Inverse Bayesian Optimization: Learning Human Acquisition Functions in an Exploration vs Exploitation Search Task

Nathan Sandholtz[*], Yohsuke Miyamoto[†], Luke Bornn[‡] and Maurice A. Smith[§]

Abstract. This paper introduces a probabilistic framework to estimate parameters of an acquisition function given observed human behavior that can be modeled as a collection of sample paths from a Bayesian optimization procedure. The methodology involves defining a likelihood on observed human behavior from an optimization task, where the likelihood is parameterized by a Bayesian optimization subroutine governed by an unknown acquisition function. This structure enables us to make inference on a subject's acquisition function while allowing their behavior to deviate around the solution to the Bayesian optimization subroutine. To test our methods, we designed a sequential optimization task which forced subjects to balance exploration and exploitation in search of an invisible target location. Applying our proposed methods to the resulting data, we find that many subjects tend to exhibit exploration preferences beyond that of standard acquisition functions to capture. Guided by the model discrepancies, we augment the candidate acquisition functions to yield a superior fit to the human behavior in this task.

Keywords: Bayesian optimization, directional statistics, exploration vs. exploitation, human cognition, inverse optimization, lab experiment, probabilistic models.

1 Introduction

The problem we study in this paper is a type of supervised learning problem: given observed human behavior from a sequential optimization task, we aim to create a model that accurately describes how they will behave on future rounds of the task. However, unlike most supervised learning problems, we explicitly assume that the observed behavior in each round of the task represents an approximate solution to a latent optimization problem faced by the subject. Not only do we want to incorporate this optimization subroutine into our model, but inferring the criteria governing the subject's optimization strategy is actually our primary interest. This type of analysis can be termed an *inverse optimization*.

A notable gap within the inverse optimization literature is the lack of statistical treatment of its methodology. Only very recently have statistical concepts such as consistency

---

(Aswani et al., 2018), goodness of fit (Chan et al., 2019), and robustness (Ghobadi et al., 2018; Esfahani et al., 2018; Shahmoradi and Lee, 2019) been treated in methodological papers. This is surprising because many optimization problems are assumed to be carried out imperfectly, thereby introducing noise to the data. In such cases, inverse optimization becomes an inference problem well-suited to probabilistic modeling. A key contribution of our work is the formulation of an inverse decision problem via a probabilistic model, which provides a principled framework to quantify both the variability with which the subject performs the optimization as well as the uncertainty around the parameter estimates governing their underlying optimization strategy.

We illustrate our methods by analyzing human decision-making behavior from a sequential optimization task developed in collaboration with the Neuromotor Control Lab at Harvard University. As will be further explained in Section 2, the task was specifically designed to present subjects with a conflict between exploration and exploitation on the second move of the task. Our goal is to make inference on how the subjects strategically balance exploration and exploitation based on their behavior on this second move.

In order to make inference on the subjects' strategies, we require a model for their optimization procedure. As will be detailed in Section 3, our model framework draws on Bayesian optimization, which is a sequential model-based approach to maximize (or minimize) an unknown objective function (Jones et al., 1998; Shahriari et al., 2015).[1] The core idea of Bayesian optimization is to characterize the uncertainty about an unknown objective function with a statistical model, termed a *surrogate*, then to strategically synthesize the model uncertainty via an *acquisition function* in order to determine promising new locations to sample. We adopt this terminology in our paper, as the fundamental components of our optimization model follow those of Bayesian optimization.

Within this context, our goal becomes to characterize and estimate subjects' acquisition functions (i.e. their exploration/exploitation preferences) based on their observed behavior. The novelty of our method is that we define a likelihood on a subject's observed search paths parameterized by an underlying Bayesian optimization subroutine. This structure enables us to make inference on the acquisition function governing the subject's optimization subroutine while allowing their behavior to deviate around the corresponding solution. By using probabilistic models of this form, both the variability with which subjects optimize as well as the uncertainty around their estimated exploration/exploitation preferences can be quantified.

We find that a wide range of acquisition functions are exhibited across the subjects in our study, but probability of improvement (PI) and upper confidence bound (UCB) functions tend to fit the majority of subjects best. Unlike previous studies of this problem, we are able to provide credible intervals to characterize the uncertainty in the estimated acquisition functions. Finally, there are many subjects for which none of the candidate acquisition functions we initially consider provide a satisfactory model of their behavior. Guided by the model discrepancies for these subjects, we propose

---

[1]Bayesian optimization is closely related to (or synonymous with, in some cases) sequential design of experiments, efficient global optimization, adaptive sampling, active learning, optimal search, and hyperparameter optimization.

an augmentation to the acquisition functions to construct a superior model of human optimization behavior in this task.

1.1 Related work and contributions

A number of recent papers investigate the correspondence between ML algorithms and actual human behavior for various decision-making processes (Borji and Itti, 2013; Wilson et al., 2015; Schulz et al., 2015; Wu et al., 2018; Plonsky et al., 2019; Gershman, 2019; Candelieri et al., 2020). Within this context, Borji and Itti (2013) compare several ML techniques against observed human behavior in a set of sequential optimization tasks that require subjects to make choices along one continuous dimension, finding that Bayesian optimization algorithms best approximate the human behavior. Recent papers expand their work to tasks that require discrete choices (Wu et al., 2018) and tasks where the decision space is continuous in two dimensions (Candelieri et al., 2020). These studies essentially explore the inverse problem of Bayesian optimization. Our research builds on these studies by introducing a probabilistic framework for the inverse problem, which we introduce in Section 4.

More broadly, our work contributes to the literature on inverse optimization (Ahuja and Orlin, 2001) and related problems such as inverse reinforcement learning (Ng et al., 2000) and inverse decision theory (Swartz et al., 2006). Li (2020) is particularly relevant to our project, which uses inverse optimization to learn a convex function representative of a subject's risk preferences. The estimand in our project—the acquisition function of a Bayesian optimization—is analogous to the risk preferences estimated in Li's paper.

1.2 Outline

The rest of this paper is outlined as follows. In Section 2, we describe the search task and explain how it forces subjects to make an exploration vs exploitation trade-off on the second move of the task. In Section 3, we frame the search task as a Bayesian optimization procedure and show how the forward optimization proceeds under three acquisition function classes. Section 4 introduces the inverse problem and illustrates our proposed solution framework based on the second move of the search task. We also augment the acquisition functions to enable a more accurate fit to the human tendencies we observe. In Section 5 we present our results and explore relationships between the estimated acquisition functions and subjects' corresponding performance in the task. Section 6 summarizes our contributions and suggests directions for future work.

All of our data and code to reproduce our results is hosted publicly at the first author's GitHub page: https://github.com/nsandholtz/hotspot_paper.

2 Search task description

In collaboration with the Neuromotor Control Lab at Harvard University, we designed a "hotspot" search task to present subjects with exploration vs exploitation conflicts

while searching for an unknown optimal location. On each round, subjects searched for an invisible target location (a "hotspot") randomly placed on a 9-inch diameter circular task-region shown to them on a computer monitor. Rounds consisted of 3 to 10 moves, where the number of moves was determined randomly according to a uniform distribution and the number was unbeknownst to the subject until the end of the round.

Within a round, a "move" entails sampling a location on the task region (i.e. clicking on it) after which a numerical score is immediately displayed to the user proportional to the click location's proximity to the hotspot. Specifically, the reward received on move t is a deterministic function of the distance from the click location to the hotspot:

$$\begin{aligned} r_t &= f(d(\mathbf{m}_t)) \\ &= r_0 + k \times (d(\mathbf{m}_0) - d(\mathbf{m}_t)), \end{aligned} \tag{2.1}$$

where r_0 denotes the initial score, $d(\mathbf{m}_t)$ denotes the Euclidean distance (in pixels) of move $\mathbf{m}_t$ (a two-dimensional coordinate) to the hotspot, and k denotes the reward scale (i.e. the reward's sensitivity to hotspot distance). The reward scale k is randomly generated for each new round of the task according to a Uniform$[0, 5/3]$ points/pixel distribution and is opaque to the subject. This prevents subjects from gaining information about the objective function along the orthogonal direction of their first move, thus ensuring that the second move presents a tradeoff between exploration and exploitation.[2]

To minimize effects of the task-region boundary guiding subject search behavior, the task always began at the center of the task region. We refer to this initial starting point as "move 0". Each subsequent move was constrained to be within a small circular region (0.2 inch radius) around the previous move. Subjects performed many rounds of this task; across the 28 subjects who participated, the minimum number of rounds played was 228 and the maximum was 716. Panel (a) of Figure 1 displays an example round of the experiment.

We are primarily interested in the second move (move 2) of the task because this move provides the most information about a subject's exploration vs. exploitation preferences. The first move (move 1) a subject makes is virtually random since they have no information about the direction of the hotspot upon beginning each round, as illustrated in panel (b) of Figure 1. After receiving feedback from their first move, the subject acquires information about the gradient of the objective function *but only along a single direction*. This poses a conflict between exploration and exploitation on the second move: continuing along the direction of their first move represents pure exploitation of their current knowledge, while any deviation from this direction represents some degree of exploration. Thus the combination of a subject's first and second moves provides direct insight into how they balance exploration and exploitation, as illustrated in panel (c) of Figure 1.

---

[2]If the reward scale was identical from round to round, subjects could solve for the gradient information along the direction orthogonal to their first move despite only having two responses from the objective, thus giving them a near-complete characterization of the local reward surface on their second move. By randomizing the reward scale on each round, we ensure that subjects have uncertainty about the gradient information orthogonal to their first move. This argument implicitly assumes that we can linearize the response surface in the local region defined by the move 1 boundary. We justify this assumption in Section 3.
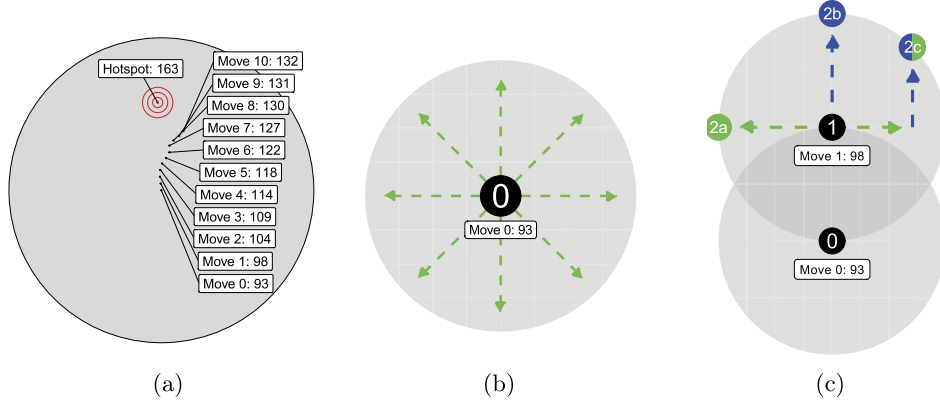
Figure 1: (a): An example round of the hotspot task. The red target shows the invisible hotspot location and the dots track the subject's search path. Each move's score is shown in a text box. The score at the hotspot is the score the subject would be given if they sampled the hotspot. (b): Initial state (move 0) for the example round shown in (a), zoomed in to the move 1 click-region. Moving in any direction represents exploration, as shown by the dashed green lines. (c): The subject moves to the upper edge of the move 1 click-region, receiving a score of 98. This creates a conflict between exploration (green) and exploitation (blue) on move 2. Moving perpendicular to direction of move 1 represents pure exploration (move 2a), while moving along the same direction of move 1 represents pure exploitation (move 2b). Any move between these extremes represents a combination of exploration and exploitation (move 2c).

In the next section we frame this task as a Bayesian optimization procedure. Ultimately, our goal is to characterize and estimate subjects' exploration/exploitation preferences via the acquisition function of the Bayesian optimization model of their behavior on the first and second moves of the task.

3   Optimization framework

Mathematically, the subjects' goal is to maximize an unknown objective function f

$$\max_{\mathbf{m} \in \mathcal{M}} \quad f(\mathbf{m}) \tag{3.1}$$

where $\mathbf{m}$ is a two-dimensional location and $\mathcal{M}$ is the circular task region defined on $\mathbb{R}^2$. However, as each move is constrained to be within a small region about the location of their previous move, the problem can be more precisely characterized as

$$\begin{aligned} \max_{\mathbf{m}_k \in \mathcal{M}_k} \quad & f(\mathbf{m}_k) \\ \text{s.t.} \quad & \mathbf{m}_0 = (0,0) \\ & \mathbf{m}_i \in \mathcal{M}_i, \quad i \in 1, \ldots, k \end{aligned} \tag{3.2}$$

where $\mathcal{M}_i$ is the local click-region surrounding $\mathbf{m}_{i-1}$ as described in Section 2, and k is randomly selected from the uniform distribution on the integers from 3 to 10.

In order to make inference on how the subject's solve (3.2), we model their optimization procedure as a Bayesian optimization (Borji and Itti, 2013). This entails characterizing the uncertainty about the unknown objective function with a surrogate, then strategically synthesizing the surrogate uncertainty via an acquisition function in order to determine promising new locations to sample. We define this framework in the context of the first and second moves of the task, as we exclusively consider these moves in the inverse problem. Appendix A in the supplemental material to this paper (Sandholtz et al., 2022) explains why moves 3 and up provide comparatively little information about the subjects' acquisition preferences.

3.1 Choosing a surrogate

A 'surrogate' is simply a term for a statistical model with an emphasis on pragmatism and prediction rather than interpretability and identification (Gramacy, 2020). Bayesian optimization applications typically use Gaussian process (GP) surrogates, which is a highly flexible class of models that can easily be updated as new samples from the objective are obtained.

In our case, GP surrogates would inaccurately specify the subjects' beliefs about the objective given their prior knowledge of the reward structure, and would therefore invalidate any inference about their acquisition function. Subjects know a priori that there is an optimum somewhere on the circular task-region and that the rewards decrease uniformly and monotonically from this hotspot with distance. This reward structure is illustrated in Figure 2, which shows the objective function from the example round shown previously.

As illustrated in the left plot, globally the objective function is conical. However, the experiments were designed to yield a surface that is approximately linear in the *local* region around each individual move, as shown in the right panel. Given the subjects' prior awareness of the reward structure, in addition to the approximate linearity in the neighborhood defined by a move's click-region, we assume that the subjects use a linear model as a surrogate of the objective:

$$\hat{f}_1(\mathbf{m}_1) = r_0 + \mathbf{m}_1^\intercal \boldsymbol{\beta} + \epsilon_1 \tag{3.3}$$
$$\epsilon_1 \sim \mathcal{N}(0, \sigma_s^2),$$

where $\mathbf{m}_1^\intercal = (x_1, y_1)$, $\boldsymbol{\beta} = \begin{pmatrix} \beta^x \\ \beta^y \end{pmatrix}$ represents the reward gradient with respect to the Cartesian plane and ϵ_1 represents a deviation from the surrogate to the true objective, which we assume to be Gaussian. We fix σ_s at a tiny value ($\sigma_s = 0.01$) because the deviations from the linear model to the objective are negligible in the local region about $\mathbf{m}_1$.[3]

---

[3]In order to have an analytic update for the surrogate after each move, we use a Gaussian distribution to model the error term ϵ_1. While this greatly speeds up the inference when solving the inverse problem, it is actually a misspecification—the errors are not Gaussian. Conditional on feedback from the first move, the model given by (3.3) is effectively a tangent plane of the underlying conical objective, hence the ϵ_1 are almost exclusively non-negative. The mean of $\hat{\epsilon}_1$ across all rounds of the experiment is 0.006.
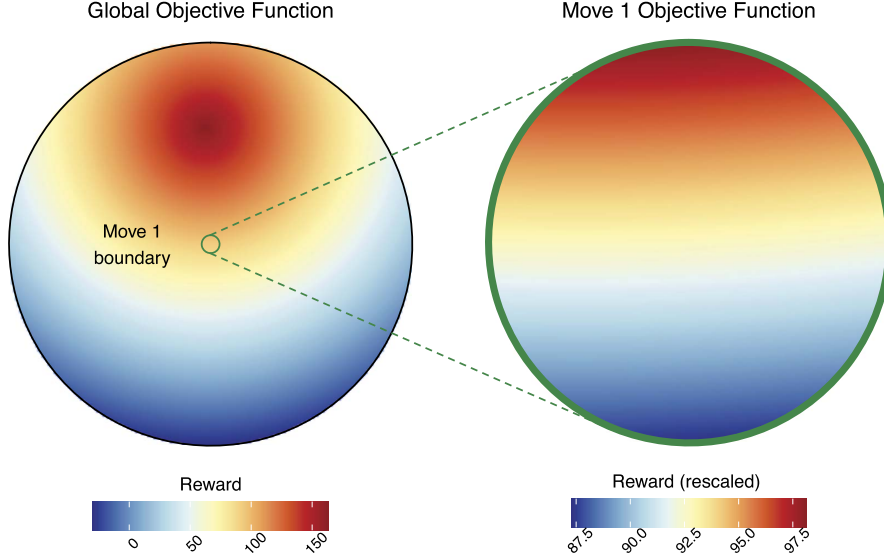
Figure 2: *Left*: The global objective function for the example round shown in Figure 1. *Right*: The objective function zoomed in to the click-region of the first move.

The approximate linearity of the objective function in the click-region around each move is an important feature of the experiment, as it is what allows us to assume that the directions of exploration and exploitation are orthogonal on move 2. We note, however, that the fidelity of the linear approximation to f declines as the distance to the hotspot decreases. This fact required a careful design of the experiment; we wanted the majority of clicks to take place far from the hotspot to make the linear approximation more accurate, but we also wanted the hotspot to be reachable so that subjects performed the optimization seriously. Balancing these competing desires, we selected the number of moves per round and the size of the local click-region such that it would be impossible to cover the distance to the hotspot in approximately 90% of all rounds. We omitted rounds where the hotspot was reachable in our analysis.

3.2 Updating the surrogate

As subjects sample new locations and receive additional feedback, their uncertainty about the objective function diminishes. In the context of Bayesian optimization, this corresponds to updating the surrogate as new information is gained about the objective function.

We apply a Bayesian framework for the surrogate inference. Since the hotspot is randomly placed over the task-region, we assume an isotropic, zero-mean Gaussian prior for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} \sim \mathcal{N}_2\left(\boldsymbol{\mu}_0 = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right), \boldsymbol{\Sigma}_0 = \sigma_\beta^2 \mathbf{I}\right), \tag{3.4}$$
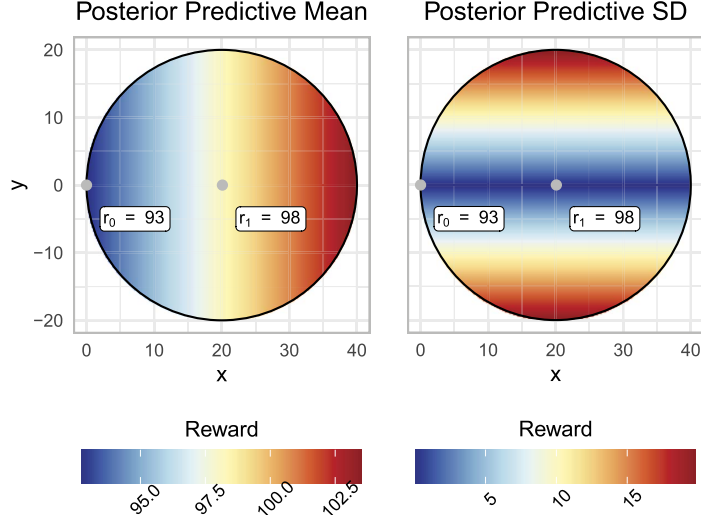
Figure 3: The move 2 posterior predictive distribution of $\hat{f}_1$ given $r_0 = 93$ and $r_1 = 98$. The left and right plots show the posterior predictive mean and standard deviation surfaces respectively, projected onto a fine grid over the move 2 click-region.

where σ_β^2 is a hyperparameter that we select to give a weakly-informative prior on the mean. Let the observed search path after move 1 be denoted as $\mathcal{D}_1 = \{\mathbf{M}_1, \mathbf{r}_1\}$, where $\mathbf{M}_1$ is a 2×2 matrix with the first row equal to $(0,0)$ and the second row equal to $\mathbf{m}_1^\mathsf{T}$, and $\mathbf{r}_1$ is a length-2 column vector of observed rewards on moves 0 and 1. Given $\mathcal{D}_1$, (3.4) can be updated yielding a posterior distribution on $\boldsymbol{\beta}$. The prior is conjugate and the posterior can be computed analytically (Taboga, 2021, Online appendix):

$$\boldsymbol{\beta}|\mathcal{D}_1 \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{3.5}$$

$$\boldsymbol{\Sigma} = \Big(\frac{1}{\sigma_s^2}\mathbf{M}_1^\mathsf{T}\mathbf{M}_1 + \boldsymbol{\Sigma}_0^{-1}\Big)^{-1}, \tag{3.6}$$

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}\Big(\frac{1}{\sigma_s^2}\mathbf{M}_1^\mathsf{T}(\mathbf{r}_1 - r_0) + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\Big). \tag{3.7}$$

The posterior predictive distribution of the surrogate for move 2 is given by:

$$\hat{f}_1(\mathbf{m}_2|\mathcal{D}_1) \sim \mathcal{N}(\mathbf{m}_2^\mathsf{T}\boldsymbol{\mu} + r_0, \ \mathbf{m}_2^\mathsf{T}\boldsymbol{\Sigma}\mathbf{m}_2 + \sigma_s^2), \tag{3.8}$$

where $\mathbf{m}_2$ is a potential move 2 location and all other terms are as defined previously (Taboga, 2021, Online appendix). Figure 3 shows the updated surrogate predictive distribution for an example round after move 1.

3.3 Choosing an acquisition function

The updated surrogate can be utilized to obtain promising new locations to sample. This process is operationalized through the acquisition function. The acquisition function u

is a function of a proposed location $\mathbf{m}_2$ and the surrogate after move 1 (which itself is a function of $\mathcal{D}_1$). Acquisition functions are typically constructed such that high values of the function correspond to *potentially* high values of the objective, either because the predicted mean is high, the uncertainty is high, or both (Brochu et al., 2010). The experimenter maximizes this function over $\mathcal{M}_2$, the set of potential move 2 locations, thus obtaining a new location to sample:

$$\mathbf{m}_2 = \underset{\mathbf{m} \in \mathcal{M}_2}{\arg\max} \ u(\mathbf{m}|\hat{f}_1, \mathcal{D}_1). \tag{3.9}$$

A host of acquisition functions have been proposed in the literature (see Shahriari et al. (2015) for a review). While any acquisition function that selects new locations deterministically is a viable candidate for the methods we introduce in this paper, we restrict our analysis to probability of improvement (PI), expected improvement (EI), and an upper confidence bound (UCB) criterion since these functions are well-known among Bayesian optimization practitioners and have analytic solutions under Gaussian models (Jones et al., 1998; Srinivas et al., 2009). Mathematically, these are defined by

$$u_{\text{PI}}(\mathbf{m}|\hat{f}_1, \xi_{\text{PI}}, \mathcal{D}_1) = \mathbb{P}\big(\hat{f}_1(\mathbf{m}|\mathcal{D}_1) \geq f(\mathbf{m}_1^+) + \xi_{\text{PI}}\big) \tag{3.10}$$

$$u_{\text{EI}}(\mathbf{m}|\hat{f}_1, \xi_{\text{EI}}, \mathcal{D}_1) = \mathbb{E}[\max\big(0, \hat{f}_1(\mathbf{m}|\mathcal{D}_1) - (f(\mathbf{m}_1^+) + \xi_{\text{EI}}))] \tag{3.11}$$

$$u_{\text{UCB}}(\mathbf{m}|\hat{f}_1, p, \mathcal{D}_1) = \text{infimum}\{\mathbf{m} : p \leq \hat{F}_1(\mathbf{m}|\mathcal{D}_1)\}, \tag{3.12}$$

where $p > 0.5$, $\mathbf{m}_1^+$ is the best location observed so far over the two existing samples (i.e. $\mathbf{m}_1^+ = \text{argmax}_{\mathbf{m} \in \mathbf{M}_1} f(\mathbf{m})$), and $\hat{F}_1$ is the cumulative distribution function of $\hat{f}_1$.

Each of the acquisition functions in (3.10)-(3.12) depend on an additional parameter which controls the premium on exploration. Acquiring via PI (Kushner, 1964) results in the location that most confidently predicts an increase in the response but ignores improvements less than ξ_{PI}. EI (Mockus et al., 1978) considers the magnitude of improvement at a particular location, where ξ_{EI} controls a tradeoff between exploration (high ξ_{EI}) and exploitation (low ξ_{EI}) (Lizotte, 2008). UCB acquisition functions (Cox and John, 1992; Srinivas et al., 2009) select new locations to sample based on optimistic estimates of the resulting reward at each location. Higher values of p encourage more exploration.[4]

Figure 4 shows the acquisition surfaces for the PI, EI, and UCB functions defined in (3.10)-(3.12) in the context of the example round shown in Figure 1. We reoriented the data so that move 1 falls along the horizontal-axis for visual clarity. The $\arg\max(s)$ of each surface is denoted by a green star.

Notice that each method prescribes different move 2 locations; PI recommends pure exploitation, while EI and UCB recommend almost pure exploration. Of course, this is not always the case. Depending on the change in score from move 0 to move 1,

---

[4]The UCB acquisition function is usually defined in terms of a mean function and covariance function since the surrogate is a Gaussian process in most Bayesian optimization applications. We define it via the quantile function in order to have an upper and lower bound on the parameter p, which simplifies the methodology in Section 4.
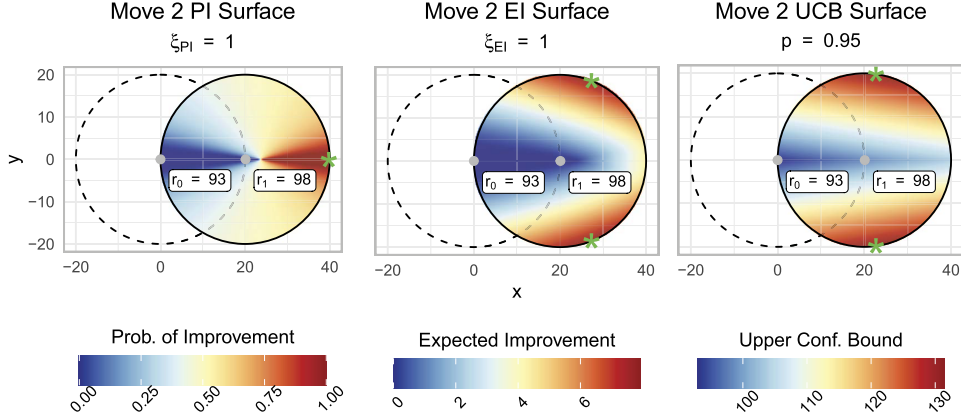
Figure 4: Acquisition surfaces for the PI, EI, and UCB functions defined in (3.10)-(3.12) given $r_0 = 93$ and $r_1 = 98$. Each surface is on a different scale: the PI surface (left) is on the probability scale, EI (middle) is in terms of points over 99, and UCB (right) is in terms of the 95th percentile of the surrogate. The click-region boundary is shown by the black circle encompassing the colored surfaces and the $\arg\max$(s) of each surface is denoted by a green star.

in addition to the exploration parameter values (ξ_{PI}, ξ_{EI}, and p), the optimal locations returned by the acquisition functions can vary substantially. Also note that each surface is symmetric across the exploitation axis (i.e. $y = 0$). This is a feature of the subject having gained information exclusively about a single direction after the first move. Due to this phenomenon, the surfaces may be bimodal or unimodal.

Because we use a linear model as the surrogate for the objective function, the optimal location to sample on move 2 (as determined by (3.9)) always falls on the click-region boundary, regardless of the change in score on move 1. This harmonizes with the subjects' actual behavior; in 95% of the rounds, subjects moved to the click-region boundary on move 2.[5] This effectively allows us to reduce the $\arg\max$ search from two dimensions to one dimension. Leveraging polar coordinates, we can fix the radius of the move 2 location at the click-region boundary and limit the $\arg\max$ search solely to θ_2, the angle of move 2 relative to the direction of move 1:

$$\underset{\mathbf{m} \in \mathcal{M}_2}{\arg\max} \ u(\mathbf{m}|\hat{f}_1, \mathcal{D}_1) \longrightarrow \underset{\theta \in (-\pi, \pi]}{\arg\max} u(\theta|\hat{f}_1, \mathcal{D}_1).$$

By reducing the dimensionality of the optimization we can illustrate acquisition values over the entire range of $\Delta r_1 = r_1 - r_0$, the change in score after a subject's first move, as shown in Figure 5. The three plots show acquisition values over a fine grid of $(\Delta r_1, \theta_2)$ pairs for the same acquisition functions shown in Figure 4. In each plot,

---

[5]The search task was programmed to snap the cursor to the edge of the boundary if the mouse exceeded it, which made it easy for subjects to make moves along the boundary.
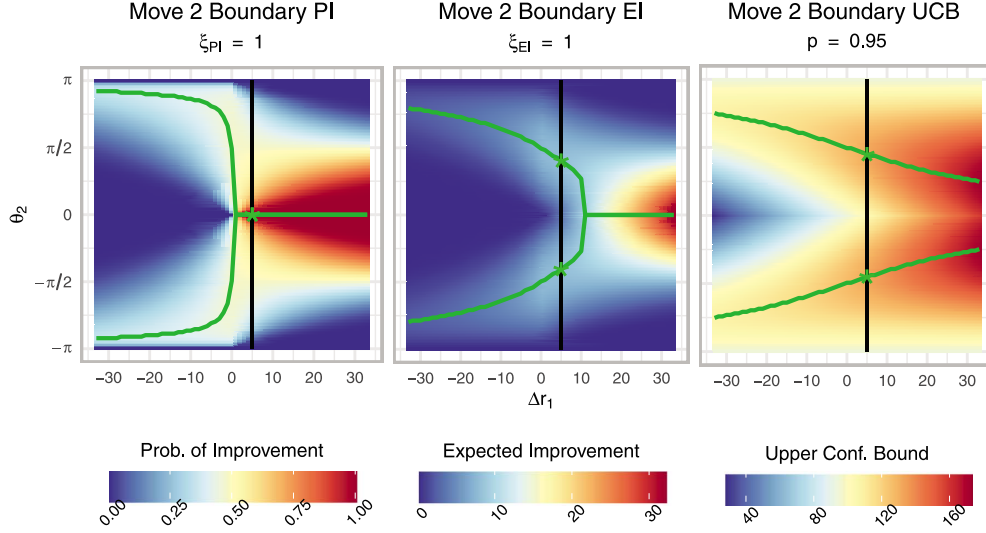
Figure 5: Click-region boundary acquisition values over the range of possible Δr_1 values for three sample acquisition functions. The horizontal axis denotes Δr_1, the change in reward from $\mathbf{m}_0$ to $\mathbf{m}_1$. The vertical axis shows θ_2, the angle of the second move relative to the first move. Color indicates the acquisition function value for any given $(\Delta r_1, \theta_2)$ pair. In each plot, the green curve denotes the angle that yields the maximum of the boundary acquisition values as a function of Δr_1. The vertical black lines at $\Delta r_1 = 5$ correspond to the circular black lines denoting the click-region boundaries in Figure 4. Similarly, the green stars correspond to the stars in Figure 4.

the green curves denote the angles that yield the maximum of the acquisition values as a function of Δr_1. Note that the move 2 locations are assumed to be made on the click-region boundary.

Going forward, we will refer to an *acquisition curve* as the curve defined by the optimal θ_2 as a function of Δr_1 for a given acquisition function. Figure 6 shows various acquisition curves from the PI, EI, and UCB acquisition families. Depending on the acquisition family and exploration parameter, the acquisition curves can vary substantially.

Our goal is not to determine an optimal strategy, or acquisition function, for this task. Rather, using Bayesian optimization as a model of each subject's optimization procedure, we aim to learn subject-specific acquisition functions guided by their observed behavior in the experiment.

Before proceeding to Section 4, which formally introduces the inverse problem, we pause to comment on the unconventional nature of the Bayesian optimization framework we have developed thus far. The reward structure and transparent experimental setup (which allows subjects to understand the reward surface in relatively few moves) were carefully designed to allow us to assume that subjects' beliefs about the reward
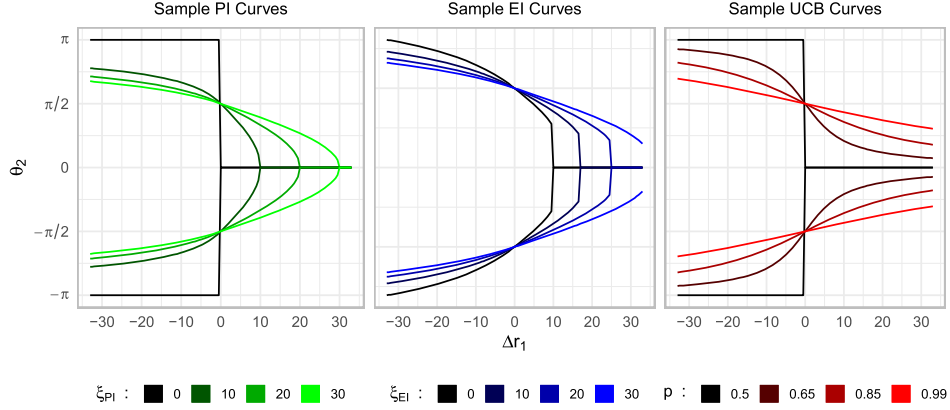
Figure 6: Acquisition curves for various exploration parameter values of the corresponding family—PI (left), EI (middle), and UCB (right).

4   Learning human acquisition functions

In this section we propose a method to estimate a subject's unknown acquisition function u given their observed decisions on move 2 assuming that these decisions are approximately optimal solutions to (3.9) with respect to u. We say 'approximately optimal' because the subjects' decisions exhibit considerable variability, as illustrated in Figure 7 which shows the move 2 behavior for three subjects in our study as a function of Δr_1.[6]

This variability could arise from a number of factors. For one, the modeling described in Section 3 happens subconsciously; subjects "update" their uncertainty about the objective function intuitively, which introduces human error. Noise could also be due to variation in a subject's ability to click exactly where they intend to, changes to their sampling strategy, or sloppiness from performing the task rapidly.

Formally, given a collection of N move 2 angles $\boldsymbol{\theta_2} = (\theta_2^1, \ldots, \theta_2^N)$ that were taken after receiving feedback from the corresponding move 1 (as represented by $\boldsymbol{\Delta r_1} = (\Delta r_1^1, \ldots, \Delta r_1^N)$), the inverse problem we consider is to identify an acquisition function $u \in \mathcal{U}$ that minimizes a given loss function $\ell(\cdot, \cdot)$

$$\min_{u \in \mathcal{U}} \ell(\boldsymbol{\theta_2}, \boldsymbol{\theta_2^*}(\boldsymbol{\Delta r_1}, u)), \tag{4.1}$$

where $\boldsymbol{\theta_2^*}(\boldsymbol{\Delta r_1}, u) = (\theta_2^*(\Delta r_1^1, u), \ldots, \theta_2^*(\Delta r_1^N, u))$ is the vector of optimal decisions

---

[6]Figure 13 in Appendix E of the supplemental material shows the (x,y) coordinates of these same data.
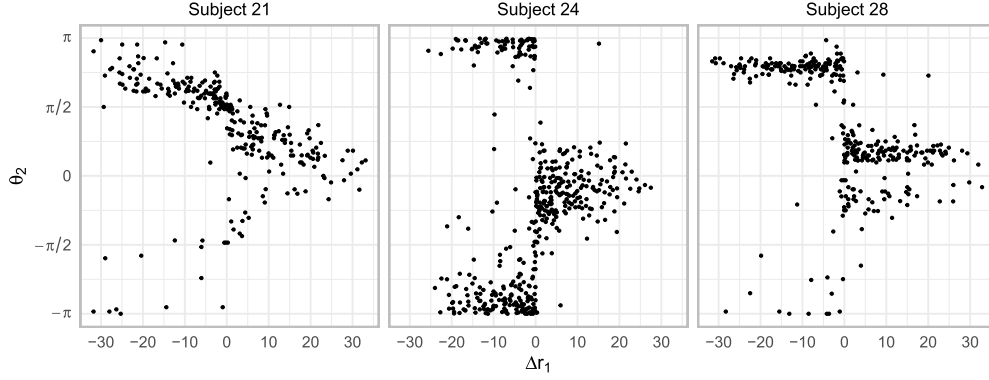
Figure 7: Move 2 behavior for subjects 21 (left), 24 (middle), and 28 (right). Each scatterplot shows the $(\Delta r_1, \theta_2)$ pairs for the subject's rounds of the task.

with respect to u conditional on $\mathbf{\Delta r_1}$, with components defined as

$$\theta_2^*(\Delta r_1^i, u) = \underset{\theta \in (-\pi, \pi]}{\arg\max} \, u(\theta | \Delta r_1^i, \hat{f}_1) \tag{4.2}$$

for $i = 1, \ldots, N$. The function class $\mathcal{U}$ contains candidate acquisition functions that summarize the value of a particular move 2 angle for a given Δr_1 value and surrogate $\hat{f}_1$.

Conceptually, solving (4.1) can be understood as finding acquisition curves (see Figure 6) that best fit the humans' observed optimization behavior (see Figure 7). As such, the problem becomes one of inference rather than optimization. The optimization has already occurred—we want to infer *how* each subject optimized.

4.1 A probabilistic solution framework

We assume subjects' optimization behavior is characterized by an average strategy but that their behavior deviates randomly around this strategy from round to round. Under this assumption, we propose a probabilistic solution framework for (4.1) as follows:

1. Assume a likelihood for $(\theta_2^i | \Delta r_1^i, \hat{f}_1, u)$, parameterized such that the mode of the distribution equals the arg max of the acquisition function u.

2. Select a set of candidate acquisition functions $\mathcal{U}$ as potential characterizations of the optimizer's preferred strategy and set a prior distribution over $\mathcal{U}$.

3. Compute the posterior probability of each candidate acquisition function $u \in \mathcal{U}$:

$$p(u | \boldsymbol{\theta_2}, \mathbf{\Delta r_1}, \hat{f}_1) \propto \prod_{i=1}^{N} \ell \left(\theta_2^i \, \middle| \, \underset{\theta \in (-\pi, \pi]}{\arg\max} \, u(\theta | \Delta r_1^i, \hat{f}_1) \right) p(u). \tag{4.3}$$

where $\ell(\cdot | \cdot)$ is the likelihood from step 1, and $p(\cdot)$ is the prior from step 2.

We employ a Bayesian framework in our approach, but the procedure could be carried out using other estimators (e.g. maximum likelihood). We now explain how we implement each of these steps in context of the hotspot search task.

Choosing the likelihood

To model the symmetric bimodal nature of our data, we assume θ_2^i follows a two-component wrapped Cauchy mixture distribution:[7]

$$\ell(\theta_2^i|w, \Delta r_1^i, u, \hat{f}_1, \gamma) = w \times h^-(\theta_2^i|\Delta r_1^i, u, \hat{f}_1, \gamma) + (1-w) \times h^+(\theta_2^i|\Delta r_1^i, u, \hat{f}_1, \gamma), \ (4.4)$$

where

$$h^-(\theta_2^i|\Delta r_1^i, u, \hat{f}_1, \gamma) = \sum_{k=-\infty}^{\infty} \frac{\gamma}{\pi(\gamma^2 + (\theta_2^i - \arg\max_{\theta\in(-\pi,0]} u(\theta|\Delta r_1^i, \hat{f}_1) + 2\pi k)^2)}, \qquad (4.5)$$

$$h^+(\theta_2^i|\Delta r_1^i, u, \hat{f}_1, \gamma) = \sum_{k=-\infty}^{\infty} \frac{\gamma}{\pi(\gamma^2 + (\theta_2^i - \arg\max_{\theta\in(0,\pi]} u(\theta|\Delta r_1^i, \hat{f}_1) + 2\pi k)^2)}. \qquad (4.6)$$

The location of each mixture component in (4.4) is constrained to be at the argmax(s) defined by $\arg\max_{\theta\in(-\pi,\pi]} u(\theta|\Delta r_1^i, \hat{f}_1)$. The parameter γ measures a subject's variation about their acquisition curve and the mixture weight w governs a subject's propensity to favor exploring to the right over exploring to the left. We use the wrapped Cauchy distribution because it is more robust to outliers, which are fairly common across the subjects' move 2 behavior.

Together, (4.4)-(4.6) yield a likelihood that adheres to the defining features of our data. It is not only defined on the proper support for θ_2 (i.e. $(-\pi, \pi]$), it is also guaranteed to be symmetric over the axis of exploitation (i.e. $\theta_2 = 0$); since the argmaxes defined in the denominators of (4.5) and (4.6) will always be reflected about $\theta = 0$, the corresponding mixture distribution will also be symmetric about $\theta = 0$.

The fact that we parameterize the likelihood such that the mode equals the arg max of the acquisition function is what renders this an 'inverse optimization' method. The novelty is that we couch this restriction within a likelihood, thereby allowing subject behavior to deviate around an 'optimal' strategy (with respect to a given acquisition function u) according to the specified probability distribution. This combination distinguishes our approach from both from other probabilistic models and other inverse optimization methods.

Setting priors

We select the PI, EI, and UCB acquisition families defined in (3.10)-(3.12) as the set of candidate acquisition functions $\mathcal{U}$. Note that our method does not preclude using other

---

[7]When only a single value defines $\arg\max_{\theta\in(-\pi,\pi]} u(\theta|\Delta r_1^i, \hat{f}_1)$, the likelihood simplifies to a single wrapped Cauchy distribution.

acquisition families—any acquisition function can be included in the candidate set $\mathcal{U}$. Each acquisition function $u \in \mathcal{U}$ is assumed to have equal prior probability, thus we place a uniform prior over the parameters indexing the acquisition function space: ξ_{PI}, ξ_{EI}, p. We also put a uniform prior distribution over γ, the scale parameter governing the variability in each subject's optimization behavior. Appendix B in the supplemental material to this paper contains additional details on prior specification.

Approximating the posterior

The existence of the $\arg\max$ in the parameterization of (4.4)-(4.6) creates an additional optimization step when evaluating (4.3). To ease the computational burden posed by this additional optimization subroutine, we precompute a fine grid of acquisition curves over a plausible range of values for each candidate family and restrict our inference to this discrete set. We also precompute the maximum likelihood estimate of the mixture weight w using the EM algorithm (Dempster et al., 1977). We can then compute (4.3) for each $u \in \mathcal{U}$ by multiplying its prior probability by the joint likelihood of $\boldsymbol{\theta_2}$ (plugging in the precomputed $\hat{w}$) and the prior over γ.

Figure 8 shows the $(\Delta r_1, \theta_2)$ pairs for subjects 21, 24, and 28 overlaid with their corresponding maximum a posteriori (MAP) acquisition curves in color. The light-gray shaded areas show approximate 95% highest posterior density (HPD) prediction regions, while the dark-gray regions denote 95% HPD credible regions on the acquisition curves.
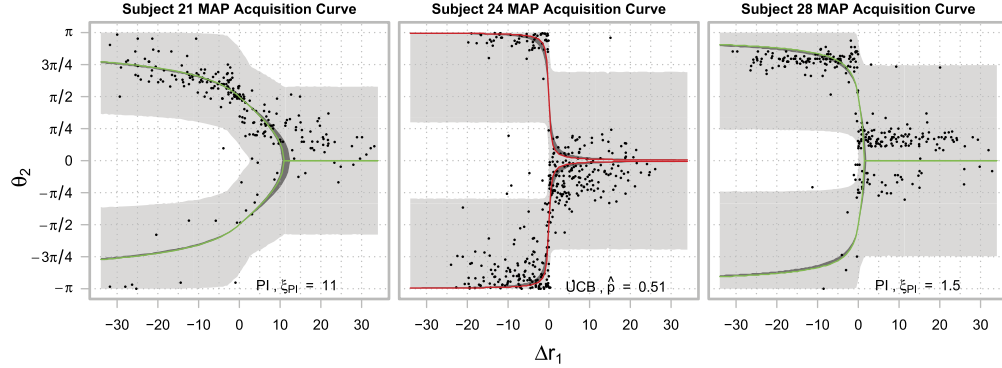
Figure 8: $(\Delta r_1, \theta_2)$ pairs for subjects 21 (left), 24 (middle), and 28 (right) overlaid with each subject's corresponding MAP acquisition curve. In each plot, text in the lower-right corner shows the estimated exploration parameter for the subject's MAP acquisition function. Around each curve is an approximate 95% HPD credible region on the acquisition curve (dark-gray) and 95% HPD prediction region (light-gray).

4.2   Incorporating human tendencies

On visual inspection of Figure 8, the MAP acquisition curves for subjects 21 and 24 fit their data fairly well but best fit curve for subject 28 fits poorly. This is because there

are no candidates in $\mathcal{U}$ that prescribe such a drastic change in strategy from negative Δr_1 to positive Δr_1 values while maintaining highly exploratory preferences. The set of acquisition functions as currently defined cannot yield acquisition curves with this shape. Consequently, there is no parameter setting of ξ_{PI}, ξ_{EI}, or p which renders subject 28's average behavior as optimal.

In order to allow this type of behavior to be rendered optimal, we propose augmenting the acquisition functions by an additional parameter τ. For a given acquisition function u, we define augmented acquisition function $\widetilde{u}$ as

$$\widetilde{u}(\theta_2|\Delta r_1, \hat{f}_1, \tau) = \begin{cases} u(\theta_2|\Delta r_1, \hat{f}_1) & \text{if } (|\theta_2| > \tau) \cap (\Delta r_1 \geq 0), \\ u(\theta_2|\Delta r_1, \hat{f}_1) & \text{if } (|\theta_2| < \pi - \tau) \cap (\Delta r_1 < 0), \\ \min_{\theta_2, \Delta r_1} \left(u(\theta_2|\Delta r_1, \hat{f}_1) \right) & \text{otherwise,} \end{cases} \quad (4.7)$$

where $\tau \in [0, \pi/2]$. Equation (4.7) defines the augmented value to be the minimum acquisition value (with respect to θ_2 and Δr_1) for all move 2 angles that aren't at least τ-radians exploratory, either in the forward or backward directions. If a given move 2 angle is greater than τ and less than $\pi - \tau$ (in absolute value), the move is "sufficiently exploratory" and the acquisition value is unchanged.

Figure 9 illustrates how this augmentation changes the acquisition surfaces and corresponding acquisition curves for the three acquisition functions represented previously in Figure 5.
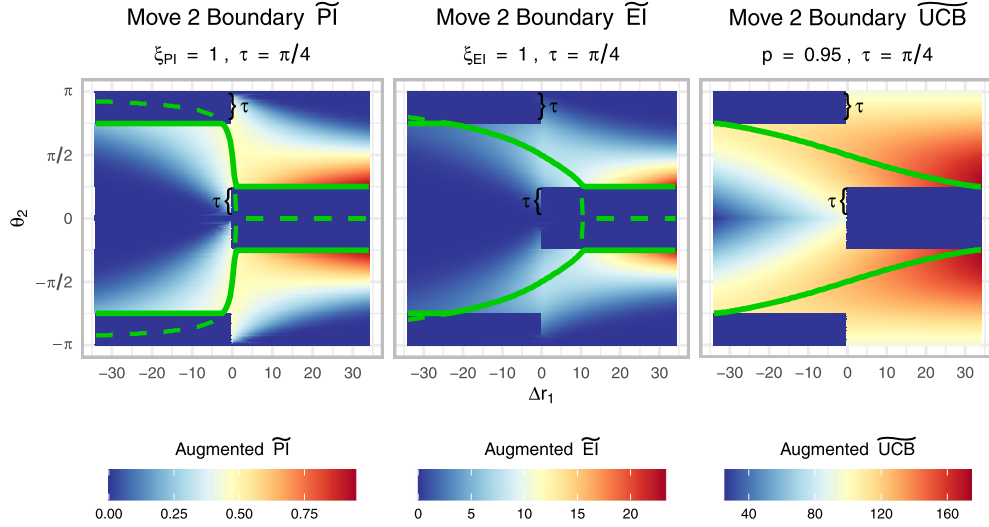
Figure 9: Click-region boundary augmented acquisition values over the range of possible Δr_1 values for three sample acquisition functions. The solid green curves denote the angles that yield the maximum of $\widetilde{u}(\theta_2|\Delta r_1, \hat{f}_1, \tau)$, while the dashed green curves show the angles that yield the maximum of u without being augmented by τ (i.e. $\arg\max u(\theta_2|\Delta r_1, \hat{f}_1)$).

Another feature we account for is the human tendency to react differently for positive vs. negative feedback (Tversky and Kahneman, 1979). Some subjects exhibit different exploration tendencies depending on whether they get a negative or positive change in score on their first move. In order to allow for this type of behavior we modify (4.7) to allow different values of τ depending on whether Δr_1 is positive or negative:

$$\widetilde{u}^{\pm}(\theta_2|\Delta r_1, \hat{f}_1, \tau^+, \tau^-) = \begin{cases} u(\theta_2|\Delta r_1, \hat{f}_1) & \text{if } (|\theta_2| > \tau^+) \cap (\Delta r_1 \geq 0), \\ u(\theta_2|\Delta r_1, \hat{f}_1) & \text{if } (|\theta_2| < \pi - \tau^-) \cap (\Delta r_1 < 0), \\ \min_{\theta_2, \Delta r_1} (u(\theta_2|\Delta r_1, \hat{f}_1)) & \text{otherwise.} \end{cases}$$

(4.8)

The practical effect of (4.7)-(4.8) is that they allow the optimization criteria to be based solely on exploration. While the parameters in the unmodified PI, EI, and UCB acquisition families allow the optimizer to *balance* exploration vs. exploitation differently when synthesizing the uncertainty in $\hat{f}$, they do not enable the optimizer to let exploration *completely dominate* exploitation. The augmented acquisition functions allow exploration to trump exploitation, regardless of $\hat{f}_1$. We observe this type of behavior by many subjects in our study.

Figure 10 shows the data for the same subjects as before, this time overlaid with the fitted models using the augmented acquisition function in (4.8). Visually, the fit appears to be superior.
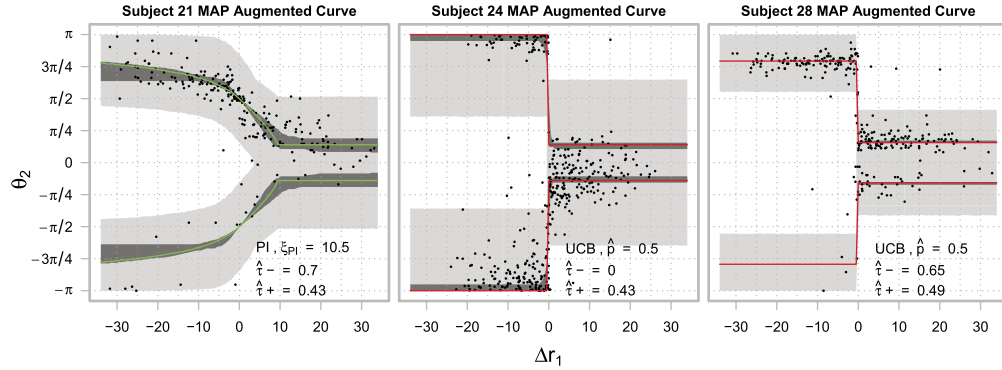
Figure 10: All pairs of $(\Delta r_1, \theta_2)$ data for subjects 21 (left), 24 (middle), and 28 (right). Each scatterplot is overlaid with the subject's corresponding MAP augmented acquisition curve in color. Green denotes PI and red denotes UCB. Around each curve is an approximate 95% HPD credible interval on the acquisition curve (dark gray) and 95% HPD prediction interval (light gray). Point estimates for the parameters governing each subject's acquisition function are listed in the lower right corner.

4.3   Model validation

We fit the acquisition models in (3.9), (4.7), and (4.8) to each subject's data following the procedure outlined in Section 4.1. For the augmented models we estimate the τ parameters as well. As with the other exploration parameters, we set discrete uniform priors on τ, τ^+, and τ^-. Table 1 shows out-of-sample log-likelihoods (using an 80%/20% train/test split of the data) for each model across the subjects in our study.

	Acquisition model				Acquisition model		
Subject	u	$\widetilde{u}$	$\widetilde{u}^{\pm}$	Subject	u	$\widetilde{u}$	$\widetilde{u}^{\pm}$
1	**-72.20**	**-72.20**	**-72.20**	15	-71.97	-70.48	**-68.85**
2	-72.84	-72.95	**-67.33**	16	-55.97	-55.97	**-35.07**
3	-70.23	-70.77	**-63.85**	17	-52.69	-47.06	**-36.55**
4	-74.76	-74.76	**-70.62**	18	-103.88	-99.60	**-78.09**
5	**-42.42**	**-42.42**	**-42.42**	19	-139.76	-138.81	**-115.09**
6	-47.16	3.37	**4.65**	20	**-26.53**	**-26.53**	-27.06
7	-73.80	-68.52	**-67.30**	21	-50.09	**-41.22**	-41.77
8	-77.99	-77.99	**-70.73**	22	-78.73	-77.45	**-75.40**
9	-34.44	-34.75	**-31.41**	23	-122.13	-122.13	**-122.12**
10	-68.14	-32.20	**-21.49**	24	-90.92	-87.96	**-81.29**
11	**-96.68**	**-96.68**	**-96.68**	25	-36.28	-24.00	**-18.03**
12	-75.46	-49.71	**-47.55**	26	**-80.13**	**-80.13**	-83.02
13	-95.25	-95.25	**-95.24**	27	**-69.83**	-74.28	-75.18
14	-77.52	-79.58	**-74.29**	28	-78.66	**-33.51**	-35.21
				Total	-2036.46	-1843.54	**-1709.19**

Table 1: Out-of-sample log-likelihoods for (4.4) fit using u, $\widetilde{u}$, and $\widetilde{u}^{\pm}$ for all 28 subjects in our study. The model(s) with the greatest log-likelihood is shown in bold for each subject. The bottom row shows the sum of the log-likelihoods across all subjects for each model.

The $\widetilde{u}^{\pm}$ model provides the best fit on aggregate, though there are some subjects for which the additional flexibility offered by the τ parameters does not substantially improve the fit. Unless otherwise specified, from this point onward all model references utilize the augmented acquisition function in (4.8).

We also have included a simulation study in Appendix C of the supplemental material to validate our methodology in a controlled setting. The study demonstrates that our method accurately recovers the true underlying acquisition function under different acquisition function parameters and observation sample sizes. It also explores how our method performs when the true acquisition function class is omitted from the inference acquisition function class.

5   Results

Table 2 shows each subject's MAP estimates of the $\widetilde{u}^{\pm}$ parameters from (4.8) in columns 1-3. Columns 4 and 5 show their estimates of γ and w, respectively. Figure 14 in Appendix E of the supplemental material plots the data, MAP augmented acquisition curves, and 95% HPD intervals for all 28 subjects.

	Fitted Model					Task Performance		
Subject	$\hat{\xi}_{\mathrm{PI}}/\hat{\xi}_{\mathrm{EI}}/\hat{p}$	$\hat{\tau^{-}}$	$\hat{\tau^{+}}$	$\hat{\gamma}$	$\hat{w}$	$\overline{\Delta r_2}$	$\overline{\Delta r_3}$	$\sum_{i=2}^{10}\Delta r_i$
1	$\hat{p}=0.50$	0.00	0.00	0.22	0.50	7.84	9.14	53.33
2	$\hat{p}=0.50$	0.16	0.32	0.13	0.52	8.78	9.66	54.07
3	$\hat{p}=0.50$	0.00	0.43	0.19	0.44	7.09	9.74	55.00
4	$\hat{p}=0.50$	0.22	0.00	0.11	0.50	8.74	9.99	57.37
5	$\hat{p}=0.50$	0.00	0.00	0.18	0.50	7.71	9.48	57.30
6	$\hat{p}=0.50$	1.57	1.57	0.06	0.92	0.12	13.79	61.17
7	$\hat{p}=0.50$	0.92	0.76	0.18	0.60	5.98	13.10	62.64
8	$\hat{p}=0.50$	0.22	0.00	0.09	0.51	7.89	8.86	47.85
9	$\hat{p}=0.50$	0.00	0.16	0.10	0.48	9.47	9.58	53.78
10	$\hat{p}=0.50$	0.81	0.92	0.09	0.72	6.25	11.39	51.06
11	$\hat{p}=0.50$	0.00	0.00	0.31	0.50	6.43	8.84	50.15
12	$\hat{p}=0.50$	0.76	0.65	0.11	0.59	7.22	13.07	66.12
13	$\hat{p}=0.52$	0.00	0.00	0.33	0.50	7.00	11.61	56.77
14	$\hat{p}=0.50$	0.22	0.76	0.23	0.55	5.54	10.11	49.29
15	$\hat{p}=0.50$	0.32	0.60	0.29	0.45	7.07	8.11	46.09
16	$\hat{p}=0.50$	0.00	0.60	0.13	0.36	9.98	13.40	72.27
17	$\hat{p}=0.50$	0.11	0.43	0.11	0.50	8.95	9.38	59.03
18	$\hat{p}=0.50$	1.57	0.54	0.27	0.46	3.63	10.22	50.49
19	$\hat{p}=0.50$	1.30	0.70	0.22	0.45	4.09	9.53	47.42
20	$\hat{p}=0.50$	0.16	0.00	0.14	0.50	9.13	10.53	56.24
21	$\hat{\xi}_{\mathrm{PI}}=10.5$	0.70	0.43	0.15	0.87	6.19	15.16	66.54
22	$\hat{p}=0.50$	0.22	0.27	0.19	0.53	8.37	10.74	61.12
23	$\hat{p}=0.50$	0.00	0.00	0.37	0.50	7.01	9.49	48.60
24	$\hat{p}=0.50$	0.00	0.43	0.24	0.41	8.83	10.95	59.73
25	$\hat{p}=0.50$	0.81	0.97	0.07	0.10	6.93	10.65	50.55
26	$\hat{p}=0.50$	0.22	0.00	0.18	0.48	8.44	9.66	58.67
27	$\hat{\xi}_{\mathrm{PI}}=7$	0.97	0.87	0.29	0.32	4.06	12.31	55.49
28	$\hat{p}=0.50$	0.65	0.49	0.10	0.83	6.99	14.37	70.59

Table 2: Each subject's estimated fit to (4.4) using the augmented acquisition formulation in (4.8). Columns 1-5 show MAP parameter estimates for the acquisition family shape parameter $(\xi_{\mathrm{PI}}/\xi_{\mathrm{EI}}/p)$, as well as estimates for τ^{-}, τ^{+}, γ, and w. Columns 6-8 show corresponding measures of subject performance in the search task, where $\overline{\Delta r_i}$ denotes the average change in score after a subject's ith move.

One observation that immediately jumps out from the table is that the vast majority of subjects are best represented by the UCB acquisition function with $p = 0.5$.[8] Before augmenting the acquisition function, this would imply that most subjects are purely exploitative on their second move, however, after augmenting the acquisition function this interpretation no longer holds. Depending on a subject's estimated values of τ^- and τ^+, subjects can exhibit highly exploratory behavior despite having $\hat{p} = 0.5$. Under the augmented acquisition function, the parameters $\xi_{\mathrm{PI}}/\xi_{\mathrm{EI}}/p$ denote the *shape* of the curve more than the magnitude of their exploration preferences. Instead, the values of τ^- and τ^+ primarily explain a subject's exploration vs. exploitation preferences in the augmented formulation.

The estimated values of τ^- and τ^+ in Table 2 show that while many subjects exhibit strong exploratory preferences, there is not a universal pattern exhibited across all of the subjects. Many subjects appear highly exploratory regardless of the change in score on move 1 (subjects 6, 7, 10, 12, 25, and 27) while even more subjects are always exploitative (subjects 1, 4, 8, 9, 11, 13, 20, 23, and 26). Some subjects had significantly stronger exploratory preferences when their first move improved their score (subjects 14 and 16) while others were more likely to explore when their first move yielded a negative change in their score (subjects 18 and 19).

Another behavioral tendency we can assess is the subjects' preferences toward either turning left or right on move 2 (relative to the direction of their first move). Based on the estimates of w shown in Table 2, we find that most subjects do not exhibit a strong bias for one direction over the other, but some subjects do have a strong preference to turn right after move 1 (subjects 6, 10, 21, 28) and one subject has a strong preference to turn left (subject 25).

While the focus of our paper is on making inference on human acquisition strategies when facing an exploration vs exploitation conflict, there are a few relationships between the estimated acquisition functions and subjects' corresponding performance in the task which we will briefly discuss. First, there is a strong relationship between high exploitation preferences and success in the early moves of the task. Letting $\widehat{\overline{\tau}}_j = (\widehat{\tau}_j^- + \widehat{\tau}_j^+)/2$ and $\overline{\Delta r_{2j}}$ denote the average change in score after the second move across all of subject j's rounds of the task, the correlation between $\widehat{\overline{\tau}}$ and $\overline{\Delta r_2}$ is -0.81. This is not surprising—subjects who explore on their second move naturally get a lower average score on this move. However, this exploratory sacrifice is balanced by the ability to better exploit the resulting information on the remaining moves in the round, as reflected in the correlation($\widehat{\overline{\tau}}$, $\overline{\Delta r_3}$) = 0.49. Though the relationship is not equally strong, high explorers perform better on average on moves 3 and up.

Ultimately, we care most about the total score improvement on all informed moves: $\sum_{i=2}^{10} \Delta r_i$. We do not find a significant relationship between the estimated acquisition function parameters and this measure. However, a strong relationship exists between $\sum_{i=2}^{10} \Delta r_i$ and $\sum_{i=4}^{10} |\Delta \theta_i|$, where $\Delta \theta_i = \theta_i - \theta_{i-1}$. This measures how much "zig-zag"

---

[8]The $p = 0.5$ acquisition curve is equivalent to that of $\xi_{\mathrm{PI}} = 0$, hence the posterior mass is equally distributed between these two models. We chose to report the UCB parameter since this family yielded the best fit for these subjects prior to augmenting the acquisition function.

a subject exhibits after having the necessary information to estimate a planar gradient over the click-region. As explained in Appendix A of the supplemental material, given a linear model for the surrogate, the Bayesian optimization algorithm will no longer yield drastic changes in direction after the 3rd move. By contrast, most subjects showed significantly positive values of $\sum_{i=4}^{10} |\Delta\theta_i|$. This could indicate preferences for continued exploration, but it could also result from other unintentional sources of sampling variation (e.g. sloppiness due to the speed with which subjects performed the task). Not surprisingly, the degree of "zig-zag" is related to the aggregate performance measure: correlation($\sum_{i=4}^{10} |\Delta\theta_i|, \sum_{i=2}^{10} \Delta r_i$) = -0.47. Subjects who kept a more constant direction after their 3rd move performed better in the task.

6 Conclusion

This work introduces a probabilistic framework for inverse Bayesian optimization and shows how it can be implemented through a lab-based sequential optimization experiment. Our methodology involves defining a likelihood on a subject's observed behavior from a search task, where the likelihood is parameterized by a Bayesian optimization subroutine governed by an acquisition function with parameters of its own. This novel structure enables us to make inference on the subject's acquisition function while allowing their behavior to deviate around the solution to the underlying Bayesian optimization subroutine. This allows us to simultaneously make probabilistic inference about the subject's acquisition function in addition to providing a predictive model of their behavior. The combination of an optimization subroutine within a likelihood-based predictive model distinguishes our approach from other inverse optimization methods.

There are numerous directions that could be explored in future work. In this paper, we base our inference on the 2nd acquisition conditional on the first, but in theory all of a subject's decisions (acquisitions) in a given sample path could be leveraged when making this inference. How to incorporate all of this information into the inference on the acquisition function is a promising area of further study. In addition, our work focuses on a highly customized Bayesian optimization framework. A more general treatment of the inverse Bayesian optimization problem (including GP surrogates, higher dimensionality, a richer class of acquisition functions, etc.) represents a rich and challenging area of future work.

Finally, our study supports findings in human cognition. Specifically, we find that subjects exhibit a wide array of acquisition preferences, but that nearly all of them exhibit exploration tendencies beyond the ability of standard acquisition functions to capture. As in Wu et al. (2018), additional exploratory modifications must be included in the acquisition model in order to accurately represent the observed human behavior. Consistent with Borji and Itti (2013), Wu et al. (2018), and Candelieri et al. (2020), we find that UCB acquisition functions best represent the search strategies of the majority of subjects in our study.

Supplementary Material

Supplementary material of "Inverse Bayesian Optimization: Learning Human Acquisition Functions in an Exploration vs Exploitation Search Task"
(DOI: 10.1214/21-BA1303SUPP; .pdf). The supplemental material contains five appendices. Appendix A details why moves 3 and up provide comparatively little information (relative to move 2) about the subjects' acquisition preferences. Appendix B contains the prior specification for the model parameters in (4.4)-(4.8). Appendix C provides a simulation study which demonstrates that our method can recover a latent acquisition function used to generate θ_2 (move 2 angles) conditional on randomly generated values of Δr_1 (changes in score from move 0 to move 1). It also explores how our method performs when the true generating acquisition function class is omitted from the inference acquisition function class. Appendix D is a link to a public github repository which contains all of our data and R code to reproduce our results. Appendix E contains two additional figures.

References

Ahuja, R. K. and Orlin, J. B. (2001). "Inverse optimization." *Operations Research*, 49(5): 771–783. MR1860425. doi: https://doi.org/10.1287/opre.49.5.771.10607. 3

Aswani, A., Shen, Z.-J., and Siddiq, A. (2018). "Inverse optimization with noisy data." *Operations Research*, 66(3): 870–892. MR3826860. doi: https://doi.org/10.1287/opre.2017.1705. 2

Borji, A. and Itti, L. (2013). "Bayesian optimization explains human active search." In *Advances in neural information processing systems*, 55–63. 3, 6, 21

Brochu, E., Cora, V. M., and De Freitas, N. (2010). "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning." *arXiv preprint arXiv:1012.2599*. 9

Candelieri, A., Perego, R., Giordani, I., Ponti, A., and Archetti, F. (2020). "Modelling Human Active Search in Optimizing Black-box Functions." *arXiv preprint arXiv:2003.04275*. 3, 21

Chan, T. C., Lee, T., and Terekhov, D. (2019). "Inverse optimization: Closed-form solutions, geometry, and goodness of fit." *Management Science*, 65(3): 1115–1135. 2

Cox, D. D. and John, S. (1992). "A statistical method for global optimization." In *[Proceedings] 1992 IEEE International Conference on Systems, Man, and Cybernetics*, 1241–1246. IEEE. MR1033666. doi: https://doi.org/10.1109/21.44040. 9

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1): 1–22. MR0501537. 15

Esfahani, P. M., Shafieezadeh-Abadeh, S., Hanasusanto, G. A., and Kuhn, D. (2018). "Data-driven inverse optimization with imperfect information." *Mathemat-

ical Programming, 167(1): 191–234. MR3748350. doi: https://doi.org/10.1007/s10107-017-1216-6. 2

Gershman, S. J. (2019). "Uncertainty and exploration." *Decision*, 6(3): 277. MR3866028. doi: https://doi.org/10.1162/neco_a_01023. 3

Ghobadi, K., Lee, T., Mahmoudzadeh, H., and Terekhov, D. (2018). "Robust inverse optimization." *Operations Research Letters*, 46(3): 339–344. MR3801810. doi: https://doi.org/10.1016/j.orl.2018.03.007. 2

Gramacy, R. B. (2020). *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. CRC Press. MR4283556. doi: https://doi.org/10.1201/9780367815493. 6

Jones, D. R., Schonlau, M., and Welch, W. J. (1998). "Efficient global optimization of expensive black-box functions." *Journal of Global optimization*, 13(4): 455–492. MR1673460. doi: https://doi.org/10.1023/A:1008306431147. 2, 9

Kushner, H. J. (1964). "A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise." *Journal of Basic Engineering*, 97–106. 9

Li, J. (2020). "Inverse Optimization of Convex Risk Functions." *Available at SSRN 3697392*. 3

Lizotte, D. J. (2008). *Practical Bayesian optimization*. University of Alberta. 9

Mockus, J., Tiesis, V., and Zilinskas, A. (1978). "The Application of Bayesian Methods for Seeking the Extremum." *Towards Global Optimization*, 2(117-129): 2. MR0471305. 9

Ng, A. Y., Russell, S. J., et al. (2000). "Algorithms for inverse reinforcement learning." In *Icml*, volume 1, 2. 3

Plonsky, O., Apel, R., Ert, E., Tennenholtz, M., Bourgin, D., Peterson, J. C., Reichman, D., Griffiths, T. L., Russell, S. J., Carter, E. C., et al. (2019). "Predicting human decisions with behavioral theories and machine learning." *arXiv preprint arXiv:1904.06866*. 3

Sandholtz, N., Miyamoto, Y., Bornn, L., and Smith, M. A. (2022). "Supplementary material of "Inverse Bayesian Optimization: Learning Human Acquisition Functions in an Exploration vs Exploitation Search Task"." *Bayesian Analysis*. doi: https://doi.org/10.1214/21-BA1303SUPP. 6

Schulz, E., Tenenbaum, J. B., Reshef, D. N., Speekenbrink, M., and Gershman, S. (2015). "Assessing the Perceived Predictability of Functions." In *CogSci*. Citeseer. 3

Shahmoradi, Z. and Lee, T. (2019). "Quantile Inverse Optimization: Improving Stability in Inverse Linear Programming." *arXiv preprint arXiv:1908.02376*. 2

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. 2, 9

Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. (2009). "Gaussian process optimization in the bandit setting: No regret and experimental design." *arXiv preprint arXiv:0912.3995*. MR2952544. doi: https://doi.org/10.1109/TIT.2011.2182033. 9

Swartz, R. J., Cox, D. D., Cantor, S. B., Davies, K., and Follen, M. (2006). "Inverse decision theory: characterizing losses for a decision rule with applications in cervical cancer screening." *Journal of the American Statistical Association*, 101(473): 1–8. MR2252428. doi: https://doi.org/10.1198/016214505000000998. 3

Taboga, M. (2021). *Lectures on probability theory and mathematical statistics*. Kindle Direct Publishing. 8

Tversky, A. and Kahneman, D. (1979). "Prospect theory: An analysis of decision under risk." *Econometrica*, 47(2): 263–291. MR3618580. doi: https://doi.org/10.2307/1914185. 17

Wilson, A. G., Dann, C., Lucas, C., and Xing, E. P. (2015). "The human kernel." In *Advances in neural information processing systems*, 2854–2862. 3

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., and Meder, B. (2018). "Generalization guides human exploration in vast decision spaces." *Nature human behaviour*, 2(12): 915–924. 3, 21

Acknowledgments