

DETC2018-85460

HOW DO DESIGNERS CHOOSE AMONG MULTIPLE NOISY INFORMATION SOURCES IN ENGINEERING DESIGN OPTIMIZATION? AN EXPERIMENTAL STUDY

Ashish M. Chaudhari *

Graduate Research Assistant
School of Mechanical Engineering
Purdue University
West Lafayette, IN

Ilias Billionis

Assistant Professor
School of Mechanical Engineering
Purdue University
West Lafayette, IN

Jitesh H. Panchal

Associate Professor
School of Mechanical Engineering
Purdue University
West Lafayette, IN

ABSTRACT

Designers make process-level decisions to (i) select designs for performance evaluation, (ii) select information source, and (iii) decide whether to stop design exploration. These decisions are influenced by problem-related factors, such as costs and uncertainty in information sources, and budget constraints for design evaluations. The objective of this paper is to analyze individuals' strategies for making process-level decisions under the availability of noisy information sources of different cost and uncertainty, and limited budget. Our approach involves a) conducting a behavioral experiment with an engineering optimization task to collect data on subjects' decision strategies, b) eliciting their decision strategies using a survey, and c) performing a descriptive analysis to compare elicited strategies and observations from the data. We observe that subjects use specific criteria such as fixed values of attributes, highest prediction of performance, highest uncertainty in performance, and attribute thresholds when making decisions of interest. When subjects have higher budget, they are less likely to evaluate points having highest prediction of performance, and more likely to evaluate points having highest uncertainty in performance. Further, subjects conduct expensive evaluations even when their decisions have not sufficiently converged to the region of maximum performance in the design space and improvements from additional cheap evaluations are large. The implications of the results in identifying deviations from optimal strategies and structuring decisions for further model development are discussed.

1 INTRODUCTION

In decision-based design, the engineering design process is perceived as a decision making process where different design options are identified, options are evaluated sequentially, and the design option with highest performance is selected [1]. Under this framework, designers make process-level decisions such as 1) selecting a design at every iteration to evaluate performance, 2) selecting a model or information source for performance evaluation, and 3) deciding whether the best performance is achieved.

External factors related to the design problem, budget constraints, and incentives to designers influence process-level decisions. These decisions are inherently dependent upon problem-specific factors such as uncertainty in performance, and costs of evaluation for different information sources. The number of design evaluations are limited by budget constraints [2,3]. In such scenarios, the objective of maximizing utility is satisfied by maximizing the profit from the final design. The higher the resources spent on evaluations to get the same final performance, the lower is the profit. Therefore, design decisions are made under the constraints of limited budget and a decision maker's goal is to maximize the design performance while minimizing the cost of design evaluations.

Many theoretical studies have proposed optimal strategies for the above decisions [2,4,5]. However, there is a lack of understanding about what strategies individuals actually follow in the design process. Some strategies are computationally intensive, and therefore unlikely to represent actual human decisions. Also, there is a lack of understanding about how individuals' strategies

*Corresponding author: chaudha5@purdue.edu

change when the budget is fixed, or designers' payoff is tied to their performance and how much budget they save. Towards addressing this gap, our objective in this paper is *to understand what strategies individuals use to make design decisions with information sources of different cost and noise, and how budget constraints affect these decisions.*

1.1 Relevant Literature

Numerous optimization methods for optimal design have been developed in literature [6, 7, 8]. In recent studies in design literature, optimal strategies based on meta-modeling of performance observations have been proposed [4, 9, 10, 5, 11]. The optimal decisions are defined as picking alternatives in design space that maximize the expectation of payoff. However, human decision making literature has found that thinking ahead to predict expectation of utilities is a cognitively difficult task [12, 13]. Humans may be unable to follow normative models of optimal strategies due to their computationally intensive nature.

Prior work on human strategies in design has compared human subjects' decisions against proposed rational strategies based on the expected utility theory and game theory [14, 15]. Studies have estimated human strategies in terms of the parameters of Bayesian optimization (BO). They found that the estimated strategies closely represent actual decisions [16] and can provide better outcomes than default BO algorithms [17].

An alternative approach to studying human decision making is to analyze decision criteria or heuristic rules that humans use to make decisions based on information processing [18, 19]. This approach conceptualizes humans as information processors who search for information and integrate acquired information to identify problem structure. Based on the inference from the acquired information, humans use simple rules to select search strategy and decide when to stop information search [20]. In the literature, models of explicit rules used by humans to stop acquiring information have been studied. Common stopping rules identified are stopping on the basis of the first cue found (take-the-best), list of criteria satisfied (mental List), the key amount accumulated (magnitude threshold), and problem structure identified (representational stability) [21, 12]. Further, common rules for choosing among alternatives are to choose the best alternative (dominance rule), exclude alternatives that do not satisfy a criterion (elimination by aspects rule), select the alternative which has the highest utility in all attributes (addition of utilities rule), etc. [22].

1.2 Approach

In this study, we employ a behavioral experiment to assess the effects of budget constraints on designers' decisions. The behavioral experiment provides a way to control external factors that may influence designers' decisions [23]. In an experimental task, various external influences can be mitigated while main-

taining realism by using humans as designers and real money as incentives [24]. Based on this methodology, we (i) collect data on human subjects' decision strategies using a behavioral experiment with engineering optimization task, (ii) elicit their decision strategies using a survey in the experiment, and (iii) perform a descriptive analysis comparing elicited strategies with observations from the experimental data.

The experimental task used in the behavioral experiment simulates iterative evaluations in engineering design optimization in a manner consistent with the decision-based design process. In this task, designers have control over different decisions and strategies they want to use. We control how each individual processes the acquired information by showing them acquired samples and corresponding surrogate models at each instance they make a decision. The task involves a simulated problem, and not a real problem for the reasons that: 1) subjects' knowledge of the problem does not interfere with the experimental task, 2) design space is well defined, and 3) design costs, uncertainty, and budget can be controlled to observe their effects on subjects' decisions. The experimental task is still a design optimization task because finding the best solution requires evaluations of different designs, assessing their values, and deciding the best design [1]. Also, despite a well-defined design space, the best design is not known until an information search, processing of the acquired information, and stopping are completed [19].

The rest of the paper is outlined as follow. Section 2 details the experimental task used in the behavioral experiment and its underlying assumptions. Section 3 describes the behavioral experiment which uses the experimental task. Section 4 lists strategies subjects used and corresponding decision models identified. Section 5 present descriptive analysis of the experimental data using decision models from Section 4. Section 6 summarizes the conclusions from this study.

2 AN EXPERIMENTAL TASK REPRESENTING TYPICAL DESIGN OPTIMIZATION PROBLEMS

The experimental task represents a hypothetical design scenario of a *designer* who makes design decisions, and a *statistician* who develops a surrogate model for processing the acquired information. The designer receives the payment as a function of outcomes of her decisions. The roles of designer and statistician are separated to maintain uniformity in how all designers process acquired information. Also, in the actual experiment, a subject acts as a designer while the back-end of the user interface acts as a statistician.

2.1 The Designer

Consider a designer designing a component by evaluating performance at various points in the design space. The designer's objective is to find a design that maximizes the performance.

We represent the design space as $\mathcal{X}$ and a point in the design space as x . Let denote the actual performance at x by $f(x)$ which is unknown to the designer. The designer's task is to find $x^* = \arg \max_{x \in \mathcal{X}} f(x)$.

Assumption 1 (Continuous design space) The design performance $f(x)$ is a one-dimensional continuous function of a single design parameter x . The design space $\mathcal{X}$ is a finite interval, and the designer can evaluate $f(x)$ at any point $x \in \mathcal{X}$.

Assumption 2 (Sequential design process) We assume that the designer evaluates multiple designs sequentially. Each evaluation takes one unit of time to run. However, the designer may take varying units of time to choose design point x in the design space. But, once a design is chosen to be evaluated, the designer may not begin another evaluation until the previous evaluation is complete.

Assumption 3 (Multiple Noisy Information Sources) The designer may evaluate performance at any design using either *cheap simulations*, or *expensive physical prototypes*. We refer to them as information sources. We assume that both information sources provide noisy observations. In simulations, observations may be noisy due to approximations such as discretization of design space, computational limitations, and errors from theoretical inadequacy. In physical prototypes, the noise may arise from manufacturing defects, or machining tolerances when preparing test specimen. Therefore, the evaluation at design point x generates an observation y which may be different from the true $f(x)$. We assume that observation y is distributed with fixed variance about $f(x)$.

The observation from information sources are distributed as a normal distribution about the true performance function.

$$y_a|x \sim \mathcal{N}(f(x), \sigma_a^2) \quad (1)$$

where y_a is an observation at design point x from information source a . The variance of observation σ_a^2 is known to the designer. Let us assume that the cheap simulations are denoted with $a = l$, and the expensive prototypes with $a = h$. σ_l^2 and σ_h^2 are respective noise variances. By the assumption, $\sigma_l^2 > \sigma_h^2$.

Assumption 4 (Known costs) The evaluation of the performance of any design costs a fixed amount for either information source known apriori to the designer. The total cost is the sum of costs for all evaluations. If c_l and c_h are the costs of a simulation and a physical prototype respectively, then $c_l < c_h$.

2.2 The Statistician

For reasons such as evaluations are expensive, time consuming, and performance function is rugged, a statistician accompanies the designer whose task is to characterize a meta-model of the true performance function. At the time of any evaluation, the statistician co-worker has access to all previous de-

sign points, observations, and noisy information sources. After $n \in \mathbb{N}^+$ evaluations, n_l is the number of observations from cheap simulations and n_h is the number of observations from expensive prototypes, such that $n = n_l + n_h$. We assume that designs $x_a = \{x_a^{(1)}, x_a^{(2)}, \dots, x_a^{(n_a)}\}$ and corresponding observations $y_a = \{y_a^{(1)}, y_a^{(2)}, \dots, y_a^{(n_a)}\}$ are known for both information sources $a = l$ and $a = h$.

Since the engineer and the statistician do not know $f(x)$, we assume that they assign zero mean Gaussian prior on $f(x)$,

$$f \sim GP(0, k) \quad (2)$$

where $k(x, x') = v \exp\left\{-\frac{1}{2} \frac{(x - x')^2}{l^2}\right\}$, is a radial basis covariance function with parameters lengthscale l and variance v .

The predictive probability that y is the true performance at x conditional on all of the available data is

$$p(y|x, \mathcal{D}) = \mathcal{N}(y|\mu_n(x), \sigma_n^2(x)), \quad (3)$$

where $\mathcal{D} = \{(x_l, y_l), (x_h, y_h)\}$. $\mu(x) = k(x, X)(K + \Sigma)^{-1}Y$ and $\sigma^2(x) = k(x, x) - k(x, X)(K + \Sigma)^{-1}k(X, x)$ are the predictive mean and predictive variance functions respectively. X and Y represent the matrices containing simulations and prototype inputs and function outputs, respectively, $k(x, X)$ is the cross-covariance between x and X , K is the covariance at X , and Σ is a diagonal matrix with the first n_l elements of the diagonal equal to σ_l^2 , and the remaining n_h elements equal to σ_h^2 . For example, the predictive distribution of y over all x 's, given $l = 2$, $v = 600$, $\sigma_l = 10$, and $\sigma_h = 0.1$, is shown in Figure 1 in terms of the predictive mean and 95th percentile bounds.

Assumption 5 (Surrogate model) The statistician provides a surrogate model to the designer in terms of the mean estimate of true performance, and the 5th and 95th percentile bounds along with it. It is assumed that the designer understands this information.

2.3 Performance-based Payment

Assumption 6 (Fixed budget) The total budget B assigned to the designer for performance evaluations is fixed. The designer can spend at most B on all evaluations. The designer may stop before exhausting the entire budget B . Therefore, the total cost incurred, $b_n = c_l n_l + c_h n_h$, is less than B .

The designer receives a fixed salary for her effort. Additionally, she may receive bonuses proportional to her best physical prototype, and the amount of budget she saved. At least one physical prototype is required because the noise in simulations is large but the noise in physical prototypes is small.

Assumption 7 (Performance-based payment) The engineer's payment is performance-based that includes a fixed salary,

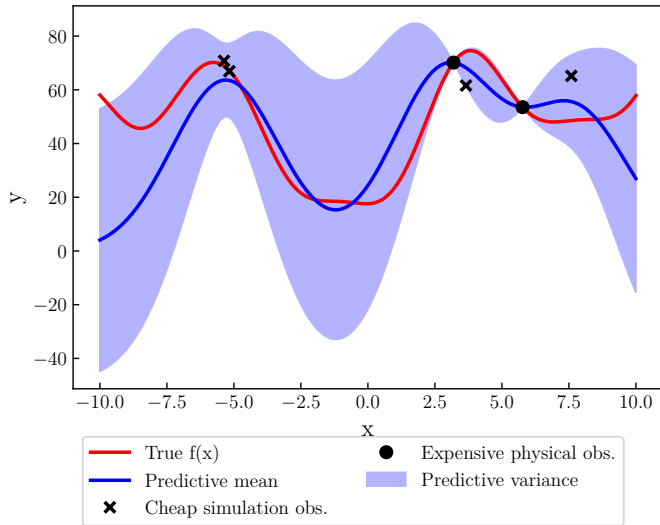


FIGURE 1. The predictive distribution of function values given observations from cheap simulations and expensive physical prototypes. Assumed that $l = 2$, $v = 600$, $\sigma_l = 10$, and $\sigma_h = 0.1$.

and bonuses based on her best performance and the budget she saves. If she stops after n observations with observations $\mathcal{D}$ where $n = \{1, 2, 3, \dots\}$, then her payment I_n is:

$$I_n(\mathcal{D}, B) = I_0 + \delta(n_h \geq 1) \{H_f(y_h^*) + H_b(B - b_n)\} \quad (4)$$

where I_0 is a fixed salary. H_f and H_b , respectively, are bonuses from the best performance and remaining budget. $y_h^* = \max_{1 \leq i \leq n_h} y_h^{(i)}$ being the best physical prototype. And, $\delta(n_h \geq 1)$ is the Kronecker delta function which is 1 if $n_h \geq 1$ and 0 otherwise.

3 BEHAVIORAL EXPERIMENT TO GATHER DATA ABOUT SUBJECTS' DECISION STRATEGIES

A total of 63 student subjects were recruited from the introductory machine design class (ME352: Machine Design I) in the School of Mechanical Engineering, Purdue University. The participation was voluntary and was not considered towards students' grades. All subjects were asked to stay until at least 45 minutes even if they finished early so that other students completing their tasks were not disturbed. Subjects were paid \$15 on average.

3.1 Tasks and Treatments

In the experiment, the subjects performed 18 runs of the experimental task described in Section 2 each with different performance function $f(x)$. A run of the experimental task was called

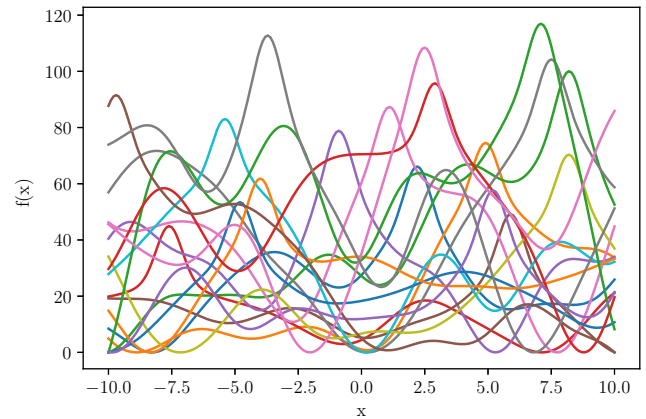


FIGURE 2. The plot of 18 unknown functions used in the experiment.

a period. The objective in each period was to find a maximum of the unknown function $f(x)$. Each period involved multiple sequential iterations, and at each iteration, every subject was required to 1) choose a value for a design parameter x , 2) select a model to evaluate $f(x)$ at x , i.e., either simulation (model l) or physical prototype (model h), and 3) decide whether to stop evaluations. Some parameters of the experimental task were fixed at $c_l = 2$, $c_h = 8$, $\sigma_l = 10$, $\sigma_h = 0$, and $\mathcal{X} = [-10, 10]$.

18 distinct functions were used in the periods (shown in Figure 2). They were randomly generated prior to the experiment and fixed for all subjects. The order in which these 18 functions were assigned to different periods was randomized to minimize interactions between functions and treatments.

The experiment was divided into three treatments¹.

1. *Trial* (2 periods): The first treatment involved two trial periods to help the subjects get familiar with the user interface before starting the actual experiment. The outcomes of these functions were not considered towards subjects' payment.
2. *Use-or-lose budget* (9 periods): For this treatment, a subject was allocated a fixed budget per period. Any remaining budget after evaluations were stopped was discarded and not added to the subject's payment. In this treatment, a subject evaluated 9 different unknown functions in 9 different periods, with 3 functions each for three categories of fixed budget (B) per period equal to 20, 40 and 60.
3. *Save-remaining budget* (9 periods): In this treatment, a subject evaluated 9 different unknown functions in 9 different periods, with 3 functions each for the three categories of fixed budget per period equal to 20, 40 and 60. Any re-

¹We have also gathered data about subjects' risk preferences in an additional treatment using a set of lottery questions. Its analysis, however, will be reported elsewhere.

maintaining budget after evaluations were stopped was added to the subject's payment.

At the end of the above three treatments, the subjects were presented with a survey on their computer screens. In the survey, the subjects gave responses to three questions asking them to list strategy or strategies they used when 1) choosing next x , b) choosing between cheap simulations and expensive prototypes, and 3) deciding to stop function evaluations.

The order of treatments was varied across the subjects to adjust for *order effects*. After the trial treatment, use-or-lose and save-remaining budget treatments were conducted in four different orders: 1) *Use-or-lose*: budgets 20, 40, 60; *Save-remaining*: budgets of 20, 40, 60; 2) *Use-or-lose*: budgets of 60, 40, 20; *Save-remaining*: budgets of 60, 40, 20; 3) *Save-remaining*: budgets of 20, 40, 60; *Use-or-lose*: budgets of 20, 40, 60; 4) *Save-remaining*: budgets of 60, 40, 20; *Use-or-lose*: budgets of 60, 40, 20.

3.2 Payment Rule

The payment was calculated according to Eq. 4. The fixed salary for every subject was 5 USD ($I_o = 5$). A subject's performance in a period was defined as a function of his/her best physical prototype in that period, called *gross payoff*. If y_n^* was the best observation in physical prototypes, then the gross payoff H_f was calculated as,

$$H_f(y_n^*) = 100 - (f_{max} - y_n^*) \quad (5)$$

where f_{max} was the true maximum value of a given unknown function. The gross payoff was revealed after all evaluations were ended for the period. For periods in the save-remaining treatment, any remaining budget was added to the gross payoff such that $H_b = B - b_n$. For periods in the use-or-lose budget treatment, $H_b = 0$.

A subject's bonus payment, additional to the fixed salary, was his/her net payoffs ($H_f(y_n^*) + H_b$) in two randomly selected periods, one each in use-or-lose budget and save-remaining budget treatments. The payment was based on random periods to encourage subjects to put their best effort in every period. Also, the payment was revealed at the end of the experiment, mandating participation in all parts to receive any payment. This rule minimized effects on future effort due to winnings from previous periods called *wealth effects*. It also reduced the *selection bias* that could discourage participation in future treatments once the payment from previous treatments was received.

4 DECISION MODELS IDENTIFIED FROM THE SURVEY RESPONSES

For a comparison to subjects' actual strategies, we first provide a brief overview of optimal strategies of designers' decisions

described in past literature.

4.1 Optimal Strategies for Designers' Decisions

The optimal strategies specify decisions that maximize the expectation of payoff. For the decision to choose next x , the optimal strategy is to select x where the expected improvement (EI) in the maximum of predictive mean is the highest [25, 10]. This strategy is however independent of fixed budget and encodes next x only based on previous observations. For the decision to choose information source, the optimal strategy is to pick the source which has the highest *value of information* (maximum of EI in the current function maximum minus the cost of additional evaluation using the source) [4]. For our case, both information sources have the true performance as mean (Eq. 1). As a result, the EI from an additional evaluation using either of the sources is the same. Therefore, the optimal strategy is to only select cheap simulations, and perform one expensive prototype at the end when the EI is small. For the decision to stop, the optimal strategy is to stop when the value of information from an additional evaluation is negative for either of the information sources [4, 14].

4.2 Decision Models and Corresponding Survey Responses

In this section, we identify decision models for the three decisions of interest from the survey responses.

For the **decision to choose next x** , few examples of responses to the survey which asked subjects to list their strategies are as follows: (1) "I (chose x in) intervals of 2 and then built a physical prototype on the highest peak.", (2) "-5,0,5 were usually always my first guesses to get a general shape of the entire graph..." (3) "I sampled 5 points -8 -4 0 4 8 for fasted relatively detailed general view.", and 4) "I picked x values around large uncertainty boundaries to reduce them".

Based on the survey responses, we identify three decision models for the decision to choose x . Subjects, in reality, may use a combination of the following strategies, as evident from some of their responses. However, we identify distinct rules that may be common.

1. *Fixed Interval Model*: According to this model, points are sampled at fixed intervals across the design space.
2. *Dominant Prediction Model*: According to this model, points are sampled where the prediction of the function is the highest. That is, the next x is selected where the predictive mean function $\mu_n(x)$ (Eq. 3) is maximum during that sample.
3. *Highest Uncertainty Model*: In this model, the next design is evaluated at the design point where the prediction of uncertainty is the highest. The next point is chosen at x where the predictive variance function $\sigma_n^2(x)$ (Eq. 3) is maximum.

For the **decision to choose an information source**, examples of survey responses are: (1) “I would fill in some gaps ..., until I had 8 or 16 or 24 currency left to make prototypes.”, (2) “...once high values were found, I built prototypes...”, (3) “First run (a simulation) until I had an image of what the function may look like—Once I knew what the image might look like I guessed with (a prototype) where I thought the maximum of the function would be”, and (4) “Once the computer simulations kept resulting in consistent values, I picked the one that appeared to be the highest along the predicted graph to find x .”

From the survey responses, we identify three decision models for choosing an information source. In all three models, we assume that subjects begin using cheap simulations, and shift to using expensive prototypes.

1. *Fixed Sample Model*: According to this model, cheap simulations are used for a fixed number of samples, expensive physical prototypes are used thereafter. The number of samples for both simulations and prototypes are fixed before the search begins.
2. *Fixed Maxima-Region Entropy Model*: In this model, we assume that subjects use the acquired information to decide whether the region of function maximum has been sufficiently identified. Once the information entropy of such region reduces to a fixed value, subjects start evaluations using physical prototypes. Evidently, subjects in survey responses mention that they use physical prototype when “high values are found” or “it appeared that the maximum is around”.
3. *Fixed Expected Improvement Model*: In this model, cheap simulations are used for initial samples. With more number of cheap simulations, the expectation of improvement in the current best observation reduces. When this expectation reduces to some fixed values, expensive prototypes are used thereafter.

For the **decision to stop evaluations**, common survey responses are: (1) “I stopped after I saw the (best performance) go up to a high number or stop changing when I added more prototypes.”, (2) “I stopped evaluations when my physical prototype seemed to reach the local maximum of the area I was searching.”, (3) “(I stopped when) marginal gain from cost of new prototype to potential gain in the final result (was low.)” (4) “When saving budget however I tried to get the most out of about half the budget”

We identify the following four decision models for the decision to stop evaluations based on survey responses.

1. *Dominant Prototype Model*: Evaluations are stopped if the best observation from physical prototypes is higher than the predictive mean over the entire design space, i.e., stop if $y_n^* > \max_x \mu_n(x)$ (Eq. 3).
2. *Constant Maxima-Region Entropy Model*: Here, evaluations are stopped when the region of maximum is sufficiently

identified. Specifically, when the information entropy of the region of function maximum predicted by the surrogate model reduces to a constant value, i.e., evaluations are stopped when additional evaluations are not expected to reduce entropy further.

3. *Fixed Budget Model*: In this model, evaluations are stopped after spending a fixed amount. The amount to spend is a priori fixed and can be less than the total budget.
4. *Constant Expected Improvement Model*: In this model, when the predictive mean function (Eq. 3) is not expected to improve further by additional samples, evaluations are stopped.

5 DESCRIPTIVE ANALYSIS OF THE EXPERIMENTAL DATA

In this section, we analyze how decisions of all subjects in aggregate are affected by the total fixed budget, and incentive to save budget. For this, we quantify all subjects’ decisions in different treatments into metrics that specify closeness of their decisions to different models identified in Section 4. The effects of interest are observed by analyzing the variation in values of closeness metrics across different treatments. The findings of this descriptive analysis are summarized in Table 1.

Mathematically, the decision to choose next x after n^{th} evaluations is quantified by $x^{(n+1)} \in \mathcal{X}$, the decision to choose information source by a_{n+1} which is 1 if expensive prototype is selected and 0 otherwise, and decision to stop by s_n which is 1 if stopped, and 0 otherwise. Each decision model depends upon a single criterion such as information entropy, fixed number of samples, expected improvement, etc. Let us denote this criterion for each decision model by statistic M . To observe the effect of budget and incentives, we need to observe how $x^{(n+1)}$, a_{n+1} , and s_n change as M is changed in different treatments. Therefore, for every model, we categorize M into different intervals and estimate the values of closeness metrics for each interval of M . The closeness metrics for different decisions are defined as follows.

1. *Closeness ratio (z_M)*: If a model describes choosing x where statistic $M(x)$ is maximum, and a subject’s choice of x at sample number n is $x^{(n+1)}$, then the closeness of the subject’s decisions to the model is $z_M(x^{(n+1)}) = \frac{M(x^{(n+1)})}{\max_x M(x)}$.
2. *Probability of choosing expensive prototype (p_H)*: a_{n+1} is assumed to have a Bernoulli distribution with success probability equal to the probability of choosing physical prototype p_H . For any interval of M , multiple data points for a_{n+1} are available from decisions of multiple subjects in multiple periods. Therefore, we use *Wilson Score Interval* [26] to estimate the mean and error bounds for p_H in every interval of M .
3. *Probability of stopping (p_S)*: The decision to stop s_n is

TABLE 1. The decision models of subjects' strategies based on survey responses, and corresponding observations from the descriptive analysis.

Model of strategy	# Subjects who indicated this model	Key observations based on subjects' decisions	Effect of increasing total budget	Effect of saving remaining budget
<i>Decision to choose x</i>				
a) Fixed Interval Model	43	Common partitions for the first 5 samples: $(-10, -9, -7.5, -5, -2.5, 0, 2.5, 5, 7.5, 10)$	Intervals are finer.	Intervals are coarse.
b) Dominant Prediction Model	17	Used in late stages of evaluations.	Dominant predictions are chosen with lesser probability.	Dominant predictions are chosen with higher probability only at budget 60.
c) Highest Uncertainty Model	7	Used in early stages of search.	Highest uncertainty points are chosen with more probability.	Unobserved due to high variance in observations.
<i>Decision to choose between two information sources</i>				
a) Fixed Sample Model	10	Common sequences for a) budget 20: $(6l, 1h)$, b) budget 40: $(9l, 1h)$, c) budget 60: $(10l, 5h)$.	The fixed number for both simulations and prototypes increases.	Subjects use more simulations, and less prototypes.
b) Fixed Maxima-Region Entropy Model	21	Fixed entropy for a) budget 20: $[0, 2)$, b) budgets 40 & 60: $[0, 1)$	Prototypes are chosen with higher probability.	At a given budget, prototypes are chosen with lower probability.
d) Fixed Expected Improvement (EI) Rule	4	Fixed value not observed	The probability of using an expensive prototype increases.	The prob. of choosing prototype decreases, but only at budget 60.
<i>Decision to stop design evaluations</i>				
a) Dominant Prototype Model	10	Common differences for a) budget 20 and 40: $[-20, 0)$, b) budget 60: $[0, 20]$.	The difference (best prototype minus maximum of predictive mean) at stopping increases.	The difference reduces.
b) Constant Maxima-Region Entropy	13	Constant entropy value < 1	The probability of stopping at any entropy reduces.	No effects observed.
c) Fixed Budget Model	5	Fixed values for a) budget 20: 20, b) budget 40: $(20, 40]$, c) budget 60: $[20, 40)$ or $(40, 50]$.	No effect when using the entire budget. The probability of stopping increases.	The probability of stopping increases, but only for budgets 40 & 60.
d) Constant Expected Improvement Model	9	Constant EI value < 2	The probability of stopping reduces.	The probability of stopping increases only for budgets 40 & 60.

modeled as a Bernoulli distribution with success probability equal to the probability of stopping (p_S). Also, for any interval of M , multiple data points for s_n are available from decisions of multiple subjects and multiple periods. Therefore, we estimate a confidence interval for p_S from these decisions for every category of M using the *Wilson Score Interval* [26].

Overall, each subject sampled 165 points on average in all 18 periods. Therefore, for 63 subjects, a total of about 10395 data points on every decision are available. In the next section, we compare the estimations of M for each decision model are different in different treatments.

5.1 Decision to Choose x

5.1.1 Fixed Interval Model From the observation of Figure 3, the subjects sample most number of times at $(x =$

$0, 5, -5, -10$ in decreasing frequency for the 1st sample. They sample at $5, -5, 0$ for the 2nd sample in decreasing frequency, and at $0, -5, 5$ for the 3rd sample in decreasing frequency. The distribution of sampled point at the 5th sample is closer to the uniform distribution. Out of the 12 possible sequences, the following 5 sequences are more likely for the first three samples: $(0, 5, -5)$, $(-5, 5, 0)$, $(-10, 5, 0)$, $(-10, 5, -5)$, $(-10, -5, 0)$, since 0 has a small frequency to appear on the 2nd sample, and 5 has a small frequency to appear on the 3rd sample. Some subjects in the survey mentioned that they use finer intervals when the fixed budget is higher. This trend is not observed in Figure 3.

5.1.2 Dominant Prediction Model Here, $M(x) = \mu_n(x)$, the prediction mean function over the design space $\mathcal{X}$. As observed from Figure 4, subjects increasingly sample points closer to the highest predictive mean during later stages of evaluations. The mean and median values of z_{μ_n} increase as the num-

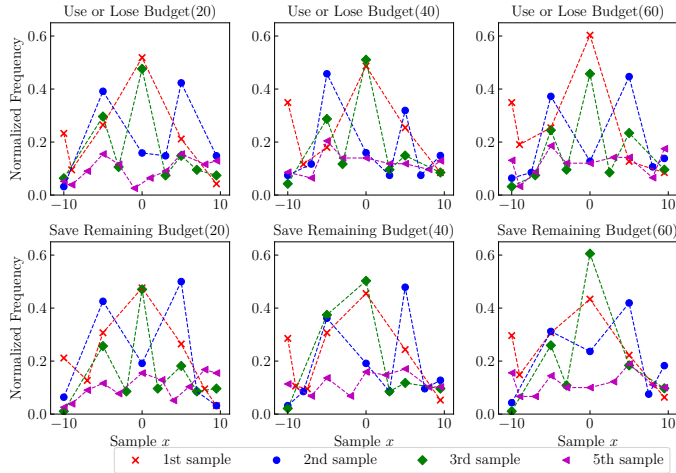


FIGURE 3. Dominant peaks in frequency of subjects' choices of x at t^{th} sample for $t = 1, 2, 3$, and 5 (Fixed Interval Model). The number of (evenly-spaced) bins used is 20. Frequencies are normalized with respect to the number of total samples. Peaks with frequency less than 10^{th} fraction of the maximum frequency are ignored.

ber of samples n increases. Further, the fixed budget affects the above behavior. In the *Use-or-lose Budget* treatments, subjects begin using this model early as the fixed budget available decreases. For example in Figure 4, compare the mean value of z_{μ_n} (equal to 0.8) at interval $n \in (6, 7)$ for total budget 20, and that of (equal to 0.55) at the same interval for total budget 40. In the *Save-remaining Budget* treatments, similar trends are observed except for between the fixed budgets of 40 and 60 (compare at $n \in (10, 11)$). This may be because subjects stop early in *Save-remaining Budget (60)* treatment in expectation of saving larger budget.

The group of subjects who indicated in the survey that they used this strategy have higher z_{μ_n} values than the aggregate of all subjects.

5.1.3 Highest Uncertainty Model For this model, $M(x) = \sigma_n^2(x)$, the predictive variance function. From Figure 5, we observe that subjects sample near the point of highest predictive variance during the initial stages of search but move away from this strategy as more samples are accumulated². Subjects use this model more frequently when the fixed budget is higher. For instance, in the *Use-or-lose Budget* treatments, the mean of $z_{\sigma_n^2}$ at $n \in (6, 7)$ increases as the fixed budget increases. This trend is observed for intervals $(10, 11)$ and $(14, 15)$ of n between total

²Ignore intervals $(10, 11)$ and $(14, 15)$ in the treatments of total budget 20 and 40 respectively. Subjects who sampled this many points could do so because they did not sample any physical observations. They are included in the plot to convey the existence of such data points.

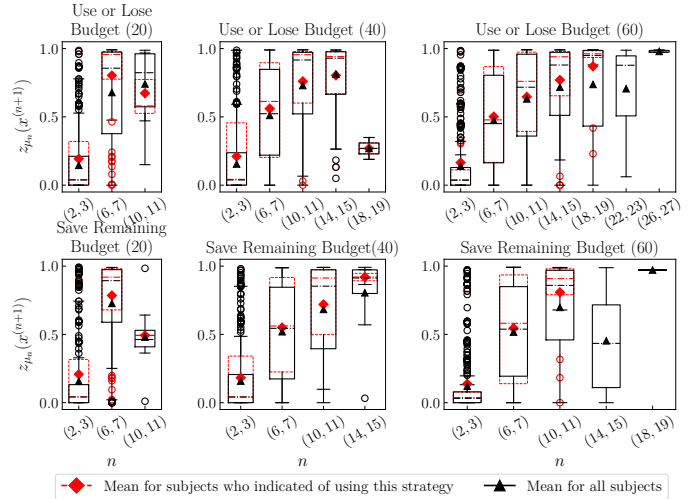


FIGURE 4. The estimates of closeness ratio z_{μ_n} for Dominant Prediction Model at different intervals of sample number n .

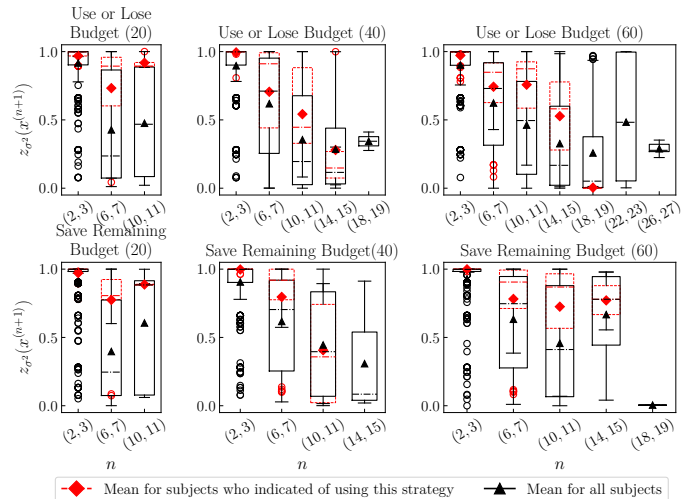


FIGURE 5. The estimates of closeness ratio $z_{\sigma_n^2}$ for Highest Uncertainty Model at different intervals of sample number n .

budgets of 40 and 60. In the *Save-remaining Budget* treatments, the mean of $z_{\sigma_n^2}$ increases for interval $(6, 7)$ between budgets 20 and 40. But, this trend cannot be observed between budgets 40 and 60 due to large variance of $z_{\sigma_n^2}$ values.

5.2 Decision to Choose a Information Source

5.2.1 Fixed Sample Model

For this model, $M = n$. The amount of fixed budget affects the numbers of cheap simulations and physical prototypes used by subjects (see Figure 6).

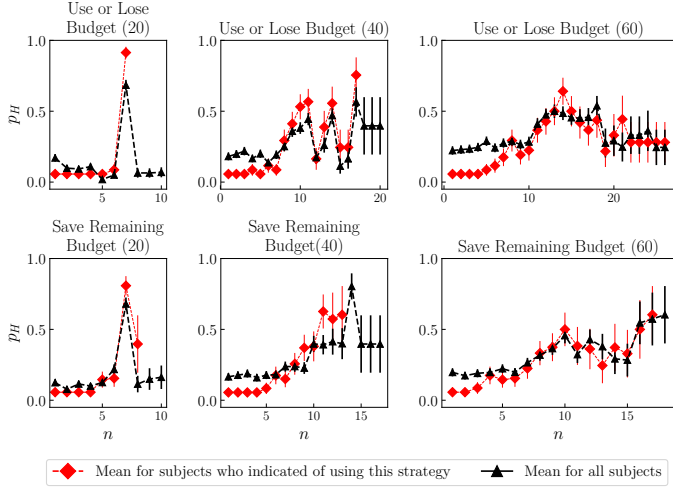


FIGURE 6. Estimations of probability of choosing expensive prototype p_H for Fixed Sample Model in different intervals of sample number.

According to the peaks observed in Figure 6, the most frequent strategies for budget 20 is $(7l, 1h)$, i.e., 7 cheap simulations and then one expensive physical prototype. For the treatments with budget $B = 40$, all strategies are equally distributed. Additionally however, for $B = 60$, the relatively sharper jump at $n = 10$ and a plateau thereafter suggest that strategy $(10l, 5h)$ may be prominent. In the *Save-remaining Budget* treatments, likely strategies are a) budget 20: $(6l, 1h)$, b) budget 40: $(9l, 1h)$, $(9l, 2h)$, $(10l, 1h)$, $(11l, 1h)$, $(12l, 1h)$ and $(13l, 1h)$.

The group of subjects who said that they used this model in the survey use this model more closely. At initial samples, the estimations of p_H for this group is smaller than that of all subjects combined together. Also, their probabilities suggest sharper rise at some fixed sample numbers.

5.2.2 Fixed Maxima-Region Entropy Model To quantify whether the region of function maximum has been identified, we use the entropy of distribution over $x_n^* = \arg \max_x f_n(x)$ as the statistic M . Here, $f_n(x)$ is a sample function from the predictive distribution (Eq. 3). To find the entropy of distribution over x_n^* , denoted by $H(x_n^*)$, we sample 500 functions, identify x_n^* where a sampled function is maximum for all sampled functions, and estimate the entropy of the distribution of identified x_n^* s. Therefore, $M(x) = H(x_n^*)$, where $H(X) = -\sum_i^N q(x_i) \log q(x_i)$ represents the entropy of distribution $q(X)$ over $X = \{x_1, x_2, \dots, x_N\}$.

As observed from Figure 7, subjects use physical prototypes even when the entropy of the region of maxima is reduced, especially at high budgets. The values of p_H are greater than 0.25

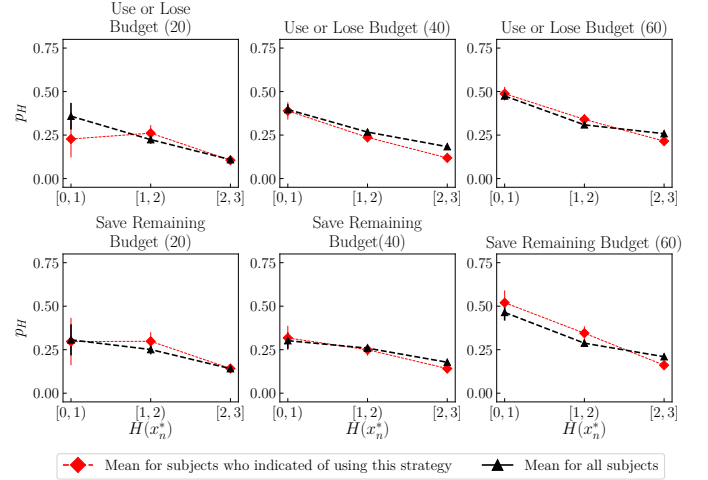


FIGURE 7. Estimations of probability of choosing expensive prototype p_H for Fixed Maxima-Region Entropy Model in different intervals of entropy.

for all intervals of the entropy. The fixed budget affects decisions in this model. In the *Use-or-lose Budget* treatments, the value of p_H at any interval of entropy increases as fixed budget increases from 20 to 40, while it remains the same going from budget 40 to 60. The similar trend is observed for the *Save-remaining Budgets* treatments.

5.2.3 Fixed Expected Improvement Model For this model, $M = \max_x EI_n(x)$. Here, $EI_n(x)$ is the expected improvement in the maximum of predictive mean function. Because the samples from simulations are noisy, the current maximum is defined as the maximum of predictive mean, i.e., $\mu_n^* = \max_{1 \leq i \leq n} \mu_n(x_i)$. Let $\Delta_n(x)$ be the positive difference between predictive mean at x and the best estimate so far. That is, $\Delta_n(x) = \max\{\mu_n(x) - \mu_n^*, 0\}$. Then, the expected improvement in μ_n^* by doing an additional evaluation at $x \in \mathcal{X}$ is calculated as [25],

$$EI_n(x) = \Delta_n(x) \Phi\left(\frac{\Delta_n(x)}{\sigma_n(x)}\right) + \sigma_n(x) \phi\left(\frac{\Delta_n(x)}{\sigma_n(x)}\right) \quad (6)$$

where $\Phi(x)$ and $\phi(x)$, respectively, are the cumulative distribution function and the probability density function of standard normal distribution.

The results from Figure 8 suggest that subjects evaluate using expensive prototypes even at higher values of the highest expected improvement (EI) $\max EI_n(x)$. The probability of picking expensive prototype is high ($p_H > 0.2$) for all intervals of $\max EI_n$ in all treatments. The change from using cheap simulations to ex-

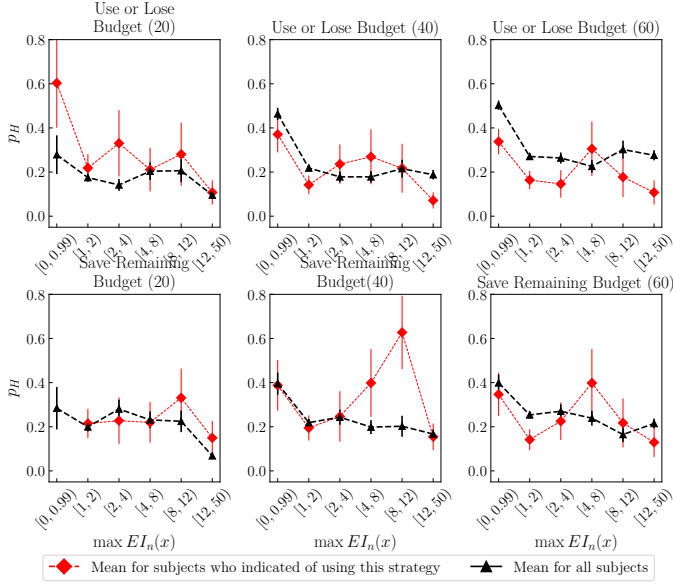


FIGURE 8. Estimations of probability of choosing expensive prototype p_H for Fixed Expected Improvement (EI) Model in different intervals of EI.

pensive prototypes is only significant at very small values of EI when no additional evaluation is preferred. The estimations of p_H for the group of subjects who indicated that they used this strategy have high variance. Therefore, the observations of their strategies do not offer more insights.

5.3 Decision to Stop

5.3.1 Dominant Prototype Model : For this model, we define $M = \max_{1 \leq i \leq n_h} y_h^{(i)} - \max_x \mu_n(x)$. In effect, when M increases to a certain value, search is stopped.

We observe that p_S increases as subjects' best physical observation is closer to the maximum of predictive mean function. Also, the amount of fixed budget affects subjects' probability of stopping. At smaller budget, subjects stop when the difference $\Delta_n(x)$ is large. For example, for $\Delta_n(x) \in [-20, 0)$ in *Use-or-lose Budget* treatments, $p_S (= 0.4)$ at budget 20 is smaller than $p_S (= 0.1)$ at budget 60.

5.3.2 Constant Maxima-Region Entropy Model

In this model, the entropy is defined and calculated in the same way as in Section 5.2.2. That is, the statistic M is equal to $H(x_n^*)$.

From Figure 9, we observe that the subjects are more likely to stop evaluations at lower entropy of maxima region when the budget is small. p_S for entropy in $[0, 1)$ is about 0.5 at budget 20, and reduces to 0.25 for budget 60. However, saving budget does not affect this behavior, since p_S for same budgets in *Use-or-lose*

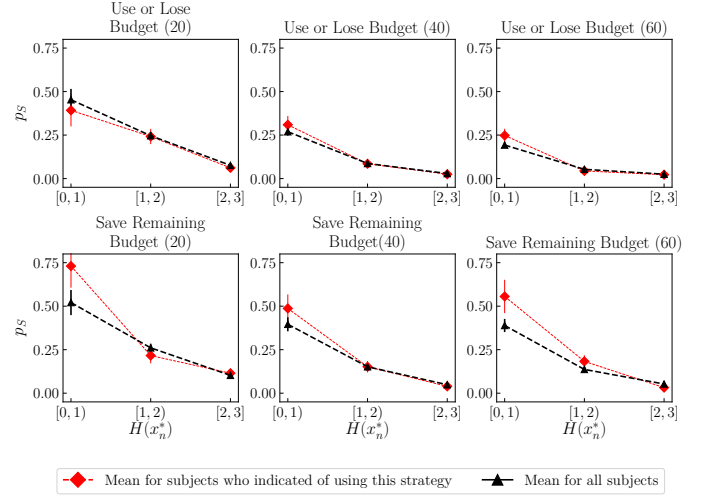


FIGURE 9. Estimations of probability of stopping p_S for Constant Maxima-Region Entropy Model in different intervals of entropy.

Budget and *Save-remaining Budget* are the same.

5.3.3 Fixed Budget Model : For this model, the statistic M is equal to b_n , the total cost of evaluations incurred. $b_n = n_l c_l + n_h c_h$.

Subjects use the entire budget in all budget treatments of *Use-or-lose Budget*, and the *Save-remaining Budget* treatment of budget 20. In other *Save-remaining Budget* treatments, subjects' probability of stopping increasing as the total budget spent or cost increases. However, when subjects have already spent large budget, they more likely use the entire remaining budget. For example, in *Save-remaining Budget* treatment of total budget 60, p_S reduces for $b_l \in [50, 60)$, and spikes at $b_l = 60$ where the search ends.

5.3.4 Constant Expected Improvement Model

For this model, the expected improvement is defined in the same manner as in section 5.2.3. And, $M = \max_x EI_n(x)$.

We observe that subjects use this strategy with higher probability at smaller budgets of *Use-or-lose Budget* treatments, and in *Save-remaining Budget* treatments. For example, p_S between $[0, 2)$ is highest for budget 20 compared to budgets 40 and 60. Further, in the *Save-remaining Budget* treatments, this probability increases for EI between $[0, 2)$ at all budgets compared to their budget counterparts in *Use-or-lose Budget* treatments. In all treatments, the constant expected improvement value when stopping is less than 1.

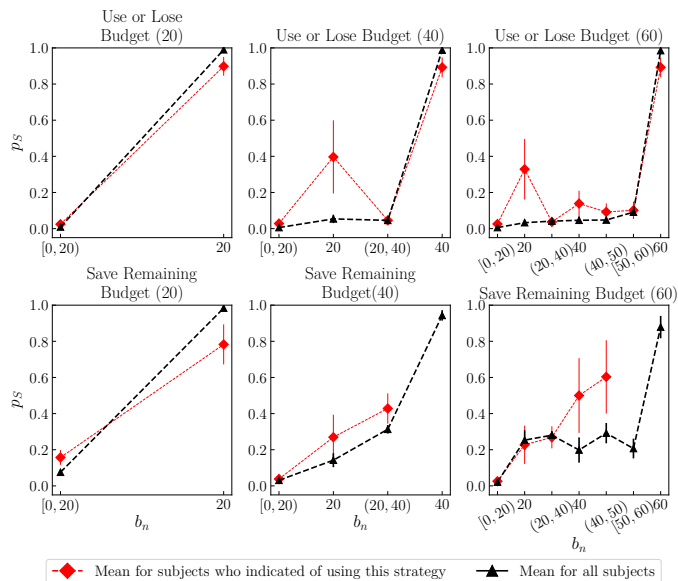


FIGURE 10. Estimations of probability of stopping p_s at given amount of budget spent in different treatments.

6 CONCLUSIONS

In decision-based design, designers make decisions about choosing design points, evaluating performance at those points using multiple information sources, ranking them based on pay-offs, and picking the best design point [1]. These decisions are influenced by the constraints of fixed budget and minimizing spending on performance evaluations. In order to study how humans make these decisions under such constraints, we conduct a controlled experimental task for engineering design optimization. Using a survey in the experiment, we identify that subjects combine different criteria when making sequential decisions. We quantify these criteria into decision models using different statistics of the surrogate model, and descriptively analyze effects of budget constraints (1).

The descriptive analysis of the experimental data draws several conclusions on the effects of budget constraints on designers' strategies:

1. Subjects use explicit criteria to make various decisions. These criteria are consistent with the decision-making literature [12, 21], and are based on prior assumptions or information acquired during design evaluations. These criteria are (a) fixed values of attributes (Sections 5.1.1, 5.3.3), e.g., specific design points, or amount of budget to spend, (b) dominance (Section 5.1.2) such as choosing highest prediction, or highest prototype, (c) stability of representation (Section 5.3.2) such as low information entropy, or uncertainty, and (d) difference threshold (Section 5.3.4) such as expected improvement.

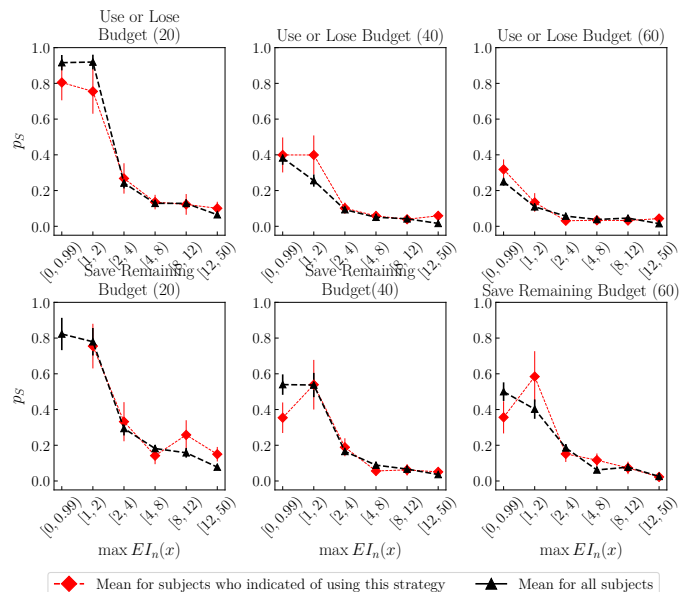


FIGURE 11. Estimations of probability of stopping p_s for Constant Expected Improvement (EI) Model in different intervals of EI.

2. Some of the above criteria fix decisions before the search begins, e.g., evaluating specific points defines the decision to choose x . In other criteria, decisions are made sequentially, e.g., additional evaluations at new points are performed if the region of performance maxima is unclear. This behavior correlates with the previous research on testing in product design which suggest that evaluations are pre-planned when they are expensive [27].
3. When subjects have a higher fixed budget allotted for design evaluations, they are less likely to evaluate points having highest prediction of performance (Section 5.1.2), and more likely to evaluate points having highest uncertainty in performance (Section 5.1.3). Subjects sample more number of expensive prototypes in general and start using expensive prototypes even when the region of the best design is not well identified (Section 5.2.2), and the expected improvement in performance is large (Section 5.2.3). At higher fixed budget, subjects explore more designs, and at the end of the process, the performance of the best expensive prototype design is higher (Section 5.3.1).
4. When subjects are incentivized to save budget, their strategies are affected in the same manner as when the total fixed budget is decreased. However, the effect is only prominent when the fixed budget is high.
5. Subjects who explicitly state strategies they used for making decisions follow those strategies more closely than the aggregate of all subjects.

Our findings have implications in design research for understanding deviations of designers' decisions from optimal strategies. Optimally, cheap simulations should be used until the very last evaluation when the expected improvement (EI) in function maximum is small. However, subjects in our experiment use expensive prototypes even when EI values are large (Figure 8). Subjects use more number of expensive evaluations when the fixed budget is higher (see Figure 10). Further, evaluations should be stopped when the EI from an additional evaluation is smaller than the cost of evaluation. However, subjects do not stop evaluations at small EI values when they have a larger budget to spend. (see Figure 11).

Additionally, our findings provide some directions for structuring various designers' decisions and defining theoretical models of these decisions. The effects of budget constraints and incentives observed in our study necessitate these factors to be incorporated in theoretical models. For example, in the decision to choose next design point, subjects use high uncertainty points more frequently than high mean prediction points as fixed budget increases. For the decision choose an information source, subjects decide the numbers of cheap and expensive evaluations they want to perform a priori. For the decision to stop evaluations, subjects decide the portion of budget they want to spend a priori.

In future work, more rigorous analysis of the experimental data will be performed at the level of an individual subject. Every subject's strategies will be analyzed individually to investigate which decision models are more likely under different treatments. It is possible that some subjects' strategies consistently deviate from the aggregate strategies discussed in this paper. Or, individuals use particular models over others. Such analysis will be performed using estimations of likelihood probabilities for different decision models.

Further research is needed to test the generality of these results to realistic design settings where factors such as designer's expertise, problem domain, non-monetary incentives, and organization practice can affect subjects' strategies. For example, designers' long experience of solving particular domain problems may stir them toward evaluating particular designs. Tests of generality can either be performed using field experiments, or observational studies on real design and systems engineering organizations.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support from the US National Science Foundation (NSF) CMMI through grants # 1400050 and # 1662230.

REFERENCES

- [1] Hazelrigg, G. A., 1998. "A framework for decision-based engineering design". *Journal of mechanical design*, **120**(4), pp. 653–658.
- [2] Freriks, H., Heemels, W., Muller, G., and Sandee, J., 2006. "On the systematic use of budget-based design". In IN-COSE International Symposium, Vol. 16, Wiley Online Library, pp. 788–803.
- [3] Larson, W. J., and Wertz, J. R., 1992. Space mission analysis and design. Tech. Rep. No. DOE/NE/32145-T1, Microcosm, Inc., Torrance, CA (USA).
- [4] Moore, R. A., Romero, D. A., and Paredis, C. J., 2014. "Value-based global optimization". *Journal of Mechanical Design*, **136**(4), p. 041003.
- [5] Xiong, Y., Chen, W., and Tsui, K.-L., 2008. "A new variable-fidelity optimization framework based on model fusion and objective-oriented sequential sampling". *Journal of Mechanical Design*, **130**(11), pp. 111401–111401–9.
- [6] Papalambros, P. Y., and Wilde, D. J., 2000. *Principles of optimal design: modeling and computation*. Cambridge university press, New York, NY.
- [7] Deb, K., 2012. *Optimization for engineering design: Algorithms and examples*. PHI Learning Pvt. Ltd., New Delhi, India.
- [8] Ravindran, A., Reklaitis, G. V., and Ragsdell, K. M., 2006. *Engineering optimization: methods and applications*. John Wiley & Sons, Hoboken, New Jersey.
- [9] Wang, G. G., and Shan, S., 2007. "Review of metamodeling techniques in support of engineering design optimization". *Journal of Mechanical design*, **129**(4), pp. 370–380.
- [10] Pandita, P., Bilonis, I., and Panchal, J., 2016. "Extending expected improvement for high-dimensional stochastic optimization of expensive black-box functions". *Journal of Mechanical Design*, **138**(11), pp. 111412–111412–8.
- [11] Liu, H., Ong, Y.-S., and Cai, J., 2018. "A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design". *Structural and Multidisciplinary Optimization*, **57**(1), pp. 393–416.
- [12] Browne, G. J., Pitts, M. G., and Wetherbe, J. C., 2007. "Cognitive stopping rules for terminating information search in online tasks". *MIS Quarterly*, **31**(1), pp. 89–104.
- [13] Busemeyer, J. R., and Rapoport, A., 1988. "Psychological models of deferred decision making". *Journal of Mathematical Psychology*, **32**(2), pp. 91–134.
- [14] Panchal, J. H., Sha, Z., and Kannan, K. N., 2017. "Understanding design decisions under competition using games with information acquisition and a behavioral experiment". *Journal of Mechanical Design*, **139**(9), pp. 091402–091402–12.
- [15] Sha, Z., Kannan, K. N., and Panchal, J. H., 2015. "Behavioral experimentation and game theory in engineering

- systems design”. *Journal of Mechanical Design*, **137**(5), pp. 051405–051405–10.
- [16] Borji, A., and Itti, L., 2013. “Bayesian optimization explains human active search”. In *Advances in neural information processing systems*, Lake Tahoe, Nevada, pp. 55–63.
 - [17] Sexton, T., and Ren, M. Y., 2017. “Learning an optimization algorithm through human design iterations”. *Journal of Mechanical Design*, **139**(10), pp. 101404–101404–10.
 - [18] Newell, B., and Bröder, A., 2008. “Cognitive processes, models and metaphors in decision research”. *Judgment and Decision Making*, **3**(3), p. 195.
 - [19] Simon, H. A., and Newell, A., 1971. “Human problem solving: The state of the theory in 1970.”. *American Psychologist*, **26**(2), p. 145.
 - [20] Glckner, A., and Betsch, T., 2008. “Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making”. *SSRN Online Library*.
 - [21] Browne, G. J., and Pitts, M. G., 2004. “Stopping rule use during information search in design problems”. *Organizational Behavior and Human Decision Processes*, **95**(2), pp. 208–224.
 - [22] Montgomery, H., 1983. “Decision rules and the search for a dominance structure: Towards a process model of decision making”. *Advances in psychology*, **14**, pp. 343–369.
 - [23] Panchal, J. H., and Szajnfarder, Z., 2017. “Experiments in systems engineering and design research”. *Systems Engineering*, **20**(6), pp. 529–541.
 - [24] 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.
 - [25] Jones, D. R., Schonlau, M., and Welch, W. J., 1998. “Efficient global optimization of expensive black-box functions”. *Journal of Global optimization*, **13**(4), pp. 455–492.
 - [26] Dunnigan, K., 2008. “Confidence interval calculation for binomial proportions”. In *MWSUG Conference*, Indianapolis, IN.
 - [27] Tahera, K., Earl, C., and Eckert, C., 2012. “The role of testing in the engineering product development process”. In *Proceedings of TMCE 2012*, Karlsruhe, Germany.