Inverse Bayesian Optimization: Learning Human Search Strategies in a Sequential Optimization Task

Nathan Sandholtz∗ Yohsuke Miyamoto† Luke Bornn‡ and Maurice Smith§

Abstract. Bayesian optimization is a popular algorithm for sequential optimization of a latent objective function when sampling from the objective is costly. The search path of the algorithm is governed by the acquisition function, which defines the agent’s search strategy. Conceptually, the acquisition function characterizes how the optimizer balances exploration and exploitation when searching for the optimum of the latent objective. In this paper, we explore the inverse problem of Bayesian optimization; we seek to estimate the agent’s latent acquisition function based on observed search paths. We introduce a probabilistic solution framework for the inverse problem which provides a principled framework to quantify both the variability with which the agent performs the optimization task as well as the uncertainty around their estimated acquisition function.

We illustrate our methods by analyzing human behavior from an experiment which was designed to force subjects to balance exploration and exploitation in search of an invisible target location. We find that while most subjects demonstrate clear trends in their search behavior, there is significant variation around these trends from round to round. A wide range of search strategies are exhibited across the subjects in our study, but upper confidence bound acquisition functions offer the best fit for the majority of subjects. Finally, some subjects do not map well to any of the acquisition functions we initially consider; these 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 exploration vs exploitation tradeoff is a dilemma faced in optimization tasks when the objective function is only partially known: as long as portions of the decision space remain unexplored, agents must choose between maximizing immediate reward (exploitation), or minimizing objective uncertainty (exploration) thereby enabling potentially greater gains in future exploitation. For decades, psychologists and social scientists...
have studied how humans balance exploration and exploitation when making decisions under uncertainty (March, 1991; Cohen et al., 2007). Machine learning (ML) researchers have studied this topic as well; 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 the human player to make choices along one continuous dimension. They find that Bayesian optimization algorithms best approximate human behavior in the experiments they conducted. 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). Borji and Itti (2013), Wu et al. (2018), and Candelieri et al. (2020) can be viewed as studies of the inverse problem of Bayesian optimization; given optimization sequences exhibited by an agent, they estimate the latent strategy or algorithm that generated the sequence.

Our paper adds to this body of work by introducing a probabilistic framework for inverse Bayesian optimization. The novelty of our method involves defining a likelihood on an agent’s observed search paths parameterized by an optimization subroutine with latent parameters of its own. This allows us to make probabilistic inferences about the agent’s latent acquisition function (i.e. exploration/exploitation search strategy), in addition to providing a predictive model of their behavior. By using a probabilistic model of this form, both the variability with which the agent optimizes as well as the uncertainty around their estimated exploration/exploitation preferences can be quantified.

We illustrate our methods by analyzing human decision-making behavior from an experimental sequential optimization task we designed to present subjects with a conflict between exploration and exploitation. We find that a wide range of search strategies are exhibited across the subjects in our study, but upper confidence bound (UCB) acquisition 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 an augmentation to the acquisition functions to construct a superior model of human optimization behavior in this task.

1.1 Related work

The problem we study in this paper is a type of supervised learning problem: given an agent’s observed search paths from a sequential optimization task, we want to create a model that accurately predicts how they will behave on future rounds of the task. However, unlike most supervised learning problems, we explicitly assume that each sample in the observed paths is a solution to a latent optimization problem faced by the agent. Not only do we want to incorporate this optimization subroutine into our model,
but inferring the criteria governing the agent’s search strategy is actually our primary interest. We term this type of problem—where decisions are assumed to be optimal and the goal is to infer the latent criteria governing the decision maker’s optimization behavior—an inverse decision problem.

Inverse decision problems have received attention in statistics and related fields: namely, inverse decision theory, inverse reinforcement learning, and inverse optimization. Inverse decision theory seeks to characterize the loss function under which a given decision rule is optimal (Swartz et al., 2006). Similarly, inverse reinforcement learning is the process of inferring the reward function of a Markov decision process based on observed behavior that presumably optimizes it (Ng et al., 2000).

Inverse optimization is a methodology to estimate parameters of a constrained optimization model (typically either the cost function or the constraints) given a set of observed decisions, such that the given decisions are optimal or near-optimal (Ahuja and Orlin, 2001). Li (2020) is particularly relevant to our project, which uses inverse optimization to learn a convex function representative of an agent’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. Additionally, the sequential nature of the learning problem we study relates to online learning in inverse optimization (Bärmann et al., 2018; Dong et al., 2018; Dong and Zeng, 2020) and inverse Markov decision processes (Erkin et al., 2010).

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 decision problems become inference problems, well-suited to be approached with probabilistic models. 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 agent performs the optimization as well as the uncertainty around the parameter estimates governing the latent optimization subroutine (e.g. the acquisition function).

Finally, our work is closely related to Borji and Itti (2013), Wu et al. (2018), and Candelieri et al. (2020), which utilize lab experiments to estimate human optimization strategies from a set of ML algorithms/models. We likewise create a lab experiment to test how humans perform an optimization task; however, our focus is on formalizing the statistical inference component rather than discovering novel relationships between human cognition and ML algorithms.

1.2 Outline

The rest of this paper is outlined as follows. In Section 2, we describe the search task we designed and how it forces subjects to make exploration vs exploitation trade-offs
Inverse Bayesian Optimization

while optimizing. In Section 3, we frame the search task as a Bayesian optimization procedure. We propose a linear model as the surrogate of the objective function and show how Bayesian optimization proceeds under this model. In Section 4, we introduce the inverse problem and illustrate our proposed solution framework in the context of the search task. In Section 5, we augment the acquisition functions to enable a more accurate fit to the human tendencies we observe. After validating this model, we explore relationships between the estimated acquisition functions and subject’s corresponding performance in the task. We summarize our contributions and findings in Section 6.

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 a latent bliss point. 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. 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. On each round, the reward scale was randomized and was opaque to the subject.1

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” despite the fact that subjects had no choice as to the location of this initial move. Each subsequent move (i.e. moves 1 and up) 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. 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. After receiving feedback from their first move the subject acquires information about the gradient of the objective function but only along a single direction.2 This poses a conflict be-

1The reward scale was randomized such that the maximum score at the hotspot varied uniformly between 0 and 1000.
2If the reward gradient (i.e. a location’s sensitivity to hotspot distance) 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 gradient 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 2.
Figure 1: An example round of the hotspot task. The red target shows the invisible hotspot location and the dots track the subject’s search path. The score at the hotspot is the score the subject would be given if they sampled the hotspot. Subjects always begin the task in the center of the click region (move 0, shown in blue).

between exploitation and exploration 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 Figure 2.

In the next section we show how this task can be framed as a Bayesian optimization procedure. Ultimately, our goal is to estimate subjects’ latent acquisition preferences (i.e. how they balance exploration and exploitation) using their observed behavior in this task.
Figure 2: *Left:* Initial state for the example round shown in Figure 1. Moving in any direction represents exploration, as shown by the dashed green lines. *Right:* The subject moves to the upper edge of the move 1 click-boundary, 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).

3 Bayesian Optimization Framework

Bayesian optimization is an algorithm for sequential optimization of a latent objective function with expensive trials (Jones et al., 1998; Shahriari et al., 2015). The core idea of the method is to characterize the uncertainty about the latent objective with a statistical model, termed a *surrogate*, then to strategically synthesize the model uncertainty in order to determine promising new locations to sample. The process of synthesizing the uncertainty about the objective function is governed by the *acquisition function*. Conceptually, the acquisition function defines how the optimizer balances exploration and exploitation when searching for the optimum of the latent objective. In this section we detail the Bayesian optimization procedure in context of the hotspot search task.

3.1 Choosing a surrogate

A ‘surrogate’ is simply a term for a statistical model, but it implies an emphasis on pragmatism and prediction rather than interpretability and identification (Gramacy, 2019). 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.
In most Bayesian optimization applications, the surrogate is a Gaussian process. In our case, subjects know there is an optimum somewhere on the circular task-region and that the rewards (i.e. objective) decrease uniformly and monotonically from this hotspot with distance. Specifically, for a given round of the task the reward received on move $t$ is a deterministic function of the distance from the click location to the hotspot:

$$ r_t = f(d_t) $$
$$ = r_0 + k(d_0 - d_t), $$

where $r_0$ denotes the initial score, $d_t$ denotes the euclidean distance (in pixels) of the newly sampled location to the hotspot, $d_0$ denotes the distance of the starting location to the hotspot, and $k$ denotes the reward gradient (i.e. the reward’s sensitivity to hotspot distance). As alluded to in Section 2, $k$ is randomly generated for each new round of the task according to a Uniform[0, 5/3] points/pixel distribution. This prevents subjects from gaining information about the objective function along the orthogonal direction of the first move, thus ensuring that the second move presents a tradeoff between exploration and exploitation.

Figure 3 shows the objective function $f$ for the example round shown in Figure 1. As illustrated in left panel, 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 panel on the right.

![Figure 3](image-url)

**Figure 3:** *Left:* The global objective function for the example round shown in Figure 1. *Right:* The objective function within the click-region of the first move (i.e. zoomed in to the green circle in the left plot).

Applying the Bayesian optimization paradigm, we assume subjects intuitively construct a surrogate of the objective function as they perform this task. Since the variables that actually govern the reward on each move (i.e. $k$ and $d_t$) are not known to the subject, we model the surrogate as a function of a move’s $x, y$ coordinates:

$$ r_t = \hat{f}_t(x_t, y_t), $$
where \( \hat{f}_t \) denotes the surrogate at move \( t \). Given that the objective function is approximately linear 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:

\[
\hat{f}_t(m_t) = r_0 + m_t^T \beta + \epsilon_t \tag{3.2}
\]

\( \epsilon_t \sim N(0, \sigma_s^2) \),

where \( m_t^T = (x_t, y_t) \), \( \beta = \begin{pmatrix} \beta_x \\ \beta_y \end{pmatrix} \) represents the reward gradient with respect to the Cartesian plane, and \( \epsilon_t \) represents deviations from the surrogate to the true objective. We fix \( \sigma_s \) 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 \( m_t \).

4 In order to have an analytic update for the surrogate after each move, we use a Gaussian distribution to model the error term \( \epsilon_t \). 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 (3.2) is effectively a tangent plane of the underlying conical objective, hence the \( \epsilon_t \) are almost exclusively non-negative. The mean of \( \epsilon_t \) across all rounds of the experiment is 0.006.

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. We decided to select the number of moves and the size of the click-region on each move such that it would be impossible to cover the distance to the hotspot in approximately 90% of all rounds, and 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, the surrogate is updated to reflect the new information 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 \( \beta \):

\[
\beta \sim N_2(\mu_0 = (0, 0), \Sigma_0 = \sigma_\beta^2 I), \tag{3.3}
\]

where \( \sigma_\beta^2 \) is a hyperparameter that we select to give a weakly-informative prior on the mean. Let the observed search path at move \( t \) be denoted as \( D_t = \{M_t, r_t\} \), where \( M_t \) is a \((t+1) \times 2\) matrix with row \( t+1 \) equal to \( m_t^T \), and \( r_t \) is a \((t+1)\)-length column vector of observed rewards on moves 0 through \( t \). Given \( D_t \), (3.3) can be updated yielding a posterior distribution on \( \beta \). The prior is conjugate and the posterior can be computed
N. Sandholtz et al.

analytically (Box and Tiao, 2011):

\[
\beta | D_t \sim N_2(\mu, \Sigma) \quad (3.4)
\]

\[
\Sigma = \left( \frac{1}{\sigma^2} M_t^T M_t + \Sigma_0^{-1} \right)^{-1}, \quad (3.5)
\]

\[
\mu = \Sigma \left( \frac{1}{\sigma^2} M_t^T (r_t - r_0) + \Sigma_0^{-1} \mu_0 \right). \quad (3.6)
\]

The posterior predictive distribution of the surrogate is given by:

\[
\hat{f}_t(m_{t+1} | D_t) \sim N(m_{t+1} \mu + r_0, m_{t+1} \Sigma m_{t+1} + \sigma_s^2), \quad (3.7)
\]

where \( m_{t+1} \) is a potential location for the next move and all other terms are as defined previously. Figure 13 in the appendix shows an example of the surrogate predictive distribution on move 2.

### 3.3 Choosing an acquisition function

After modeling the objective function via the surrogate, we utilize the model to obtain a promising new location to sample. This process is operationalized through the acquisition function. The acquisition function \( u \) is a function of a proposed location \( m_{t+1} \) and the surrogate \( \hat{f}_t \) (which itself is a function of \( D_t \)). 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 \( M_{t+1} \), the set of potential locations for their next move, thus obtaining a new location to sample:

\[
m_{t+1} = \arg \max_{m \in M_{t+1}} u(m | \hat{f}_t, D_t) \quad (3.8)
\]

After the objective function is sampled at the resulting location, the surrogate is updated and the process is repeated until some stopping criterion is reached, which in our case is when the final move for a given round is reached. Algorithm 1 details these steps in pseudo-code.

---

**Algorithm 1:** Bayesian optimization for hotspot search task

**Input:** \( D_0, \hat{f}_0, u \)

**Output:** \( m_T^+ = \arg \max_{m \in M_T} f(m) \)

1. for \( t = 0, \ldots, T - 1 \) do
   2. determine new location: \( m_{t+1} = \arg \max_{m \in M_{t+1}} u(m | \hat{f}_t, D_t) \);
   3. sample objective: \( r_{t+1} = f(m_{t+1}) \);
   4. augment data: \( D_{t+1} = \{D_t, (m_{t+1}, r_{t+1})\} \);
   5. update surrogate: \( \hat{f}_{t+1} = \hat{f}_t | D_{t+1} \)
2. end

7. return \( m_T^+ = \arg \max_{m \in M_T} f(m) \)
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

\[
\begin{align*}
    u_{PI}(\mathbf{m} | \hat{f}_t, \xi_{PI}, D_t) &= \mathbb{P}(\hat{f}_t(\mathbf{m} | D_t) \geq f(\mathbf{m}^+_t) + \xi_{PI}) \quad (3.9) \\
    u_{EI}(\mathbf{m} | \hat{f}_t, \xi_{EI}, D_t) &= \mathbb{E}[\max(0, \hat{f}_t(\mathbf{m} | D_t) - (f(\mathbf{m}^+_t) + \xi_{EI}))] \quad (3.10) \\
    u_{UCB}(\mathbf{m} | \hat{f}_t, p, D_t) &= \inf \{ \mathbf{m} : p \leq \hat{F}_t(\mathbf{m} | D_t) \}, \quad (3.11)
\end{align*}
\]

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

Each of the acquisition functions in (3.9)-(3.11) 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 \( \xi_{PI} \). EI (Mockus et al., 1978) considers the magnitude of improvement at a particular location, where \( \xi_{EI} \) controls a tradeoff between exploration (high \( \xi_{EI} \)) and exploitation (low \( \xi_{EI} \)) (Lizotte, 2008). Intuitively, one can think of EI as related to PI in that the function weighs a location’s probability of improvement by the amount of improvement associated with the location. Thus compared to the PI acquisition function, EI is more likely to select a location associated with low probability of improvement if the low probability is offset by the location’s potential for high improvement. The improvement term in EI can be exponentiated by an additional parameter to put even greater preferential weight on exploration (Schonlau et al., 1998). 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.\(^5\)

Figure 4 shows the acquisition surfaces for the PI, EI, and UCB functions defined in (3.9)-(3.11) 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.

\(^5\)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.
Notice that each method prescribes different move 2 locations (i.e. green stars): 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, in addition to the exploration parameter values ($\xi_{PI}$, $\xi_{EI}$, and $p$), the optimal location to sample can vary substantially. Also note that each surface is symmetric across the exploitation axis. 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 response surface, the optimal location to sample on move 2 (i.e. the arg max as determined by (3.8)) 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. 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 $\theta_2$, the angle of move 2 relative to the direction of move 1:

$$\arg\max_{m\in M_2} u(m|f_1, D_1) \rightarrow \arg\max_{\theta \in [-\pi, \pi]} u(\theta|f_1, D_1).$$

By reducing the dimensionality of the optimization we can illustrate acquisition values over the entire range of $\Delta r_1 = r_1 - r_0$, 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. Note that the move 2 locations are assumed to be made on the click-region boundary.

---

$^6$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.
Inverse Bayesian Optimization

Figure 5: Click-region boundary acquisition values over the range of possible $\Delta r_1$ values for three sample acquisition functions. The horizontal axes denote $\Delta r_1$, the change in reward from $m_0$ to $m_1$. The vertical axes show $\theta_2$, the angle of the second move relative to the first move. Move 2 is assumed to be made on the click-region boundary. 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 $\Delta r_1$. 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.

In each plot, the green curves denote the angle that yields the maximum of the acquisition values as a function of $\Delta r_1$. Going forward, we will refer to an acquisition curve as the curve defined by the optimal $\theta_2$ as a function of $\Delta r_1$ for a given acquisition function. As evident from the figure, depending on the acquisition function, the prescribed move 2 strategies (i.e. acquisition curves) can be quite different.

Our goal is not to find an optimal strategy, or acquisition function, for this task. Rather, assuming that subjects use Bayesian optimization intuitively as an optimization framework (Borji and Itti, 2013; Wu et al., 2018), we want to estimate the latent acquisition functions employed by the subjects in the experiment. Section 4 formally introduces this problem.

3.4 Moves 3 and up

Before moving to the next section, we pause to consider moves 3 and up under the methodology described above. Since the surrogate we use is a linear model in two dimensions, once a subject has made their second move they have a near-complete characterization of the objective function. After receiving feedback from move 2, subjects can approximate the objective function with high fidelity using the plane defined by moves 0, 1, and 2. In theory, as long as their second move is not along the same direction as move 1, they can solve for the direction of the hotspot.

This phenomenon is illustrated in Figure 6. These plots are the same as those in Figure 4, only here we plot the move 3 acquisition surfaces instead of the move 2 surfaces. In each case, move 2 was selected at the optimum of the previous iteration of the process.
(i.e. the green stars in Figure 4). The green stars denote the arg max of each surface, while the pink dots show the true direction of the hotspot. In every case, the green stars are either on top of or very close to the pink dots.

Figure 6: Move 3 acquisition surfaces for three sample acquisition functions. In each plot, \( r_0 = 93 \) and \( r_1 = 98 \), but the \( r_2 \) values are different because each \( \mathbf{m}_2 \) was selected at the optimum of the previous iteration of the Bayesian optimization process (i.e. the green stars in Figure 4). The arg max of each surface is denoted by a green star while the pink dots show the true direction of the hotspot.

While it is possible to learn the direction of the hotspot after move 2, whether subjects actually do learn the optimal direction is a separate question. Figure 7 shows directional histograms of subject behavior on moves 1, 2, and 3 relative to the direction of the hotspot (aggregated over all subjects).

Figure 7: Histograms of subjects’ move angles relative to hotspot direction on moves 1, 2, and 3 for all rounds.

On move 1, the angles are uniform on the circle because subjects have no directional information on this move. The move 2 angles are directionally informed, but subjects must move off the axis of the informed direction in order to gain information about the orthogonal gradient. This can be seen in the move 2 histogram by the modes over \( \pm \pi/4 \).
The move 3 histogram sheds light on how well subjects synthesize the information from moves 1 and 2 in order to learn the direction of the hotspot. As shown by the variance around 0 in the move 3 histogram, subjects do not always synthesize this information perfectly, but the mode angle is toward the direction of the hotspot. On moves 4 through 10 (not shown), the mode remains centered around the direction of the hotspot while the variance around the mode decreases slightly on each subsequent move. Since subject behavior essentially amounts to a noisy walk around the optimal direction after move 2, moves 3 and up provide comparatively little information about the subjects’ acquisition preferences. For this reason we do not analyze these moves in our paper.

4 Inverse Bayesian Optimization

Each sequence of clicks in a given round of the task can be viewed as a subject’s realized optimization routine generated according to their latent acquisition preferences. We leverage the Bayesian optimization framework as a model of subject behavior in this task, which provides a methodology to quantify and communicate how a subject balances exploration and exploitation. We assume subjects’ optimization behavior is characterized by an average strategy but that their behavior deviates around this strategy from round to round. This variability could arise from a number of factors. For one, the modeling described in Section 3 happens subconsciously; subjects must update their surrogate model intuitively, which introduces human error. Noise could also be due to variation in the subject’s ability to click exactly where they intend to, changes to their sampling strategy, or sloppiness from performing the task rapidly.

Our goal is to estimate each subject’s mean acquisition function in a probabilistic framework. As such, the problem becomes one of inference rather than optimization. The optimization routine has already occurred—we want to infer how each subject optimized. We term this problem ‘inverse Bayesian optimization’ (IBO). In our framing, IBO can be understood as finding the Bayesian optimization trajectories (see Figure 8b) that best describe humans’ observed optimization paths (Figure 8a).

The scatter plots in (a) show three subjects’ click behavior on move 2 conditional on feedback from move 1, where each dot represents a separate round of the task. The plots in (b) show various acquisition curves from the three acquisition families described in Section 2. IBO can be viewed as finding the curves from (b) that best fit the subjects’ data in (a).

In this section we propose a framework for solving the inverse Bayesian optimization problem in the context of the hotspot search task. This can be approached from two different perspectives—in Section 4.1, we consider the problem when Bayesian optimization is known to be explicitly used in the forward optimization. In Section 4.2, we assume the forward optimization routine is unknown but that it resembles a Bayesian optimization procedure. This second case is more representative of our hotspot experiments.

Figure 12 in Appendix B shows these data in Cartesian coordinates.
Figure 8: (a) Move 2 behavior for subjects 21 (left), 24 (middle), and 28 (right). In each scatterplot, dots represent the $(\Delta r_1, \theta_2)$ pairs for the subject’s rounds of the task. (b) Candidate acquisition curves under various exploration parameter values for the corresponding family—PI (left), EI (middle), and UCB (right).

### 4.1 Known forward optimization method

The first setting we consider is when Bayesian optimization is known to be explicitly used in the forward optimization but that the acquisition function $u^*$ that was employed is latent. This would be the case if we had given subjects access to statistical software and tasked them to code a Bayesian optimization routine. Under this setting, the inverse problem (i.e. finding the function $u^*$ used in the forward optimization) can be solved exactly, but a few conditions must exist in order to guarantee this outcome.

First, we must know the surrogate function $f_0$ used to initialize the optimization and the mode of inference for updating the surrogate. Next, the forward optimization must be carried out perfectly—the objective function must be sampled precisely at the locations given by the latent acquisition function $u^*$. Finally, we must know a priori a set of candidate acquisition functions $\mathcal{U}$ from which the true acquisition function $u^*$
Inverse Bayesian Optimization

is a member. Under these conditions, we can solve the inverse problem by sequentially updating the surrogate at each step $t$, then pruning the set of candidate acquisition functions $U$ to those that yield $m_{t+1}$ until we reach the final step $T$ of the search path. The result is $U_{T-1}$, which ideally is a singleton set containing $u^*$, the true acquisition function employed by the agent. Algorithm 2 details this process in pseudo code.

Algorithm 2: Inverse Bayesian optimization under perfect acquisition

| Input: $\mathcal{D}_T$, $\hat{f}_0$, $U$ |
|------------------------------------------|
| Output: $U_{T-1}$                        |

1. $U_0 \leftarrow U$;
2. for $t = 1, \ldots, T - 1$ do
   3. $\hat{f}_t = \hat{f}_{t-1}|\mathcal{D}_t$;
   4. $U_t = \{ u \in U_{t-1} \text{ s.t. } m_{t+1} = \arg \max_{m \in \mathcal{M}_t} u(m|\mathcal{D}_t, \hat{f}_t) \}$;
5. end
6. return $U_{T-1}$ where $u^* \in U_{T-1}$

We note that there are cases in which this algorithm could fail despite knowing Bayesian optimization to be utilized in the forward optimization. This could happen if the trajectories were corrupted by measurement error, or if the sampling was carried out imperfectly. These type of situations would be ideally situated to be approached using the methods we propose in the next subsection.

4.2 Unknown forward optimization method

While Algorithm 2 will yield the exact acquisition function employed by an agent (or set of functions all producing the same search path), the required conditions may not hold in practical applications. For many sequential decision problems, the strategy underlying a decision-maker’s optimization procedure is unknown. In these situations, Bayesian optimization can be used as a model for their latent optimization routine, and inverse Bayesian optimization provides a way to characterize their strategy. In the case of our search task experiment, subjects were not asked to define their strategy beforehand, they were simply instructed to get as close to the hotspot as possible. However, doing this implicitly required them to optimize under uncertainty. When people must intuitively balance exploration and exploitation in an optimization task and the preferences governing their decisions are unknown (possibly even to the optimizers themselves), we propose IBO as a methodology to characterize and estimate their latent preferences.

Under scenarios of this nature, we expect that human error will corrupt the optimization, particularly if the objective is defined over a continuous space. We therefore propose a probabilistic framework for IBO under this type of setting as laid out in the following steps:
1. Choose a surrogate $\hat{f}$ to model the optimizer’s beliefs about the objective.

2. Select a set of candidate acquisition functions $\mathcal{U}$ as potential characterizations of the optimizer’s latent acquisition preferences.

3. Assume a likelihood for $(m_{t+1} | D_t, u)$, parameterized such that the mode of the distribution equals the arg max of the latent acquisition function $u$.

4. Select a prior distribution over $\mathcal{U}$.

5. For each step $t \in \{1, \ldots, T - 1\}$, compute the updated surrogate and the posterior probability of each candidate acquisition function $u \in \mathcal{U}$:

$$
\hat{f}_t = \hat{f}_{t-1} | D_t \quad P(u|m_{t+1}, D_t, \hat{f}_t) \propto P(m_{t+1} | \text{arg max } m | D_t, \hat{f}_t) P(u). 
$$

We assume the only available input in this process is the observed data $D_T$. Also, while we employ a Bayesian framework in our approach to IBO, the procedure could be carried out using other estimators (e.g. maximum likelihood). We will now explain how we implement each of these steps in context of the hotspot search task. In order to estimate each subject’s acquisition function, we follow the steps listed above, but we consider only the first and second moves of each round (see Section 3.4).

For steps 1 & 2, we model the subject’s beliefs about the latent objective with the linear model defined in (3.2) and we use the PI, EI, and UCB acquisition families defined in (3.9)-(3.11) as the set of candidate acquisition functions $\mathcal{U}$. Note that our method does not preclude using other acquisition families—any acquisition function can be included in the candidate set $\mathcal{U}$.

For step 3, we select a likelihood for $\theta_2$ conditional on $\Delta r_1$, $\hat{f}_1$, and $u$. Importantly, we parameterize the likelihood such that the mode equals the arg max of the acquisition function (i.e. the mode must be an acquisition curve as defined by (3.8)). By so doing, we implicitly assume that each subject behaves according to a specific strategy on the mode but that their behavior randomly deviates around this strategy according to a probability distribution. Essentially, we require an optimization subproblem to exist inside the likelihood of the observed decisions. This requirement distinguishes our approach from other probabilistic models.

In order to model the symmetric bimodal nature of these data, we assume $\theta_2$ follows a reflected wrapped Cauchy distribution:

$$
g(\theta_2 | \Delta r_1, \hat{f}_1, \gamma) = \frac{1}{2} h(\theta_2 | \Delta r_1, \hat{f}_1, \gamma) + \frac{1}{2} h(-\theta_2 | \Delta r_1, \hat{f}_1, \gamma),
$$

$$
h(\theta_2 | \Delta r_1, \hat{f}_1, \gamma) = \sum_{k=-\infty}^{\infty} \frac{\gamma}{\pi(\gamma^2 + (\theta_2 - \mu + 2\pi k)^2)},
$$

$$
\mu = \begin{cases} 
\text{arg max } u(\theta_2 | \Delta r_1, \hat{f}_1) & \text{if } \theta_2 \in (-\pi, 0], \\
\text{arg max } u(\theta_2 | \Delta r_1, \hat{f}_1) & \text{if } \theta_2 \in (0, \pi].
\end{cases}
$$
Conceptually, (4.2) alters the standard wrapped Cauchy distribution (Mardia and Jupp, 2009) by reflecting the distribution about $\theta_2 = 0$. Since subjects only have information about the objective along a single direction after their first move, the information they gain by moving off this direction is equivalent in expectation regardless of whether they angle left or angle right. This leads to bimodal move 2 angles for most subjects, as exhibited by Subject 24 in Figure 8.

The wrapped Cauchy distribution is defined in (4.3), where $\gamma$ and $\mu$ are scale and location parameters respectively. In the context of the task, $\gamma$ measures a subject’s variation about their acquisition curve, which itself is denoted by $\mu$. We use the wrapped Cauchy distribution because it is more robust to outliers, which are fairly common across many subjects’ move 2 behavior (see Figure14(5,6),(997,991) in Appendix B for illustration). Finally, (4.4) ensures that the mode of the likelihood is given by the arg max of the latent acquisition function $u$. It is defined piecewise in order to properly identify which mode to use when computing the likelihood for any given $\theta_2$.

Together, (4.2)-(4.4) yield a likelihood that adheres to the defining features of our data; it is not only defined on the proper support for $\theta_2$ (i.e. $(-\pi, \pi]$), it is also guaranteed to be symmetric over the axis of exploitation (i.e. $\theta_2 = 0$). We utilize the circular R package (Agostinelli and Lund, 2013) and the wrapped R package (Nadarajah and Zhang, 2017) to compute (4.2)-(4.4).

Next, for step 4 we assume that each acquisition function $u \in U$ has equal prior probability, thus we assume a uniform prior over the parameters indexing the acquisition function space: $\xi_{\text{EI}}, \xi_{\text{PI}}, \rho$. We also put a uniform prior distribution over $\gamma$, the scale parameter governing the variability in each subject’s optimization behavior. In step 5, we update the surrogate for $t = 1$, then compute (4.1) for each $u \in U$ by multiplying its prior probability by the likelihood. This yields a posterior distribution over $U$, which allows us to characterize the uncertainty in our estimates of each subject’s acquisition function. Note that the existence of the arg max in the parameterization of (4.2)-(4.4) creates an additional optimization step when evaluating (4.1). To ease the computational burden posed by this additional optimization step, 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. Figure 9 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 gray-shaded areas show approximate 95% highest posterior density (HPD) prediction regions, while the red-shaded regions denote 95% HPD credible regions on the acquisition function.

See Appendix A for additional details on prior specification.
Figure 9: \((\Delta r_1, \theta_2)\) 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 interval on the acquisition function (red) and 95% HPD prediction interval (gray).

On visual inspection, 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 \(U\) that prescribe such a drastic change in strategy from negative \(\Delta r_1\) to positive \(\Delta r_1\) values while maintaining highly exploratory preferences. In the next section we propose an augmentation to the candidate acquisition functions guided by discrepancies such as these.

## 5 Incorporating Human Tendencies

Many subjects in our experiment tend to exhibit exploration tendencies beyond what the candidate acquisition functions can adequately capture, particularly for highly negative and highly positive values of \(\Delta r_1\). In this section we propose ways to augment the acquisition functions defined in Section 4 to better model the behavior of the subjects.

### 5.1 Exploratory heuristics

Consider Subject 28’s behavior in Figure 8a. The general shape of their data looks like a compressed version of the PI acquisition curve with a low value of \(\xi_{p1}\). As shown by the black curve in the left plot of Figure 8b, the \(\xi_{p1} = 0\) acquisition curve changes abruptly when going from positive to negative values of \(\Delta r_1\). Subject 28 likewise exhibits an abrupt change in strategy when going from positive to negative values of \(\Delta r_1\), but their strategy remains significantly exploratory for both positive and negative \(\Delta r_1\), whereas the strategy implied by \(\xi_{p1} = 0\) is completely exploitative.

The set of acquisition functions as currently defined cannot exhibit this type of shape. This means that there is no parameter setting of \(\xi_{p1}, \xi_{e1},\) or \(p\) which renders Subject 28’s behavior as optimal. In order to allow these types of exploratory tenden-
cies to be rendered optimal, we propose augmenting the acquisition functions by an additional exploration threshold parameter $\tau$. For a given acquisition function $u$, we define augmented acquisition function $\tilde{u}$ as

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

(5.1)

where $\tau \in [0, \pi/2]$. Eq. (5.1) defines the acquisition value to be the minimum possible acquisition value for all move 2 angles that aren’t at least $\tau$-radians exploratory, either in the forward or backward directions. If a given move 2 angle is greater than $\tau$ and less then $\pi - \tau$ (in absolute value), the move is “sufficiently exploratory” and the acquisition value is unchanged. Figure 10 illustrates how this augmentation changes the acquisition surface and the corresponding acquisition curve when $\xi_{PI} = 1$.

The practical effect of (5.1) is it allows 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. Eq. (5.1) allows exploration to trump exploitation, regardless of $\hat{f}$. We observe this type of behavior by many subjects in our study.
5.2 Prospect theory

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 (5.1) to allow different values of $\tau$ depending on whether $\Delta r_1$ is positive or negative:

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

(5.2)

5.3 Validation and fit

We fit the augmented acquisition models in (5.1) and (5.2) to each subject’s data following the same procedure as outlined in Section 4.2, only for these models we estimate the exploratory heuristic parameters as well. As with the other exploration parameters, we set discrete uniform priors on $\tau$, $\tau^+$, and $\tau^-$. Table 1 shows the out-of-sample log-likelihoods (using an 80%/20% train/test split of the data) for each model across the subjects in our study.

Table 1: Out-of-sample log-likelihoods for the model defined in (4.2)-(4.4) fit using $u$, $\tilde{u}$, and $\tilde{u}^\pm$ 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.

| Subject | Acquisition model | Acquisition model |
|---------|-------------------|-------------------|
|         | $u$               | $u$               | $u^\pm$ |
| 1       | -56.54            | -**42.34**        | -42.34 |
| 2       | -52.73            | -44.54            | -38.78 |
| 3       | -56.81            | -50.60            | -49.05 |
| 4       | -72.54            | -59.21            | -50.91 |
| 5       | -44.19            | -**36.72**        | -36.75 |
| 6       | -30.95            | 19.06             | **20.40** |
| 7       | -42.68            | -33.41            | -31.79 |
| 8       | -77.48            | -59.64            | -45.18 |
| 9       | -11.51            | -2.98             | **1.44** |
| 10      | -38.51            | -5.53             | **-0.23** |
| 11      | -93.98            | -**84.90**        | -84.90 |
| 12      | -37.81            | -21.40            | -19.64 |
| 13      | -87.28            | -**87.20**        | -87.20 |
| 14      | -66.48            | -62.23            | -60.97 |
| 15      | -57.14            | -44.61            | **-42.73** |
| 16      | -41.98            | -36.45            | -11.11 |
| 17      | -43.74            | -35.54            | -18.03 |
| 18      | -36.59            | -31.79            | -19.64 |
| 19      | -57.73            | -44.54            | -38.78 |
| 20      | -23.98            | **-13.23**        | -13.23 |
| 21      | 2 -28.34           | -32.20            | -32.20 |
| 22      | -60.11            | -44.76            | **-42.90** |
| 23      | -116.49           | -**93.25**        | -93.59 |
| 24      | -78.89            | -65.93            | -64.37 |
| 25      | -28.16            | -19.67            | -10.52 |
| 26      | -64.04            | -64.91            | -65.31 |
| 27      | -55.19            | -55.39            | -56.58 |
| 28      | -56.16            | -17.10            | **-15.97** |
| Total   | -1591.67          | -1245.57          | **-1093.16** |
The $\tilde{u}^{\pm}$ model provides the best fit on aggregate, though there are some subjects for which the additional flexibility offered by the $\tau$ parameters does not substantially improve the fit. Unless otherwise specified, from this point onward all model references utilize the augmented acquisition function in (5.2).

Table 2 shows each subject’s MAP estimates of the $\tilde{u}^{\pm}$ parameters and $\gamma$ from (4.2)-(4.4). 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 $p = 0.5$. 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 $\tau^-$ and $\tau^+$, subjects can exhibit highly exploratory behavior despite having $\tilde{p} = 0.5$. Under the augmented acquisition function, the parameters $\xi_{u1}/\xi_{u1}/p$ denote the shape of the curve more than their exploration vs. exploitation tendencies. The values of $\tau^-$ and $\tau^+$ primarily explain a subject’s exploration vs. exploitation preferences in the augmented formulation.

Figure 11 shows the data for the same subjects as in Figures 8 & 9 overlaid with the MAP augmented acquisition curves, 95% HPD credible intervals (color), and 95% HPD prediction intervals (gray). As shown by the credible intervals, our estimates of the acquisition functions are quite precise, but the round-to-round variation that subject’s exhibit around their acquisition function is comparatively large. These type of insights are possible due to the probabilistic modeling framework we use. Figure 14 in Appendix B shows the data, MAP augmented acquisition curves, and 95% HPD intervals for all subjects in the study.

Figure 11: 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 augmented acquisition function (color) and 95% HPD prediction interval (gray). Point estimates for the parameters governing each subject’s $\tilde{u}^{\pm}$ are listed in the lower right corner.

$^9$Note that $p = 0.5$ acquisition curve is equivalent to that of $\xi_{p1} = 0$, 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.
Table 2: Each subject’s estimated fit to (4.2)-(4.4) using the augmented acquisition formulation \( \hat{u}^+ \) for the mean. Columns 2-5 show MAP parameter estimates for the acquisition family shape parameter \((\hat{\xi}_{PL}/\hat{\xi}_{EI}/\hat{p})\), as well as estimates for \(\tau^-, \tau^+\), and \(\gamma\). Columns 6-8 show corresponding measures of subject performance in the search task: \(\Delta r_i\) denotes the average change in score after a subject’s \(i\)th move.

| Subject | Fitted Model | Task Performance |
|---------|--------------|------------------|
|         | \(\hat{p} = 0.50\) | \(\hat{\xi}_{PL}/\hat{\xi}_{EI}/\hat{p}\) | \(\tau^-\) | \(\tau^+\) | \(\gamma\) | \(\Delta r_2\) | \(\Delta r_3\) | \(\sum_{i=2}^{10} \Delta r_i\) |
| 1       | \(0.50\) \(0.16\) \(0.16\) \(0.17\) | 7.84 | 9.14 | 53.33 |
| 2       | \(0.50\) \(0.16\) \(0.31\) \(0.12\) | 8.78 | 9.66 | 54.07 |
| 3       | \(0.50\) \(0.16\) \(0.42\) \(0.14\) | 7.09 | 9.74 | 55.00 |
| 4       | \(0.50\) \(0.21\) \(0.05\) \(0.09\) | 8.74 | 9.99 | 57.37 |
| 5       | \(0.50\) \(0.16\) \(0.05\) \(0.17\) | 7.71 | 9.48 | 57.30 |
| 6       | \(0.52\) \(1.52\) \(1.57\) \(0.06\) | 0.12 | 13.79 | 61.17 |
| 7       | \(0.50\) \(0.89\) \(0.79\) \(0.17\) | 5.98 | 13.10 | 62.64 |
| 8       | \(0.50\) \(0.26\) \(0.05\) \(0.09\) | 7.89 | 8.86 | 47.85 |
| 9       | \(0.50\) \(0.10\) \(0.16\) \(0.09\) | 9.47 | 9.58 | 53.78 |
| 10      | \(0.50\) \(0.79\) \(0.94\) \(0.09\) | 6.25 | 11.39 | 51.06 |
| 11      | \(0.50\) \(0.21\) \(0.21\) \(0.20\) | 6.43 | 8.84 | 50.15 |
| 12      | \(0.50\) \(0.73\) \(0.63\) \(0.12\) | 7.22 | 13.07 | 66.12 |
| 13      | \(0.53\) \(0.10\) \(0.10\) \(0.28\) | 7.00 | 11.61 | 56.77 |
| 14      | \(0.50\) \(0.21\) \(0.79\) \(0.22\) | 5.54 | 10.11 | 49.29 |
| 15      | \(0.50\) \(0.37\) \(0.58\) \(0.28\) | 7.07 | 8.11 | 46.09 |
| 16      | \(0.50\) \(0.10\) \(0.63\) \(0.12\) | 9.98 | 13.40 | 72.27 |
| 17      | \(0.50\) \(0.10\) \(0.42\) \(0.12\) | 8.95 | 9.38 | 59.03 |
| 18      | \(0.50\) \(1.57\) \(0.52\) \(0.25\) | 3.63 | 10.22 | 50.49 |
| 19      | \(0.50\) \(1.31\) \(0.73\) \(0.22\) | 4.09 | 9.53 | 47.42 |
| 20      | \(0.50\) \(0.16\) \(0.16\) \(0.12\) | 9.13 | 10.53 | 56.24 |
| 21      | \(\hat{\xi}_{PL} = 10\) \(0.68\) \(0.42\) \(0.14\) | 6.19 | 15.16 | 66.54 |
| 22      | \(\hat{\xi}_{PL} = 0.50\) \(0.21\) \(0.31\) \(0.17\) | 8.37 | 10.74 | 61.12 |
| 23      | \(\hat{\xi}_{PL} = 0.50\) \(0.31\) \(0.26\) \(0.25\) | 7.01 | 9.49 | 48.60 |
| 24      | \(\hat{\xi}_{PL} = 0.50\) \(0.31\) \(0.37\) \(0.22\) | 8.83 | 10.95 | 59.73 |
| 25      | \(\hat{\xi}_{PL} = 0.84\) \(0.99\) \(0.09\) \(0.09\) | 6.93 | 10.65 | 50.55 |
| 26      | \(\hat{\xi}_{PL} = 0.50\) \(0.21\) \(0.16\) \(0.14\) | 8.44 | 9.66 | 58.67 |
| 27      | \(\hat{\xi}_{PL} = 7\) \(0.99\) \(0.84\) \(0.28\) | 4.06 | 12.31 | 55.49 |
| 28      | \(\hat{\xi}_{PL} = 0.50\) \(0.68\) \(0.52\) \(0.12\) | 6.99 | 14.37 | 70.59 |

5.4 Task performance

While the focus of our paper is on characterizing and estimating the latent acquisition functions employed by the subjects in this task, 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 behavior and success in the early moves of the task. Letting \( \hat{\tau}_j = (\hat{\tau}_j^+ + \hat{\tau}_j^-)/2 \) and \( \Delta r_j \) denote the average change in score after the second move across all of subject \( j \)'s rounds of the task, the correlation between \( \hat{\tau} \) and \( \Delta r_j \) is -0.85. This is not surprising—subjects who explore their second move 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 \( \hat{\tau}_j, \Delta r_j = 0.47 \). 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 \( \sum_{i=4}^{10} |\Delta \theta_i| \), which is a measure of how much “zig-zag” a subject exhibits after having the necessary information to estimate a planar gradient over the click-region. As explained in Section 3.4, 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. The methodology involves defining a likelihood on the observed trajectories, parameterized by an optimization subroutine with latent parameters of its own. This allows us to make probabilistic inference about the subjects’ search strategies in addition to providing a predictive model of behavior. Also, by using probabilistic models, both the variability with which an agent optimizes as well as the uncertainty around their acquisition function estimates can be quantified.

Our study also 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. Finally, we find evidence that humans respond to differently to positive versus negative rewards.
References

Agostinelli, C. and Lund, U. (2013). “R package ‘circular’: Circular Statistics (version 0.4-7).” URL https://r-forge.r-project.org/projects/circular.

Ahuja, R. K. and Orlin, J. B. (2001). “Inverse optimization.” Operations Research, 49(5): 771–783.

Aswani, A., Shen, Z.-J., and Siddiq, A. (2018). “Inverse optimization with noisy data.” Operations Research, 66(3): 870–892.

Bärmann, A., Martin, A., Pokutta, S., and Schneider, O. (2018). “An online-learning approach to inverse optimization.” arXiv preprint arXiv:1810.12997.

Borji, A. and Itti, L. (2013). “Bayesian optimization explains human active search.” In Advances in neural information processing systems, 55–63.

Box, G. E. and Tiao, G. C. (2011). Bayesian inference in statistical analysis, volume 40. John Wiley & Sons.

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.

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.

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.

Cohen, J. D., McClure, S. M., and Yu, A. J. (2007). “Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration.” Philosophical Transactions of the Royal Society B: Biological Sciences, 362(1481): 933–942.

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.

Dong, C., Chen, Y., and Zeng, B. (2018). “Generalized inverse optimization through online learning.” arXiv preprint arXiv:1810.01920.

Dong, C. and Zeng, B. (2020). “Inverse Multiobjective Optimization Through Online Learning.” arXiv preprint arXiv:2010.06140.

Erkin, Z., Bailey, M. D., Maillart, L. M., Schaefer, A. J., and Roberts, M. S. (2010). “Eliciting patients’ revealed preferences: An inverse Markov decision process approach.” Decision Analysis, 7(4): 358–365.

Esfahani, P. M., Shafiezadeh-Abadeh, S., Hanasusanto, G. A., and Kuhn, D. (2018). “Data-driven inverse optimization with imperfect information.” Mathematical Programming, 167(1): 191–234.
Gershman, S. J. (2019). “Uncertainty and exploration.” *Decision*, 6(3): 277. 2

Ghobadi, K., Lee, T., Mahmoudzadeh, H., and Terekhov, D. (2018). “Robust inverse optimization.” *Operations Research Letters*, 46(3): 339–344. 3

Gramacy, R. B. (2020). *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. CRC Press. 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. 6, 10

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. 10

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

Lizotte, D. J. (2008). *Practical bayesian optimization*. University of Alberta. 10

March, J. G. (1991). “Exploration and exploitation in organizational learning.” *Organization science*, 2(1): 71–87. 2

Mardia, K. V. and Jupp, P. E. (2009). *Directional statistics*, volume 494. John Wiley & Sons. 18

Mockus, J., Tiesis, V., and Zilinskas, A. (1978). “Toward global optimization, volume 2, chapter bayesian methods for seeking the extremum.” 10

Nadarajah, S. and Zhang, Y. (2017). “Wrapped: An R package for circular data.” *Plos one*, 12(12): e0188512. 18

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*. 2

Schonlau, M., Welch, W. J., and Jones, D. R. (1998). “Global versus local search in constrained optimization of computer models.” *Lecture Notes-Monograph Series*, 11–25. 10

Schulz, E., Tenenbaum, J. B., Reshef, D. N., Speekenbrink, M., and Gershman, S. (2015). “Assessing the Perceived Predictability of Functions.” In *CogSci*. Citeseer. 2

Shahmoradi, Z. and Lee, T. (2019). “Quantile Inverse Optimization: Improving Stability in Inverse Linear Programming.” *arXiv preprint arXiv:1908.02376*. 3

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. 6, 10
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*. 10

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. 3

Tversky, A. and Kahneman, D. (1979). “Prospect theory: An analysis of decision under risk.” *Econometrica*, 47(2): 263–291. 21

Wilson, A. G., Dann, C., Lucas, C., and Xing, E. P. (2015). “The human kernel.” In *Advances in neural information processing systems*, 2854–2862. 2

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. 2, 3, 12, 24

Acknowledgments

We would like to thank Derek Bingham, Nasrin Yousefi, and Ian Zhu for helpful comments and suggestions on the paper.
Appendix A: IBO prior specification

We use discrete uniform priors on the parameters defining the likelihood in (4.2)-(4.4). In the notation that follows, \( \mathcal{U}_n\{a, b\} \) denotes a discrete uniform distribution spanning \( a \) to \( b \) over \( n \) equally spaced intervals.

\[
\begin{align*}
\xi_{\theta 1}, \xi_{\theta 1} &\sim \mathcal{U}_{30}\{0, 30\} \\
p &\sim \mathcal{U}_{30}\{0.5, 0.99\} \\
\gamma &\sim \mathcal{U}_{30}\{0.01, \pi/4\} \\
\tau, \tau^-, \tau^+ &\sim \mathcal{U}_{30}\{0, \pi/2\}
\end{align*}
\] (A.1) (A.2) (A.3) (A.4)

Appendix B: Additional figures

Figure 12: Move 2 \( x, y \) pairs for three subjects in the experiment. On each plot, the large black dots labeled 0 and 1 show the starting location and first move location, respectively (the first move is shown along the \( x \)-axis in order to illustrate trends in the move 2 behavior). The smaller colored dots show the locations of the subjects’ second moves relative to their first moves. The color gradient denotes \( \Delta r_1 \).

Figure 13: The posterior predictive distribution of \( \tilde{f}_t \) 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.
Figure 14: \((\Delta r, \theta_2)\) pairs for each subject in the study, overlaid with the subject’s MAP augmented acquisition curve, 95% HPD credible intervals on the acquisition curve, and 95% HPD prediction intervals. Green denotes PI and red denotes UCB.