

Descriptive Models of Sequential Decisions in Engineering Design: An Experimental Study

Ashish M. Chaudhari

School of Mechanical Engineering,
Purdue University,
West Lafayette, IN 47907
e-mail: chaudha5@purdue.edu

Ilias Bilonis

Assistant Professor
School of Mechanical Engineering,
Purdue University,
West Lafayette, IN 47907
e-mail: ibilon@purdue.edu

Jitesh H. Panchal¹

Associate Professor
School of Mechanical Engineering,
Purdue University,
West Lafayette, IN 47907
e-mail: panchal@purdue.edu

Engineering design involves information acquisition decisions such as selecting designs in the design space for testing, selecting information sources, and deciding when to stop design exploration. Existing literature has established normative models for these decisions, but there is lack of knowledge about how human designers make these decisions and which strategies they use. This knowledge is important for accurately modeling design decisions, identifying sources of inefficiencies, and improving the design process. Therefore, the primary objective in this study is to identify models that provide the best description of a designer's information acquisition decisions when multiple information sources are present and the total budget is limited. We conduct a controlled human subject experiment with two independent variables: the amount of fixed budget and the monetary incentive proportional to the saved budget. By using the experimental observations, we perform Bayesian model comparison on various simple heuristic models and expected utility (EU)-based models. As expected, the subjects' decisions are better represented by the heuristic models than the EU-based models. While the EU-based models result in better net payoff, the heuristic models used by the subjects generate better design performance. The net payoff using heuristic models is closer to the EU-based models in experimental treatments where the budget is low and there is incentive for saving the budget. This indicates the potential for nudging designers' decisions toward maximizing the net payoff by setting the fixed budget at low values and providing monetary incentives proportional to saved budget. [DOI: 10.1115/1.4045605]

Keywords: engineering design decisions, sequential information acquisition, heuristics, decision theory, design automation, design process

1 Introduction

The engineering design process is recognized as an iterative decision-making process, which proceeds in stages, with each stage marked by decisions such as material selection, design evaluation, manufacturing process selection, etc. [1]. A particular class of design decisions is information acquisition decisions, which includes decisions such as whether to gain more information about a concept, whether to execute a simulation, how much to refine a model, and which model to choose. The understanding of how a designer makes information acquisition decisions is important for improving the outcomes of design processes, e.g., effectively managing the trade-off between the design performance and the cost of information gathering. This understanding can allow researchers to predict the outcomes of engineering design and systems engineering processes (e.g., Refs. [2,3]), to identify human-related sources of inefficiencies such as cognitive biases, and to find ways to reduce inefficiencies.

Despite extensive research on decision-making in design, there is a lack of quantitative descriptive models of information acquisition decisions that represent how humans actually make such decisions in engineering design and systems engineering. Existing studies in design decision-making have focused on normative frameworks (e.g., Refs. [4–7]), such as the expected utility (EU) theory [8], which assumes that designers are rational decision makers. However, it is well known that humans do not necessarily follow the normative models of decision-making [9–11]. Researchers in cognitive psychology and behavioral economics have

developed various descriptive models of human decision makers [12,13]. Examples of these descriptive models include bounded rationality-based models [14], fast and frugal heuristics [13], models based on deviations from rationality [12], and cognitive architecture-based models [15]. Although existing descriptive models are alternatives to the normative models, they do not account for the nuances of information acquisition decisions in engineering design. For example, engineering design decisions require comparisons between multiple information sources (e.g., simulation and physical prototypes), and they are constrained by budget, time, and resources. Hence, the primary objective of this paper is to identify models that provide the best description of a designer's information acquisition decisions when multiple information sources are present and the total budget is limited.

The approach followed in this paper consists of (i) designing a simple, but nontrivial, experimental task representative of the sequential information acquisition process, (ii) collecting experimental evidence on individuals' decisions, (iii) formulating alternative models of decision strategies, and (iv) performing Bayesian model comparison for identifying the best-fit models and estimating the posterior distribution over the model parameters for quantification of treatment effects (see Fig. 1). A within-subject controlled experiment is useful for evidence collection as it mitigates the influence of external factors [16] and maintains a degree of realism by using humans as designers and real money as incentives [17]. A controlled experiment, as opposed to protocol analysis, elicits decision data in a nonintrusive manner. The decision models used in the paper incorporate strategies ranging from simple heuristics to expected utility-based judgments. This approach is adaptable to different design situations by changing the experiment task and associated candidate models.

To bound the scope of the paper, we make a few simplifying assumptions. The design task in the experiment consists of a one-dimensional parametric design problem with a continuous design

¹Corresponding author.

Contributed by the Design Automation Committee of ASME for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received February 4, 2019; final manuscript received November 12, 2019; published online December 5, 2019. Assoc. Editor: Harrison M. Kim.

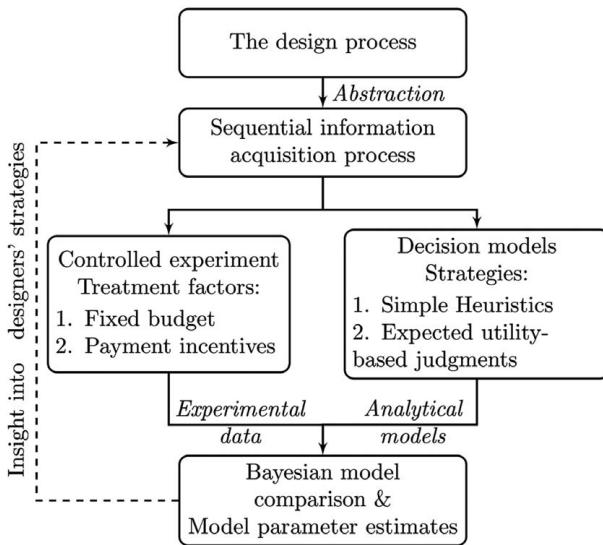


Fig. 1 An overview of the research approach

space. A designer's problem-specific domain knowledge does not have significant bearing on their sequential information acquisition decisions, and they adhere to the same decision strategy until stopping. The analysis approach and the models of the individual decisions are, however, general and can be applied to more complex design decision-making problems. Despite the simplifications, the experimental task represents a surrogate problem that embodies certain characteristics of real design problems. It requires evaluation of different designs using multiple information sources, assessing their values, and deciding the best design [1], all under a fixed budget. The best design is not known until information search, processing of the acquired information, and stopping are completed [18].

The primary contribution of this paper is an approach that combines computational models of decision-making with behavioral experiments to understand human decision-making behavior in design under uncertainty. The paper points to specific heuristic models that describe the subjects' information acquisition decisions better than the counterpart expected utility-based models. Another contribution is the insights about nudging toward cost-effectiveness by fixing the budget at low values and/or using an incentive to reduce budget spending. Systems engineers can leverage these insights for balancing the trade-off between performance and design evaluation costs in their design processes.

The paper is organized as follows. Section 2 introduces the *sequential information acquisition process*, which we use as a basis for the design of the experiment task. Section 3 provides a rationale for various descriptive decision models. Section 4 presents the experimental treatments, subject population, and payment

structure. Sections 5 and 6 provide the results and discuss their implications in the engineering design context.

2 Information Acquisition in Engineering Design

We begin with an abstraction of the design process called a sequential information acquisition and decision-making process [19], where design is considered a problem-solving activity with known design parameters and evaluation criteria but unknown mapping between the two.

2.1 Sequential Information Acquisition Decisions. A designer's objective is to find the design parameter values that maximize the performance (see Fig. 2). To achieve this objective, the designer performs iterative evaluations of the design performance. At each iteration, the designer makes the following three decisions: (1) a decision to choose next design, (2) a decision to choose an information source for performance evaluation, and (3) a decision of whether to stop evaluations.

The process is constrained by a fixed budget, which limits the number of design evaluations. The budget type may be financial (e.g., fixed cash or capital) or technical (e.g., fixed computational resources or energy) [20]. There are many examples of such a design situation. For example, in the control problem for a room heating system, a designer finds the temperature set point that minimizes energy consumption while maintaining thermal comfort [21]. Another example is the design of superconducting materials such as $\text{Cu}_x\text{Bi}_2\text{Se}_3$, where a designer finds the dopant composition (x) that maximizes superconductivity through a series of magnetization experiments [22].

The process of iterative design evaluations, here referred to as *information acquisition*, is typically performed with the help of multiple prototypes that serve as information sources with different costs and uncertainty. Various front-end methods for quantifying the uncertainty associated with information sources are available, e.g., probability distribution fitting on performance data, Delphi approach to elicit expert knowledge, and evidence theory or information gap theory to model information deficit [23].

Based on the preceding specifications, we model the case of multiple information sources as follows. Set \mathcal{X} denotes the design space, and $x \in \mathcal{X}$ denotes a point in the design space. The *performance function* $f(x)$ is a scalar function of the design, i.e., $f: \mathcal{X} \rightarrow \mathbb{R}$. However, the value $f(x)$ is not directly observable. A designer can obtain information about $f(x)$ through the query of a costly and uncertain information source. We assume that the designer has access to $M \geq 1$ such information sources. The information source labeled by m in $\{1, \dots, M\}$ has a cost $c_m \geq 0$. When this information source is evaluated at a point x , it reports a performance measurement $y = f(x) + \epsilon_m$, where ϵ_m is a random variable representing the measurement uncertainty.

Considering that the designer performs information acquisition sequentially, at each step of the process, the designer chooses a

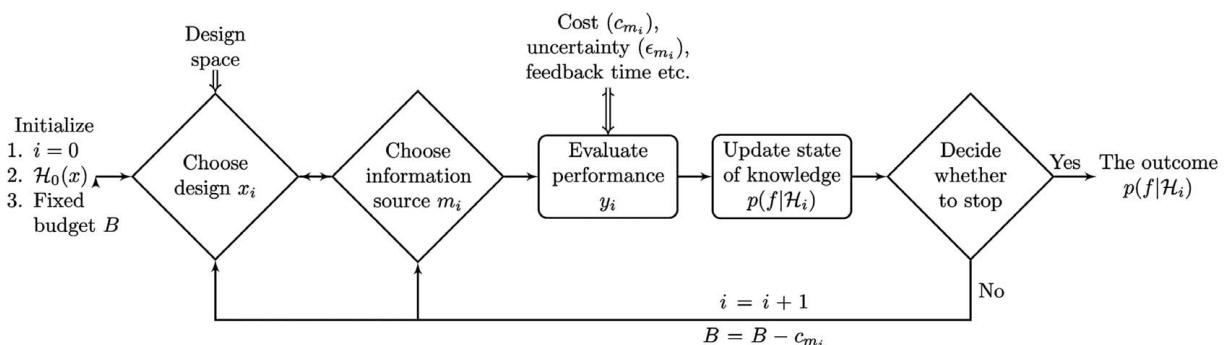


Fig. 2 Sequential information acquisition process with multiple uncertain information sources and fixed budget

design point and an information source to query based on their current state of knowledge about the performance function and the measurement uncertainties. Mathematically, at iteration i , the designer evaluates the information source m_i at a point $x_i \in \mathcal{X}$ using the observed history of evaluations $\mathcal{H}_{i-1} = \{(x_j, m_j, y_j)\}_{j=1}^{i-1}$, where $\mathcal{H}_0 = \emptyset$ along with any prior beliefs (see Fig. 2). For a parallel design procedure, in contrast, the designer would query multiple pairs of information sources and designs at each step without incorporating learning derived from the current knowledge of performance observations. Although the design process may comprise parallel and sequential queries [24,25], we focus on sequential queries as a first step toward modeling information acquisition in engineering design.

After observing the performance at the end of each step, the designer decides whether to continue or stop the evaluations. If the decision is to continue, the designer evaluates the performance function at a new point. The designer cannot perform an additional evaluation if the cost of querying the information source is larger than the available budget amount. If the decision is to stop, the outcome of the design process is the most recent state of knowledge about f , denoted by the probability measure $p(f|\mathcal{H}_i)$.

2.2 Design of the Experimental Task. To operationalize the sequential information acquisition process, we designed an experiment task using specific assumptions about the nature of the design space, information sources and fixed budget. The task involves a *subject* making design decisions, and a *user interface* processing acquired information in the back-end to display the state of knowledge about the design performance in a visual format. The roles of the subject and the user interface are separated to maintain uniformity in how different subjects process the acquired information. The incentives are proportional to the outcomes of the process to motivate the subject to maximize the design performance.

ASSUMPTION 1 (Continuous design space). *The design performance $f(x)$ is a scalar continuous function of a single design parameter x .*

ASSUMPTION 2 (Sequential evaluation process). *Subjects evaluate multiple designs sequentially. Each evaluation takes one unit of time to run, during which the subject may not begin another evaluation.*

ASSUMPTION 3 (Multiple uncertain information sources). *Subjects evaluate the performance using either a low-fidelity or a high-fidelity information source ($M=2$).*

We denote the low-fidelity information source by m 1 or “L” and the high-fidelity information source by m 2 or “H.” We denote the total number of low (high)-fidelity observations at step i by $n_{i,L}$ ($n_{i,H}$). It is $n_{i,L} = \sum_{j=1}^i 1_{\{L\}}(m_j)$ ($n_{i,H} = \sum_{j=1}^i 1_{\{H\}}(m_j)$), where $1_A(\cdot)$ is the indicator function of the set A . An example of a low-fidelity source is a stochastic computer-based simulation with large uncertainty due to approximations such as discretization of the design space, computational limitations, and errors from theoretical inadequacy. An example of a high-fidelity source is a physical prototype with relatively low uncertainty due to manufacturing defects or machining tolerances when preparing a test specimen. Simulations and prototype tests in this scenario assume aleatory uncertainty, in that they generate different observations from different evaluations of the same design point.

ASSUMPTION 4 (Gaussian measurement uncertainty). *The measurement process is modeled as a Gaussian random variable centered at the true (but unknown) performance function.*

That is, the measurement y conditioned on the design x and the information source m is as follows:

$$y|x, m \sim \mathcal{N}(f(x), v_m^2) \quad (1)$$

The noise variance v_m^2 is constant for each source m and is known to the designer. By definition, we have $v_L^2 > v_H^2$.

ASSUMPTION 5 (Known costs). *The evaluation of the performance at any design point costs a fixed amount, which is known a priori to*

the designer. The cost in this scenario is tied to the definition of budget, which may mean financial or technical resources.

If c_L and c_H are the costs of the low-fidelity and high-fidelity observations, respectively, then $c_L < c_H$.

Since the true performance function $f(x)$ is unknown, we assume that the user assigns a zero mean Gaussian prior on $f(x)$,

$$f|\mathcal{H}_0 \sim p(f|\mathcal{H}_0) := \text{GP}(0, k) \quad (2)$$

where $k(x, x') = v_0 \exp\left\{-\frac{1}{2}((x - x')^2/\ell^2)\right\}$ is a squared exponential covariance function with parameters length scale $\ell > 0$ and variance $v_0 > 0$. After making i observations, the state of the knowledge changes to

$$f|\mathcal{H}_i \sim p(f|\mathcal{H}_i) := \text{GP}(\mu_i, k_i) \quad (3)$$

where $\mu_i(x) = k(x, x_{1:i})(K_i + V_i)^{-1}y_{1:i}$ and $k_i(x, x') = k(x, x') - k(x, x_{1:i})(K_i + V_i)^{-1}k(x_{1:i}, x')$ are the posterior mean and posterior covariance functions, respectively. Here, we have defined $x_{1:i} = (x_1, \dots, x_i)$ to be the vector of the first i designs and $y_{1:i}$ to be the vector of the corresponding measurements. The $1 \times i$ matrix $k(x, x_{1:i})$ is the cross-covariance between x and $x_{1:i}$, $K_i = k(x_{1:i}, x_{1:i})$ is the $i \times i$ covariance matrix of $x_{1:i}$, Σ is an $i \times i$ diagonal matrix with elements, and $V_i = \text{diag}(v_{m_1}^2, \dots, v_{m_i}^2)$. Then, the point predictive probability of $f(x)$ is as follows:

$$f(x)|\mathcal{H}_i \sim \mathcal{N}(\mu_i(x), \sigma_i^2(x)) \quad (4)$$

where the posterior predictive variance is $\sigma_i^2(x) = k_i(x, x)$. We assume that ℓ and v_0 are constants and independent of i .

ASSUMPTION 6 (Visualization of the state of knowledge). *The user interface visualizes the state of knowledge $p(f|\mathcal{H}_i)$ by displaying the mean estimate of true performance, the 5th and 95th percentiles. This way the acquired information gets processed and visualized in the same manner for all subjects, and observing the subjects' decisions remains the focus of the experiment task.*

ASSUMPTION 7 (Fixed budget). *Each subject has a total budget of B for performance evaluations. A subject may stop before exhausting the entire budget B , so the total cost incurred, $C_i = c_L n_{i,L} + c_H n_{i,H}$, is less than or equal to B .*

ASSUMPTION 8 (Performance-based payment). *The subject's payment includes a fixed payment, a bonus proportional to the best high-fidelity observation, and a bonus proportional to the budget saved.*

The payment after i iterations is as follows:

$$I(\mathcal{H}_i) := I_0 + 1_{[1, \infty)}(n_H)[G(y_i^*) + H(B - C_i)] \quad (5)$$

where we term G as the gross payoff, which is proportional to the best high-fidelity measurement $y_i^* = \max_{j:1 \leq j \leq i, m_j=H} y_j$. Bonus H is a function of the remaining budget $B - C_i$. At least one high-fidelity measurement is required to receive bonuses because high-fidelity measurements are valued more than low-fidelity observations. Then, the *net payoff* is defined as the difference between the gross payoff and cost incurred so far,

$$N(\mathcal{H}_i) := G(y_i^*) - C_i \quad (6)$$

3 Formulation of Descriptive Decision Models

The descriptive models of information acquisition decisions fall on a spectrum with *expected utility theory* on one end and *simple heuristics* on the other end. Models closer to the EU theory embody rational judgments such as where the expectation of information gain is maximum and whether maximum performance has been achieved. They are based on criteria such as the probability of improvement (PI) [26], the expected improvement (EI) [27], the expected conditional improvement (ECI) [28], and the maxima-region entropy. Models closer to simple heuristics use cues from the environment (e.g., user interface in the experiment). Examples of such cues are predictive mean, variance, remaining

budget, and the number of evaluations. Simple heuristic models of the information acquisition decisions include the upper confidence bound (UCB) [29], conditional UCB (CUCB), the fixed sample number (FSN), the fixed remaining budget (FRB), and the dominant physical prototype (DPP). Table 1 lists descriptive models used in this paper and their underlying type (expected utility-based or simple heuristic). These models are inspired from the literature, observations of past experiments [30], and survey responses in the experiment detailed in Ref. [31].

The definition of a descriptive decision model involves two stages, (i) formulating a decision strategy as an acquisition function or a feature of observed history and (ii) modeling deviation from the strategy using a likelihood function. Acquisition functions and features are deterministic models that predict decisions for a given decision strategy, while likelihood functions, with their model parameters, impose a layer of uncertainty around those predictions. Such a construct assumes that designers are likely to make errors and deviate from predicted decisions, irrespective of whether their underlying strategies are EU based or heuristic based. For the EU-based models, the assumption of probabilistic decisions mirrors the limited cognitive ability of designers to make accurate decisions even though their judgments may be aligned with rational judgments.

3.1 Modeling the Decision to Choose the Next Design. In selecting x_{i+1} for the $(i+1)$ th iteration, an acquisition function assigns a value to every $x \in \mathcal{X}$ based on the observed history \mathcal{H}_i . The acquisition function, denoted by $\chi_i(x, \mathcal{H}_i, \psi)$, has a set of designer-specific parameters ψ . It also depends on the number of iterations i , which represents the possibility that a strategy is likely to adapt as more evaluations are accumulated, e.g., exploring the design space during initial stages while exploiting regions of high performing designs at later stages. In a deterministic setting, the best design to pick is the one that maximizes the acquisition function. For a probabilistic setting, however, we define the probability of picking x_{i+1} using Boltzmann-like likelihood function:

$$p(x_{i+1} | \chi_i, \theta_d, \mathcal{H}_i) \propto \exp\{\gamma \chi_i(x_{i+1}, \mathcal{H}_i, \psi)\} \quad (7)$$

where $\theta_d = \{\gamma, \psi\}$ collectively denotes all model parameters. This likelihood function ensures that parts of the design space with

high acquisition function values are preferred over other parts and that points with the same value of acquisition function have the same probability of being selected. The rate parameter $\gamma \geq 0$ is associated with the sensitivity of probability density to changes in the acquisition function. In one extreme, as $\gamma \rightarrow \infty$, the model becomes equivalent to the deterministic one as the likelihood function collapses to a Dirac delta centered at the maximum of the acquisition function. At the other extreme, as $\gamma \rightarrow 0^+$, the likelihood function becomes uniform over \mathcal{X} . Different decision models based on different acquisition functions are presented later.

3.1.1 Upper Confidence Bound. The upper confidence bound model promotes the exploration of design space during the initial iterations and exploitation of peak performance areas during later iterations [29]. Design points with high values of $\mu_i(x) + \alpha_i \sigma_i(x)$, where α_i is an exploration parameter, are preferred. The corresponding information acquisition function is as follows:

$$\chi_i(x, \mathcal{H}_i, \psi) = \mu_i(x) + a e^{-bi} \sigma_i(x) \quad (8)$$

where $\psi = \{a, b\}$ are the model parameters. The parameter $\alpha_i = a e^{-bi}$ is a decreasing exponential function of i because the model represents the strategy to explore first and exploit later.

3.1.2 Probability of Improvement. The acquisition function for this model is proportional to the probability that the next evaluation at x will generate higher performance than the current best performance [26]. We take the current best to be the maximum of predictive means at past design evaluations, $\mu_i^* = \max_{1 \leq j \leq i} \mu_i(x_j)$, because of the uncertain information sources. Let $\Delta_i(x) = \max\{f(x) - \mu_i^*, 0\}$ be the improvement at x . Given that the state of knowledge about performance $f(x)$ is distributed as Eq. (4), the parameter-free information acquisition function is as follows:

$$\chi_i(x, \mathcal{H}_i) = \mathbb{P}[\Delta_i(x) > 0 | \mathcal{H}_i] = 1 - \Phi\left(\frac{\mu_i(x) - \mu_i^*}{\sigma_i(x)}\right) \quad (9)$$

where Φ is the cumulative density function of the standard normal distribution.

3.1.3 Expected Improvement. The expected improvement model prefers design points with high expectation of improvement

Table 1 Models for the information acquisition decisions and their underlying type (simple heuristic or expected utility (EU) based)

Decision model	Underlying strategy	Model type
1. Decision to choose next design		
a. UCB [29]	Explore design space during initial iterations while exploit during later iterations.	Simple heuristic
b. Probability of improvement (PI) [26]	Selection probability proportional to PI value.	EU based
c. EI [27]	Selection probability proportional to EI value.	EU based
d. ECI [28]	Selection probability inversely proportional to ECI value.	EU based
2. Decision to choose information source		
a. FSN	Test high-fidelity source after a fixed number of samples.	Simple heuristic
b. FRB	Test high-fidelity source if the remaining budget is smaller than a fixed value.	Simple heuristic
c. Fixed maximum-region entropy (FME)	Test high-fidelity source if the information entropy of the location of maximum is smaller than a fixed value.	Simple heuristic
d. Fixed expected conditional improvement (FECI)	Test high-fidelity source when the difference between EI from one step and EI from two steps is smaller than a fixed value.	EU based
3. Decision to stop		
a. FSN	Stop after a fixed number of samples.	Simple heuristic
b. FRB	Stop after a fixed amount of budget is remaining.	Simple heuristic
c. DPP	Stop when the best high-fidelity measurement minus the largest predictive mean is smaller than a fixed value.	Simple heuristic
d. FME	Stop after the entropy of the location of maximum is smaller than a fixed value.	EU based
e. FEI [32]	Stop after EI is below a fixed value.	EU based

relative to μ_i^* from a single additional design evaluation [27]. Its acquisition function is as follows:

$$\begin{aligned}\chi_i(x, \mathcal{H}_i) &= \mathbb{E}[\Delta_i(x)|\mathcal{H}_i] \\ &= (\mu_i(x) - \mu_i^*)\Phi\left(\frac{\mu_i(x) - \mu_i^*}{\sigma_i(x)}\right) \\ &\quad + \sigma_i(x)\phi\left(\frac{\mu_i(x) - \mu_i^*}{\sigma_i(x)}\right)\end{aligned}\quad (10)$$

where the expectation is over the state of knowledge at step i and Φ is the cumulative distribution function of the standard normal distribution.

3.1.4 Expected Conditional Improvement. This model formulates the strategy of looking two steps ahead and minimizing the aggregate expectation of conditional improvement from the second step after the next one [28]. Suppose that we are making a hypothetical observation using the hypothetical next-step design, x , and the hypothetical next choice of information source, m . The hypothetical point predictive density is $\mathcal{N}(\mu_{i|x,m}(\tilde{x}), \sigma_{i|x,m}^2(\tilde{x}))$. Assume that this density has the mean equal to the current mean,

$$\mu_{i|x,m}(\tilde{x}) = \mu_i(\tilde{x}) \quad (11)$$

which is justifiable because the predictive density of $f(\tilde{x}|x, m)$ is unchanged without observing the outcome y at $\{x, m\}$. Given that y at $\{x, m\}$ has been added to the observations, the hypothetical predictive variance is as follows:

$$\sigma_{i|x,m}^2(\tilde{x}) = \sigma_{i+1|x,m,y}^2(\tilde{x}) = \sigma_i^2(\tilde{x}) - \frac{k_i^2(\tilde{x}, x)}{k_i(x, x) + v_m^2} \quad (12)$$

Note that $\sigma_{i|x,m}^2(\tilde{x})$ is equal to $\sigma_{i+1|x,m,y}^2(\tilde{x})$ because $\sigma_{i+1|x,m,y}^2(\tilde{x})$ is in fact independent of y .

Then, $\Delta_i(\tilde{x}|x, m) = \max\{f(\tilde{x}|x, m) - \mu_i, 0\}$ is called the conditional improvement at \tilde{x} . If the current maximum is taken as the maximum predictive mean over \mathcal{X} , $\hat{\mu}_i = \max_{\tilde{x} \in \mathcal{X}} \mu_i(\tilde{x})$, the following monotonicity condition is true for all $\tilde{x} \in \mathcal{X}$ [28]:

$$\mathbb{E}[\Delta_i(\tilde{x}|x, m)|\mathcal{H}_i] \leq \mathbb{E}[\Delta_i(\tilde{x})|\mathcal{H}_i] \quad (13)$$

where the expected conditional improvement, $\mathbb{E}[\Delta_i(\tilde{x}|x, m)|\mathcal{H}_i]$, is calculated by substituting the hypothetical mean $\mu_{i|x,m}(\tilde{x})$ and variance $\sigma_{i|x,m}^2(\tilde{x})$ in Eq. (10). This means that the improvement potential of any design \tilde{x} reduces or remains as is after adding the design x . If the selected x is influential at reference point \tilde{x} , then the conditional improvement at \tilde{x} is small. Then, the most influential design is the one that minimizes the aggregate conditional improvement over all $\tilde{x} \in \mathcal{X}$ (or maximize the negative of the same statistic). Accordingly, we define the acquisition function as the integrated expected conditional improvement:

$$\chi_i(x, \mathcal{H}_i|m) = - \int_{\mathcal{X}} \mathbb{E}[\Delta(\tilde{x}|x, m)|\mathcal{H}_i] d\tilde{x} \quad (14)$$

In the analysis, we take m to be the low-fidelity source, “L.”

3.2 Models of the Decision to Choose an Information Source. Different attributes of the observed history \mathcal{H}_i influence the decision to choose among information sources. Apart from past observations, these attributes can be derived quantities such as the sequence of different information sources used, the frequency of a particular source, the total cost, etc. Formally, we refer to a mapping between the observed history to some attribute as a feature function. A feature function (or simply feature) incorporates the observed history into the decision models. Given that multiple history attributes may influence decisions, a decision strategy is specified in terms of a weighted sum of multiple independent features. The decision models defined on the basis of feature functions are *threshold based*, i.e., a decision is made based on whether the

weighted sum of features is greater or less than a threshold value. Mathematically, we characterize a particular strategy using R_m independent features denoted by $g_{m,1}(\mathcal{H}_i), \dots, g_{m,R_m}(\mathcal{H}_i)$, with $w_{m,1}, \dots, w_{m,R_m}$ as the weight parameters. The likelihood of choosing the high-fidelity information source ($m_{i+1} = 2$) is defined using the sigmoid function as follows:

$$p(m_{i+1} = H|\theta_m, \mathcal{H}_i) = \text{sigm}\left(\sum_{r=1}^{R_m} w_{m,r} g_{m,r}(\mathcal{H}_i)\right) \quad (15)$$

where $\text{sigm}(\lambda) = (1 + \exp\{-\lambda\})^{-1}$ is the sigmoid function and $\theta_m = \{w_{m,1:R_m}\}$ are model parameters. The weight parameter $w_{m,r}$ is positive or negative, respectively, based on whether an increase in $g_{m,r}(\mathcal{H}_i)$ increases or reduces the selection probability of the high-fidelity source. The likelihood of selecting the low-fidelity information source is $p(m_{i+1} = L|\theta_m, \mathcal{H}_i) = 1 - p(m_{i+1} = H|\theta_m, \mathcal{H}_i)$.

In a threshold-based decision model, we always include a constant, negative basis function $g_{m,1}(\mathcal{H}_i) = -1$ because the *difference* between the weighted sum and a threshold determines the decision strategy. Furthermore, we assume that each designer's strategy relies upon a single element of history, and the decision model uses two features ($R_m = 2$), one more in addition to the constant one. This assumption reflects that people's cognitive ability is limited, and they do not consider all the relevant information while making decisions [33].

3.2.1 Fixed Sample Number. In this model, the low-fidelity information source is used for a fixed number of samples, and the expensive high-fidelity source is used thereafter. The feature used for this model is as follows:

$$g_{m,1,2}(\mathcal{H}_i) = i \quad (16)$$

3.2.2 Fixed Remaining Budget. In this model, the remaining budget determines the choice between the two information sources. Low-fidelity observations, if any, are collected during initial iterations until a fixed amount of remaining budget is left, and high-fidelity observations are collected thereafter.

The feature for this model is as follows:

$$g_{m,1,2}(\mathcal{H}_i) = B - C_i \quad (17)$$

where cost $C_i = n_{i,LC} + n_{i,HCH}$ is a function of observed history \mathcal{H}_i .

3.2.3 Fixed Maximum-Region Entropy. A strategy is based on the judgment of whether the region of function maximum has been sufficiently identified given the observed history. It is assumed that once the information entropy of the posterior probability density of the performance maximum reduces to a fixed value, the designer starts design evaluations using the high-fidelity information source.

To define the entropy, let X^* be the random variable representing the location of the maximum, i.e., $X^* = X^*[f] = \arg \max_{x \in \mathcal{X}} f(x)$. The posterior probability density of X^* is given by:

$$p(x^*|\mathcal{H}_i) = \mathbb{E}\left[\delta\left(x^* - \arg \max_{x \in \mathcal{X}} f(x)\right)|\mathcal{H}_i\right] \quad (18)$$

The entropy of this distribution is as follows:

$$\begin{aligned}g_{m,1,2}(\mathcal{H}_i) &= S(\mathcal{H}_i) := -\mathbb{E}[\log p(X^*|\mathcal{H}_i)|\mathcal{H}_i] \\ &= - \int_{x^* \in \mathcal{X}} \log p(x^*|\mathcal{H}_i) p(x^*|\mathcal{H}_i) dx^*\end{aligned}\quad (19)$$

We estimate the entropy numerically using the maxima of 500 functions sampled from the posterior GP, see Eq. (3). Specifically, we estimate $p(x^*|\mathcal{H}_i)$ by building the histogram of the sampled maxima and then performing the integration of Eq. (19) numerically.

3.2.4 Fixed Expected Conditional Improvement. In this model, two possible actions are compared at each iteration for the selection

between information sources, (i) evaluate using the expensive high-fidelity source or (ii) evaluate using the cheap low-fidelity source first and then using the expensive high-fidelity source next. The first action is desirable when only a small improvement in performance is expected, whereas the second action is desirable when large improvements are possible and exploration of the design space is beneficial. Accordingly, this model compares the maximum expected improvement from a single physical prototype ($\max_{x \in \mathcal{X}} \mathbb{E}[\Delta_i(x) | \mathcal{H}_i]$) to that from one computer simulation and one physical prototype ($\max_{x, \tilde{x} \in \mathcal{X}} (\mathbb{E}[\Delta_i(x) | \mathcal{H}_i] + \mathbb{E}[\Delta_i(\tilde{x} | x, L) | \mathcal{H}_i])$). The cost difference between the two actions, i.e., the cost of one computer simulation is absorbed in the constant feature. The feature function for choosing the first action over the second one is as follows:

$$g_{m,i,2}(\mathcal{H}_i) = \max_{x, \tilde{x} \in \mathcal{X}} (\mathbb{E}[\Delta_i(x) | \mathcal{H}_i] + \mathbb{E}[\Delta_i(\tilde{x} | x, L) | \mathcal{H}_i]) - \max_{x \in \mathcal{X}} \mathbb{E}[\Delta_i(x) | \mathcal{H}_i] \quad (20)$$

3.3 Models of the Decision to Stop. The strategies for deciding whether to stop are modeled using basis functions in the same manner as the decision to choose information source. Assume that a strategy for stopping after i iterations is dependent on R_s features of history \mathcal{H}_i and characterized by basis functions $g_{s,i,1}(\mathcal{H}_i), \dots, g_{s,i,R_s}(\mathcal{H}_i)$. Then, the likelihood of stopping is defined using a sigmoid function as follows:

$$p(s_i = 1 | \theta_s, \mathcal{H}_i) = \text{sigm}\left(\sum_{r=1}^{R_s} w_{s,r} g_{s,i,r}(\mathcal{H}_i)\right) \quad (21)$$

where $\theta_s = \{w_{s,1:R_s}\}$ are designer-specific model parameters. The weight parameter $w_{s,r}$ can be positive or negative depending on whether an increase in $g_{s,i,r}(\mathcal{H}_i)$, respectively, increases or reduces the probability of stopping. Again, we take $g_{s,i,1}(\mathcal{H}_i) = -1$, and we assume that the designer only relies upon a single feature of history while deciding whether to stop. Thus, $R_s = 2$ for all the models.

3.3.1 Fixed Sample Number. According to this model, the total number of iterations performed is fixed. Accordingly, the feature function is as follows:

$$g_{s,i,2}(\mathcal{H}_i) = i \quad (22)$$

3.3.2 Fixed Remaining Budget. In this model, we assume that design evaluations are stopped when the remaining budget $B - C_i$ reduces to a fixed value. The basis function has the following form:

$$g_{s,i,2}(\mathcal{H}_i) = B - C_i \quad (23)$$

3.3.3 Dominant Physical Prototype. Design evaluations are stopped when the difference between the largest high-fidelity measurement and the maximum of predictive mean is smaller than a fixed value. We define the corresponding feature as follows:

$$g_{s,i,2}(\mathcal{H}_i) = \max_{x \in \mathcal{X}} \mu_i(x) - \max_{j: 1 \leq j \leq i, m_j = H} y_j \quad (24)$$

3.3.4 Fixed Expected Improvement. In this model, we assume that design evaluations are stopped if the maximum expected improvement from the next sample is small. This strategy fits into Eq. (21) by defining the feature function as follows:

$$g_{s,i,2}(\mathcal{H}_i) = \mathbb{E}[\Delta_i(x) | \mathcal{H}_i] \quad (25)$$

where the expected improvement is calculated in the same way as given in Eq. (10).

3.4 Conditional Decisions to Choose the Next Design and to Choose the Information Source. The decisions of choosing the next design and choosing information source can be

interdependent. For example, a designer may use the low-fidelity information source to evaluate design points with large uncertainty (for exploration) and the high-fidelity information source to evaluate design points closer to regions of large performance and relatively low uncertainty (for exploitation). We call this model the CUCB model.

For the case when selecting x_{i+1} is conditional on the choice of information source m_{i+1} , the joint probability of both the decisions is $p(x_{i+1}, m_{i+1} | \mathcal{H}_i) = p(m_{i+1} | \mathcal{H}_i)p(x_{i+1} | m_{i+1}, \mathcal{H}_i)$, where $p(x_{i+1} | m_{i+1}, \mathcal{H}_i)$ is the conditional likelihood function for the decision to choose next design,

$$p(x_{i+1} | m_{i+1}, \mathcal{H}_i) \propto \exp\{\gamma'[\mu_i(x_{i+1}) + \alpha' \sigma_i(x_{i+1})]\} \quad (26)$$

The exploration parameter α' is $a' e^{-b' i}$ for $m_{i+1} = L$ and 0 for $m_{i+1} = H$, where a' and b' are positive parameters. The conditional likelihood function is dependent on m_{i+1} and independent of its probability $p(m_{i+1} | \mathcal{H}_i)$, which can take any of the forms given in Sec. 3.2.

The joint probability can be alternatively written as follows: $p(x_{i+1}, m_{i+1} | \mathcal{H}_i) = p(x_{i+1} | \mathcal{H}_i)p(m_{i+1} | x_{i+1}, \mathcal{H}_i)$. Here, the likelihood of choosing the high-fidelity source ($m_{i+1} = H$) conditional on x_{i+1} is as follows:

$$p(m_{i+1} = H | x_{i+1}, \mathcal{H}_i) = \text{sigm}\left(\sum_{r=1}^{R_m} w_{m,r} g_{m,i,r}(x_{i+1}, \mathcal{H}_i)\right) \quad (27)$$

where $w_{m,r}$ are model parameters. In a manner similar to the models in Sec. 3.2, we fix $R_m = 2$ and $g_{m,i,1} = -1$. Then, the feature that specifies the dependence between the decisions is as follows:

$$g_{m,i,2}(x_{i+1}, \mathcal{H}_i) = \max_{x \in \mathcal{X}} \mu_i(x) - \mu_i(x_{i+1}) \quad (28)$$

The above formulation supports picking the high-fidelity information source for designs that have high expectation of performance.

4 A Controlled Experiment to Elicit Design Decisions

For parameter estimation and comparison of the decision models, we gathered data from decisions by conducting an experiment with the experimental task presented in Sec. 2.2.

4.1 Subjects, Treatments, and Payment. A total of 63 student subjects were recruited from an introductory undergraduate level machine design course. The participation was voluntary and was not considered toward students' grades.

Each subject performed 18 runs of the experimental task with a distinct unknown performance function. A run of the experimental task is called *a period*. The objective in each period was to find the maximum of an unknown function. For every iteration in each period, subjects made three decisions, (i) decision to choose x , (ii) decision to choose an information source, and (iii) decision about whether to stop. The 18 distinct functions were randomly generated prior to the experiment and were fixed for all subjects. The assignment of these functions to periods was randomized for each subject to minimize potential confounding between functions and treatments. Some parameters of the experimental task were fixed. In particular, the design evaluation costs were $c_L = 2$, $c_H = 8$, the measurement variance were $v_L = 10$, $v_H = 0.0$, the design space was $\mathcal{X} = [-10, 10]$, and the fixed minimum payment $\mathcal{I}_0 = \$5$. The user interface is shown in Fig. 3. For the ease of understanding of the subjects, we termed the low-fidelity information source as a computer simulation, and the high-fidelity information source as a physical prototype.

The experiment was divided into three parts:

- (1) *Trial part* (two periods): The first part involved two trial periods to help the subjects get familiarized with the user interface before starting the actual experiment. The outcomes of these functions were not considered toward the subjects' payment.

Condition: Use or Lose Budget, Total budget: 40

Function: 5

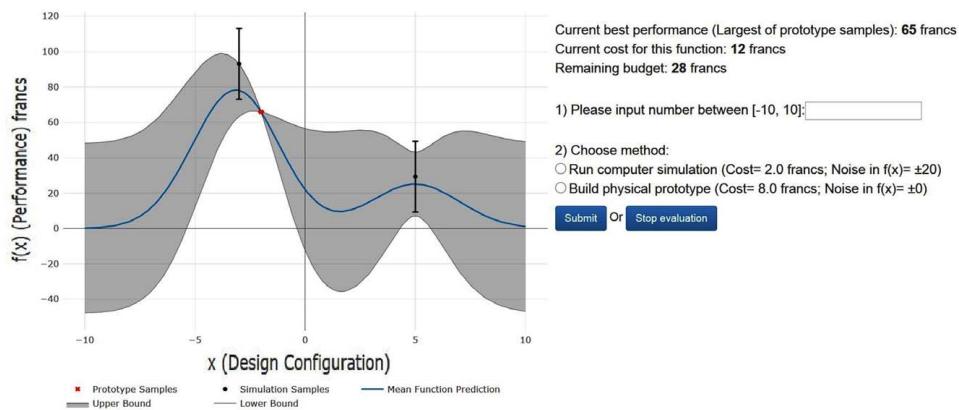


Fig. 3 A screenshot of the user interface, developed using oTree [34], shows the information provided to the subjects during the experiment

- (2) *Use-it-or-lose-it part* (nine periods): For this part, the subjects were allocated a fixed budget per period. Any remaining budget was discarded and not added to the subject's payment. In this part, a subject evaluated nine unknown functions in nine periods, with three functions each for three treatments of fixed budget per period: (i) treatment $T1 : B=20$, (ii) treatment $T2 : B=40$, and (iii) treatment $T3 : B=60$.
- (3) *Save-remaining-budget part* (nine periods): For this part, the subjects were allocated a fixed budget per period and any remaining budget at the end of every period was added to payment as a bonus H_b . Subjects evaluated nine unknown functions in nine periods, with three functions each for the three treatments of fixed budget per period: (i) treatment $T4 : B=20$, (ii) treatment $T5 : B=40$, and (iii) treatment $T6 : B=60$.

At the end of the above three parts (six treatments), the subjects completed a survey on their computer screen where they responded to three questions asking them to list the strategies they used for the three decisions.

The order of treatments was varied across the subjects to control for *order effects* [35]. The four different orders of six treatments were as follows: (i) $T1 - T2 - T3 - T4 - T5 - T6$, (ii) $T3 - T2 - T1 - T6 - T5 - T4$, (iii) $T4 - T5 - T6 - T1 - T2 - T3$, and (iv) $T6 - T5 - T4 - T3 - T2 - T1$.

If y_i^* was the best observation in high-fidelity observations, then the gross payoff was calculated as follows:

$$G(y_i^*) = 100 - (f_{\max} - y_i^*) \text{ francs} \quad (29)$$

where f_{\max} was the true maximum value of a given unknown function and franc was the experimental currency unit. The gross payoff for a period was revealed only after stopping. For the periods in save-remaining-budget treatments, any remaining budget was added to the gross payoff, i.e., $H(B - C_i) = B - C_i$. For the periods in use-it-or-lose-it treatments, $H(B - C_i)$ equalled 0. We converted the total payment ($G(y_i^*) + H(B - C_i)$) from two randomly selected periods into equivalent US dollars for payment so as to encourage subjects to put their best effort in each period. The bonus payment and the conversion rate between francs and US dollars was revealed at the end of the experiment to mandate participation in all parts to receive any payment. This rule was aimed at minimizing the *wealth effect* that influences the future effort once winnings from the previous periods are revealed. The rule also reduces the *selection bias*, which

discourages participation in future treatments once the payment from previous treatments is received.

4.2 Data Acquisition. We collected data on the choice of design point, $x_i \in \mathcal{X}$, the choice of information source, $m_i \in \{L, H\}$, and the choice of stopping, s_i which is 0 if the subject stopped after the iteration or 1 otherwise. We also recorded the related quantities such as gross payoff, fixed budget, functional performances, and the index of iteration i associated with every evaluation. In addition, we recorded the text of the subjects' survey responses.

The descriptive statistics for different treatments are shown in Table 2. With the increase in the amount of fixed budget, the subjects performed more iterations and evaluated more number of high-fidelity information sources. When the incentive to save budget was in place, the number of iterations reduced. Higher number of iterations resulted in better performance on average as well as higher costs, as shown in Fig. 4, but the net payoff (performance minus cost) was lower. Net payoff increased with the incentive-to-save-budget. These observations highlight the usefulness of high fixed budget for improving design performance and that of the incentive-to-save-budget for reducing spending and improving net payoff.

Table 2 Mean and standard deviation of some basic attributes from the dataset

Attribute	Treatments					
	T1	T2	T3	T4	T5	T6
Number of subjects	63	63	63	63	63	63
Total number of decisions	1292	2100	2602	1155	1608	1743
Number of iterations per period	6.8 (±1.7)	11.1 (±3.5)	13.8 (±4.7)	6.1 (±1.6)	8.5 (±2.9)	9.2 (±3.0)
Number of high-fidelity information sources per period	1.0 (±0.6)	2.7 (±1.2)	4.3 (±1.8)	1.0 (±0.5)	1.8 (±1.1)	2.3 (±1.7)
Duration of iteration (s)	11.5 (±8.7)	11.1 (±6.6)	11.6 (±7.6)	11.5 (±8.1)	10.6 (±6.1)	11.1 (±9.3)
Duration of period (s)	76.7 (±30.8)	123.6 (±54.2)	159.3 (±70.0)	68.9 (±28.4)	89.4 (±40.6)	101.9 (±50.3)

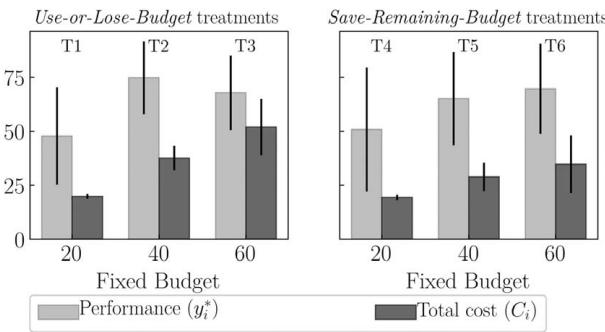


Fig. 4 The subjects' actual performance and total cost after stopping different treatments

5 Bayesian Model Comparison of the Descriptive Decision Models

In this section, we describe the *variational Bayes* approach to find approximations to posterior distributions of the model parameters and to estimate lower bounds to marginal log-likelihoods of the decision models conditional on the experimental data [36]. The variational Bayes approach is useful in complex stochastic models where analytical forms of posterior distributions are intractable. The model evidence lower bound (ELBO), that is, an approximation to the expectation of log posterior probability of the data, quantifies the support for a model, i.e., the accuracy with which a model represents the experimental data. We denote ELBO, say for model M_j , as \mathcal{L}_j .

It is assumed that all models are equally likely to represent the data a priori. To facilitate more intuitive explanation of the parameter estimates in the results, we transform the likelihood functions in Eqs. (15), (21), and (27) and take the weight parameters $w_{m,2}$ and $w_{s,2}$ out of the summation. This way the threshold parameters are given by $w_{m,1}/w_{m,2}$ and $w_{s,1}/w_{s,2}$.

A hierarchical form of a decision model is fitted to the experimental data for each treatment group separately, as described in Fig. 5. Hyperparameters are parameters of the prior distributions over model parameters. They characterize the group-level preferences of the subject population. The prior distributions of model parameters are functions of samples from hyperpriors, that is, priors over the hyperparameters. Subject-specific model parameters

are independent samples from these prior distributions. With this setup, the group-level treatment effects and individualized treatment effects are implied from the posterior distributions of hyperparameters and models parameters, respectively. See Table 3 for the hyperpriors and the prior distributions.

Before performing model training, we partition the experimental data into training and test datasets. The training dataset is drawn from randomly selected 15 periods and used to estimate the model parameters. The test dataset consisting of the remaining three periods is used for estimating the posterior predictive accuracy of the decision models. The predictive accuracy on the test data is given by the computed *log pointwise predictive density* as follows [37]:

$$\text{Computed } lppd = \sum_{n=1}^{\hat{N}} \log \left(\frac{1}{S} \sum_{s=1}^S p(z_n | \theta^s) \right) \quad (30)$$

where $\theta^s \sim p_{post}(\theta)$ for $s = 1, 2, \dots, S$ are posterior samples of given model parameters, z_n is a test data point, and \hat{N} is the number of test data points. A non-Bayesian measure, called *accuracy score*, is also calculated for the decision models of choosing information source and stopping which have discrete alternatives. If $z_n^{(p)}$ is a prediction corresponding to a data point z_n , then the accuracy score is $(1/\hat{N}) \sum_{n=1}^{\hat{N}} \delta_{\{z_n\}}(z_n^{(p)})$.

The posterior distribution approximations and ELBO for decision models were estimated using automatic differentiation variational inference [38] algorithm in PyMC3 module of PYTHON [39]. This algorithm was run for 50,000 iterations, and among those, last 5000 iterations were used to calculate the average ELBO.

5.1 Estimates of Model Evidence Lower Bound. The positive values of estimates of model ELBOs relative to random sampling $\mathcal{L}_j - \mathcal{L}_{rand}$ in Fig. 6 highlight that the predictions of the decision models are more accurate than random predictions. The ELBO of random sampling \mathcal{L}_{rand} for the decision to choose next design is $N \log(\frac{1}{20})$ assuming a uniform distribution function over the design space $[-10, 10]$ and N is the training data size. It is $N \log(0.5)$ for the decisions to choose an information source and to stop assuming the probability of 0.5 for each of the two alternatives in both the decisions. We are able to compare the support for any two models, say j_1 and j_2 , by comparing $\mathcal{L}_{j_1} - \mathcal{L}_{rand}$ and $\mathcal{L}_{j_2} - \mathcal{L}_{rand}$, because \mathcal{L}_{rand} remains constant in a given treatment.

RESULT 1. *For the decision to choose the next design, the UCB model and the CUCB model have the highest ELBOs.*

From Result 1 and Fig. 6, we conclude that exploration during initial iterations while exploitation during later iterations, captured by UCB and CUCB models, is the most likely strategy for choosing the next design point. At low-budget treatments $T1$ and $T4$, the ELBO of the CUCB model is the highest, suggesting that the selection of a design point and an information source are interdependent. The subjects use low-fidelity observations for exploration and high-fidelity observations for exploitation.

RESULT 2. *For the decision to select an information source, the FSN model and the CUCB model have higher ELBOs than the other alternative models. The FSN model has the highest ELBO at low budget, whereas the CUCB model has the highest ELBO at medium and high budgets.*

Selecting the first high-fidelity observation after a fixed number of iterations is the most likely strategy at low budget. However, with higher total budget, subjects also rely on the predictive mean to select whether to choose the high-fidelity information source. Evaluating high-fidelity observations at locations closest to the highest predictive mean is the most likely strategy for medium and high budgets.

RESULT 3. *For the decision to stop, the FRB model has the highest ELBO in all treatments, except in medium- and high-budget treatments of "save-remaining-budget" part where the DPP model has the ELBO similar to that of the FRB model.*

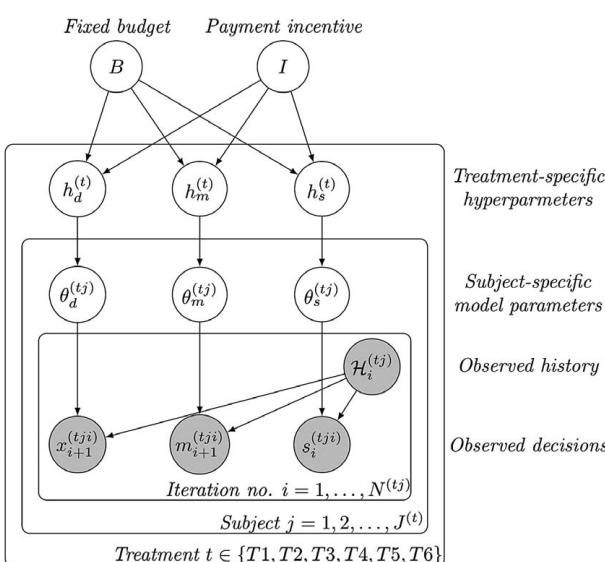


Fig. 5 A graphical representation of the hierarchical decision models. A hollow circular node denotes random parameters and a shaded circular node denotes observations.

Table 3 Uninformative hyperpriors and priors of model parameters for the hierarchical decision models

	Decision to choose $x_{i+1}^{(ji)}$	Decision to choose $m_{i+1}^{(ji)}$	Decision to choose $s_{i+1}^{(ji)}$
Hyperpriors over hyperparameters $h_d^{(t)}, h_m^{(t)}, h_s^{(t)}$	$\alpha_a^{(t)}, \alpha_b^{(t)}, \alpha_\gamma^{(t)}, \beta_\gamma^{(t)} \sim \text{Gamma}(1, 1)$ $\alpha_{w_{m1}/w_{m2}}^{(t)}, \beta_{w_{m1}/w_{m2}}^{(t)}, \alpha_{w_{m,2}}^{(t)}, \beta_{w_{m,2}}^{(t)} \sim \text{Gamma}(1, 1)$ $\alpha_{w_{s1}/w_{s2}}^{(t)} \sim \text{Normal}(1, 1)$ for CUCB	$\alpha_{w_{m1}/w_{m2}}^{(t)}, \beta_{w_{m1}/w_{m2}}^{(t)}, \alpha_{w_{m,2}}^{(t)}, \beta_{w_{m,2}}^{(t)} \sim \text{Gamma}(1, 1)$ $-w_{m,2}^{(t)} \sim \text{Gamma}(\alpha_{w_{m,2}}^{(t)}, \beta_{w_{m,2}}^{(t)})$ for FME, FRB, FECI, CUCB	$\alpha_{w_{s1}/w_{s2}}^{(t)}, \beta_{w_{s1}/w_{s2}}^{(t)}, \alpha_{w_{s,2}}^{(t)}, \beta_{w_{s,2}}^{(t)} \sim \text{Gamma}(1, 1)$ $\alpha_{w_{s1}/w_{s2}}^{(t)} \sim \text{Normal}(0, 1)$ for DPP
Priors over model parameters $\theta_d^{(ji)}, \theta_m^{(ji)}, \theta_s^{(ji)}$	$a^{(ji)} \sim \text{Gamma}(\alpha_a^{(t)}, 5)$ $b^{(ji)} \sim \text{Gamma}(\alpha_b^{(t)}, 5)$ $\gamma^{(ji)} \sim \text{Gamma}(\alpha_\gamma^{(t)}, \beta_\gamma^{(t)})$	$-w_{m,2}^{(t)} \sim \text{Gamma}(\alpha_{w_{m,2}}^{(t)}, \beta_{w_{m,2}}^{(t)})$ for FME, FRB, FECI, CUCB $w_{m,2}^{(t)} \sim \text{Gamma}(\alpha_{w_{m,2}}^{(t)}, \beta_{w_{m,2}}^{(t)})$ for FSN $\frac{w_{m,1}^{(t)}}{w_{m,2}^{(t)}} \sim \text{Gamma}(\alpha_{w_{m,2}}^{(t)}, \beta_{w_{m,2}}^{(t)})$ for FME, FECI, CUCB $\frac{w_{m,1}^{(t)}}{w_{m,2}^{(t)}} \sim \text{Normal}(\alpha_{w_{m,2}}^{(t)}, \beta_{w_{m,2}}^{(t)})$ for FSN, FRB	$-w_{s,2}^{(t)} \sim \text{Gamma}(\alpha_{w_{s,2}}^{(t)}, \beta_{w_{s,2}}^{(t)})$ for FME, FRB, FECI $w_{s,2}^{(t)} \sim \text{Gamma}(\alpha_{w_{s,2}}^{(t)}, \beta_{w_{s,2}}^{(t)})$ FSN, DPP $\frac{w_{s,1}^{(t)}}{w_{s,2}^{(t)}} \sim \text{Gamma}(\alpha_{w_{s,2}}^{(t)}, \beta_{w_{s,2}}^{(t)})$ for FME, FEI $\frac{w_{s,1}^{(t)}}{w_{s,2}^{(t)}} \sim \text{Normal}(\alpha_{w_{s,2}}^{(t)}, \beta_{w_{s,2}}^{(t)})$ for FSN, DPP, FRB

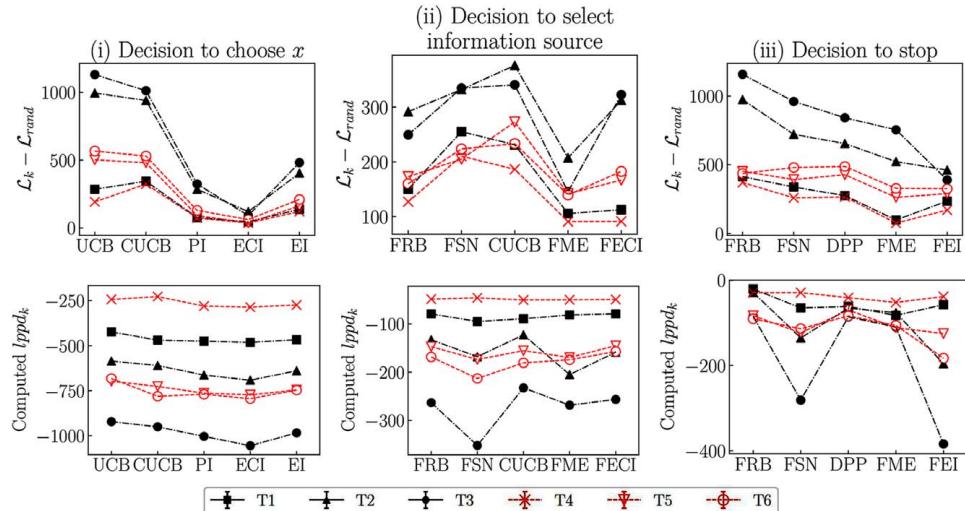


Fig. 6 Model evidence lower bound ($\mathcal{L}_k - \mathcal{L}_{\text{rand}}$) relative to random selection from training data and computed log pointwise posterior density (computed lppd_k) from the test data in different treatments

According to the results in Fig. 6, the subjects stopped after exhausting the entire or part of the available budget in treatments T_1, T_2, T_3 , and T_4 . However, at medium and high budget in the “save-remaining-budget” part (treatments T_5 and T_6), the subjects stopped when the existing best performance from high-fidelity observations was closer to the highest mean prediction of the performance.

Results 1–3 also hold true for the test dataset based on computed lppd metrics in Fig. 6 and prediction accuracy scores in Fig. 7. Both

the larger computed lppd and larger accuracy score imply better support for a model. The differences in prediction accuracy from different models are substantial given that lppd is defined in the logarithmic scale.

5.2 Effects of Fixed Budget and Payment Incentives. The amount of fixed budget and the incentive-to-save-budget affect the posterior distributions of model parameters and, by implication, the subjects’ strategies for information acquisition decisions. The exceptions are the rate parameter γ for the EI, PI, and ECI models, where γ remains largely constant. Small γ means that the decision is random according to the model, whereas a large γ means the information acquisition function is closely followed. Posterior distributions suggest that the UCB and CUCB models have high γ ’s, which is consistent with the ELBO results. Appendix A.2 presents the posterior distributions of the model parameters. Specific observations about the subjects’ behaviors are as follows.

RESULT 4. Exploration of design space increases with the increase in fixed budget.

As observed in Fig. 8, the mean posterior estimates of the exploration scale α increase with fixed budget between treatments. Note that the mean posterior estimates of α are similar at medium and high budget (treatments T_5 and T_6) in “save-remaining-budget” part of the experiment. This implies that there is a reduction in

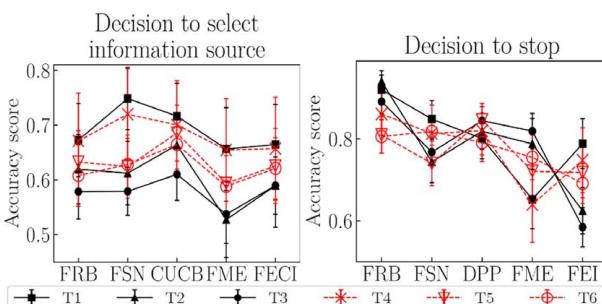


Fig. 7 Predictive accuracy scores of the decision models on the test data

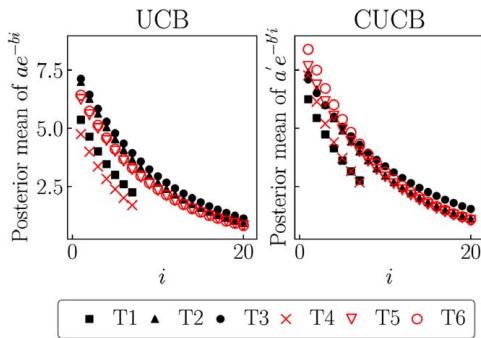


Fig. 8 Mean posterior estimations of the exploration scales α and α' for the decision to choose the next design

exploration when the subjects have an incentive to reduce budget spending and when large savings are possible.

RESULT 5. *There is an increase in the probability of selecting high-fidelity information source as the fixed budget increases, except for medium and high budgets in the “save-remaining budget” part.*

We observe that the posterior distribution of the threshold parameter $w_{m,1}/w_{m,2}$ for the FSN model increases with the increasing fixed budget and decreases with the incentive-to-save budget. These posterior values on average equal the number of high-fidelity information sources provided in Table 2. The subjects choose a single high-fidelity observation after six samples at low-budget treatments (treatments T1 and T4). The average number of high-fidelity observations increases from 2.6 in treatment T2 to 4.2 in treatment T3.

RESULT 6. *The probability of stopping early at high values of remaining budget increases with an increase in fixed budget and with the incentive-to-save budget.*

The result follows from the posterior distribution of model parameters in the FRB model. Figure 9 shows the posterior probability of stopping as a function of the remaining budget from the FRB model. The FRB model's threshold parameter estimate for the “use-it-or-lose-it budget” part (mean posterior $w_{s,1}/w_{s,2} \approx 1$) is smaller than the cost of one high-fidelity observation ($c_L = 2$), which implies that the subjects stop after exhausting almost the entire fixed budget. On the other hand, in “save-remaining budget” part, the mean estimates of $w_{s,1}/w_{s,2}$ increase as the fixed budget increases.

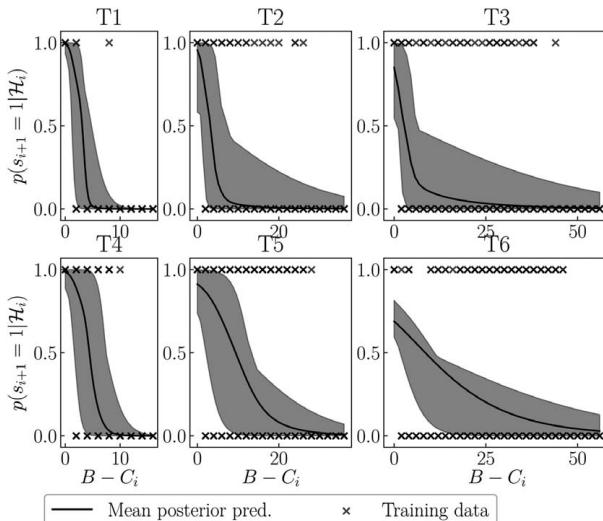


Fig. 9 Mean posterior probability of stopping for all subjects as a function of the remaining budget under the FRB model

6 Discussion

6.1 Accuracy of the Models of Designers' Decisions. The results indicate that the simple heuristic models represent designers' decisions in the sequential information acquisition process more accurately than the expected utility-based models. No single model captures all strategies exactly; however, the heuristic models with the highest ELBOs provide most accurate approximations to the subjects' strategies. As a result of accurate predictions of the information acquisition decisions, the heuristic models also predict the performance more accurately than the expected utility-based models, e.g., EI, FECI, and FEI. To verify the results, we performed 150 simulation runs for the sequential information acquisition process using both a triplet of highest ELBO heuristic models and a triplet of the expected utility-based models. At each iteration i of a run, we quantified the current belief about the design performance using normalization of the highest predictive mean $\max_{j=1 \geq j \leq i} \mu_j(x_j)$. A comparison of the predictions of these quantities with their actual values in the test dataset in Fig. 10 confirms that the highest ELBO heuristic triplet has better predictive strength than the expected-utility based triplet.

The heuristic models remain more likely to represent subjects' decisions if the assumptions about the prior state of knowledge are changed. For example, when Gaussian priors with means 30 and 50 are implemented as the prior state of knowledge instead of the zero mean Gaussian prior in Eq. (4), the ELBO of the CUCB model still remains higher than that of the EI-based model for choosing next the design. The ELBOs of the CUCB model in treatment T1 for means 0, 30, and 50, respectively, are 1783, 1722, and 1771, whereas those of the EI model are 1977, 1826, and 1850. A possible reason for the high ELBO of the CUCB model is the interdependence between the decisions of choosing the next design and choosing an information source. Such interdependence is inevitable as the subjects have few available cues for most of the decisions [33].

6.2 Implications for Engineering Design. On the objective functions in the test dataset, the heuristic triplet model generates higher design performance (gross payoff) than the EU-based triplet model. Figure 11 plots the gross payoff (Eq. (5)) for both the heuristic triplet and the EU-based triplet models. We observe that the EU-based triplet has poorer gross performance, especially in treatment T3. The gross performance of the heuristic triplet model improves with the high fixed budget and without the incentive for saving budget. That is because with the high fixed budget and without the incentive for saving budget, the heuristic model

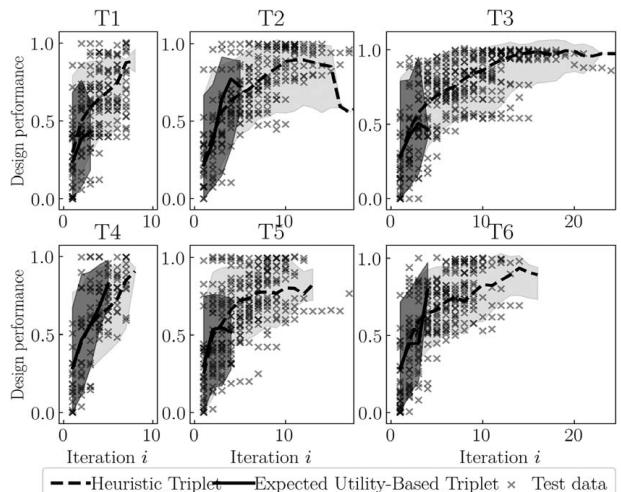


Fig. 10 Normalized design performance quantified as highest predictive mean

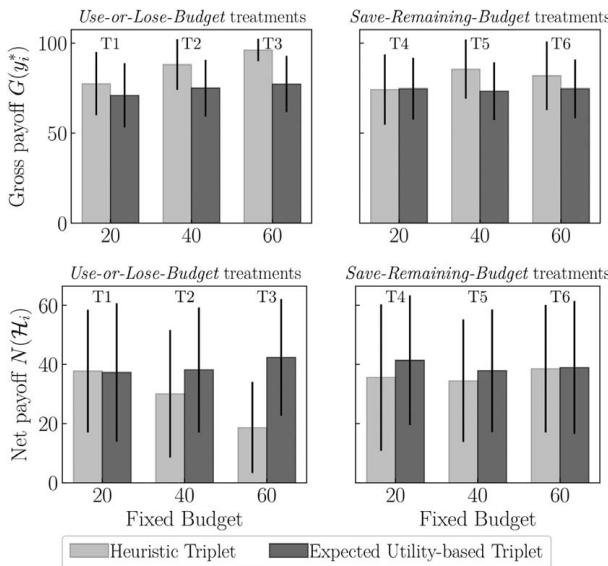


Fig. 11 The gross payoff $G(H_i)$ and the net payoff $N(H_i)$ after stopping for heuristic triplet model and the expected utility-based triplet model. The gross payoff is a function of the best design performance y_i^* calculated according to Eq. (29). The net payoff is the gross payoff minus total cost, i.e., $G(H_i) - C_i$.

triplet completes larger number of iterations, conducts more design exploration, and receives better performance. Figure 12 provides evidence for this explanation. Therefore, if the goal is to maximize the gross design performance, the system designers should allow designers to spend a large fixed budget without any incentive to reduce spending. Under this incentive structure, there is a greater chance of finding the best design that maximizes the performance.

Despite the large gross payoff and the closeness to human design decisions, the heuristic models are less efficient in terms of the net payoff. More iterations of the heuristic triplet model result in higher total cost and therefore reduce the net payoff, i.e., the difference in the achieved performance and the total cost incurred until stopping, as shown in Fig. 11. We observe that the EU-based triplet model provides higher net payoff on average than the heuristic triplet models in all treatments. The difference in average net payoffs from the heuristic triplet model and the EU-based triplet model reduces with the decreasing fixed budget and with the incentive to save budget.

If the goal is to maximize the net payoff, system designers should restrict the amount of fixed budget or implement monetary incentives proportional to the saved budget. The latter option is more viable than the former if the appropriate amount of the

fixed budget cannot be determined. Under monetary incentives for reducing spending, not only are the designers more likely to maximize the net payoff but also their decisions are more likely to be aligned with the expected utility-based models. The prediction accuracy score of the FEI model on the test data is larger in treatment T_6 with the incentive-to-save budget than in treatment T_3 without such an incentive. The same holds true for the training data where the FEI model's accuracy scores in treatments T_3 and T_6 are, respectively, 0.71 ± 0.02 and 0.76 ± 0.024 (p -value < 0.0001). Note that we implemented the incentive-to-save-budget by paying to the subjects the entire remaining budget they saved. However, its effects may likely be obtained by paying a smaller amount proportional to the remaining budget. This is because people have the comparative view of monetary benefits and prefer avoiding losses to acquiring equivalent gains; for them, failing to receive even a small potential benefit is a lost opportunity [9].

The implication for our understanding of human decision-making in engineering design is that *designers may be more attentive to the design performance than the cost of design evaluations or the relative difference in two*. Possible explanations of this observed gap may include high cognitive load associated with processing predictive uncertainty and estimating utility of the next design in relation to the cost of evaluation. It is also likely that the subjects are driven by intrinsic factors such as satisfaction from finding the best design and delivering the best outcomes for a given task. Further research is needed to determine the root causes for this observed trend.

7 Conclusion

In this paper, we present an approach that combines computational modeling and behavioral experiments to quantify designers' decision strategies during the sequential information acquisition process. By using Bayesian inference, we observe that the heuristic models provide the best descriptions of subjects' strategies for making information acquisition decisions. The subjects rely on simple cues accessible via graphical interfaces for making the most of the information acquisition decisions. This reliance on simple cues for making design decisions may be attributed to the relatively smaller cognitive effort involved in using simple cues. Moreover, the subjects' decisions are affected by the amount of fixed budget and incentives to save budget. For example, the subjects select design points close to the highest upper confidence bound (UCB and CUCB models) when seeking to maximum design performance. The subjects mostly select a fixed number of low-fidelity and a fixed number of high-fidelity observations (FSN model) at low budget. At large budget, they query the low-fidelity source for evaluating high uncertainty regions (exploration in CUCB model) and the high-fidelity source for low uncertainty regions (exploitation in CUCB model). For stopping evaluations, the subjects exhaust entire or a fixed fraction of the fixed budget (FRB model), unless they are incentivized to save budget in which case they stop if the current best performance is marginally better than the mean of the predicted performance (DPP model).

The insights from the analysis have implications for engineering design research and practice. With the models that incorporate simple heuristics, researchers can quantify design performance in terms of designers' decision strategies, as illustrated in Fig. 10. The applications of this include design crowdsourcing where game-theoretic models lack design process models [2,40] and the agent-based models of engineering systems design where characterization of quality as a function of designer effort is difficult to achieve [3]. Furthermore, system engineers and managers can set the fixed budget at low values or provide monetary incentives for reducing spending to nudge a designer's decisions toward EU-based strategies, which are efficient for maximizing net payoff (design performance minus cost of evaluation).

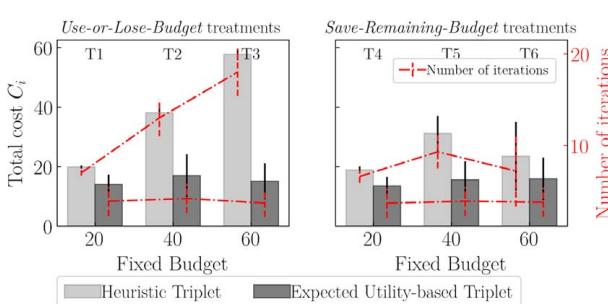


Fig. 12 Total cost C_i and the number of iterations i after stopping for the heuristic triplet model and the expected utility-based triplet model

The methodology used for eliciting decisions and estimating decision strategies is particularly suited for the embodiment phase of the design process. There is a need for further research to establish generalizability across contexts, problems, and populations. Some assumptions require further validation, e.g., Assumption 6, which states based on the existing studies in Refs. [41–43] that the Gaussian processes closely represent human information processing. Because this study utilizes student population and a short-run decision-making process, more empirical evidence is required to establish the generalizability of the results to engineers as designers and long-term design processes. Future descriptive modeling efforts need to account for context-dependent design situations, where decision strategies depend on the availability of problem-specific information or the lack thereof [44], an acceptable quantification of predictive uncertainty is absent [45], and the mapping between resources expended and the value of prototypes created varies across disciplines and knowledge domains [46]. Such design situations should include multiple objectives and/or multidimensional design parameters.

Acknowledgment

The authors gratefully acknowledge the financial support from the US National Science Foundation (NSF) CMMI (Grant No. 1662230).

Appendix

A.1 Design Performance Functions. Design performance functions are shown in Fig. 13.

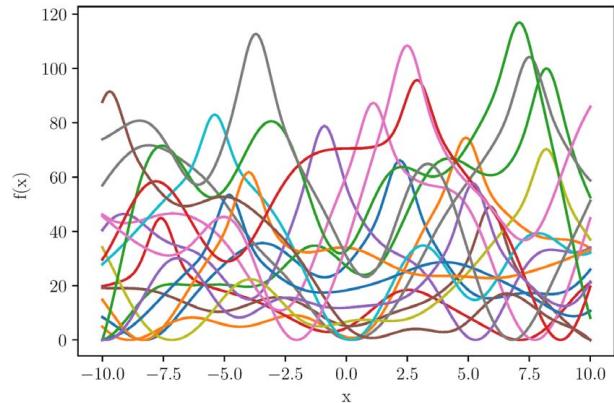


Fig. 13 Eighteen unknown design performance functions used in the experiment

A.2 Posterior Distributions of the Model Parameters. Posterior distributions of the model parameters are presented in Table 4.

Table 4 Mean and standard deviation (in parenthesis) of the posterior distribution for the subject-specific model parameters

Model	Model parameters																	
	Decision to choose design x												Decision to choose an information source					
	γ or γ'						a or a'						b or b'					
Model	$T1$	$T2$	$T3$	$T4$	$T5$	$T6$	$T1$	$T2$	$T3$	$T4$	$T5$	$T6$	$T1$	$T2$	$T3$	$T4$	$T5$	$T6$
UCB	12.87 (1.37)	12.08 (0.65)	10.0 (0.46)	11.18 (2.38)	11.62 (0.96)	10.87 (1.08)	5.37 (0.13)	6.99 (0.14)	7.12 (0.14)	4.75 (0.12)	6.24 (0.14)	6.42 (0.15)	0.15 (0.03)	0.11 (0.02)	0.1 (0.03)	0.17 (0.03)	0.11 (0.03)	0.11 (0.03)
CUCB	12.81 (1.26)	9.47 (0.75)	6.96 (0.45)	12.18 (1.11)	9.58 (0.79)	9.75 (0.84)	5.72 (0.14)	7.21 (0.15)	7.8 (0.16)	6.11 (0.14)	7.14 (0.16)	7.23 (0.16)	0.13 (0.05)	0.12 (0.05)	0.1 (0.05)	0.16 (0.04)	0.12 (0.13)	0.12 (0.05)
PI	1.47 (0.15)	1.58 (0.11)	1.58 (0.1)	1.56 (0.17)	1.32 (0.14)	1.3 (0.14)	—	—	—	—	—	—	—	—	—	—	—	—
EI	1.92 (0.17)	1.9 (0.11)	1.95 (0.11)	1.88 (0.16)	1.72 (0.15)	1.65 (0.13)	—	—	—	—	—	—	—	—	—	—	—	—
ECI	1.4 (0.14)	1.32 (0.1)	1.08 (0.08)	1.35 (0.14)	1.17 (0.12)	1.19 (0.12)	—	—	—	—	—	—	—	—	—	—	—	—
Decision to stop																		
FSN	$w_{m,2}$						$w_{m,1}/w_{m,2}$						—				—	
FRB	0.2 (0.02)	0.07 (0.0)	0.03 (0.0)	0.17 (0.01)	0.06 (0.0)	0.03 (0.0)	2.79 (0.25)	3.22 (0.27)	4.03 (0.29)	2.79 (0.25)	2.66 (0.29)	2.74 (0.3)	—	—	—	—	—	—
FME	1.17 (0.08)	0.95 (0.06)	0.73 (0.05)	1.05 (0.07)	1.03 (0.06)	0.93 (0.06)	0.56 (0.08)	0.83 (0.1)	0.96 (0.11)	0.52 (0.08)	0.65 (0.09)	0.56 (0.09)	—	—	—	—	—	—
FECI	1.93 (0.12)	2.81 (0.13)	3.09 (0.17)	1.69 (0.11)	1.94 (0.11)	1.92 (0.11)	0.34 (0.05)	0.52 (0.03)	0.58 (0.03)	0.34 (0.05)	0.41 (0.05)	0.39 (0.04)	—	—	—	—	—	—
CUCB	0.57 (0.11)	0.18 (0.02)	0.16 (0.02)	0.24 (0.03)	0.23 (0.03)	0.19 (0.02)	1.48 (0.15)	1.15 (0.25)	2.28 (0.29)	6.48 (0.3)	6.05 (0.34)	5.99 (0.39)	—	—	—	—	—	—
Decision to choose an information source																		
FSN	$w_{s,2}$						$w_{s,1}/w_{s,2}$						—				—	
FRB	2.71 (0.33)	1.71 (0.18)	1.06 (0.13)	1.86 (0.2)	0.33 (0.03)	0.1 (0.01)	2.86 (0.19)	3.3 (0.2)	3.29 (0.24)	4.35 (0.19)	9.59 (0.36)	9.01 (0.5)	—	—	—	—	—	—
DPP	0.52 (0.12)	0.25 (0.03)	0.42 (0.04)	0.27 (0.06)	0.41 (0.05)	0.41 (0.19)	—2.2 (0.16)	1.35 (0.11)	2.21 (0.11)	—2.61 (0.23)	—0.74 (0.19)	—0.16 (0.21)	—	—	—	—	—	—
FME	1.16 (0.07)	1.6 (0.06)	1.79 (0.06)	1.03 (0.07)	1.39 (0.07)	1.45 (0.06)	0.41 (0.07)	0.25 (0.04)	0.2 (0.04)	0.44 (0.07)	0.32 (0.05)	0.29 (0.05)	—	—	—	—	—	—
FEI	1.83 (0.17)	1.42 (0.1)	1.01 (0.08)	0.92 (0.07)	1.26 (0.1)	1.52 (0.14)	2.07 (0.11)	1.97 (0.12)	2.05 (0.19)	1.82 (0.12)	1.99 (0.11)	1.93 (0.11)	—	—	—	—	—	—

Note: Columns $T1$, $T2$, $T3$, $T4$, $T5$, and $T6$ denote different experiment treatments.

References

[1] Hazelrigg, G. A., 1998, "A Framework for Decision-Based Engineering Design," *ASME J. Mech. Des.*, **120**(4), pp. 653–658.

[2] Chaudhari, A. M., Sha, Z., and Panchal, J. H., 2018, "Analyzing Participant Behaviors in Design Crowdsourcing Contests Using Causal Inference on Field Data," *ASME J. Mech. Des.*, **140**(9), p. 091401.

[3] Safarkhani, S., Bilionis, I., and Panchal, J. H., 2018, "Understanding the Effect of Task Complexity and Problem-Solving Skills on the Design Performance of Agents in Systems Engineering," 44th Design Automation Conference, Quebec City, Canada, Aug. 26–29, Vol. 2A, p. V02AT03A060.

[4] Hazelrigg, G. A., 1999, "An Axiomatic Framework for Engineering Design," *ASME J. Mech. Des.*, **121**(3), pp. 342–347.

[5] Marston, M., Allen, J. K., and Mistree, F., 2000, "The Decision Support Problem Technique: Integrating Descriptive and Normative Approaches in Decision Based Design," *Eng. Valuation Cost Anal.*, **3**(2), pp. 107–129.

[6] Wassenaar, H. J., and Chen, W., 2003, "An Approach to Decision-Based Design With Discrete Choice Analysis for Demand Modeling," *ASME J. Mech. Des.*, **125**(3), pp. 490–497.

[7] Moore, R. A., Romero, D. A., and Paredis, C. J., 2014, "Value-Based Global Optimization," *ASME J. Mech. Des.*, **136**(4), p. 041003.

[8] Fishburn, P. C., 1988, "Expected Utility: An Anniversary and a New Era," *J. Risk Uncertainty*, **1**(3), pp. 267–283.

[9] Kahneman, D., and Tversky, A., 1979, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, **47**(2), pp. 263–292.

[10] Tebay, R., Atherton, J., and Wearne, S., 1984, "Mechanical Engineering Design Decisions: Instances of Practice Compared With Theory," *Proc. Inst. Mech. Eng., Part B Manage. Eng. Manuf.*, **198**(2), pp. 87–96.

[11] Finger, S., and Dixon, J. R., 1989, "A Review of Research in Mechanical Engineering Design. Part I: Descriptive, Prescriptive, and Computer-Based Models of Design Processes," *Res. Eng. Des.*, **1**(1), pp. 51–67.

[12] Tversky, A., and Kahneman, D., 1992, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *J. Risk Uncertainty*, **5**(4), pp. 297–323.

[13] Gigerenzer, G., and Gaismaier, W., 2011, "Heuristic Decision Making," *Annu. Rev. Psychol.*, **62**, pp. 451–482.

[14] Simon, H. A., 1955, "A Behavioral Model of Rational Choice," *Q. J. Econ.*, **69**(1), p. 99.

[15] Gonzalez, C., Lerch, J. F., and Lebriere, C., 2003, "Instance-Based Learning in Dynamic Decision Making," *Cognit. Sci.*, **27**(4), pp. 591–635.

[16] Panchal, J. H., and Szajnfarber, Z., 2017, "Experiments in Systems Engineering and Design Research," *Syst. Eng.*, **20**(6), pp. 529–541.

[17] Falk, A., and Heckman, J. J., 2009, "Lab Experiments Are a Major Source of Knowledge in the Social Sciences," *Science*, **326**(5952), pp. 535–538.

[18] Simon, H. A., and Newell, A., 1971, "Human Problem Solving: The State of the Theory in 1970," *Am. Psychologist*, **26**(2), p. 145.

[19] Shergadwala, M., Bilionis, I., Kannan, K., and Panchal, J. H., 2018, "Quantifying the Impact of Domain Knowledge and Problem Framing on Sequential Decisions in Engineering Design," *ASME J. Mech. Des.*, **140**(10), p. 101402.

[20] Freriks, H., Heemels, W., Muller, G., and Sandee, J., 2006, "On the Systematic Use of Budget-Based Design," INCOSE International Symposium, Orlando, FL, July 8–14, Vol. 16, Wiley Online Library, pp. 788–803.

[21] Moroşan, P.-D., Bourdais, R., Dumur, D., and Buisson, J., 2010, "Building Temperature Regulation Using a Distributed Model Predictive Control," *Energy Build.*, **42**(9), pp. 1445–1452.

[22] Hor, Y. S., Williams, A. J., Checkelsky, J. G., Roushan, P., Seo, J., Xu, Q., Zandbergen, H. W., Yazdani, A., Ong, N., and Cava, R. J., 2010, "Superconductivity in $Cu_3Bi_2Se_3$ and Its Implications for Pairing in the Undoped Topological Insulator," *Phys. Rev. Lett.*, **104**(5), p. 057001.

[23] Nikolaidis, E., Ghiocei, D. M., and Singhal, S., 2004, *Engineering Design Reliability Handbook*, CRC Press, Boca Raton, FL.

[24] Loch, C. H., Terwiesch, C., and Thomke, S., 2001, "Parallel and Sequential Testing of Design Alternatives," *Manage. Sci.*, **47**(5), pp. 663–678.

[25] Dahan, E., and Mendelson, H., 1997, "Optimal Sequential and Parallel Prototyping Policies," Innovation in Technology Management, The Key to Global Leadership, PICMET, Portland, OR, July 31, IEEE.

[26] Jones, D. R., 2001, "A Taxonomy of Global Optimization Methods Based on Response Surfaces," *J. Global Optim.*, **21**(4), pp. 345–383.

[27] Jones, D. R., Schonlau, M., and Welch, W. J., 1998, "Efficient Global Optimization of Expensive Black-Box Functions," *J. Global Optim.*, **13**(4), pp. 455–492.

[28] Gramacy, R. B., and Lee, H. K. H., 2011, "Optimization Under Unknown Constraints," *Bayesian Statistics* 9, J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West, eds., Oxford University Press, Oxford, pp. 229–256.

[29] Auer, P., 2002, "Using Confidence Bounds for Exploitation-Exploration Trade-Offs," *J. Mach. Learn. Res.*, **3**, pp. 397–422.

[30] Chaudhari, A. M., and Panchal, J. H., 2019, "An Experimental Study of Human Decisions in Sequential Information Acquisition in Design: Impact of Cost and Task Complexity," Research Into Design for a Connected World, Bangalore, India, Vol. 134, pp. 321–332.

[31] Chaudhari, A. M., Bilionis, I., and Panchal, J. H., 2018, "How Do Designers Choose Among Multiple Noisy Information Sources in Engineering Design Optimization? An Experimental Study," 44th Design Automation Conference, Quebec City, Canada, Aug. 26–29, Vol. 2A, p. V02AT03A021.

[32] Panchal, J. H., Sha, Z., and Kannan, K. N., 2017, "Understanding Design Decisions Under Competition Using Games With Information Acquisition and a Behavioral Experiment," *ASME J. Mech. Des.*, **139**(9), p. 091402.

[33] Shafir, E., and Tversky, A., 1992, "Thinking Through Uncertainty: Nonconsequential Reasoning and Choice," *Cognit. Psychol.*, **24**(4), pp. 449–474.

[34] Chen, D. L., Schonger, M., and Wickens, C., 2016, "oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments," *J. Behav. Exp. Finance*, **9**, pp. 88–97.

[35] Strack, F., 1992, "Order Effects in Survey Research: Activation and Information Functions of Preceding Questions," *Context Effects in Social and Psychological Research*, N. Schwarz, and S. Sudman, eds., Springer, New York, pp. 23–34.

[36] Kingma, D. P., and Welling, M., 2013, "Auto-Encoding Variational Bayes," arXiv:1312.6114.

[37] Gelman, A., Hwang, J., and Vehtari, A., 2014, "Understanding Predictive Information Criteria for Bayesian Models," *Stat. Comput.*, **24**(6), pp. 997–1016.

[38] Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., and Blei, D. M., 2017, "Automatic Differentiation Variational Inference," *J. Mach. Learn. Res.*, **18**(1), pp. 430–474.

[39] Salvatier, J., Wiecki, T. V., and Fonnesbeck, C., 2016, "Probabilistic Programming in Python Using PyMC3," *PeerJ Comput. Sci.*, **2**, p. e55.

[40] Sha, Z., Chaudhari, A. M., and Panchal, J. H., 2019, "Modeling Participation Behaviors in Design Crowdsourcing Using a Bipartite Network-Based Approach," *J. Comput. Inf. Sci. Eng.*, **19**(3), p. 031010.

[41] Borji, A., and Itti, L., 2013, "Bayesian Optimization Explains Human Active Search," Advances in Neural Information Processing Systems, Lake Tahoe, NV, Dec. 5–10, pp. 55–63.

[42] Griffiths, T. L., Lucas, C., Williams, J., and Kalish, M. L., 2009, "Modeling Human Function Learning With Gaussian Processes," *Advances in Neural Information Processing Systems*, Vancouver, BC, pp. 553–560.

[43] Lucas, C. G., Griffiths, T. L., Williams, J. J., and Kalish, M. L., 2015, "A Rational Model of Function Learning," *Psychon. Bull. Rev.*, **22**(5), pp. 1193–1215.

[44] Binder, W. R., and Paredis, C. J., 2017, "Optimization Under Uncertainty Versus Algebraic Heuristics: A Research Method for Comparing Computational Design Methods," 43rd Design Automation Conference, Cleveland, OH, Aug. 6–9, Vol. 2B, p. V02BT03A057.

[45] Sissoko, T. M., Jankovic, M., Paredis, C. J. J., and Landel, E., 2018, "An Empirical Study of a Decision-Making Process Supported by Simulation in the Automotive Industry," 30th International Conference on Design Theory and Methodology, Quebec City, Canada, Aug. 26–29.

[46] Tiong, E., Seow, O., Camburn, B., Teo, K., Silva, A., Wood, K. L., Jensen, D. D., and Yang, M. C., 2018, "The Economics and Dimensionality of Design Prototyping: Value, Time, Cost and Fidelity," *ASME J. Mech. Des.*, **143**(3), p. 031105.