{p_1^{(n)}, p_2^{(n)}\}) = \varphi(p_1^{(n)}).$$

Finally,

$$\begin{aligned} \mathbf{E}^{(n)}[\mathbf{H}(p^{(n+1)})] &\leq \mathbf{H}(p^{(n)}) - \varphi(p_1^{(n)}) \\ &\leq \mathbf{H}(p^{(n)}) - \varphi(p_{**}^{(n)}C) \\ &\leq \mathbf{H}(p^{(n)}) - 2(h(1/2) - h(a))p_{**}^{(n)}C, \end{aligned}$$

where the first line comes from Lemma B.6, the second line comes from the above analysis, and the third line is a consequence of Lemma B.7.  $\square$

We are now in position to prove Theorem 3.

**Theorem B.1** (Theorem 3). *Suppose that Assumptions 1-3 are satisfied and  $X^{(n+1)} \in \operatorname{argmax}_{X \in \mathbb{X}^d} \text{qEUBO}^{(n)}(X)$  for all  $n$ . Then,  $\mathbf{E}[f(x^*) - f(x_*^{(n)})] = o(1/n)$ .*

*Proof.* Consider the stochastic processes  $\{Z^{(n)}\}_{n=0}^\infty$  defined by

$$Z^{(n)} = \mathbf{H}(p^{(n)}) + \sum_{m=0}^{n-1} R^{(m)}, \quad n \geq 0.$$

Observe that  $Z^{(n)}$  is non-negative for all  $n$ . Moreover,

$$\begin{aligned} \mathbf{E}^{(n)}[Z^{(n+1)}] &= \mathbf{E}^{(n)}[\mathbf{H}(p^{(n+1)})] + \sum_{m=0}^n R^{(m)} \\ &\leq \mathbf{H}(p^{(n)}) - R^{(n)} + \sum_{m=0}^n R^{(m)} \\ &= \mathbf{H}(p^{(n)}) + \sum_{m=0}^{n-1} R^{(m)} \\ &= Z^{(n)}, \end{aligned}$$

where the inequality above follows from Lemma B.8. Thus,  $\{Z^{(n)}\}_{n=0}^\infty$  is a non-negative supermartingale. By Doob's martingale convergence theorem,  $\{Z^{(n)}\}_{n=0}^\infty$  converges almost surely to a random variable with finite expectation. This in turn implies that  $\sum_{n=0}^\infty \mathbf{E}[R^{(n)}] < \infty$ . Since  $R^{(n)} = 2(h(1/2) - h(a))p_{**}^{(n)}C$ , it follows that  $\sum_{n=0}^\infty \mathbf{E}[p_{**}^{(n)}] < \infty$ . Recall that  $p_{**}^{(n)} = \mathbf{P}^{(n)}(x_*^{(n)} \neq x_*)$ . By the law of the iterated expectation we obtain  $\mathbf{E}[p_{**}^{(n)}] = \mathbf{P}(x_*^{(n)} \neq x_*)$ . Hence, we have shown that  $\sum_{n=0}^\infty \mathbf{P}(x_*^{(n)} \neq x_*) < \infty$ . We deduce from this that  $\mathbf{P}(x_*^{(n)} \neq x_*) = o(1/n)$ .

Finally,

$$\begin{aligned} \mathbf{E}[f(x_*) - f(x_*^{(n)})] &= \mathbf{E}[\{f(x_*) - f(x_*^{(n)})\}^+] \\ &\leq \Delta \mathbf{P}(f(x_*) > f(x_*^{(n)})) \\ &= \Delta \mathbf{P}(x_*^{(n)} \neq x_*) \\ &= o(1/n), \end{aligned}$$

where the first and third lines hold by definition of  $x_*$ , and the second line follows from Assumption 2.  $\square$

## B.2 Proof of Theorem 4

**Theorem B.2** (Theorem 4). *There exists a problem instance (i.e.,  $\mathbb{X}$  and Bayesian prior distribution over  $f$ ) satisfying Assumptions 1-3 such that if  $X^{(n+1)} \in \operatorname{argmax}_{X \in \mathbb{X}^q} \operatorname{qEI}^{(n)}(X)$  for all  $n$ , then  $\mathbf{E}[f(x^*) - f(\hat{x}_*^{(n)})] \geq R$  for all  $n$ , for a constant  $R > 0$ .*

*Proof.* Let  $\mathbb{X} = \{1, 2, 3, 4\}$  and consider the functions  $f_i : \mathbb{X} \rightarrow \mathbb{R}$ , for  $i = 1, 2, 3, 4$ , given by  $f_i(1) = -1$  and  $f_i(2) = 0$  for all  $i$ , and

$$f_1(x) = \begin{cases} 1, & x = 3 \\ \frac{1}{2}, & x = 4 \end{cases}, \quad f_2(x) = \begin{cases} \frac{1}{2}, & x = 3 \\ 1, & x = 4 \end{cases}, \quad f_3(x) = \begin{cases} -\frac{1}{2}, & x = 3 \\ -1, & x = 4 \end{cases}, \quad f_4(x) = \begin{cases} -1, & x = 3 \\ -\frac{1}{2}, & x = 4 \end{cases}.$$

Let  $p$  be a number with  $0 < p < 1/3$  and set  $q = 1 - p$ . We consider a prior distribution on  $f$  with support  $\{f_i\}_{i=1}^4$  such that

$$p_i = \mathbf{P}(f = f_i) = \begin{cases} p/2, & i = 1, 2, \\ q/2, & i = 3, 4. \end{cases}$$

We also assume the DM's response likelihood is given by  $\mathbf{P}(r(X) = 1 \mid f(x_1) > f(x_2)) = a$  for some  $a$  such that  $1/2 < a < 1$ ,

Let  $\mathcal{D}^{(n)}$  denote the set of observations up to time  $n$  and let  $p_i^{(n)} = \mathbf{P}(f = f_i \mid \mathcal{D}^{(n)})$  for  $i = 1, 2, 3, 4$ . We let the initial data set be  $\mathcal{D}^{(0)} = \{(X^{(0)}, r^{(0)})\}$ , where  $X^{(0)} = (1, 2)$ . We will prove that the following statements are true for all  $n \geq 0$ .

1.  $p_i^{(n)} > 0$  for  $i = 1, 2, 3, 4$ .
2.  $p_1^{(n)} < \frac{1}{2}p_3^{(n)}$  and  $p_2^{(n)} < \frac{1}{2}p_4^{(n)}$ .
3.  $\operatorname{argmax}_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)] = \{2\}$ .
4.  $\operatorname{argmax}_{X \in \mathbb{X}^2} \operatorname{qEI}^{(n)}(X) = \{(3, 4)\}$ .

We prove this by induction over  $n$ . We begin by proving this for  $n = 0$ . Since  $f_i(1) < f_i(2)$  for all  $i$ , the posterior distribution on  $f$  given  $\mathcal{D}^{(0)}$  remains the same as the prior; i.e.,  $p_i^{(0)} = p_i$  for  $i = 1, 2, 3, 4$ . Using this, statements 1 and 2 can be easily verified. Now note that  $\mathbf{E}^{(0)}[f(1)] = -1$ ,  $\mathbf{E}^{(0)}[f(2)] = 0$ , and  $\mathbf{E}^{(0)}[f(3)] = \mathbf{E}^{(0)}[f(4)] = \frac{3}{2}(p - q)$ . Since  $p < q$ , it follows that  $\operatorname{argmax}_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)] = \{2\}$ ; i.e., statement 3 holds. Finally, since  $\max_{x \in \{1, 2\}} \mathbf{E}^{(0)}[f(x)] = 0$ , the qEI acquisition function at time  $n = 0$  is given by  $\operatorname{qEI}^{(0)}(X) = \mathbf{E}^{(0)}[\{\max\{f(x_1), f(x_2)\}\}^+]$ . A direct calculation can now be performed to verify that statement 4 holds. This completes the base case.

Now suppose statements 1-4 hold for some  $n \geq 0$ . Since  $X^{(n+1)} = (3, 4)$ , the posterior distribution on  $f$  given  $\mathcal{D}^{(n+1)}$  is given by

$$p_i^{(n+1)} \propto \begin{cases} p_i^{(n)} \ell, & i = 1, 3, \\ p_i^{(n)} (1 - \ell), & i = 2, 4, \end{cases}$$

where

$$\ell = a\mathbb{I}\{r^{(n+1)} = 1\} + (1 - a)\mathbb{I}\{r^{(n+1)} = 2\}.$$

Observe that  $0 < \ell < 1$  since  $0 < a < 1$ . Thus,  $\ell > 0$  and  $1 - \ell > 0$ . Since  $p_i^{(n)} > 0$  by the induction hypothesis, it follows from this that  $p_i^{(n+1)} > 0$  for  $i = 1, 2, 3, 4$ . Moreover, since  $p_i^{(n+1)} \propto p_i^{(n)} \ell$  for  $i = 1, 3$  and  $p_1^{(n)} < \frac{1}{2}p_3^{(n)}$  by the induction hypothesis, it follows that  $p_1^{(n+1)} < \frac{1}{2}p_3^{(n+1)}$ . Similarly,  $p_2^{(n+1)} < \frac{1}{2}p_4^{(n+1)}$ . Thus, statements 1 and 2 hold at time  $n + 1$ .

Now observe that

$$\begin{aligned} \mathbf{E}^{(n+1)}[f(3)] &= p_1^{(n+1)} + \frac{1}{2}p_2^{(n+1)} - \frac{1}{2}p_3^{(n+1)} - p_4^{(n+1)} \\ &= \left(p_1^{(n+1)} - \frac{1}{2}p_3^{(n+1)}\right) + \left(\frac{1}{2}p_2^{(n+1)} - p_4^{(n+1)}\right) \\ &\leq \left(p_1^{(n+1)} - \frac{1}{2}p_3^{(n+1)}\right) + \left(p_2^{(n+1)} - \frac{1}{2}p_4^{(n+1)}\right) \\ &\leq 0, \end{aligned}$$

where the last inequality holds since  $p_1^{(n+1)} < \frac{1}{2}p_3^{(n+1)}$  and  $p_2^{(n+1)} < \frac{1}{2}p_4^{(n+1)}$ . Similarly, we see that  $\mathbf{E}^{(n+1)}[f(4)] \leq 0$ . Since  $\mathbf{E}^{(n+1)}[f(1)] = -1$  and  $\mathbf{E}^{(n+1)}[f(2)] = 0$ , it follows that  $\operatorname{argmax}_{x \in \mathbb{X}} \mathbf{E}^{(n+1)}[f(x)] = \{2\}$ ; i.e., statement 3 holds at time  $n + 1$ .

Since  $\max_{x \in \mathbb{X}} \mathbf{E}^{(0)}[f(x)] = 0$ , the qEI acquisition function at time  $n + 1$  is given by  $\text{qEI}^{(n+1)}(X) = \mathbf{E}^{(n+1)}[\{\max\{f(x_1), f(x_2)\}\}^+]$ . Since  $f(1) \leq f(x)$  almost surely under the prior for all  $x \in \mathbb{X}$ , there is always a maximizer of qEI that does not contain 1. Thus, to find the maximizer of qEI, it suffices to analyse its value at the pairs (2, 3), (3, 4) and (4, 2). We have

$$\begin{aligned} \text{qEI}^{(n+1)}(2, 3) &= p_1^{(n+1)} + 1/2p_2^{(n+1)}, \\ \text{qEI}^{(n+1)}(3, 4) &= p_1^{(n+1)} + p_2^{(n+1)} \end{aligned}$$

and

$$\text{qEI}^{(n+1)}(4, 2) = 1/2p_1^{(n+1)} + p_2^{(n+1)}.$$

Since  $p_1^{(n+1)} > 0$  and  $p_2^{(n+1)} > 0$ , it follows that  $\operatorname{argmax}_{X \in \mathbb{X}^2} \text{qEI}^{(n+1)}(X) = \{(3, 4)\}$ , which concludes the proof by induction.

Finally, since  $\operatorname{argmax}_{x \in \mathbb{X}} \mathbf{E}^{(n)}[f(x)] = \{2\}$  for all  $n$ , the Bayesian simple regret of qEI is given by

$$\begin{aligned} \mathbf{E}[f(x^*) - f(2)] &= \sum_{i=1} p_i \left( \max_{x \in \mathbb{X}} f_i(x) - f_i(2) \right) \\ &= p \end{aligned}$$

for all  $n$ . □

## C ADDITIONAL EMPIRICAL RESULTS

Besides the experiments we present in the main paper, we additionally present experimental results with  $q = 4$  vs  $q = 6$  and investigate different acquisition functions' performance under various noise levels for  $q = 2$  case.

### C.1 Results for $q = 4$ vs. $q = 6$

Figure 6 shows the experiment results for the best performing three acquisition (qTS, qEI, and qEUBO) functions with  $q = 4$  and  $q = 6$  over 150 queries. As discussed in the paper,  $q = 6$  offers little improvement over  $q = 4$  for all acquisition functions.



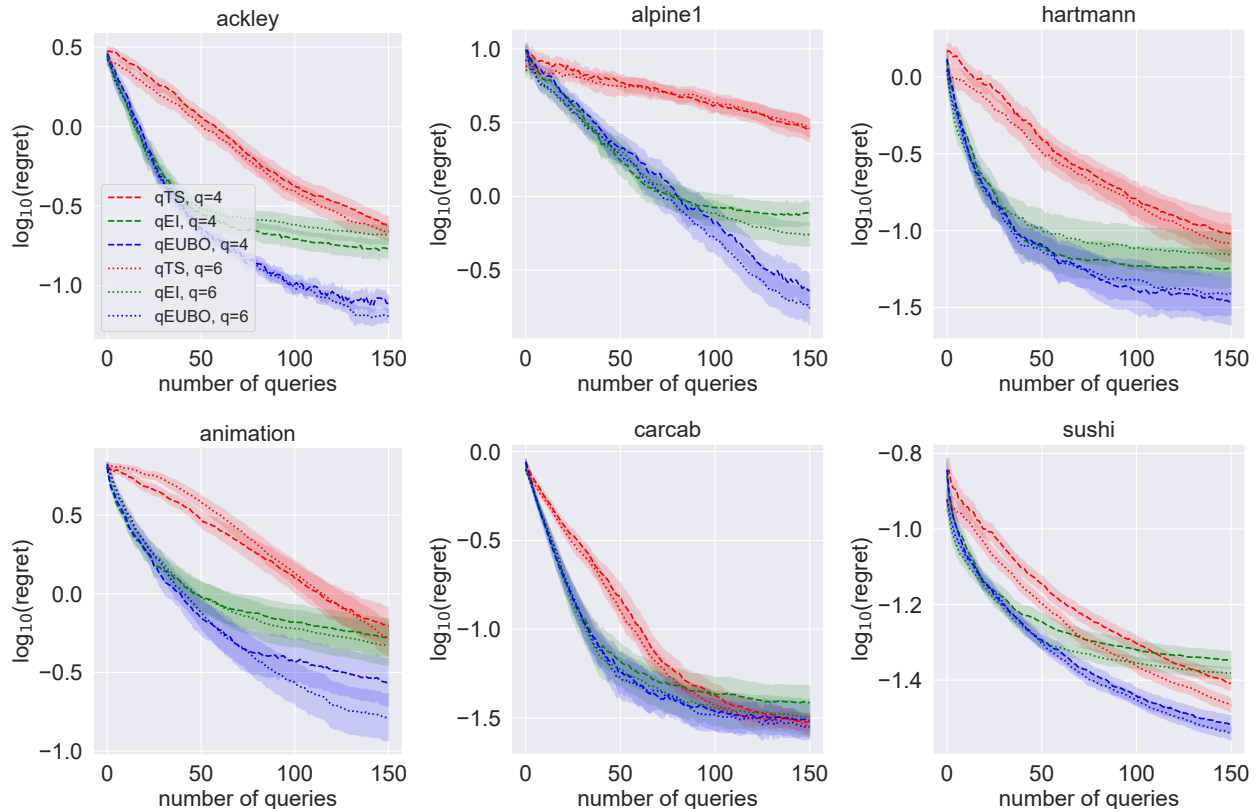


Figure 6:  $\log_{10}(\text{optimum value} - \text{objective value at the maximizer of the posterior mean})$  using moderate logistic noise and  $q = 4$  and  $q = 6$  alternatives per DM query.

## C.2 Experiments with Varying Levels of Noise

Figure 7 shows the results of three test problems (Ackley, Animation, and Sushi) by injecting varying levels of comparison noise. Left, center, and right columns show the results for low, middle, and high noise levels respectively. The noise levels are chosen so that the DM makes a comparison mistake 10, 20, and 30% of the time on average when asked to compare random pairs of points among those with the top 1% function values within the optimization domain  $\mathbb{X}$  under a logistic likelihood. Those noise levels are estimated over a large grid of random points in  $\mathbb{X}$ .

For experiments with lower noise, the performance gap between qEUBO and other baseline methods is most pronounced. This is consistent with Theorem 2, which shows that qEUBO is a better approximation of the one-step Bayes optimal policy for lower noise levels. With increasing noise, the performance of all methods decreases. However, we consistently observe superior performance of qEUBO compared to other baseline methods. These results are averaged over 50 replications for Ackley, and 100 replications for Animation and Sushi.

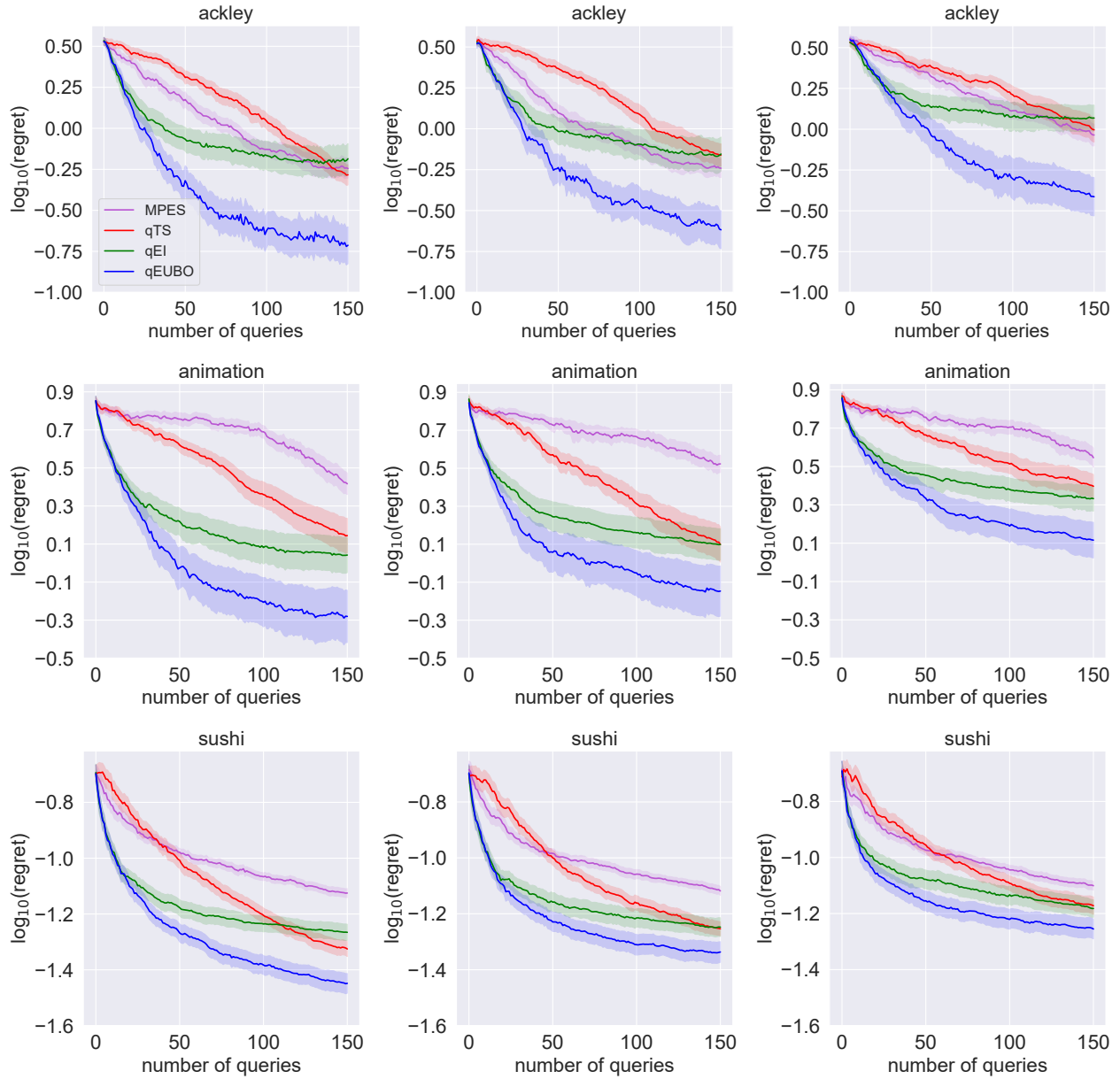


Figure 7:  $\log_{10}(\text{optimum value} - \text{objective value at the maximizer of the posterior mean})$  with  $q = 2$  alternatives per DM query and varying levels of noise. Left, center, and right columns show the results for three of our experiments with low, middle, and high noise, respectively.