INFORMATION ACQUISITION IN ENGINEERING DESIGN: DESCRIPTIVE MODELS AND BEHAVIORAL EXPERIMENTS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Ashish M. Chaudhari

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2020

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Jitesh H. Panchal, Chair
School of Mechanical Engineering
Dr. Ilias Bilionis
School of Mechanical Engineering
Dr. Karthik Kannan
Krannert School of Management
Dr. Sebastien Helie
Department of Psychological Sciences

Approved by:
Dr. Nicole Key
School of Mechanical Engineering

ACKNOWLEDGMENTS

I thank my academic advisor, Professor Jitesh Panchal, for guiding me in all aspects of the research and writing of this disseration, providing me with intellectual freedom in my work, and steering me in the right direction when needed. Prof. Panchal's support for my engagement in new ideas through discussions in research meetings as well as through my attendance at various conferences has made this research experience truly rewarding. I thank Professor Ilias Bilionis and Professor Karthik Kannan for their guidance and timely feedback in designing the experiment and building analytical models for this dissertation. I thank my committee member, Professor Sebastian Helie, for informing this work from the cognitive science perspective.

A significant portion of this dissertation relied on the communication data of the NASA Mission Design Laboratory (MDL) provided by Prof. Zoe Szajnfarber and Dr. Erica Gralla. Thank you also for guidance on the work comparing team communication patterns between NASA MDL engineers and engineering students in student design teams.

I gratefully acknowledge the support from past and current members of Design Engineering Lab at Purdue, Murtuza Shergadwala, Vikranth Reddy, Adam Dachowicz, Atharva Hans, Joseph Thekinen, Piyush Pandita, and Siva Chaduvula. Thanks for engaging in stimulating discussions that made this research happen. This work would not have been possible without financial support from the US National Science Foundation (NSF) CMMI through grants #1662230 and #1841062.

Finally, I would like to thank my friends, Kaushal Jain, Raghvendra Yasaswi, Benjamin Johnson, Naveen Madapana, Mayank Gupta, and Bhavesh Shrimali for making my time at Purdue memorable and fun. I thank my parents for sending me to the USA and believing in me throughout my graduate studies. A special thanks to Komal whose patience knows no bounds; thank you for teaching me to love and inspiring me to do my best every day.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

ABSTRACT

Chaudhari, Ashish M. Ph.D., Purdue University, August 2020. Information Acquisition in Engineering Design: Descriptive Models and Behavioral Experiments. Major Professor: Jitesh H. Panchal, School of Mechanical Engineering.

Engineering designers commonly make sequential information acquisition decisions such as selecting designs for performance evaluation, selecting information sources, deciding whom to communicate with in design teams, and deciding when to stop design exploration. There is significant literature on normative decision making for engineering design, however, there is a lack of descriptive modeling of how designers actually make information acquisition decisions. Such descriptive modeling is important for accurately modeling design decisions, identifying sources of inefficiencies, and improving the design process. To that end, the **primary research objective** of the dissertation is to understand how designers make sequential information acquisition decisions and identify models that provide the best description of a designer's decisions strategies. For gaining this understanding, **the research approach** consists of a synthesis of descriptive theories from psychological and cognitive sciences, along with empirical evidence from behavioral experiments under different design situations. Statistical Bayesian inference is used to determine how well alternate descriptive decision models describe the experimental data. This approach quantifies a designer's decision strategies through posterior parameter estimation and Bayesian model comparison.

Two research studies, presented in this dissertation, focus on assessing the effects of monetary incentives, fixed budget, type of design space exploration, and the availability of system-wide information on information acquisition decisions. The first study presented in this dissertation investigates information acquisition by an individual designer when multiple information sources are available and the total budget

is limited. The results suggest that the student subjects' decisions are better represented by the heuristic-based models than the expected utility(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 results also indicate the potential for nudging designers' decisions towards maximizing the net payoff by setting the fixed budget at low values and providing monetary incentives proportional to the saved budget.

The second study investigates information acquisition through communication. The focus is on designers' decisions about whom to communicate with, and how much to communicate when there is interdependence between subsystems being designed. This study analyzes team communication of NASA engineers at a mission design laboratory (MDL) as well as of engineering students designing a simplified automotive engine in an undergraduate classroom environment. The results indicate that the rate of interactions increases in response to the reduce in system-level design performance in both settings. Additionally, the following factors seem to positively influence communication decisions: the pairwise design interdependence, node-wise popularity (significant with NASA MDL engineers due to large team size), and pairwise reciprocity.

The dissertation work increases the knowledge about engineering design decision making in following aspects. First, individuals make information acquisition decisions using simple heuristics based on in-situ information such as available budget amount and present system performance. The proposed multi-discipline approach proves helpful for describing heuristics analytically and inferring context-specific decision strategies using statistical Bayesian inference. This work has potential application in developing decision support tools for engineering design. Second, the comparison of communication patterns between student design teams and NASA MDL teams reveals that the engine experiment preserves some but not all of the communication patterns of interest. We find that the representativeness depends not on matching subjects, tasks, and context separately, but rather on the behavior that results from

the interactions of these three dimensions. This work provides lessons for designing representative experiments in the future.

1. ELEMENTS OF INFORMATION ACQUISITION IN ENGINEERING DESIGN

Design is widely considered as a decision making process. It involves decisions such as selection of problems to be tackled, deciding product layouts, choosing information sources, selecting shape, size and material for products components, communicating with other designers and manufacturers etc. These decisions are irrevocable actions that mark substantive changes in activities that follow them [1,2]. For example, tackling selected problems and communicating with others take up resources irrespective of whether those problems are solved or not. Some decisions such as material selection can be updated in the future stages if desired outcomes are not achieved. Even then, the resources spent following any decision are irretrievable.

The amount of resources spent in the design process can be reduced by anticipating the interdependence between decisions and the outcomes of the design process. Given the importance of decisions in engineering design, many esearchers have adopted the decision based view of design, and developed methods that recognize the central role of decision making in design processes. A design methodology based on making rational decisions called **decision-based design** has been developed [3–6]. It comprises of sub-processes such as specifying criteria and constraints, generating alternatives, and choosing an acceptable alternative.

Various approaches are available within decision-based design, most of them are normative approaches. **Normative decision making** approaches are representations of the **ideal** decision strategies. They typically formulate a decision as selecting an optimal choice maximizing the expected utility for fixed performance attributes. Existing studies propose theoretical models of normative decision making in engineering design [7] with extensions for multi-fidelity design [8], high-dimensional design spaces [9–11], and multi-objective design problems [12, 13]. While the utility-maximizing approaches prescribe systematic ways of making engineering design decisions, their application in complex problems may be time-consuming and computationally resource intensive. Moreover, normative models are not defined for all design decisions, e.g., decision to communicate with others.

In complex design situations, human designers still make design decisions by learning from the outcomes of their past decisions [14, 15]. Empirical observations are a great source of knowledge for studying designers' actual decision strategies. The representation of designers' actual decision strategies is termed as **descriptive decision making.** A descriptive decision making model describes decisions with consideration of designers' biases and judgments that designers use. Descriptive decision models are more **realistic** as models of designers' design strategies than normative decisions. The **true** representation of designers' decision strategies is the *type of reasoning and judgements* designers use [16]. But designers' true reasoning and judgements are difficult to infer. Then, the observations of design decisions are useful non-intrusive options for explaining designers' decision strategies. To that effect, this dissertation's overarching goal is to develop descriptive representations of engineering design decisions which can act as lenses into the underlying types of reasoning that designers commonly use.

This chapter begins with a background and motivation for studying information acquisition in engineering design. From this background, a relevant structure of the design process under consideration is defined as *sequential* information acquisition and design decisions in this context are identified (see Section 1.1). Followed by the specification of design decisions, the questions of how to make information acquisition decisions are investigated beginning with literature review of decision-based design research in Section 1.2. This investigation leads to intellectual questions surrounding quantitative estimation of designers' subjective decision making strategies from observed contextual decisions. Section 1.3 presents the research objectives of two research studies presented in the dissertation. Connecting back with the background and motivation, Section 1.3 details the engineering and scientific significance of the two studies. Section 1.4 describes the outline of dissertation.

1.1 Design as Sequential Information Acquisition Process

1.1.1 Background

Information acquisition activities are ubiquitous in the design process. The information acquisition is not "one-shot" but rather involve repetitions. The purpose of information acquisition is to find alternatives as well as deciding which alternatives to select. For example, engineering authors [2, 7] posit that the following processes constitute information acquisition in engineering design:

1. Frame a decision to be made (How does a designer realize decision is needed?)
2. Establish criteria for comparing different alternatives. (How do these relate to higher-level goals?)
3. Evaluate the alternatives according to those criteria? (How careful should the evaluation be?)
4. Stop evaluations and choose an acceptable alternative. (What constitutes enough information to choose?)
5. Retract the selected alternative if it proves unsatisfactory and restart the process. (When should a decision be reconsidered?)

Not only do technical factors but also social factors influence the sequential information acquisition in engineering design. The technical factors such as **design problem complexity, process cost**, and **uncertainty** are important determinants of how frequently or rarely do designers perform information acquisition. The problem complexity is based on the size of the design problem, complex coupling between the sub-components, and solvability [17]. In addition to the nature of design problem, the cost and uncertainty associated with the design process determine how much information can be gathered and whether all or part of the design space is explored. In product design, the practice of prototyping is useful in learning whether design options satisfy requirements. Its outcomes, however, depend upon uncertainties associated with prototypes. For example, computer-based simulations such as in automotive crash tests are flexible while relatively expensive physical prototypes detect unanticipated phenomena [18]. Also consider VLSI semiconductor manufacturing plants where large noise in experiments due to process variability has a major influence on how efficiently designers learn from experiments [19].

Social factors are less commonly investigated as the drivers of information acquisition in engineering design. Some of the social factors are **individual designers' types of reasoning (judgement), domain knowledge**, and **type and amount of communication** between multiple designers on a team. First, individuals' reasoning or judgments can vary based on the reliance on evidence, prior belief, view of knowledge and concept of justification [16]. Figure 1.1 illustrates the view of knowledge and individuals' justification in different types of thinking type. Stages 1-3 belong to *pre-reflective thinking* level where knowledge is assumed to be certain. Stages 4-5 belong to *quasi-reflective thinking* level where uncertainty is recognized as a part of knowledge. And, Stages 6-7 are recognized as *reflective thinking* level where designers use evidence and reason comfortably in support of their judgment. Reflective designers are most favored for solving complex, ill-structured design problems. In reality, a designer's thinking process is like a wave across a mixture of stages.

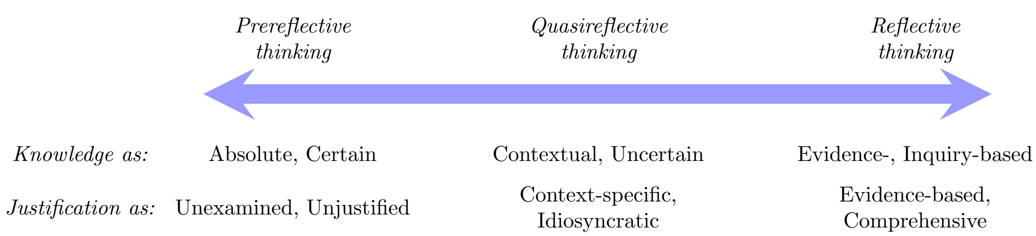


Figure 1.1. : Types of thinking and corresponding knowledge and justification in the reflective judgment model [16].

Further, the domain knowledge required to assess the sources of information acquisition is specific to the setting in which the design is being created [20]. King and Kitchener [16] show a relatively high degree of consistency in people's use of epistemic assumptions toward their reasoning about ill-structured problems. Adams et al. [21] found that the level of expertise also impacts reasoning processes in iterative design activities. They observed that senior-engineering students display broader perception of design problems, measured as the number and the breadth of issues, than the first-year engineering students.

Finally, communication with other designers an important source of information in design teams. Bucciarelli in his book, *Designing Engineers* [15], emphasizes how interactions and communication are integral to designing in engineering firms. His observations are based on the transcripts of engineers' conversations and from his personal experience working on consulting projects. Some examples of communication drivers behind information acquisition are:

1. designers resolve uncertainty through social connections, e.g., seeking advice from others and soliciting suggestions from manufacturers,
2. individuals' design decisions may stem from motivation to form alliances with others,
3. individuals plan, decide, reflect together to integrate their efforts, and
4. information is processed through pieces of paper and verbal communication.

1.1.2 Decision-based Design Framework

This dissertation relies on a decision-based design framework to study the information acquisition in engineering design. The decision-based design framework is adapted from recent efforts to put a structure on decisions in the engineering design process [22, 23]. Panchal et al. [22] first proposed a sequential information acquisition framework which involves few design decisions repeated in stages. Between any two stages of the process, a designer reflects on past outcomes and updates his/her state of belief about the performance over the problem space. The designer has control over design decisions such as selecting detailed designs to evaluate/test their values and stopping evaluations. Once the evaluations are terminated, the preferred detailed design is the one among the designs that the designer evaluates with the highest value.

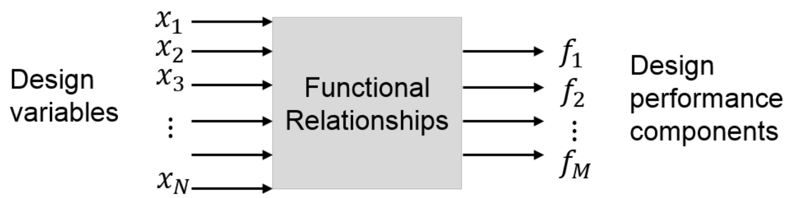


Figure 1.2. : The "black-box" nature of design problems considered in the decision-based framework of information acquisition.

In the decision-based design framework, we assume that a designer's objective is to find the design variables values that maximize given performance attributes (see Figure 3.1). To achieve this objective, the designer performs iterative evaluations of the design performance attributes. Design variables may be discrete such as material type or continuous such as length of component. Design performance attributes measure the value of detailed design in multiple dimensions. We assume that the functional relationships between design variables and the performance attributes are unknown.

The decision-based design framework involves three types of information acquisition-related decisions: (i) **design exploration**, (ii) **communication**, and (iii) **stopping**. Examples of the design exploration type decisions are selecting the number of detailed designs to evaluate their performance attributes in parallel at any given stage, choosing design points from the problem space, and selecting an information source, such as computer simulation or physical prototype, for performance evaluations. It is also possible that different information sources vary in the cost of evaluation, uncertainty of outcomes, fidelity etc.

Then, the decisions related to communication are: (i) deciding how much to communicate with others, (ii) selecting team members for communicating, and (iii) deciding which information to share with others. These communication decisions require a designer to consider trade-offs between different communication channels available. Face-to-face communication channels allows passage of richer information than text-based or numeric communication channels but face-to-face communication is time-consuming. Similarly, communication-related decisions require a designer to consider similarities between own design subcomponent and other subcomponents for deciding whom to communicate with. Talking with all team members is not feasible at it consumes time and resources and may not be beneficial for achieving better design performance.

Finally, the stopping decisions include deciding whether the best performance is achieved and deciding when to stop performance evaluations. The designer may rely on the available budget to stop, e.g., stopping if the costs of designing exceed the budget. However, when the total budget is not specified for finding a preferred detailed design, the designer must exercise judgment for deciding when to stop.

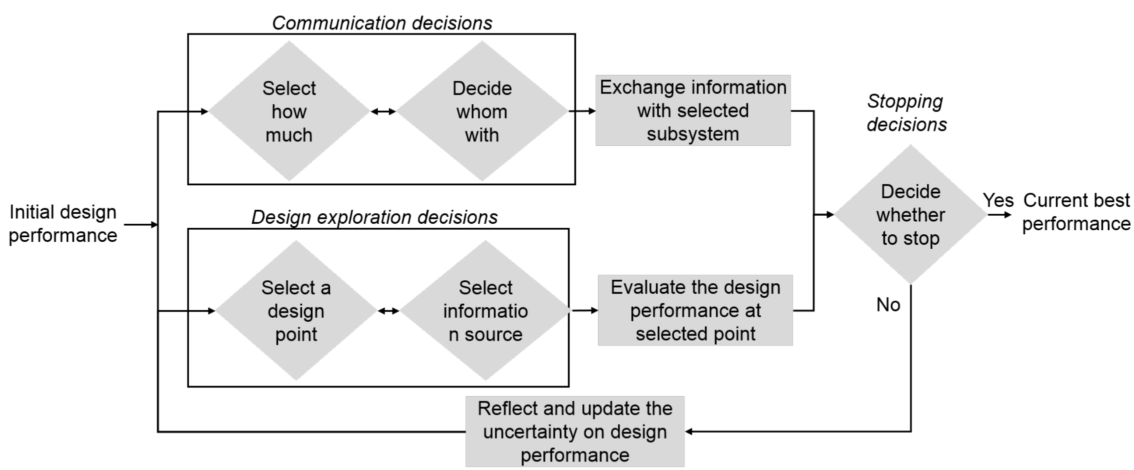


Figure 1.3. : A framework for sequential decision making in design.

In summary, the decision-based design framework assumes that functional relationships between design variables and design performance attributes are uncertain. To resolve this uncertainty, designers sequentially acquire information through design exploration and through communicating with others on same team. The decisions in such information acquisition process are made under limited resources and, therefore, they require designers balance the tradeoff between design performance and design/communication costs. The information acquisition decisions in large part are dependent on the problem complexity, process uncertainty, process costs, and social connectedness.

1.2 How Should Designers Make Decisions?

The information acquisition decisions in engineering design are difficult to make due to uncertainty in design performance, lack of information, technical errors in the design process, etc. To develop rational approaches for making the information acquisition decisions, researchers have modeled the engineering design process as a decision-making process under uncertainty.

1.2.1 Normative Frameworks of Decision-based Design

The realization of the engineering design as a decision-making process has allowed researchers to borrow analytical techniques from economics and mathematics for modeling the engineering design process. Early research has established foundation for normative decision making in engineering design where rational decisions follow the rule that the preferred decision is the option whose expectation has the highest value, or expected utility in case of uncertain value [4,24][1]. Hazelrigg [7] posits eight axioms of engineering design, *axiom of deterministic decision making, ordering of alternatives, reduction of compound lotteries, continuity, substitutability, transitivity, monotonicity*, and *reality axiom.* A valid application of a normative approach requires that all of the axioms of engineering design are satisfied.

Despite the promise of mathematical structure on the decision-making process, normative approaches suffers from validation issues, especially when modeling multi-attribute design performance under uncertainty. In engineering design, design performance levels are uncertain due to various sources of uncertainty such as measurement error, lack of information, and simplified models. The multi-attribute approaches,

[1] We will use the word "utility" as it stands for the selection criterion in the presence of uncertainty, while the word "value" is often interpreted as a selection criterion without uncertainty.

which consider two or more criteria to compare decisions alternatives, are limited in dealing with such uncertainty because assigning weights to mutually independent attributes while maintaining preference consistency are challenging [5]. The consideration of multiattribute ranking, normalization, and weighting in the multi-attribute approaches are known to invalidate the axioms of engineering design [25]. In solution, researchers have proposed using single attribute for comparing alternative designs. The net economic benefit, which is the monetary value of a product subtracted with associated costs, is thought as a single viable criterion for decision-based design under uncertainty [26]. The single-attribute approach is based on discrete choice modeling approach which requires designer choice data for quantifying uncertainties about unobserved attributes and unobserved preferences.

1.2.2 Decision-based Design under Uncertainty: Applications of Bayesian Global Optimization

In recent advances to normative decision-making approaches, the Bayesian global optimization (BGO) has been frequently employed to manage the uncertainty in the design performance. The complexity of design problems generally prohibits the exploration of an entire design space and alternatives for the best design point have to be strategically selected. A BGO approach addresses this problem by modeling the uncertainty about the design performance using probabilistic surrogate models such as Gaussian processes. Then, a typical BGO approach quantifies the relative importance of design points in the design space using one of the information acquisition functions such as expected improvement [27] and probability of improvement [28]. Finally, the selected design point is the point that maximizes the value of information acquisition function. BGO approaches have been relatively successful in addressing the information acquisition problem in engineering design [9, 10, 29]. In advancing the BGO for complex engineering design process, the value-based global optimization (VGO) approach incorporates the cost of design process into the decision making process [8]. The VGO uses the utility function as mathematical means for comparing relative economic benefit of a particular design point. In addition to the VGO, the extensions of BGO for high-dimensional design problems [12] and design problems with multiple functional requirements [13] are the state of the art.

1.3 How Do Designers Make Decisions? Research Goals and Focus

While the engineering design research has taken leaps towards subjective preference modeling and surrogate modeling in support of decision-based design framework, there is lack of investigation of descriptive decision modeling for making decisions. Similar to how designers rarely assemble preferences about multiple attributes independently, designers rarely make decisions maximizing the expected utility of selected alternatives. There is a need to extract and understand an individual-specific decision making strategies so as to create better alternatives to the expected utility-maximization strategy. Undertaking this research direction may improve the trade-off between design performance and resources spent by allowing the development of decision support tools that inform decision-making during the engineering design process.

It is well known that humans do not necessarily follow the normative models of decision making [14, 30, 31]. Researchers in cognitive psychology and behavioral economics have developed various descriptive models of human decision makers [32, 33]. Examples of these descriptive models include bounded rationality-based models [34], fast and frugal heuristics [33], models based on deviations from rationality [32], and cognitive architecture-based models [35]. These descriptive models are alternatives to the normative models, but 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.

There has been relatively modest amount of research on quantifying decision making processes of designers in engineering design. Present studies have captured the sequential decisions in the design process using Markov chains [36], mined the process heuristics using Hidden Markov Models [37], and incentives and competition using the game theory [22, 38]. The human strategies have also been estimated in terms of the hyperparameters of Bayesian Global Optimization (EGO) [39]. Others have employed controlled human experiments to study the impact of problem-related factors such as complexity and scale on human decisions They observe that experiment subjects generate lower solution quality as scale and coupling in parameteric design tasks increases [40–42]. In the studies on design teams and design performance, the research find that the coherent design communication is positively correlated with the design performance [43, 44]. This experimental design research on descriptive decision making can benefit from a systematic research approach based on computational modeling and controlled experiments, so that the descriptive decision making approaches can be applied as widely as the normative decision making approaches.

1.3.1 Intellectual Questions for Consideration

With the need for descriptive decision making research specified, the next step requires devising a research plan that is based on computational decision models, driven by the prior knowledge of the design process, and uses the contextual decision data particular to a design problem. This research plan should be able to address intellectual questions about not only the effects of external factors related such as cost, budget, and incentives on information acquisition decisions but also the procedure of data acquisition and descriptive analysis. The intellectual questions are broader than research questions in that they also reflect methodological issues in analyzing information acquisition decisions. Table 1.1 lists a sample of intellectual questions relevant to the analysis of descriptive decision making and the subsequent chapters that answer those questions.

Relevant Sections:	Chapter 2			Chapter 3			Chapter 4			Chapter 5		
Intellectual Questions	2.1	2.2	2.3	3.1	3.2	3.3	4.1	4.2	4.3	5.1	5.2	5.3
What are different types of descriptive models?	x				x			x				
How to reduce human biases in controlled experiments?		x		x			x					x
How to represent group- and individual-specific decision strategies?			x			x			x			
Which strategies do individuals use for information acquisition decisions?	x					x			x		x	
What are the effects of incentives and the problem nature?				x			x					
How does information acquisition evolve over time?						x		x	x		x	
How to make experimental insights more representative of real world?		x										x

Table 1.1. : Intellectual questions addressed in the dissertation

1.3.2 Research Objectives of the Dissertation

From vast possibilities of research directions in descriptive decision making, this dissertation focuses on specific research objectives related to the sequential information acquisition and investigates the effects of few social and technical factors on information acquisition decisions. The **primary objective in this dissertation is to identify models that provide the best description of a designer's sequential information acquisition decisions**, including decisions related design exploration, team communication and stopping. Particularly, the research questions study the effects of problem-related and process-related factors that guide the sequential information acquisition as indicated in Section 1.1.

1. How do two factors influence a designer's sequential information acquisition related to design exploration and stopping: **the monetary incentives proportional to design performance** and **the amount of available fixed budget**?
2. How do two factors influence between-designers communication during the design process: i) whether designers can search for solutions using **catalogs** with a limited selection of options (a discrete design space) or **using a simulation tool** (a continuous design space), and ii) whether designers have **access to global information about the status of the design**, such as the design variables determined by other designers, through a shared parameter database.

The *central hypothesis* behind this research is that the effects of the sequential information acquisition can be studied by quantifying individuals' decision making strategies through computational modeling and behavioral experiments. Following up on the hypothesis, the approach involves the sub-objectives of designing simple, but non-trivial, experimental tasks representative of the sequential information acquisition process, collecting experimental evidence on individuals' decisions via controlled experiments, formulating alternative models of decision strategies, and performing statistical Bayesian inference for estimating the posterior distribution over the model parameters and identifying the best-fit models (see Figure 1.4).

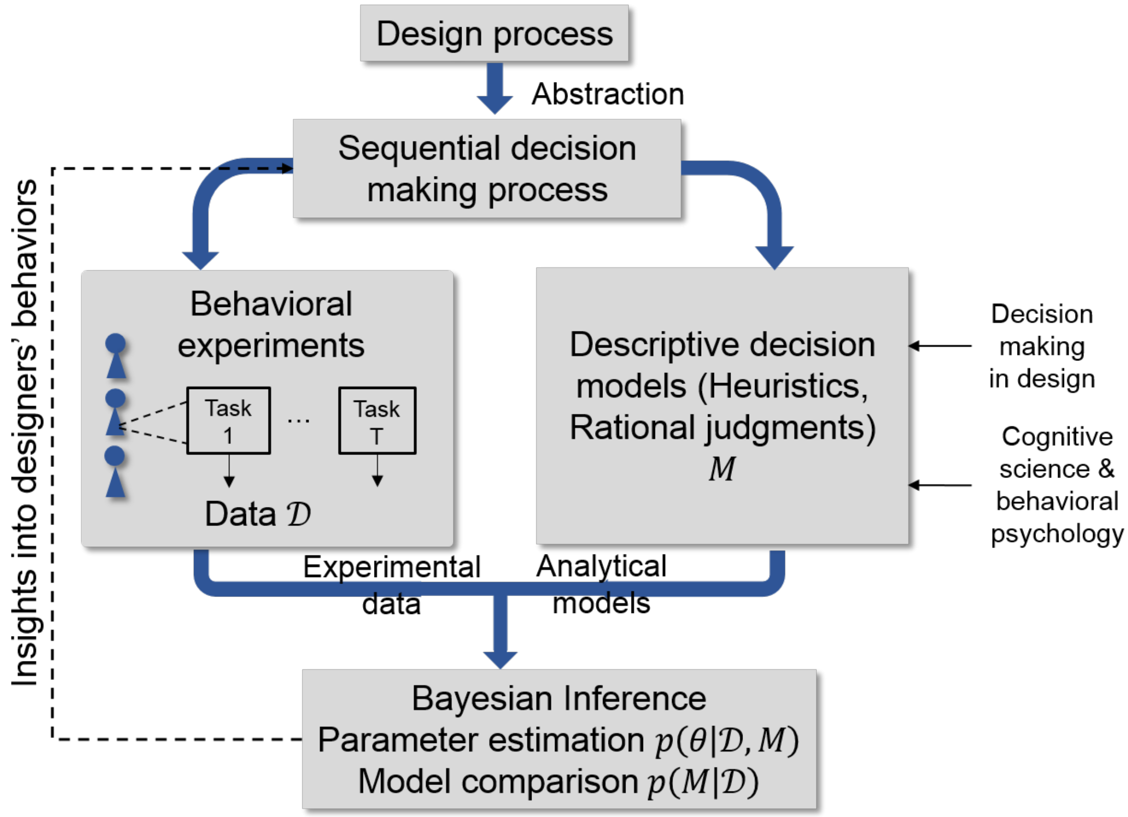


Figure 1.4. : An overview of the research approach.

1.3.3 Engineering and Scientific Relevance of the Work

The two main components of the research approach are: (i) controlled experiments and (ii) statistical Bayesian inference methodology. Both components have significance in engineering practice and scientific research related to engineering design.

Controlled experiments has scientific relevance in understanding the decision-making of designers who develop models and perform analysis under uncertainty [22]. In a multi-agency research workshop, researchers have found there is lack of knowledge about how modelers' decisions on spending effort and resources affect the resulting design performance (finding C2 in Ref. [45]). The cognitive biases and human decision making in modeling and simulation are less explored in the organizational context due the complexity of these decisions. In addressing this gap, the dissertation demonstrates through multiple experimental studies that the suggested research approach can increase the scientific knowledge of complex decision making.

Besides the relevance in understanding decision-making, the statistical Bayesian inference of descriptive decision strategies can support development of future Design Intelligence tools. The goal of such tool would be to improve engineering design decision-making (e.g. design of experiments) by data-driven support systems. The data-driven support systems would take designer decision data as input and computationally estimate the likelihood of alternative decision strategies. Based on the estimated decisions strategies from the observed history, the Design Intelligence tools then should be able to anticipate the impact of different alternatives on the design performance and suggest decision strategies for a designer to undertake in the future.

The Design Intelligence tools are similar in the purpose to Business Intelligence tools [46] but are not as widely developed. A possible reason for limited development might be that design problems, especially in system engineering, are complex and require formalized modeling to support design, analysis, verification, and validation activities. Examples of support systems tools in the context of model-based systems engineering are Cameo Systems Modeler [47], AnyLogic [48], IBM Rhapsody [49], and SysML Designer [50]. The dissertation will demonstrate how the statistical inference part of the research approach can augment and create new functionalities.

The statistical Bayesian parameter estimation and Bayesian model comparison can facilitate an accurate characterization of simulation models for the systems engineering processes. The simulation models of systems engineering processes take into account random behaviors. Such uncertainty quantification is necessary because agents (designers or system engineers) may randomly interact with each other, agents exchange information through complex interactions, and not all details of the system are modeled. When input data is available from online real-time processes or direct observations in a time study, statistical methodologies can find the best-fit model parameters for given data and additionally compare the model performance between different candidate models.

Further, researchers can quantify the design performance in terms of designers' decision strategies with the help of estimated descriptive decision models. The applications of this include design crowdsourcing where game-theoretic models lack design process models [51–53], and the agent-based models of engineering systems design where characterization of quality as a function of designer effort is difficult to achieve [54].

To envision the process of realizing the impact of descriptive decision making research, Section 1.4 presents the research roadmap undertaken in this dissertation.

1.4 The Outline of Dissertation

The dissertation uses a research roadmap presented in Figure 1.5 to formalize the research approach and to answer the research questions. The roadmap starts with **the development of design guidelines for experimental design tasks and descriptive decision models.** The dissertation synthesizes such design guidelines from the literature from experimental economics, cognitive science and applied statistics. Then, the next step in the research roadmap is to **implement the research approach on an individual's sequential information acquisition decisions regarding design exploration**. For this step, the dissertation includes a research study investigating the effects of process cost and monetary incentives on designers to select design search points, selecting an information source and deciding when to stop. The next step is to **implement the research approach on sequential information acquisition decisions in design teams** where multiple designers build separate but interdependent subsystems and interact with each other during the duration of the design process. The second research study investigates the communication decisions in design teams. The final step involves **validating the research approach to support generalization of case-specific insights**. There are two levels to experimental validity: (i) internal validity, and (ii) external validity [55]. The interval validity of the research approach can be justified by following the design guidelines for controlled experiments and by using explainable descriptive decision models. For analyzing the external validity, however, researchers need to replicate same experimental task across multiple number of subjects and across settings with different levels of realism [55]. In this dissertation, the integral validity of both the research studies are investigated through cross-validation techniques such as splitting the experimental data into training and testing data set. In addition, the external validity of the second research study on interdisciplinary communication in design teams is investigated by comparing the communication patterns between engineering student teams designing an automotive engine and NASA engineer teams designing a spacecraft system.

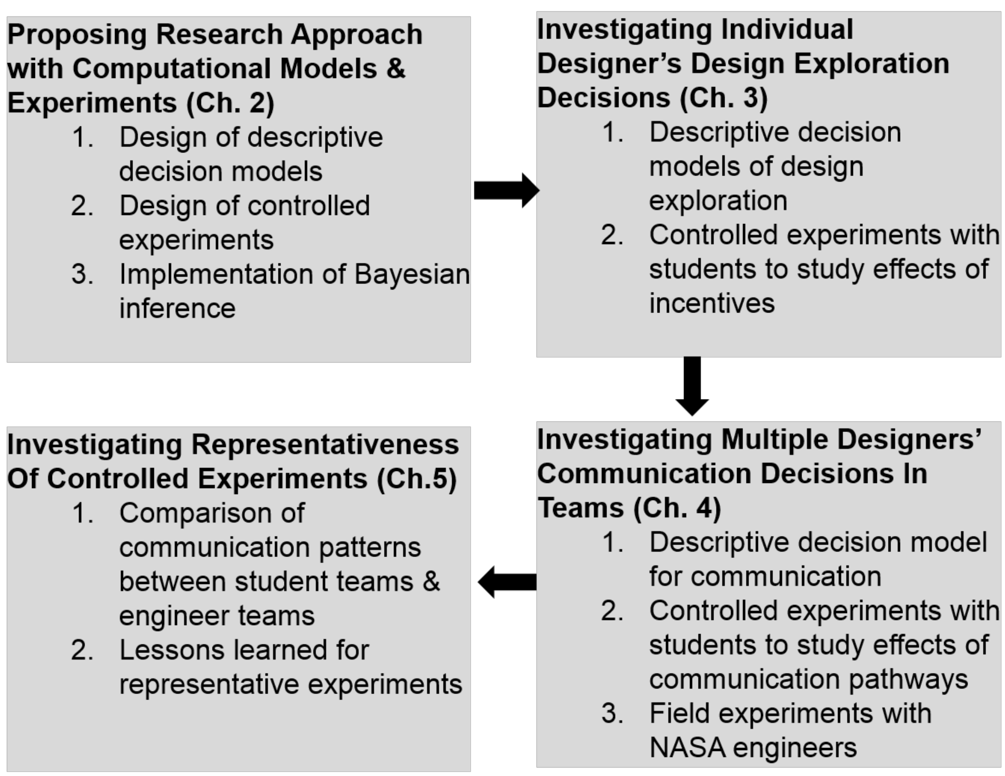


Figure 1.5. : The roadmap of the dissertation research.

The outline of the dissertation follows the research roadmap and is as following.

1. In Chapter 2, a synthesis of existing literature focuses on experimental design guidelines and descriptive decision modeling for the context of information acquisition decisions in engineering design.
2. In Chapter 3, a research study analyzes the best descriptive models for an individual designer's sequential information acquisition decisions.
3. In Chapter 4, a research study analyzes stochastic, time-series models of interdisciplinary communication in design teams. The posterior parameter estimation in these models provides insights on individuals' communication strategies.
4. In Chapter 5, a comparison of communication patterns between engineering student teams and NASA engineer teams allows the investigation of external validity and representativeness.
5. Finally, Chapter 6 presents the key conclusions of the experimental studies, lists overall contributions of the work, and summarizes some of the remaining work as future research directions.

2. DOMAIN-GENERAL APPROACH FOR DESCRIPTIVE DECISION ANALYSIS

This chapter provides technical background of computational modeling and behavioral experiments for performing descriptive decision analysis. The chapter describes a range of descriptive models, from expected utility-maximizing models to simple heuristic. On one end, expected utility-maximizing models represent rational judgements in accordance with the expected utility theory. They incorporate all information in decision making and are also called **compensatory models**. On the other hand, simple heuristic models represent simple rules that humans commonly use. Simple heuristics use selective information to make decisions and are also called **non-compensatory models** of decision making. Following up on the modeling effort, the chapter presents guidelines for selecting appropriate subject pool and designing an experimental task. Lastly, the chapter details the theoretical and practical considerations of statistical Bayesian inference for the context of descriptive decision analysis. The outline in Figure 2.1 shows different steps involved in the descriptive decision analysis approach presented in this chapter.

2.1 Formulating Descriptive Decision Models

To illustrate the process of formulating descriptive decision models, we analyze three decision making scenarios in which information acquisition decisions are performed under different sampling conditions. The focus is specifically on two types of decisions: **the decision to select a design point** from a given design space and **the decision of stopping design exploration.** The specific examples of descriptive decision models are synthesized from the existing literature. Chapter 3 and 4 build on these descriptive models to present more wide-ranging models for the infor-

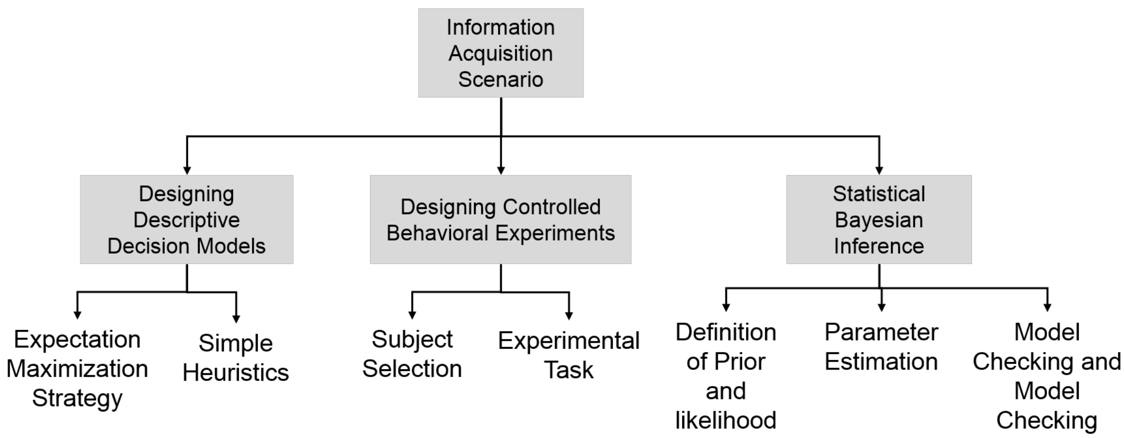


Figure 2.1. : Different elements of the descriptive decision analysis approach.

mation acquisition decisions including for the decision to communicate with others. The three decision making scenarios considered are as follows.

1. **One-shot decision making**: A decision maker evaluates the design performance only at a single design point. Zero observations of design variables and respective design performance attributes are available from the past.
2. **Sequential decision making without recall**: Prior observations of design variables and respective design performance attributes are available but an observation from the past cannot not be selected as a solution. That is, past observations are only useful for understanding of relationships between design variables and design performance attributes. Consider that $x \in \mathcal{X}$ is a design point in the space $\mathcal{X}$ and $f(x)$ is one-dimensional design performance at x. If past observations are represented by $\{x_h, f(x_h)\}_{h=1}^{i}$, then the current outcome of the sequential decision making process without recall is $y_i = f(x_i)$. Figure 2.2 presents a decision tree representation of the sequential decision making process.
3. **Sequential decision making with recall**: Prior observations of design variables and respective design performance attributes are available and one of the past observations can be selected as a solution. Under the sequential decision making process with recall, the outcome after i iterations is the best of observed design performances from the past, $y_i = \max\{f(x_1), \dots f(x_i)\}$.

Table 2.1 presents the summary of descriptive decision models for the one-shot decision making and sequential decision making scenarios.

2.1.1 Expected Utility Maximization

The fundamental premise of expected utility-based models is that design decisions should be rational, i.e., design decisions should follow the rule that the preferred decision is the option that has the highest expectation of value.

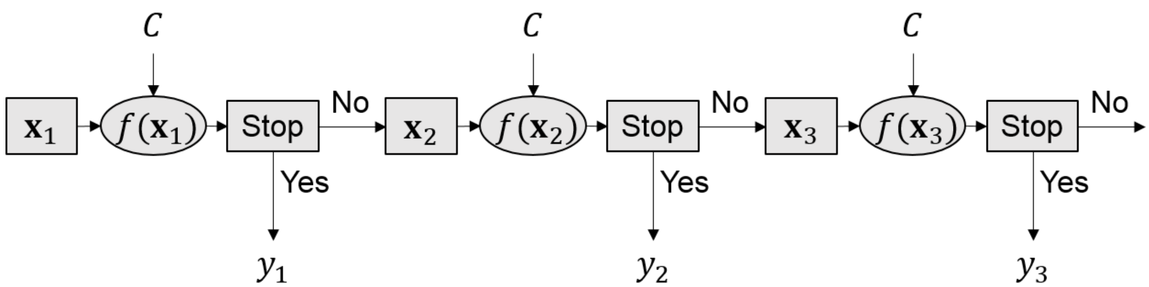


Figure 2.2. : A decision tree representation of the sequential decision making process. Inputs $\mathbf{x}_1, \mathbf{x}_2, \ldots$ are the selected design points, $f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots$ are corresponding design performances, $y_1, y_2, \ldots$ are outcomes which depend upon the design performances, and C is a fixed cost of evaluating design performance once.

Table 2.1. : The examples of descriptive decision models in one-shot decision making and sequential decision making scenarios.

Scenario	**Expectation Maximization Strategy**	**Simple Heuristics**
One-shot decision making	Select alternative that maximizes the weighted average of expected utilities (Eq.(2.3))	Dominance rule [56], Mental list rule [57], Lexicographic rule [56], Elimination by aspects [56], Take-the-best [58]
Sequential decision making without recall	Select an alternative that maximizes the expectation of gain (Eq.(2.9)) and Stop if the maximum predictive mean design performance is less than an evaluation cost (Eq.(2.10))	Representational stability rule; Magnitude threshold rule; Cutoff rule [**?**]; Successive non-candidate cutoff rule
Sequential decision making with recall	Select an alternative that maximizes the expectation of gain (Eq.(2.9)) and Stop if the maximum of maximum expectation of gain is less than an evaluation cost (Eq.(2.11))	Difference threshold rule [59]; Fixed sample rule [60]

One-shot decision making

For the case of one-shot decision making, the decisions are made in a single time period without observing consequences from similar decisions. One relevant frame-

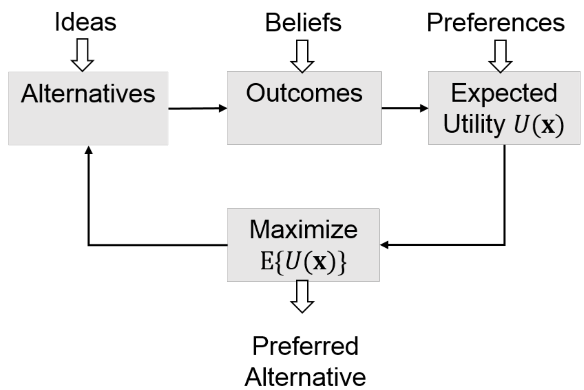


Figure 2.3. : The decision analysis framework.

work to consider such decisions is the **decision analysis framework**. Under this framework, all decisions involve three basic elements, a set of two or more alternative detailed designs (simply called alternatives) available to the decision maker, a set of beliefs that the decision maker holds with regard to the outcomes of each available alternative, and a set of preferences over all conceivable outcomes. The designer may list alternatives based on existing products, or may be expected to generate new alternatives based on ideas and prior knowledge. Similarly, the outcomes may be a priori known, or the decision maker may be expected to anticipate uncertain outcomes. If outcomes are uncertain, a decision maker is left to his/her personal beliefs about what is feasible. Finally, when multiple attributes are available, the key is to construct relationships between the overarching preference and the attributes that define individual alternatives. For example, if a designer's goal is to design an all-terrain vehicle, it will be difficult to meet requirements of high speed, long range, large payload capacity, comfortable and multiple seats, etc., as well as keep the cost of vehicle low at the same time. If the overarching goal is to achieve mobility over a highly rugged surface, the designer should prefer a vehicle with larger payload and range than one with more speed and comfort.

For **the decision to select the best alternative** from available alternatives, the suggested strategy is to evaluate each alternative's expected utility to the decision maker and pick the alternative with the largest utility. Consider that a design performance attribute f is uncertain and generates values $y_1, \dots y_S$ for a given alternative $\mathbf{x}$ with probability densities $p_1, \dots p_S$ respectively, then the expected utility of x with respect to the attribute f is:

$$\mathbb{E}\{f(\mathbf{x})\} = \frac{1}{S}\sum_{n=1}^{S} y_n p_n. \tag{2.1}$$

Further, suppose that a decision maker prefers M attributes, $f_1(\mathbf{x}), \dots, f_M(\mathbf{x})$ with preferences $\alpha_1, \dots, \alpha_M$ $(\sum_{j=1}^{M} \alpha_j = 1)$ and the attributes have respective expected utilities $E\{f_1(\mathbf{x})\}, \dots,$ $E\{f_m(\mathbf{x})\}$. Then, the expected utility of alternative $\mathbf{x}$ is the weighted average of expected utilities, and is given as:

$$\mathbb{E}\{U(\mathbf{x})\} = \sum_{j=1}^{M} \alpha_j \mathbb{E}\{f_j(\mathbf{x})\}. \tag{2.2}$$

Finally, the preferred alternative is the one that maximizes the expected utility.

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{X}} \mathbb{E}\{U(\mathbf{x})\} \tag{2.3}$$

In the above equation, the design space $\mathcal{X}$ represents the set of all available alternatives.

The decision on whether to stop design exploration without any prior information is equivalent to deciding whether to proceed with selecting *any* alternative. The suggested strategy for such decision is to continue selecting a new alternative if the highest expected utility is greater than or equal to the associated cost from time or resources to be consumed, and stop if it is not. If C is the cost associated with making a decision and $\max_{\mathbf{x}} \mathbb{E}\{U(\mathbf{x})\}$ is the highest expected utility possible, then the decision strategy is:

$$\text{Stop, if } \max_{\mathbf{x}} \mathbb{E}\{U(\mathbf{x})\} - C < 0; \text{Continue, otherwise.} \tag{2.4}$$

Sequential decision making without recall

In the sequential decision making scenario, **the decision to choose the next best design alternative or point** requires quantification of the value that each design point holds for improvement in the design performance (output) function. This quantification can be carried out in two steps, (i) quantifying the belief about design performance values over the unexplored design space, and (ii) quantifying the gain in the design performance function if it is evaluated at a selected design point.

The first step is generally performed using surrogate modeling techniques such as **Gaussian process regression.** A Gaussian process is a collection of multiple Gaussian distributions representing uncertainty in outputs over given input space [61]. A Gaussian process is a stochastic process which generate continuous output functions over the space of inputs. In the context of a design problem in Figure 1.2, the distribution of the design performance attribute (output) for given values of design variables (inputs) is a normal distribution.

When no prior observed data is available, a Gaussian process sets a prior probability measure over the output function space by specifying the mean and covariance functions ex-ante. Here, mean function $m(\mathbf{x})$ represents the expected value of design performance at $\mathbf{x} \in \mathbb{R}^N$ while the covariance function $k(\mathbf{x}, \mathbf{x}')$ models how close the design performance is for $\mathbf{x}$ and $\mathbf{x}'$ for given $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$. The covariance is mathematically defined as $k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$. A common form of covariance function is radial basis function:

$$k(\mathbf{x}, \mathbf{x}'|\mathbf{l}, \nu) = \nu e^{-\frac{1}{2}\sum_{d=1}^{N} \frac{(x_d - x'_d)^2}{l_d^2}}, \tag{2.5}$$

where n is the number of dimensions in the design space, $\mathbf{l}$ are length scale parameters of different dimensions, and ν is a variance parameter. Lengthscale parameters $\mathbf{l}$ represents the length of flatness/wiggles in the design performance function for different dimensions in $\mathbf{x}$. Variance ν is the is the noise in the values of design performance function at a particular design point. Thus, model parameters $\mathbf{l}$ and ν capture beliefs about the functional form of the design performance. In summary, the probability density over z being the actual function value at $\mathbf{x}$ is given by normal distribution with mean as $m(\mathbf{x})$ and variance $k(\mathbf{x}, \mathbf{x}|\mathbf{l}, \nu)$:

$$f(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}|\mathbf{l}, \nu)). \tag{2.6}$$

When past observations of design variable values and respective design performance are available, a Gaussian process can be used to estimate an unknown design performance function (output) conditional on the observed data. The updated model is referred to as Gaussian Regression Model. Assume that the past i observations are represented as $\mathcal{D} = \{\mathbf{x}_h, f(\mathbf{x}_h)\}_{h=1}^{i}$, then the point predictive distribution of the design performance function at any $\mathbf{x} \in \mathcal{X}$ is given as,

$$f(\mathbf{x}|\mathcal{D}, \mathbf{l}, \nu) \sim \mathcal{N}(\tilde{m}(\mathbf{x}), \tilde{\sigma}^2(\mathbf{x})). \tag{2.7}$$

Here, mean $\tilde{m}(\mathbf{x})$ and variance deviation $\tilde{\sigma}^2(\mathbf{x})$ are estimated such that the likelihood of the posterior prediction is maximized. This encodes beliefs about the design performance (output) function values after observing the data. The expressions for $\tilde{m}(\mathbf{x})$ and $\tilde{\sigma}^2(\mathbf{x})$, and related derivations are not presented here but can be found in [62, Ch.6].

The second step for modeling **the decision to choose the next design point** requires quantification of the value that each design point holds for improvement in the design performance function. The gain or loss over evaluating the design performance function at any $\mathbf{x} \in \mathcal{X}$ is quantified with respect to the current best alternative, $f^* = \max_{h=1:i} f(\mathbf{x}_h)$ as,

$$G(\mathbf{x}|\mathcal{D}, \mathbf{l}, \nu) = \max\{f(\mathbf{x}|\mathcal{D}, \mathbf{l}, \nu) - f^*, 0\}. \tag{2.8}$$

Note that the goal here is to maximize the design performance function and thus maximize the potential gain from selecting the next design point. Accordingly, the next design point $\mathbf{x}^*$ is selected such that **the expectation of gain** integrated over possible design performance values is the largest:

$$\begin{aligned} \mathrm{EI}(\mathbf{x}) &= \mathbb{E}[G(\mathbf{x}|\mathcal{D}, \mathbf{l}, \nu)] \\ &= \int_{f^*}^{\infty} (y - f^*) p(y|\mathbf{x}, \mathcal{D}, \mathbf{l}, \nu) dy \\ \mathbf{x}^* &= \arg\max_{\mathbf{x} \in \mathcal{X}} \mathrm{EI}(\mathbf{x}) \end{aligned} \tag{2.9}$$

This metric in widely known as the expected improvement metric in the Bayesian global optimization literature [27]. In the above equation, $p(y|\mathbf{x}, \mathcal{D}, \mathbf{l}, \nu)$ is the predictive probability density that the design performance at $\mathbf{x}$ is y and has a normal distribution defined by Eq.(2.7).

For **the decision to stop design exploration** under the condition of no recall, the expected utility maximization strategy suggest that stop if the maximum expected design performance is less than the associated cost of evaluating another alternative. If the design performance function has the uncertainty defined by a Gaussian process regression in Eq.(2.6), then the maximum expected design performance is the highest predictive mean over the design space $\max_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[f(\mathbf{x})] = \max_{\mathbf{x}\in\mathcal{X}} m(\mathbf{x})$. If the cost of evaluating a new design point is C, then the stopping strategy is:

$$\text{Stop, if } \max_{\mathbf{x}\in\mathcal{X}} m(\mathbf{x}) - C < 0; \text{Continue, otherwise.} \tag{2.10}$$

Sequential decision making with recall

In the scenario of sequential decision making with recall, a past observation of design point and the design performance function can be selected during any future time periods. For **the decision to choose next design point**, this scenario is equivalent of the the sequential decision without recall, since the learning from past observations is available in both scenarios. Additionally, the expected gain maximization strategy in Eq.(2.9) works on the all alternatives from the design space for selecting the next best design point.

For **the decision to stop the design exploration or not** under the condition of recall, the expected utility maximization strategy compares the current best design performance available from prior observations with the design performance possible from an additional evaluation. The evaluating the difference between the current best performance and the expectation of future design performance is equivalent to evaluating the expectation of gain in Eq.(2.9) since the the current best performance is fixed. Therefore, the stopping strategy is to stop if the expectation of gain is less than to the cost associate with an additional evaluation. If the cost of evaluating a new design point is C, then the stopping strategy is:

$$\text{Stop, if } \max_{\mathbf{x}\in\mathcal{X}} \mathrm{EI}(\mathbf{x}) - C < 0; \text{Continue, otherwise.} \tag{2.11}$$

2.1.2 Heuristic Decision Making

Heuristic decision strategies for the decision to select an alternative and deciding to stop are often correlated. This is because the simple heuristics are fast and frugal and they aim to select alternatives as early and cheaply as possible. Simple heuristics are ubiquitously used in decision making [58] and can have many different forms. This section only illustrates most relevant simple heuristics in the context of information acquisition decisions. The simple heuristics presented here are drawn from the literature, which are appropriated referenced through out the section. This dissertation work also presents additional simple heuristics for the information acquisition decisions in Chapter 3 and Chapter 4.

One shot decision making

For the decision to select a design point using one-shot decision making, some examples of the simple heuristics are as follows.

1. **Dominance Rule** [56]: Choose the design alternative which is better than other alternatives on at least one design performance attribute and not worse than others on all other attributes. If none of the alternatives satisfy this condition, then this strategy does not recommend any decision.
2. **Mental List Rule** [57]: In this variation of the dominance rule, the decision maker compares design alternatives on few preferred design performance attributes and selects the dominant alternative. The selected attributes are based on the decision maker's preferences and therefore the strategy is referred to as mental list.

3. **Lexicographic Rule** [56]: It is also possible that a decision maker uses a single attribute for comparing different alternatives and selecting the alternative that performs best on that attribute.
4. **Elimination by Aspects Rule** [56]: Exclude design alternatives that do not exceed a threshold on a preferred design performance attribute. Repeat this procedure other design performance attributes in the order of importance. Select an alternative left at end of this process. Such process will yield an alternative that meets preset thresholds on all design performance attributes. If none of the alternatives meet all thresholds, then no decision is recommended.
5. **Take-the-best Rule** [58]: When comparing two design alternatives, choose the alternative that has a better value on the most preferred design performance attribute. If both alternatives have the same value, repeat the procedure with the second design performance attribute in the order of preference. This decision making process stops as soon as the first design performance attribute that differentiates the two attributes is found.

Simple heuristic strategies typically consider no uncertainty in the design performance attributes. However, these strategies may still be applied to one-shot decisions under uncertainty by replacing the design performance attributes with their utilities as defined in Eq.(2.1). The important difference between the simple heuristic strategies and the expected utility maximization strategy from Eq.(2.3) is that the simple heuristic do not compare alternatives along all given design performance attributes. Being selective in the useful attributes and in the order in which attributes are evaluated allows the simple heuristics strategies to reduce the requirement on time and resources used. In certain circumstances, simple strategies can yield more accurate judgments than alternative strategies such as the expectation maximization that use more pieces of information. This effect is called **the-less-is-more effect** [63].

Sequential decision making without recall

In contrast to the one-shot decision making, the sequential decision making without recall queries multiple values of design variables and observes the design performance attributes sequentially one after the other. In the condition of no recall, the decision maker may not select past observations for submission without running a query. That is, each observation is only available for one time period. Once the decision maker decides to stop the exploration, he/she selects the latest observation for submission. Examples of heuristic decision making strategies for this scenario are as follows:

1. **Representational Stability Rule**: Choose an alternative that provides a greatest change in the decision maker's mental model or representation of the design variables and design performance relations. Stop after the mental model representations stops shifting. The focus is on the stability of the representation. This strategies in theory can be implemented using Gaussian processes as the representation of the design variables and design performance relations. Some research shows that the Gaussian processes represents human function learning much better than traditional optimization algorithms [64].
2. **Magnitude Threshold Rule**: Stop after the encountered design performance is greater than the pre-defined magnitude. The magnitude threshold rule only compares the present design performance with the pre-defined magnitude.
3. **Cutoff Rule** [65]: Do not stop the exploration for at least fixed k number of observations. After k observations, stop at the first observation which has design performance larger than the maximum design performance of first k unused observations. When the total number of observations are fixed, say number M, and the all observations are equally likely to produce the best alternative, the optimal number of unused observations should be $k = M/e$ [66].
4. **Candidate Count Rule**: Select the k^{th} observation irrespective of the design performance. This strategy ignores judgments about the design variables

and design performance relationships as well as it fails to maximize the design performance. However, this simplest strategy may be applicable in rare circumstances when the budget for design exploration is limited and the technical complexity is too high for systematic exploration.

5. **Successive Non-Candidate Cutoff Rule**: Select the first encountered observation after observing k consecutive observations with less than desired design performance. This dynamic decision making strategy depends on the sequence of observations encountered unlike the previous decision making strategies. Therefore, it is suitable when the properties of design point and design performance relationship change over time are expected.

Sequential decision making with recall

With the ability to retrieve previous evaluations, a decision maker has larger flexibility in selecting the best design performance for submission. Having access to past observations allows stopping early if a decision maker believes no further improvement in the design performance is possible. The decision strategies from the sequential decision making scenario without recall still apply to this scenario because a decision maker can always decide to ignore past observations. Heuristics that benefit from the access to prior observations are listed below. Under these decision strategies, the best alternative is selected at the end of the sequential decision making process from the pool of observed evaluations.

1. **Difference Threshold Rule** [59]: A decision maker sets a prior difference to gauge when the new observations produce increasingly small gains compared to the current best design performance. This set of strategies are similar to the stopping strategy with expectation maximization (Eq.(2.11)) but the monetary cost of additional observation C replaced with an individual-specific threshold value. Unlike the magnitude threshold rule, the difference threshold rule com-

pares each encountered design performance with the best design performance from previous observations.

2. **Fixed Sample Rule** [60]: Stop after querying a fixed number of design variable values. Select the observation that has the best design performance. In this strategy, a decision maker ignores the learning from past observations while deciding whether to stop or not. As a result, the fixed sample rule is an extension of one-shot decision making where more than one evaluations of design performance are allowed and the number of evaluations must be fixed and defined prior to the start any evaluation.

2.2 Designing Controlled Experiments

It is necessary to observe information acquisition decisions using controlled experiments to test different hypotheses embed in descriptive decision models about how individuals make these decisions. Things to consider while designing and executing a controlled experiment include selection of human subjects, incentivization, design of experimental design task and experimental conditions. Building on the existing literature, this section provides guidance on different elements of experiment design.

2.2.1 Subject Pool Selection: Consideration of Design Expertise and Incentives

The subject selection should be carried out such that the subjects' expertise aligns with the problem complexity and difficulty. For a general design context, Dorst [67]'s expertise framework is useful for thinking about the representativeness of subject-task interactions. Using the expertise framework, the answers to questions such as what do we mean by design activity in a model world become immediately clear once we specify subjects and tasks. Dorst categorizes such interactions into seven levels:

1. **Naive designer** makes one-off choice from available options

2. **Novice designer** follows strict rules or a formal process to meet fixed requirements given by exerts
3. **Advanced designer** adapts a formal process for considering situational aspects,
4. **Competent designer** selects relevant situation aspects and chooses a plan to achieve a goal,
5. **Expert designer** recognizes high-level patterns from years of experience and responds to a given task intuitively,
6. **Master designer** represents a new knowledge in the field,
7. **The visionary** consciously strives to extend the domain of their work.

According to the single attribute decision making approaches in engineering design [26, 68], the economic attribute is the only attribute consistent with decision analysis framework. Then, the incentives that subjects receive should be proportional to the design performance they generate and inversely proportional to the total cost of resources they spend on evaluations. Let us consider a few cases for the incentive schemes.

1. **Requirement-based incentives**: Pay a subject fixed reward, but only if they find a design point with the design performance greater than a threshold value. This is the most common form of payment in real applications. But there are two potential problems that should be addressed: (i) there should exist a design point with better properties than the pre-defined threshold; subjects may not engage in the decision making process if they believe that there is no chance that required design performance is achievable, and (ii) the requirements-based payment does not work in case of large uncertainties. This is because, out of pure chance, we may get a very large posterior mean at certain points. We should add a constraint that the incentive is to be paid only if the posterior uncertainty at this design is smaller than the measurement noise.
2. **Paying proportionally to the best design performance**: The better the discovered design, the more we pay the subject.

In both cases, there must be some penalty for delaying to report. Of course, if the cost is evaluated in the same units, this penalty can just be the current cost of information acquisition to be subtracted from the design performance.

2.2.2 Experiment Design: Consideration of Human Biases

Even with appropriate subject population and setting up incentives, safeguarding the experiment design against potential cognitive biases is important so that the behaviors resulting from such biases do not interfere with intended behaviors. A cognitive bias is systematic deviation from logical decisions in day-to-day human decision making. They are likely to seep into the information acquisition decisions and result in illogical decisions. In order to maintain the desired goals of an experiment design, the cognitive biases should be avoided using experimental “tricks”. This section reviews some tricks used to mitigate cognitive biases observed in the context of economic decisions [69, 70].

1. **Use theory to guide the experimental design**: One way to understand the causes of human behaviors in the experiments is to start with a theory. Following a theory allows to anticipate alternate behaviors and it allows to use theoretical insights for designing experiments. For example, in the context of economic decisions, a theory can be defined based on logic, probability rules, and mathematics. The predictive behaviors from such a theory will assume that a decision maker follows the rational constructs of probability and mathematics. A experiment based on such theory can be used to study the cases when rationality holds and when human decision makers deviate from rationality.
2. **Have a control group**: For every independent variable whose effect is being experimentally analyzed, include two experimental conditions, one where subjects are treated with the study variable and other where subjects are free of the study variable. Further, randomized treatments are important in experiment design to study small effects in human behaviors. Therefore, subjects

should be assigned to different experimental conditions randomly. If randomized experiment design is not possible, for example, due to a realistic, large-scale experiments, consider quasi-experimental designs. Shadish et al. [71] present quasi-experimental designs when one or both of a control group or pretest samples are lacking.

3. **Consider the order of treatment conditions**: It is advisable to repeat an experiment with all possible orders of the experimental conditions. The order in which experimental conditions are presented to subjects may have unintended effects on their behaviors. For instance, if an experiment studies the differences in learning in the design process due to two different decision support tools, then independent sessions with two different sequences of decision support tools should be implemented. Further, the correlated tasks between consecutive studies should be avoided. If two consecutive conditions require conflicting skills, e.g., using two different coding languages, then using one coding language may inhibit the skills required for using the other coding language.
4. **Provide monetary incentives at experiment's end**: Unless an experiment design involves study of incentive effects, the payment should be provided after the completion of the experiment to avoid endowment effects [72]. Paying subjects at the start or in middle of the experiment may discourage effort in subsequent tasks. Also, if the payment is proportional to the performance, it should be calculated from the performance on a randomly selected decisions. Achieving sufficient payable performance early in the experiment may make subsequent tasks unnecessary for subjects. Finally, subjects should be paid the amount commensurate of the total time they spent. Paying too small payment or delaying payment until too late after completion may prevent subjects from participating in future experiments.

2.3 Statistical Bayesian Inference

After defining alternate descriptive decision models and gathering data from a controlled experiment, the final data analysis step involves defining a Bayesian workflow for posterior estimation of model parameters. Let $\{z_s\}_{1:S}$ denote the observations of an information acquisition decision. Let $\mathcal{M}$ be one of the decision models for that information acquisition decision, and θ be the corresponding model parameters. We seek to estimate the marginal likelihood of the observations $p(z_{1:S}|\mathcal{M}_k)$ and posterior distribution over model parameters $p(\theta|z_{1:S})$.

2.3.1 Pooled vs. Hierarchical Prior Distributions

In the Bayesian workflow, the first step is to define prior distributions over model parameters. There are multiple ways to define a prior distribution such as pooled model, hierarchical model, and non-centered hierarchical model. Assume there are P individuals in the subject population and θ_p represents a model parameter capturing the strategy of individual p. In **a pooled model**, $\{\theta_p\}_{p=1:P}$ are independent samples from same distribution, say $\mathcal{D}(\gamma)$, with fixed parameters γ. Distribution $\mathcal{D}$ is the collection of all individuals' decision strategies. Fixed parameters γ mean that the decision making strategies are constant at the population level.

$$\theta_1, \ldots, \theta_P \sim \mathcal{D}(\gamma) \tag{2.12}$$

In **a hierarchical model**, the individual-specific strategies $\{\theta_p\}_{p=1:P}$ are samples from a distribution $\mathcal{D}(\gamma)$ but the distribution's parameters, i.e., hyperparameters γ are uncertain. Then, hyperparameter γ is assigned a prior distribution $\mathcal{G}(\alpha_\gamma)$, called hyperprior, which allows variations in the population-level decision strategies.

$$\begin{aligned} \gamma &\sim \mathcal{G}(\alpha_\gamma); \\ \theta_1, \ldots, \theta_P &\sim \mathcal{D}(\gamma) \end{aligned} \tag{2.13}$$

In **a non-centered hierarchical model**, the individual-specific strategies have systematic deviation from the average decision making strategies of the subject population. This allows modeling of clusters with individuals having better and worse decision making strategies than the average strategies of the population. Let μ be the population mean and σ be the population standard deviation. Let τ_p represent individual p's offset from the population mean. Since these hyperparameters are unknown, we place probability distributions over then to signify uncertainty. Let the hyperpriors over μ, σ, and τ_p be $\mathcal{G}_\mu, \mathcal{G}_\sigma$,and $\mathcal{G}_{\tau_p}$ respectively. Then, the individual specific strategy θ_p is calculated as follows:

$$\begin{aligned} \mu &\sim \mathcal{G}_\mu(\alpha_\mu); \\ \sigma &\sim \mathcal{G}_\sigma(\alpha_\sigma); \\ \tau_p &\sim \mathcal{G}_{\tau_p}(\alpha_p); \\ \theta_p &= \mu + \tau_p \sigma \end{aligned} \tag{2.14}$$

where $\alpha_{mu}, \alpha_{sigma}$ and α_p are fixed parameters of the hyperpriors.

2.3.2 Defining Likelihood Functions

The next step is define the probability of observing a particular decision given the model and its model parameters. Such probability function is also called likelihood function. A likelihood function captures the errors in an individual's information acquisition decisions. By representing decision strategies as stochastic rather than deterministic, the likelihood function allows for the possibility that an individual is can deviate from intended decision strategy. There are two types of likelihood functions often used for representing the information acquisition strategies: i) **Logistic function**, and (ii) **Softmax function**. The logistic function is suitable for modeling threshold-based decision rules where a decision is dependent whether or not greater than than a threshold value of a predictor variable. The Softmax function is suitable for modeling maximization decision strategies where a decision is to select an alternative with the maximum value of a predictor variable.

Suppose $y_{\mathcal{M}}$ is a predictor variable (e.g. design performance attribute) for given model $\mathcal{M}$. The decision under consideration involves two alternatives, for example, an individual decides whether to stop or not. Let us say the decision is $z = 1$ if the individual stops and $z = 0$ when he/she does not stop. Then, the probability of stopping is a logistic function of z given by:

$$\Pr(z = 1 | y_{\mathcal{M}}, a, b) = \frac{1}{1 + e^{a(y_{\mathcal{M}} - b)}}, \tag{2.15}$$

where a and b, respectively, are the slope and threshold parameters of the logistic function. This definition of likelihood can be extended to the multiple predictors by using the weighted average of the individual predictors as a single variable. Further, consider a situation where predictor variable $f_{\mathcal{M}}(x)$ represents the value of alternative x under model $_{\mathcal{M}}$. An individual decides to select the alternative with the best value. Then, in stochastic representation, the likelihood of selecting an alternative is proportional to its value and given by the Softmax function as:

$$\Pr(\mathbf{x}^* = \mathbf{x}) = \frac{e^{f_{\mathcal{M}}(\mathbf{x})}}{\sum_{\mathbf{x}' \in \mathcal{X}} e^{f_{\mathcal{M}}(\mathbf{x}')}}, \tag{2.16}$$

where $\mathcal{X}$ is the space of all alternatives. In case of multiple value functions, the predictor variable can be defined as the weighted average of the value functions, $U(x) = \alpha_1 f_1(\mathbf{x}) + \ldots \alpha_M f_M(\mathbf{x})$.

2.3.3 Posterior Estimation and Model Validation

Various Markov Chain Monte Carlo (MCMC) algorithms are commonly used [73] for sampling from posterior distributions. It is recommend to use *variational Bayes* methods for finding approximations to posterior distributions of the model parameters as well as estimating lower bounds to marginal log-likelihoods of the decision models conditional on the experimental data [74]. The variational Bayes approach is especially useful in complex stochastic models where analytical forms of posterior distributions are intractable. The model evidence lower bound (ELBO) quantifies the the support for a model, i.e., the accuracy with which a model represents the experimental data. We denote ELBO, say for model $\mathcal{M}$, as $\mathcal{L}$.

The goal of inference using variational Bayes is to approximate posterior distribution of model parameters θ, represented by $p(\theta|z_{1:S}, \mathcal{M})$, using a family of distributions $q(\theta|\upsilon)$ with its own variational parameters. Distributions $q(\theta|\upsilon)$ do not typically include the exact posterior distribution, however they are easy to sample from. We assume that all parameters are independent of each other. Under this assumption, $q(\theta|\upsilon)$ is factorized into distributions over individual parameters, with each distribution being a normal distribution with υ representing its mean and variance. This approximation is called the *mean field approximation.* Accordingly, rather than maximizing the marginal log-likelihood of $z_{1:S}$ directly, the inference using variational Bayes maximizes a lower bound to it, called evidence lower bound (ELBO), calculated as:

$$\mathcal{L} = \mathbb{E}_q[\log p(z_{1:S}|\mathcal{M}, \theta)] - KL[q(\theta)||p(\theta|z_{1:S})] \tag{2.17}$$

where the first term is the expectation of log-likelihood of observations with respect to distribution $q(\theta)$. The second term $KL[q(\theta)||p(\theta)] = \mathbb{E}_q\Big[\log \frac{q(\theta)}{p(\theta)}\Big]$ is the Kullback-Leibler (KL) divergence, which represents the closeness of distribution $q(\theta)$ and $p(\theta)$. This setup using KL divergence enables the independence between marginal log-likelihood $p(z_{1:S}|\mathcal{M})$ and distribution $q(\theta)$. As a result of that, maximizing the ELBO in Eq. (2.17) is equivalent to minimizing the KL divergence between $q(\theta)$ and $p(\theta|z_{1:S})$. Therefore, the closer is $q(\theta)$ to $p(\theta|z_{1:S})$, the lower is the KL divergence between the two, and the closer is the ELBO to $\log p(z_{1:S}|\mathcal{M})$, i.e., the marginal log-likelihood of observations.

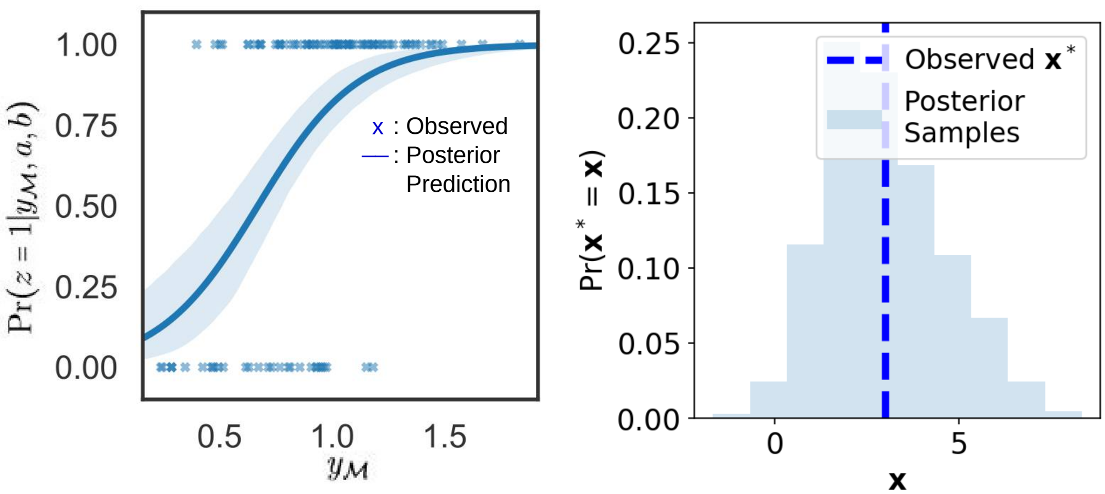
After estimating the posterior distributions of model parameters, it is important to verify how well the model with posterior parameter estimates fits the observed decision data. The literature on the topic provides formal and informal diagnostic metrics for analyzing the convergence of Markov Monte Carlo Methods (MCMC) [75]. Formal MCMC diagnostic metrics such as auto-correlation plot and Raftery-Lewis estimates [76] are important aggregate indicators of the convergence. The informal methods such as plotting the sample trace of MCMC and visualizing of posterior [77] are more straightforward approaches to investigate how the MCMC searches the model parameter space. This section presents a mix of formal and informal the model checking approaches relevant for the inference of decision strategies.

1. **Data partition into training and testing**: An important first step in model checking is to separate out some fraction of experimental subjects' decision data before running statistical Bayesian inference. The separated data is called testing data and the data used to train the model parameters is called training data. So, the model performance is checked on the testing data which is unseen by the model.

   How to partition data depends on the experimental conditions that generated the data. Typically, a large part of the total data such as $80 - 95\%$ is assigned to the training dataset and the remaining assigned to testing data. There can be different approaches to split the data into training and testing. First, the data can be divided based on individuals, so the model performance can be evaluated on new individuals' decisions. This approach however assumes that all individuals follow the same decision strategies. Second, an individual subject's data can split along independent decision samples. Very often, an experimental task on the iterative information acquisition requires an individual to repeat the same decision, thereby generating independent samples of that decision.

2. **Posterior predictive checks**: Once the posterior parameter estimation is completed, "sanity checks" can be performed by visualizing the posterior predictions of decisions along with the observed decision data on same plot. This step is typically referred to as **posterior predictive check**. It provides a visual check for how well the model fits the data. Consider the examples of posterior checking for logistic likelihood function and for a general case of continuous decision alternative in Figure 2.4.

3. **Accuracy score**: The posterior predictive checks can be further formalized by calculating comparison measures such as accuracy score. The accuracy score is the fraction of posterior predictions exactly that match with the observed data.



(a) Logistic Likelihood

(b) Softmax Likelihood

Figure 2.4. : Examples of posterior predictive checks for Bayesian models.

Let z be an observed decision and $\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_S$ be the model predictions based on posterior parameter estimates. Then, the accuracy score is defined as:

$$\text{Accuracy Score} = \frac{1}{S}\sum_{n=1}^{S} \mathbf{1}_{\hat{z}_n = z}, \tag{2.18}$$

where the indicator function $\mathbf{1}_{\hat{z}_n = z}$ is 1 if the prediction $\hat{z}_n$ equals the observed value z, and 0 otherwise. The accuracy score measure is appropriate for decisions with finite number of alternatives such as the threshold-based decision strategies. For a continuous alternative space, the probability that a predictive sample exactly equals an observed decision is always 0.

4. **Information Criteria**: The concept of accuracy score measure can be extended to the decisions with a continuous decision space by evaluating the posterior log-probability of observations using the posterior samples of model parameters. Let us consider predictive accuracy for a single data point. If $\theta_1^{\text{pos}}, \theta_2^{\text{pos}}, \ldots, \theta_S^{\text{pos}}$ represent the posterior samples of model parameters, then the expected log-probability data z is calculated as $\frac{1}{S}\sum_{n=1}^{S} \log p(z|\theta_n^{\text{pos}})$, where $p(z|\theta_n^{\text{pos}})$ is the likelihood of the decision z given model parameters θ_n^{pos}. Recall the likelihood function is defined either using logistic function (Eq.(2.15)) or softmax function (Eq.(2.16)).

   Model accuracy measures based on posterior log-probability density typically involve a correction for the number of parameters in the model. This correction term penalizes the model likelihood functions with large number of parameters since such models are likely to overfit the observed data. Common information criteria used for bias correction in practical Bayesian applications are Akaike Information Criterion [78], Deviance Information Criterion [79, 80], and, more recently, Wanatabe-Akaike Information Criterion [81]. Ref. [82] reviews these information criteria from a Bayesian perspective. The implementation of these Bayesian accuracy measures is available in PyMC3 library [83] on Python framework.

2.3.4 Model Comparison

We can formally compare how well different multiple descriptive decision models fit given data by comparing their model evidence or evidence lower bounds. Let $\{\mathcal{M}_k\}_{1:K}$ be K independent candidate models for an information acquisition decision. Let $\{\theta^{(k)}\}_{1:K}$ be the model parameters for the given models. Let $p(\theta^{(k)}|\mathcal{M}_k)$ denote the prior distribution over parameters $\theta^{(k)}$ of model $\mathcal{M}_k$. The prior probability distribution over model parameters may be selected based on expert knowledge, or assumed non-informative if no prior information is available. Let $p(\mathcal{M}_k)$ be the prior probability of model $\mathcal{M}_k$ such that $\sum_{1:K} p(\mathcal{M}_k) = 1$. If $\{z_i\}_{1:S}$ denote the observations, then the posterior probability of model $\mathcal{M}_k$ is:

$$p(\mathcal{M}_k|z_{1:S}) = \frac{p(z_{1:S}|\mathcal{M}_k)p(\mathcal{M}_k)}{p(z_{1:S})}, \tag{2.19}$$

Under the assumption that all models are equally likely to represent the observations a priori, the prior probability of model $\mathcal{M}_k$, $p(\mathcal{M}_k)$, is $\frac{1}{K}$. Since $p(\mathcal{M}_k)$ and $p(z_{1:S})$ are both constants in such a case, the posterior probability of model $\mathcal{M}_k$ can be substituted by the marginal likelihood of observations.

$$p(\mathcal{M}_k|z_{1:S}) \propto p(z_{1:S}|\mathcal{M}_k) \tag{2.20}$$

Recall, evidence lower bound (ELBO) in Eq.(2.17) estimates the logarithm of the marginal likelihood of observations.

3. SEQUENTIAL INFORMATION ACQUISITION OF AN INDIVIDUAL DESIGNER

The first research study in the dissertation analyzes the design exploration decisions of an individual designer in sequential information acquisition process. The primary research objective is to identify models that provide the best description of a designer's decisions when multiple information sources are present and the total budget is limited. This study specifically focuses on,

1. identifying designers' decision strategies, and
2. evaluating how the amount of fixed budget and payment incentives such as bonuses proportional to the budget saved affect those strategies.

This work has been published as a peer-reviewed journal paper in Ref. [84].

We begin with an abstract the design process called a sequential information acquisition process [23] where design is considered a problem solving activity with known design parameters and evaluation criteria, but unknown mapping between the two. Alike many design situations, a designer's objective in this situation is to find the design parameters values that maximize the performance (see Figure 3.1). To achieve this objective, the designer performs iterative evaluations of performance. This process is constrained by a fixed budget, which limits the number of design evaluations. In general, the budget type may be financial (e.g., fixed cash or capital), or technical (e.g., fixed computational resources, time, or energy) [85].

Many examples of this design situation exist. For example, in the control problem for a room heating system, a designer finds the temperature setpoint that minimizes energy consumption while maintaining thermal comfort [86]. Also consider the design of superconducting materials such as $Cu_xBi_2Se_3$ where a designer finds the dopant composition (x) that maximizes superconductivity through a series of magnetization experiments [87]. In the airfoil shape optimization, a designer is concerned with find-

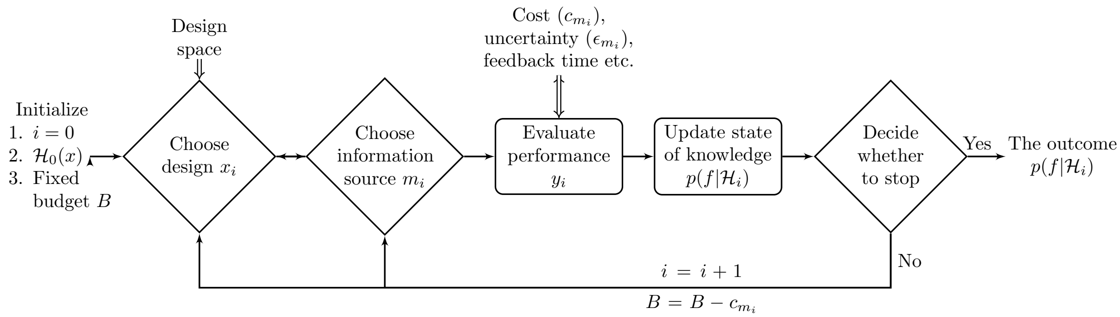


Figure 3.1. : Sequential information acquisition process with multiple uncertain information sources and fixed budget.

ing favorable values of the maximum thickness and the angle of attack that minimize the drag coefficient [29].

The process of iterative design evaluations, here referred to as *information acquisition*, is typically performed with the help of multiple prototypes (information sources) with different cost and uncertainty. The practice of prototyping is useful in learning whether or not the design options satisfy the requirements, however, this learning is hindered by the uncertainty associated with prototypes. For example, in automotive crash tests, computer-based simulations are flexible, but expensive physical prototypes are needed to detect unanticipated phenomena [18]. This uncertainty can be aleatory, see the major effect of the large noise due to process variability in VLSI semiconductor manufacturing plants [19], or epistemic, see how the lack of knowledge about the quantities of interest restricts the precise performance assessment for complex systems [88].

We model the case of multiple information sources assuming that it is possible quantify the uncertainty associated with information sources. Various front-end methods for quantifying uncertainty 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 [89].

First, we use $\mathcal{X}$ to denote the space of all possible designs and x to denote a point in the space $\mathcal{X}$. The *performance function* $f(x)$ is a scalar function of the design, i.e., $f : \mathcal{X} \to \mathbb{R}$. 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 $\{1, \ldots, M\}$ 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 modeling the measurement uncertainty.

We assume that the designer performs information acquisition sequentially. At step i the designer queries the information source m_i at a point $x_i \in \mathcal{X}$ using what they have learned from the currently observed history of evaluations $\mathcal{H}_{i-1} = \{(x_h, m_h, y_h)\}_{h=1}^{i-1}$, where $\mathcal{H}_0 = \varnothing$ along with any prior beliefs (see Figure 3.1). 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. Though the design process may comprise of parallel and sequential queries [90, 91], we focus on sequential queries as a first step towards modeling information acquisition in engineering design.

After observing performance measurement at the end of each step, the designer decides whether to continue or stop the evaluations. If they decide 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 designer decides to stop, the outcome of the design process is the most recent state of knowledge about f, denoted formally with the probability measure $p(f|\mathcal{H}_i)$.

3.1 Experiment Design

3.1.1 Objective Function Maximization Task

In the experiment task, a *designer* makes design decisions, and a *user interface* processes acquired information in the back-end to display the state of knowledge about the design performance in a visual format. The designer's objective is to find a point that maximizes the unknown design performance. The designer receives payment as a function of the outcomes of their decisions, while the user interface is representative of visual and text-based aids commonly used in design of manufacturing systems [92,93]. The roles of designer and the user interface are separated to maintain uniformity in how all designers process the acquired information.

Assumption 1 (Continuous design space). The design performance $f(x)$ is a one-dimensional continuous function of a single design parameter x.

Assumption 2 (Sequential evaluation process). The designer evaluates multiple designs sequentially. Each evaluation takes one unit of time to run, during which the designer may not begin another evaluation.

Assumption 3 (Multiple uncertain information Sources). The designer evaluates 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 with $m = 2$ or "H". We denote the total number of low (high) fidelity observations at step i by $n_{i,L}$ ($n_{i,H}$). It is $n_{i,L} = \sum_{j=1}^{i} 1_{\{L\}}(m_j)$ ($n_{i,L} = \sum_{j=1}^{i} 1_{\{H\}}(m_j)$), where $1_A(\cdot)$ is the indicator function of the set A. Of course, we have $i = n_{i,L} + n_{i,H}$. An example of a low fidelity source is a 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 test specimen.

Assumption 4 (Gaussian measurement uncertainty). The measurement process is modeled as a Gaussian distribution centered at the true (but unknown) performance function. That is, the measurement y conditioned on the design x and the information source m is:

$$y|x, m \sim \mathcal{N}(f(x), v_m^2). \tag{3.1}$$

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

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. 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 $f(x)$,

$$f|\mathcal{H}_0 \sim p(f|\mathcal{H}_0) := \mathrm{GP}(0, k), \tag{3.2}$$

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

$$f|\mathcal{H}_i \sim p(f|\mathcal{H}_i) := \mathrm{GP}(\mu_i, k_i), \tag{3.3}$$

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 $x_{1:i} = (x_1, \ldots, x_i)$ to be the vector of the first i designs, and $y_{1:i}$ 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 $x_{1:i}$, and Σ is an $i\times i$ diagonal matrix with elements, $V_i = \mathrm{diag}\left(v_{m_1}^2, \ldots, v_{m_i}^2\right)$. Then, the point predictive probability of $f(x)$ is:

$$f(x)|\mathcal{H}_i \sim \mathcal{N}\left(\mu_i(x), \sigma_i^2(x)\right), \tag{3.4}$$

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 5^{th} and 95^{th} percentiles. This way the acquired information gets processed and visualized in the same manner for all subjects and observing the subjects' decisions remains the the focus of the experiment task.

Assumption 7 (Fixed budget) The designer can spend at most B on performance evaluations. The designer may stop before exhausting the entire budget B, so the total cost incurred, $b_i = c_L n_{i,L} + c_H n_{i,H}$, is less than or equal to B.

Assumption 8 (Performance-based payment) The designer's payment includes a fixed salary I_0, and bonuses proportional to the best high fidelity observation and the budget saved. The designer payment after i evaluations is:

$$I_i = I_i(\mathcal{H}_i, B) = I_0 + 1_{[1,\infty)}(n_H)\left[H_f(y_i^*) + H_b(B - b_i)\right], \tag{3.5}$$

where H_f is the bonus from the best high fidelity measurement $y_i^* = \max_{h:1\leq h\leq i, m_h = H} y_h$, and H_b, the bonus from remaining budget $B - b_i$. At least one high fidelity measurement is required to receive bonuses because high fidelity measurements are valued more than low fidelity observations.

3.1.2 Subjects, Treatments, and Payment

A total of 63 student subjects were recruited from an introductory machine design class. The participation was voluntary and was not considered towards students' grades.

In the experiment, the subjects performed 18 runs of the experimental task each with a distinct performance function. A run of the experimental task was called *a period*. The objective in each period was to find a maximum of an unknown function. The 18 distinct functions were randomly generated prior to the experiment (see Figure 3.2) and fixed for all subjects. These functions exhibited the same level of difficulty in finding the maximum values as they were generated from a Gaussian process with fixed lengthscale parameter of the covariance function (see Eq.3.2). The

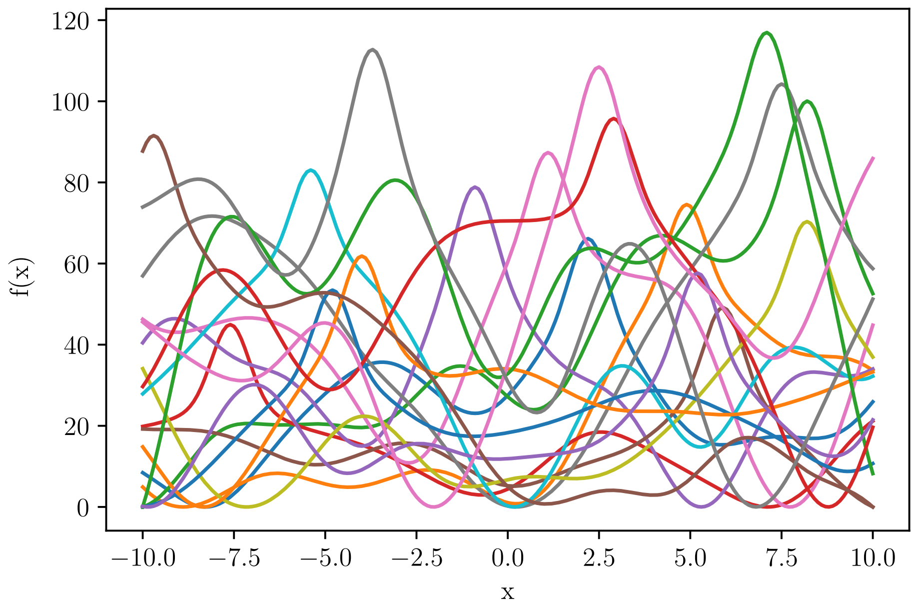


Figure 3.2. : 18 functions used as unknown design performance functions in the experiment

assignment of these functions to periods was randomized for every subject to minimize potential confounding between functions and treatments. Some parameters of the experimental task were fixed. In particular, the 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 salary $\mathcal{I}_0 = \$5$. The user interface was as shown in Figure 3.3. 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* (2 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 towards the subjects' payment.

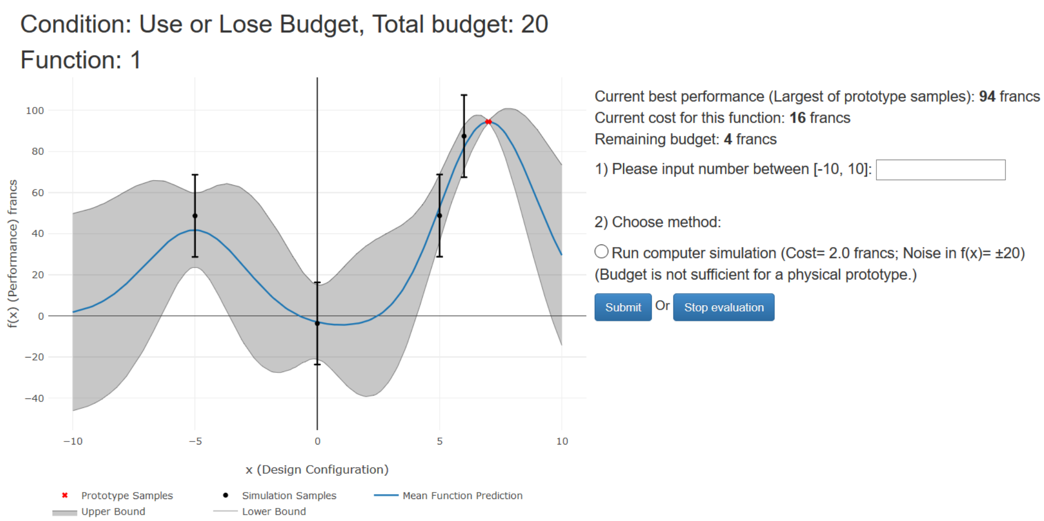


Figure 3.3. : A screenshot of the user interface developed using oTree [94].

2. *Use-it-or-lose-it part* (9 periods): For this part, a subject was 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 9 unknown functions in 9 periods, with 3 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* (9 periods): This part was similar to the use-it-or-lose-it part except any remaining budget at the end of every period was added to the subject's payment as a bonus H_b. A subject evaluated 9 unknown functions in 9 periods, with 3 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 (i) choosing the next design x, (ii) choosing between information sources, and (iii) deciding when to stop.

The order of treatments was varied across the subjects to control for *order effects* [95]. The four different orders of six treatments were: (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

$$G(y_i^*) = 100 - (f_{\max} - y_i^*) \text{ francs,} \tag{3.6}$$

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 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.

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 evaluation or 1 otherwise. We also recorded the related quantities such as gross payoff, fixed budget, functional performances, and the index of evaluation i associated with every evaluation. Additionally, we recorded the text of the subjects' survey responses. These descriptions inspired the formulation of some of the descriptive decision models given in Section 3.2.

3.2 Descriptive Models of Design Exploration Decisions

We formulate the descriptive models of information acquisition decisions based on a range of rational and heuristic strategies, which are derived from the literature and the subjects' responses to the survey questions. Existing models of rational decision making are rooted in the expected utility theory [8, 9, 11, 12] and the game theory [22, 38, 52], whereas heuristic models specify simple rules that humans use to make decisions. These rules are based on the premise that humans are cognitively limited, and therefore more likely to use a small number of available cues (criteria) to make decisions [58]. These simple rules are fast and frugal, in that they consume less time and are computationally cheaper. Examples of simple rules for stopping are "stopping if the most preferred cue is found" (take-the-best), "stopping if a fixed number of criteria are satisfied" (mental list), or "stopping if the key amount accumulated is above a threshold" (magnitude threshold) [57, 59, 60]. For selecting between alternatives, examples of simple rules are "selecting the best alternative" (dominance rule), "selecting the alternative with the highest utility" (addition of utilities rule), or "eliminating alternatives that do not satisfy a criterion" (elimination by aspects rule) [56].

Table 3.1 lists the rational and heuristic models used in this chapter. The rational models embody rational judgments about where the expectation of information gain is maximum and whether maximum performance has been achieved. They are based on quantities such as the probability of improvement (PI), the expected improvement (EI), the expected conditional improvement (ECI), and the maxima-region entropy. On the other hand, the heuristic models derive from cues available on the user interface, such as predictive mean, variance, remaining budget, and the number of evaluations. Examples of *heuristic* models are the upper confidence bound (UCB), the fixed sample number (FSN), the fixed remaining budget (FSN), and the dominant physical prototype (DPP). Readers are directed to Ref. [96] for more details on how the survey responses helped in identifying aforementioned judgments or cues.

In overview, we represent the descriptive decision models in 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 which 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 rational or heuristic-based. For rational strategies, the assumption of probabilistic decisions mirrors the limited cognitive ability of designers to make accurate predictions while having their judgments aligned with rational judgments.

3.2.1 Decision to Choose the Next Design

In selecting x_{i+1} for $(i+1)^{th}$ evaluation, 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 evaluations 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, and is given by $\max_{x \in \mathcal{X}} \chi_i(x, \mathcal{H}_i, \psi)$. For a probabilistic setting, however, we define the probability of picking x_{i+1} using Boltzmann-like likelihood function:

$$p(x_{i+1}|\chi_i, \theta_d, \mathcal{H}_i) \propto \exp\left\{\gamma \chi_i(x_{i+1}, \mathcal{H}_i, \psi)\right\}, \tag{3.7}$$

where $\theta_d = \{\gamma, \psi\}$ denotes collectively 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

Table 3.1. : A list of decision models for the information acquisition decisions

Decision model	**Underlying strategy**
1. Decision to choose x	
Upper confidence bound (UCB)	Explore design space during initial evaluations while exploit during later evaluations.
Probability of improvement (PI)	Selection probability proportional to PI value.
Expected improvement (EI)	Selection probability proportional to EI value.
Expected conditional improvement (ECI)	Selection probability inversely proportional to ECI value.
2. Decision to choose information source	
Fixed sample number (FSN)	Test high fidelity source after a fixed no. of samples.
Fixed remaining budget (FRB)	Test high fidelity source if the remaining budget is smaller than a fixed value.
Fixed maximum-region entropy (FME)	Test high fidelity source if the information entropy of the location of maximum is smaller than a fixed value.
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.
3. Decision to stop	
Fixed sample number (FSN)	Stop after a fixed no. of samples.
Fixed remaining budget (FRB)	Stop after a fixed amount of budget is remaining.
Fixed maximum-region entropy (FME)	Stop after the entropy of the location of maximum is smaller than a fixed value.
Dominant physical prototype (DPP)	Stop when the best high fidelity measurement minus the largest predictive mean is smaller than a fixed value.
Fixed expected improvement (FEI)	Stop after expected improvement (EI) is below a fixed value.

one extreme, as $\gamma \to \infty$, 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. In the other extreme, as $\gamma \to 0^+$, the likelihood function becomes uniform over $\mathcal{X}$. Different decision models based on different acquisition functions are discussed next.

Upper confidence bound

Under this model, 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:

$$\chi_i(x, \mathcal{H}_i, \psi) = \mu_i(x) + a e^{-bi} \sigma_i(x), \tag{3.8}$$

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

Probability of improvement

The acquisition function, $\chi_i(x, \mathcal{H}_i)$, for this model is proportional to the probability that the next evaluation at x will generate higher performance than the current best performance. We take the current best to be the maximum of predictive means at past design evaluations, $\mu_i^* = \max_{1 \le j \le i} \mu_i(x_j)$, 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. (3.4), the corresponding information acquisition function is:

$$\chi_i(x, \mathcal{H}_i) := \mathbb{P}\left[\Delta_i(x) > 0 | \mathcal{H}_i\right] = 1 - \Phi\left(\frac{\mu_i(x) - \mu_i^*}{\sigma_i(x)}\right), \tag{3.9}$$

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

Expected improvement

The expected improvement model prefers design points with high expectation of improvement relative to μ_i^* from a single additional evaluation [27]. Its acquisition function is:

$$\begin{aligned}\chi_i(x, \mathcal{H}_i) &= \mathbb{E}\left[\Delta_i(x) | \mathcal{H}_i\right] \\ &= (\mu_i(x) - \mu_i^*)\Phi\left(\frac{\mu_i(x) - \mu_i^*}{\sigma_i(x)}\right) \\ &\quad + \sigma_i(x)\phi\left(\frac{\mu_i(x) - \mu_i^*}{\sigma_i(x)}\right),\end{aligned} \tag{3.10}$$

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

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 [97]. 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 $\mathcal{N}(\mu_{i|x,m}(\tilde{x}), \sigma^2_{i|x,m}(\tilde{x}))$. Assume that this density has the mean equal to the current mean,

$$\mu_{i|x,m}(\tilde{x}) = \mu_i(\tilde{x}), \tag{3.11}$$

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

$$\sigma^2_{i|x,m}(\tilde{x}) = \sigma^2_{i+1|x,m,y}(\tilde{x}) = \sigma^2_i(\tilde{x}) - \frac{k_i^2(\tilde{x}, x)}{k_i(x, x) + v_m^2}. \tag{3.12}$$

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

Then, $\Delta_i(\tilde{x}|x,m) = \max\{f(\tilde{x}|x,m) - \hat{\mu_i}, 0\}$ 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 $\tilde{x} \in \mathcal{X}$ [97]:

$$\mathbb{E}\left[\Delta_i(\tilde{x}|x,m)|\mathcal{H}_i\right] \leq \mathbb{E}\left[\Delta_i(\tilde{x})|\mathcal{H}_i\right], \tag{3.13}$$

where the expected conditional improvement, $\mathbb{E}\left[\Delta_i(\tilde{x}|x,m)|\mathcal{H}_i\right]$, is calculated by substituting the hypothetical mean $\mu_{i|x,m}(\tilde{x})$ and variance $\sigma^2_{i|x,m}(\tilde{x})$ in Eq. (3.10). 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 $\tilde{x}$ will be 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:

$$\chi_i(x, \mathcal{H}_i|m) = -\int_{\mathcal{X}} \mathbb{E}\left[\Delta(\tilde{x}|x,m)|\mathcal{H}_i\right] d\tilde{x}. \tag{3.14}$$

In the analysis, we take m to be the cheap low fidelity source, "L".

3.2.2 Decision to Choose an Information Source

Different elements of the observed history $\mathcal{H}_i$ influence the decision to choose among information sources. These elements can be derived quantities such as the sequence of different information sources used, the frequency of a particular source, the total cost, etc. We incorporate such elements into decision models using feature functions. Given that multiple history elements may influence decisions, a decision strategy is specified in terms of a weighted sum of multiple independent features. The decision models based on 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 $g_{m,i,1}(\mathcal{H}_i), \ldots, g_{m,i,R}(\mathcal{H}_i)$, with $w_{m,1}, \ldots w_{m,R_m}$ as the weight parameters.

The likelihood of choosing the high fidelity information source ($m_{i+1} = 2$) is defined using the sigmoid function as:

$$p(m_{i+1} = H|\theta_m, \mathcal{H}_i) = \operatorname{sigm}\left(\sum_{r=1}^{R_m} w_{m,r} g_{m,i,r}(\mathcal{H}_i)\right), \tag{3.15}$$

where $\operatorname{sigm}(\lambda) = (1+\exp\{-\lambda\})^{-1}$ 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 $g_{m,i,r}(\mathcal{H}_i)$ 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 for choosing among information sources, we always include a constant, negative basis function $g_{m,i,1}(\mathcal{H}_i) = -1$ because the *difference* between the weighted sum and a threshold determines the decision strategy. Furthermore, we assume that every designer's strategy relies upon a single element of history, and the decision model thereof uses two features ($R_m = 2$), one more in addition to the constant one. This is an acceptable assumption under the premise that people's cognitive ability is limited and they do not consider all the relevant information while making decisions [98].

Fixed sample number

In this model, the low fidelity information source is used for a fixed number samples, and the expensive high fidelity source is used thereafter. The feature used for this model is:

$$g_{m,i,2}(\mathcal{H}_i) = i. \tag{3.16}$$

Fixed remaining budget

In this model, the remaining budget, $B - n_{i,L}c_L - n_{i,H}c_H$ determines the choice between the two information sources. Low fidelity observations, if any, are collected during initial evaluations until a fixed amount of remaining budget is left, and high fidelity observations are collected thereafter. The feature for this model is:

$$g_{m,i,2}(\mathcal{H}_i) = B - n_{i,L}c_L - n_{i,H}c_H. \tag{3.17}$$

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 evaluations using the high fidelity information source. To define the entropy, let X^* be the r.v. 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 formally by:

$$p(x^*|\mathcal{H}_i) = \mathbb{E}\left[\delta\left(x^* - \arg\max_{x\in\mathcal{X}} f(x)\right)|\mathcal{H}_i\right]. \tag{3.18}$$

The entropy of this distribution is:

$$\begin{aligned} g_{m,i,2}(\mathcal{H}_i) = S(\mathcal{H}_i) &:= \mathbb{E}\left[\log p(X^*|\mathcal{H}_i)|\mathcal{H}_i\right] \\ &= \int_{x^*\in\mathcal{X}} \log p(x^*|\mathcal{H}_i)p(x^*|\mathcal{H}_i)dx^*. \end{aligned} \tag{3.19}$$

We estimate the entropy numerically using the maxima of 500 functions sampled from the posterior GP, see Eq. (3.3). Specifically, we estimate $p(x^*|\mathcal{H}_i)$ by building the histogram of the sampled maxima and then performing the integration of Eq. (3.19) numerically.

Fixed expected conditional improvement

In this model, two possible actions are compared at every evaluation 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 ($\max_{x \in \mathcal{X}} \mathbb{E}\left[\Delta_i(x)|\mathcal{H}_i\right]$) to that from one computer simulation and one physical prototype ($\max_{x,\tilde{x} \in \mathcal{X}}(\mathbb{E}\left[\Delta_i(x)|\mathcal{H}_i\right] + \mathbb{E}\left[\Delta_i(\tilde{x}|x,L)\right])$). 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:

$$\begin{aligned} g_{m,i,2}(\mathcal{H}_i) &= \max_{x,\tilde{x} \in \mathcal{X}}(\mathbb{E}\left[\Delta_i(x)|\mathcal{H}_i\right] + \mathbb{E}\left[\Delta_i(\tilde{x}|x,L,\mathcal{H}_i)\right]) \\ &\quad - \max_{x \in \mathcal{X}} \mathbb{E}\left[\Delta_i(x)|\mathcal{H}_i\right]. \end{aligned} \tag{3.20}$$

Also consider the case when the decisions of choosing the next design and choosing information source are interdependent. For example, a designer may use the low fidelity information source for evaluating 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 conditional upper confidence bound (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 $p(x_{i+1}|m_{i+1},\mathcal{H})$ is the conditional likelihood function for the decision to choose next design,

$$p(x_{i+1}|m_{i+1},\mathcal{H}_i) \propto \exp\left\{\gamma'\left[\mu_i(x_{i+1}) + \alpha'\sigma_i(x_{i+1})\right]\right\}. \tag{3.21}$$

The exploration parameter α' is $a'e^{-b'i}$ for $m_{i+1} = L$, or 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 Section 3.2.2.

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

$$p(m_{i+1} = H|x_{i+1},\mathcal{H}_i) = \mathrm{sigm}\left(\sum_{r=1}^{R_m} w_{m,r}\, g_{m,i,r}(x_{i+1},\mathcal{H}_i)\right), \tag{3.22}$$

where $w_{m,r}$ are model parameters. In a manner similar to the models in Section 3.2.2, we fix $R_m = 2$ and $g_{m,i,1} = -1$. Then, the feature that specifies the dependence between the decisions is:

$$g_{m,i,2}(x_{i+1}, \mathcal{H}_i) = \max_{x \in \mathcal{X}} \mu_i(x) - \mu_i(x_{i+1}). \tag{3.23}$$

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

3.2.3 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 evaluations is dependent upon 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 sigmoid function as follows:

$$p(s_i = 1 | \theta_s, \mathcal{H}_i) = \mathrm{sigm}\left(\sum_{r=1}^{R_s} w_{s,r} g_{s,i,r}(\mathcal{H}_i)\right), \tag{3.24}$$

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 or not. Thus, R_s is 2 for all models.

Fixed sample number

According to this model, the total number of evaluations performed is fixed. Accordingly, the feature function is:

$$g_{s,i,2}(\mathcal{H}_i) = i. \tag{3.25}$$

Fixed remaining budget

In this model, we assume that evaluations are stopped when the remaining budget $B - n_{i,L}c_L - n_{i,H}c_H$ reduces to a fixed value. The basis function has the same following form:

$$g_{s,i,2}(\mathcal{H}_i) = B - n_{i,L}c_L - n_{i,H}c_H. \tag{3.26}$$

Dominant physical prototype

Evaluations are stopped if 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:

$$g_{s,i,2}(\mathcal{H}_i) = \max_{x\in\mathcal{X}} \mu_i(x) - \max_{j:1\leq j\leq i, m_j=H} y_j. \tag{3.27}$$

Fixed expected improvement

In this model, we assume that evaluations are stopped if the maximum expected improvement from the next sample is small. This strategy fits into Eq. (3.24) by defining the feature function as follows:

$$g_{s,i,2}(\mathcal{H}_i) = \mathbb{E}\left[\Delta_i(x)|\mathcal{H}_i\right], \tag{3.28}$$

where the expected improvement is calculated in the same as given in Eq. (3.10).

3.3 Results and Discussion

We employ the *variational Bayes* approach to find approximations to posterior distributions of the model parameters, and estimate lower bounds to marginal log-likelihoods of the decision models conditional on the experimental data [74]. The variational Bayes approach is useful in complex stochastic models where analytical forms of posterior distributions are intractable. The model evidence lower bound

(ELBO) quantifies the the support for a model, i.e., the accuracy with which a model represents the experimental data. We denote ELBO, say for model $\mathcal{M}_k$, as $\mathcal{L}_k$.

We assume 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 Eq. 3.15, 3.24, and 3.22 and take the weight parameters of the non-constant feature out of the summation. This way the thresholds are given by $\frac{w_{m,1}}{w_{m,2}}$ and $\frac{w_{s,1}}{w_{s,2}}$. The following prior distributions are placed over these parameters.

$$\begin{aligned}
\gamma, \gamma', a, b, a', b' &\sim \text{Gamma}(1,1), && \\
w_{m,2}, w_{s,2} &\sim \text{Gamma}(1,1) && \text{for FSN,} \\
-w_{m,2}, -w_{s,2} &\sim \text{Gamma}(1,1) && \text{for all other models,} \\
\tfrac{w_{m,1}}{w_{m,2}}, \tfrac{w_{s,1}}{w_{s,2}} &\sim \text{Gamma}(1,1) && \text{for FME, FECI, FEI,} \\
\tfrac{w_{m,1}}{w_{m,2}}, \tfrac{w_{s,1}}{w_{s,2}} &\sim \text{Normal}(\mu = 10, \sigma = 1) && \text{for FSN, FRB, CUCB,} \\
\tfrac{w_{s,1}}{w_{s,2}} &\sim \text{Normal}(\mu = 0, \sigma = 1) && \text{for DPP,}
\end{aligned} \tag{3.29}$$

For this analysis, we group the experimental data into different treatments. We do not address heterogeneity among designers' strategies and assume that all subjects share the same set of model parameter values within a treatment. We partition the experimental data into training dataset and test dataset. The training dataset consists of the data from randomly selected fifteen periods, and is used to estimate model parameters using Variational inference. The decision models with posterior parameter estimates are then validated using goodness-of-fit tests against the test dataset consisting of the remaining three periods.

For cross-validation, we check the posterior prediction accuracy of decision models on test dataset using Wanatabe-Akaike information criterion (WAIC) [99], which estimates the expectation of data by averaging log pointwise predictive density over the posterior distribution. A non-Bayesian measure, called *accuracy score*, is also calculated for decision models of choosing information source and stopping which have discrete alternatives. If $\hat{z}_s$ is a prediction corresponding to a data point z_n in the test dataset $z_{1:S}$, then the accuracy score is $\frac{1}{S}\sum_{n=1}^{S} 1_{\{z_n\}}(\hat{z}_n)$.

The posterior distribution approximations and evidence lower bound (ELBO) for decision models were estimated using Automatic Differentiation Variational Inference [100] algorithm in PyMC3 module of Python [73]. This algorithm was run for 20000 iterations, and among those, last 5000 iterations were used to calculate the average ELBO.

3.3.1 Models with Highest Evidence Lower Bound

To quantify the support for a model over a baseline, the ELBO for each decision models is compared against the log-likelihood of random decisions. For the decision to choose next design, the log-likelihood of random sampling ($\mathcal{L}_{rand}$) is $N\log(\frac{1}{20})$ assuming a uniform distribution function over the design space $[-10, 10]$. For the decisions to choose an information source and to stop, $\mathcal{L}_{rand}$ is $N\log(0.5)$ assuming the probability of 0.5 for each of the two alternatives in both the decisions. The positive values of log-Bayes factor $\mathcal{L}_k - \mathcal{L}_{rand}$ in Figure 3.4 highlight that the predictions of the decision models are more accurate than random predictions. We are able to compare the support for any two models, say k_1 and k_2, by comparing the differences in log-likelihoods, $\mathcal{L}_{k_1} - \mathcal{L}_{rand}$ and $\mathcal{L}_{k_2} - \mathcal{L}_{rand}$, because $\mathcal{L}_{rand}$ remains constant in a given treatment.

Result 1. For the *decision to choose the next design*, the upper confidence bound (UCB) model and the conditional upper confidence bound (CUCB) model have the highest ELBOs.

From this and Figure 3.4, we conclude that exploration during initial evaluations while exploitation during later evaluations, captured by UCB and CUCB models, is the most likely strategy for choosing the next design point. In treatments $T1$, $T4$, and $T5$, the ELBO of the CUCB model is higher than that of the UCB model suggesting low fidelity observations being used for exploration while high fidelity observations for exploitation.

Result 2. For the *decision to select an information source*, the fixed sample number (FSN) model has the highest ELBO at low budget, whereas the fixed expected conditional improvement (FECI) model has the highest ELBO at medium and high budgets. The conditional upper confidence bound (CUCB) model's ELBO in 'save-remaining-budget' part with low budget is close to that of the FSN model.

Selecting the first high fidelity observation after a fixed number of evaluations is the most likely strategy at low budget. However, with the incentive-to-save budget, subjects also rely on predictive mean to select whether to choose the high fidelity information source. On the other hand, comparing the expectation of improvement between the low fidelity and the high fidelity information sources at every iteration is the most likely strategy for medium and high budgets.

Result 3. For the *decision to stop*, the fixed remaining budget (FRB) model has the highest ELBO in all treatments, except in high budget treatment of 'save-remaining-budget' part where the dominant physical prototype (DPP) model has the highest ELBO.

According to the results in Figure 3.4, the subjects stopped after exhausting the entire or part of the available budget in treatments $T1$, $T2$, $T3$, $T4$, and $T5$. However, at high budget in the 'save-remaining-budget' part (treatment $T6$), the subjects stopped when the best performance from high fidelity observations was closer to the highest mean prediction of the performance.

Results $1, 2$, and 3 also hold true for the test dataset based on cross-validation using WAIC (see Figure 3.4). The lower WAIC implies better support for a model. The differences in prediction accuracy from different models are substantial given that WAIC is defined in logarithmic scale. For additional proof, prediction accuracy scores for discrete choices in decisions to choose information source and stop are presented in Figure 3.5.

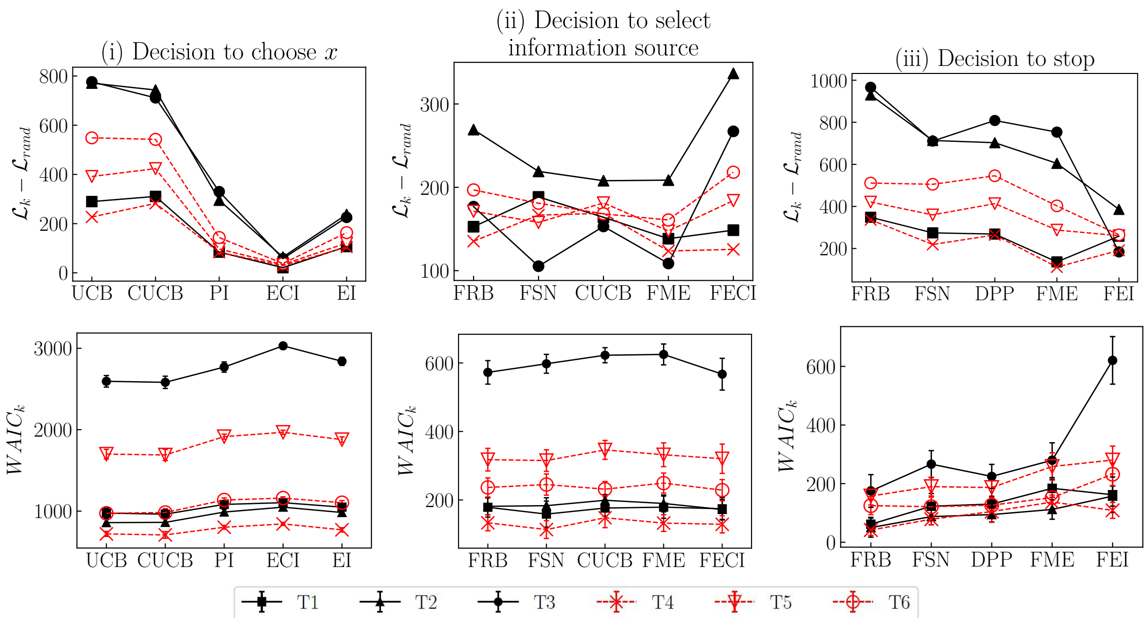


Figure 3.4. : Approximated log-Bayes factors ($\mathcal{L}_k - \mathcal{L}_{rand}$) relative to random selection from training data, and Watanabe-Akaike Information Criterion ($WAIC_k$) from test data for decision models at different treatments.

3.3.2 Effects of Fixed Budget and Payment Incentives

The amount of fixed budget as well as the incentive-to-save-budget affect the subjects' strategies for information acquisition. As fixed budget increases, the subjects evaluate more design points, whereas with the incentive-to-save-budget the number of evaluations decreases. Below, we list specific observations about the trade-off between more evaluations and saving the budget based on the posterior distributions of model parameters.

Result 4. Exploration of design space increases with the increase in fixed budget, except when fixed budget increases from medium to high in the 'save-remaining budget' part.

As observed from Figure 3.6, the mean posterior estimates of the exploration scale α increase with fixed budget between treatments $T1$, $T2$ and $T3$. However, in 'save-

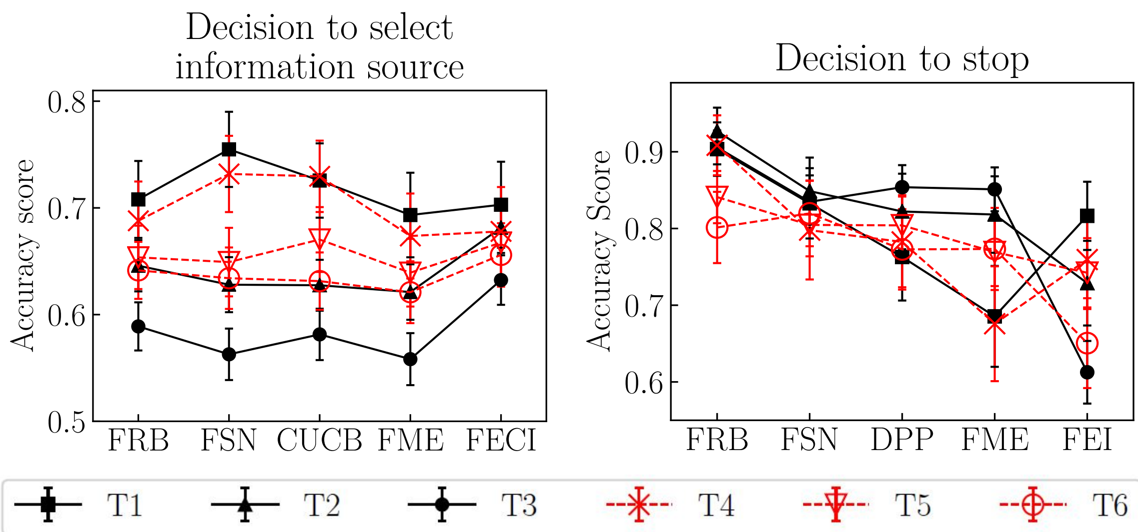


Figure 3.5. : Predictive accuracy scores of decision models on test data.

remaining-budget' part, the mean posterior estimates of α are highest at medium budget (treatment $T5$) while they decrease at high budget (treatment $T6$). Such reduction in exploration at high budget is only prominent when subjects' incentive is to save budget and when large savings are possible.

Result 5. The probability of selecting high fidelity information source increases with increase in fixed budget, except when fixed budget increases from med to high in the 'save-remaining budget' part.

Based on the posterior distribution of the threshold parameter $\frac{w_{m,1}}{w_{m,2}}$ for the FSN model, we observe that the subjects choose a single high fidelity observation after 6 samples at low budget treatments (treatments $T1$ and $T4$). The average number of high fidelity observations in the treatments $T2$ and $T3$ are 2.6 and 4.2 respectively. A larger estimate of $\frac{w_{m,1}}{w_{m,2}}$ for the FECI model implies a weaker threshold for choosing the high fidelity source over the low fidelity source, i.e, larger probability for choosing the high fidelity source. In the treatments $T2$ and $T3$, the mean posterior estimate of $\frac{w_{m,1}}{w_{m,2}}$ for the FECI model is the same (mean posterior $\frac{w_{m,1}}{w_{m,2}}$ = 0.5), whereas it increases from 0.4 to 0.5 between the treatments $T5$ and $T6$.

Result 6. The probability of stopping at high values of remaining budget increases with 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 3.7 shows the posterior probability of stopping as function of remaining budget from the FRB model. The FRB model's threshold parameter estimate for the 'use-it-or-lose-it budget' part (mean posterior $\frac{w_{s,1}}{w_{s,2}} \approx 5$) is smaller than the cost of one high fidelity observation ($c_H = 8$) 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 $\frac{w_{s,1}}{w_{s,2}}$ increases as fixed budget increases.

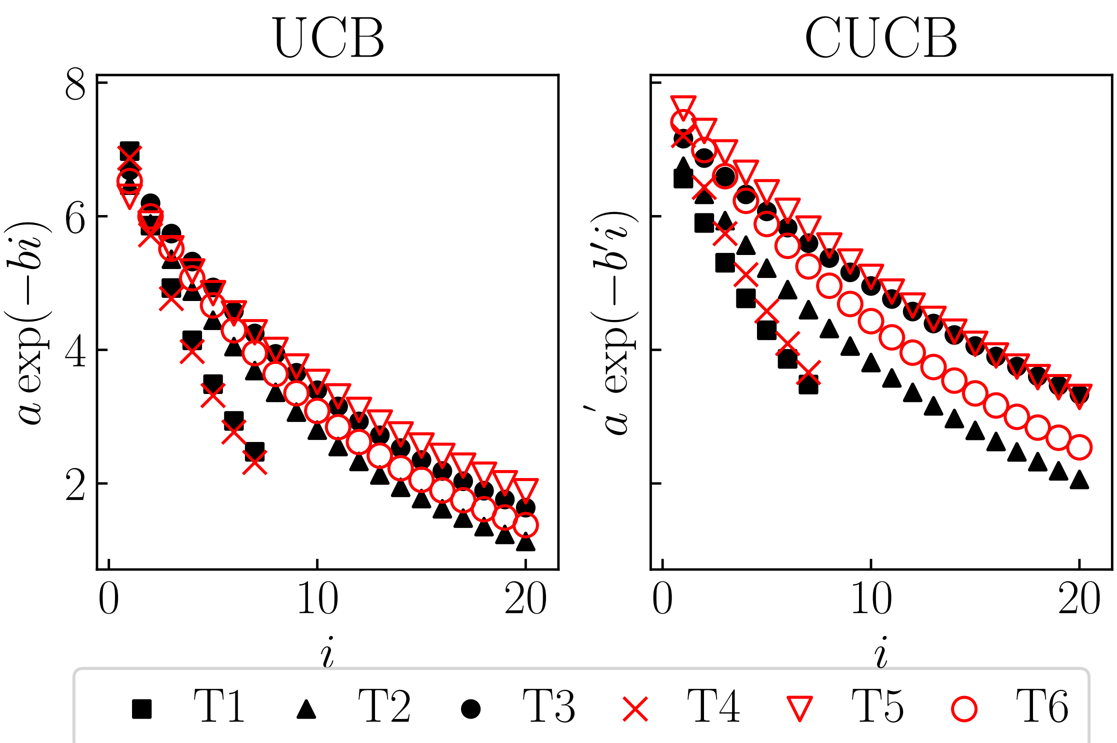


Figure 3.6. : Mean posterior estimations of the exploration scales α and α' for the the decision to choosing the next design.

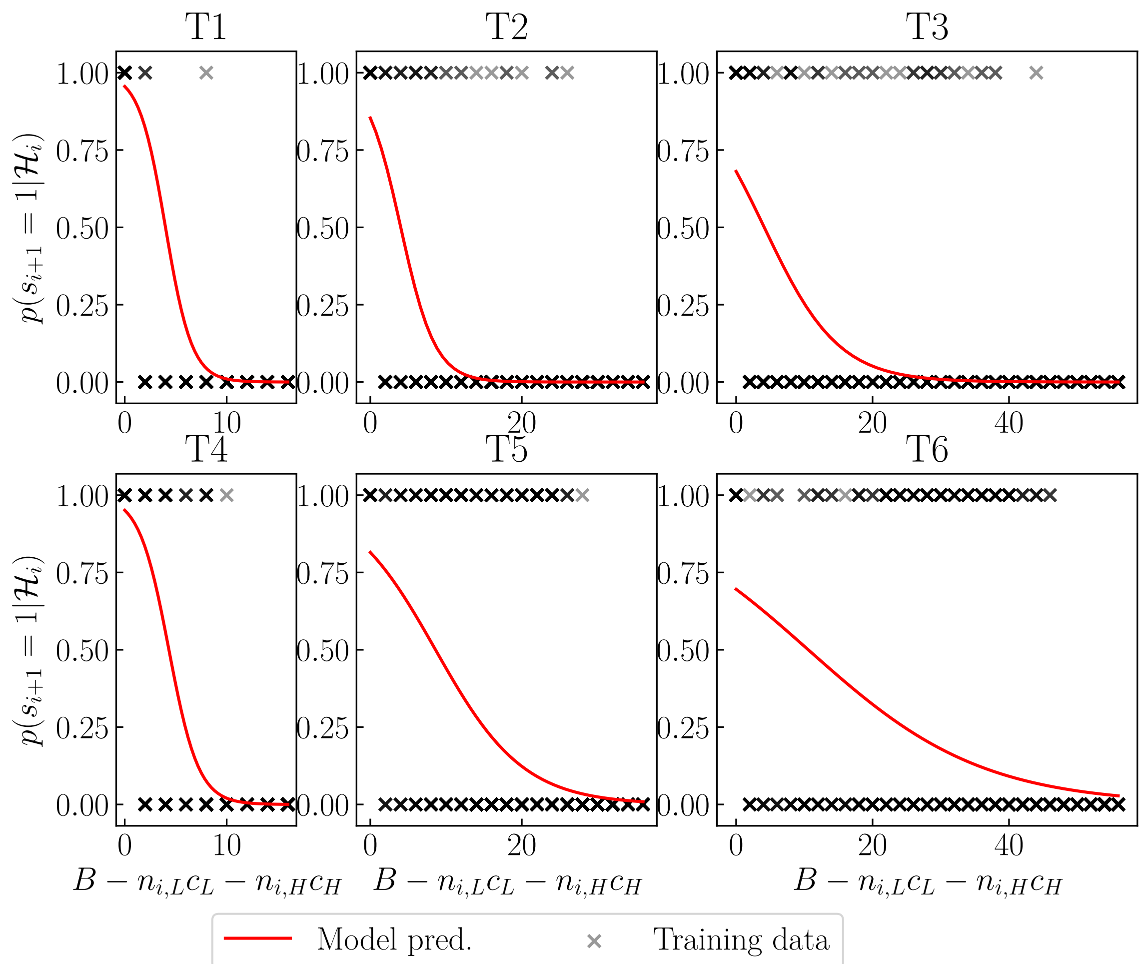


Figure 3.7. : Posterior predictions of the probability of stopping as a function of the remaining budget under the FRB model.

3.3.3 Checking Model Accuracy

The results indicate that the simple heuristic models predict 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 $\max_{j=1\geq j\leq i} \mu_i(x_j)$. A comparison of the predictions of these quantities with their actual values in the test dataset in Figure 3.8 confirms that the highest ELBO heuristic triplet has better predictive strength than the expected-utility based triplet.

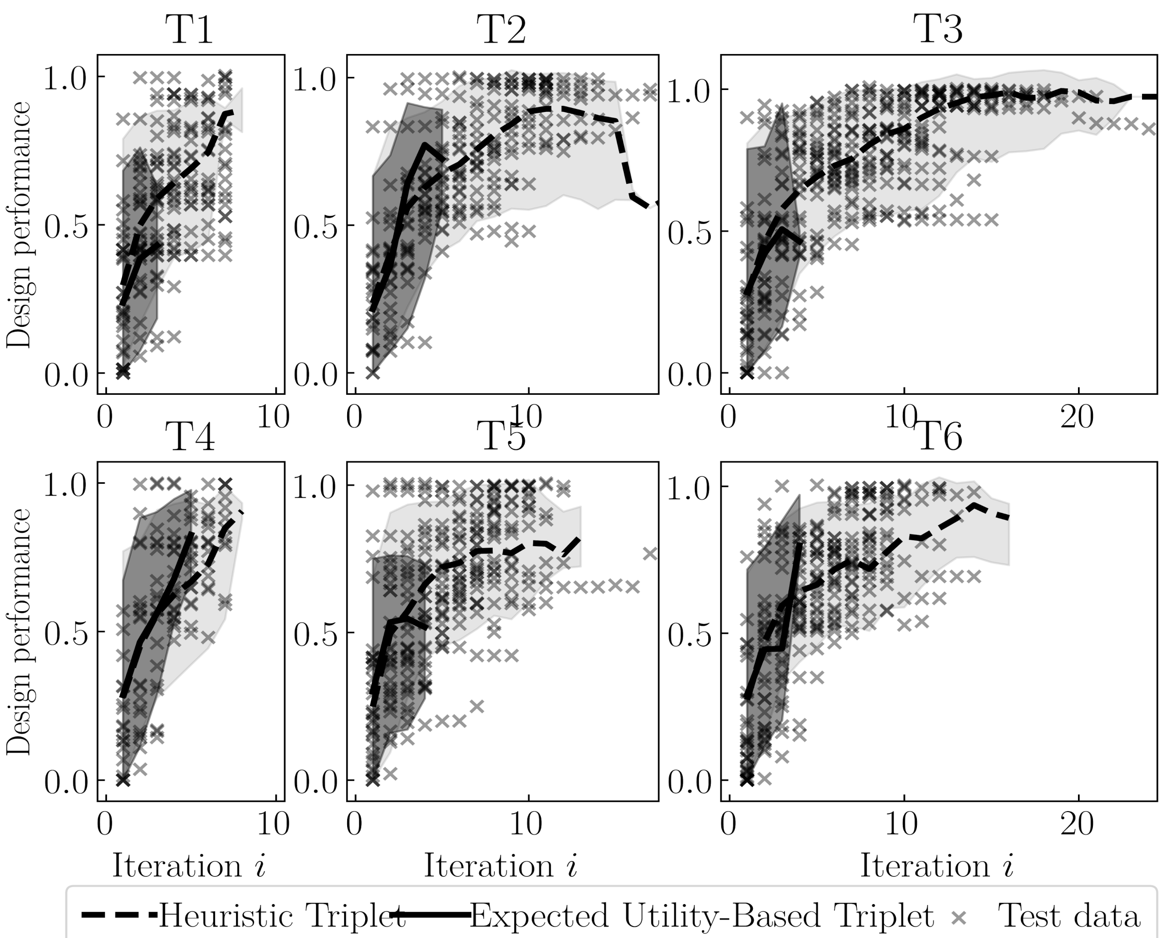


Figure 3.8. : Normalized design performance quantified as highest predictive mean.

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. (3.4), 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 [98].

3.3.4 Implications for Engineering Design

On the objective functions in the test dataset, the heuristic triplet model generates higher design performance (gross payoff) than the expected utility (EU)-based triplet model. Figure 3.9 plots the gross payoff (Eq. 3.5) 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 high fixed budget and without the incentive for saving budget. That is because with high fixed budget and without the incentive for saving budget the heuristic model triplet completes larger number of iterations, conducts more design exploration, and receives better performance. Figure 3.10 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

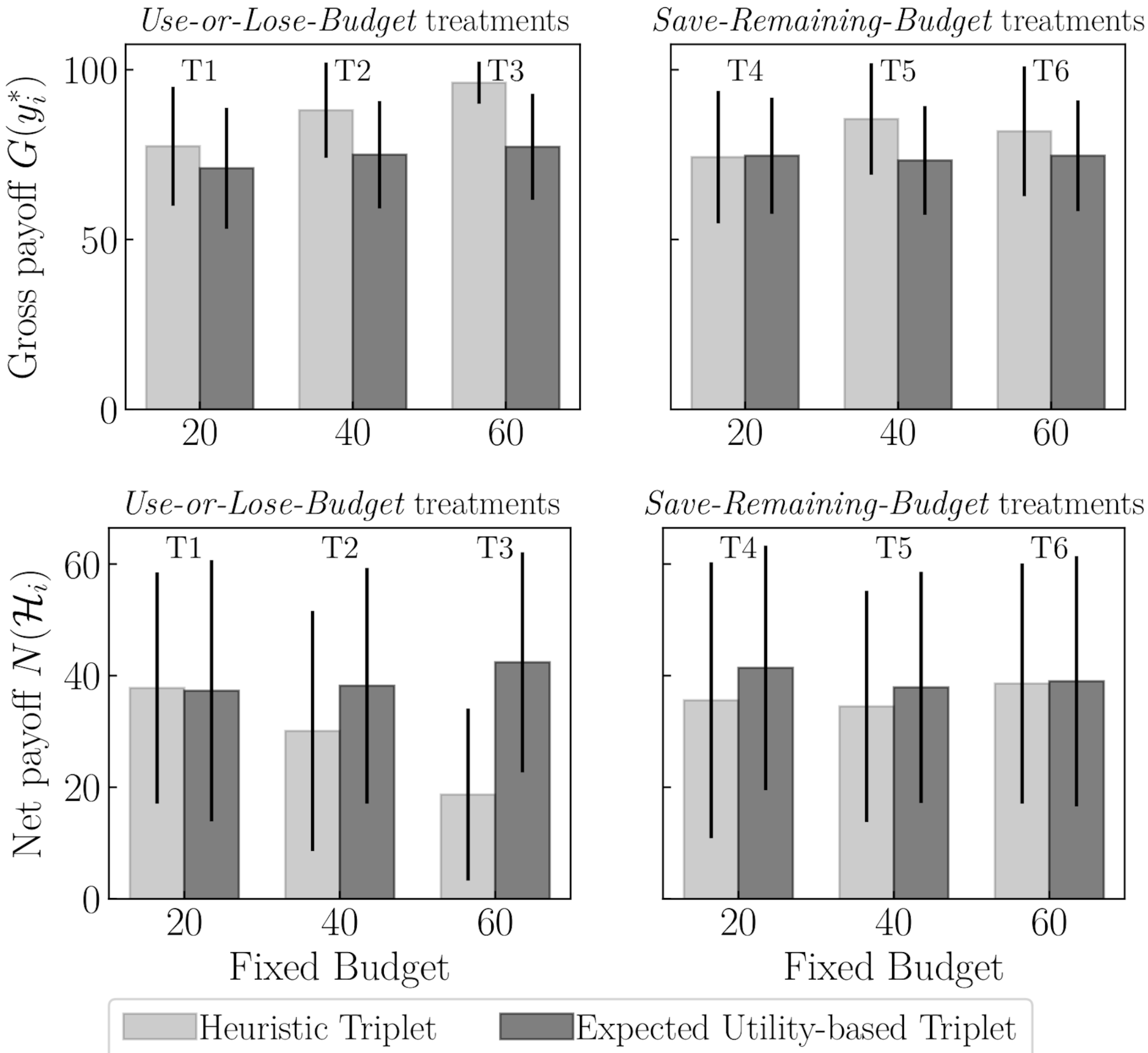


Figure 3.9. : The gross payoff $G(\mathcal{H}_i)$ and the net payoff $N(\mathcal{H}_i)$ after stopping for heuristic triplet model and the expected-utility-based triplet model. The gross payoff is a function of the best design performance y_i^* calculated according to Eqs. 3.6. The net payoff is the gross payoff minus total cost, i.e., $G(\mathcal{H}_i)$-C_i.

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 Figure 3.9. We observe that the EU-based triplet model provides higher net payoff on average than the heuristic triplet models in all treatments. The

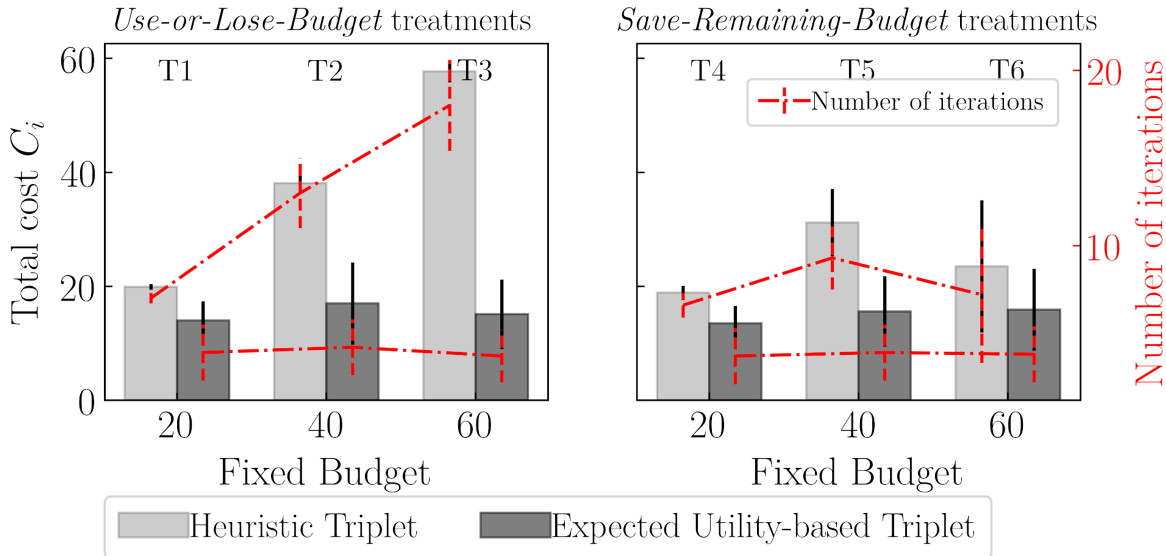


Figure 3.10. : Total cost c_i and the number of iterations i after stopping for the heuristic triplet model and the expected utility-based triplet model.

difference in average net payoffs from the heuristic triplet model and the EU-based triplet model reduces with 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 fixed budget cannot be determined. Under the 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 FEI model on the test data is larger in treatment $T6$ with the incentive-to-save budget than in treatment $T3$ without such an incentive. The same holds true for the training data where the FEI model's accuracy scores in treatments $T3$ and $T6$ 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, failing to receive even a small potential benefit is a lost opportunity. Because people are loss averse, losses loom larger than gains for them when compared against each other [30].

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.

4. SEQUENTIAL INFORMATION ACQUISITION OF MULTIPLE DESIGNERS IN COLLABORATIVE TEAM

Following on the analysis of individual designer's design exploration decisions, this chapter studies designers decisions to communicate or not and whom to communicate with. The context of the design process is a collaborative design where different designers collaborate to design a complex system with each individual designing a specific subsystem component (called discipline). This chapter analyzes the time longitudinal dynamics between team communication and system-level design performance in engineering design teams through stochastic modeling and behavioral experiment data. The hypotheses about the team communication and design performance dynamics is represented graphically in Figure 4.1. The research study has two research objectives:

1. evaluate the effects of social and technical factors such as design interdependence and social network properties on communication-related decisions, and
2. evaluate the influence of the amount of interdisciplinary communication on the design performance.

In the approach, the first step is the experiment design for collection of team communication data. The experiment design for this particular study is guided by the NASA mission design laboratory (MDL) setting for the purpose of creating a representative experiment [101, 102]. This controlled experiment mimics some features of the reference system while it defers in other features. Table 4.1 summarizes the differences and similarities in two settings. In second step, we present the descriptive decision models for communication-related decisions. Finally, these models are trained on the communication data from the controlled experiment as well as the NASA MDL setting for understanding drivers of communication decisions. We thank

Prof. Zoe Sajnfarber and Dr. Erica Gralla for providing the communication data of NASA MDL engineers.

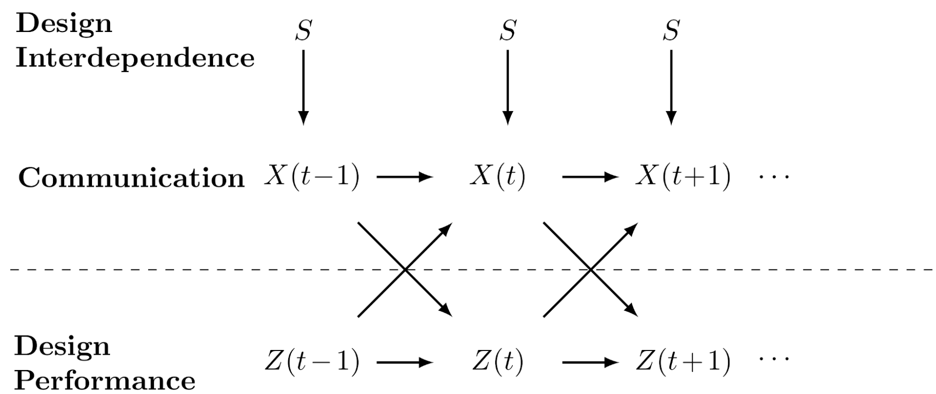


Figure 4.1. : A model of communication and design performance dynamics.

4.1 Experiment Design

This section describes the key elements of two experimental datasets, *engine design dataset* and *spacecraft design dataset*, such as the context of design task undertaken and communication pathways available, the experimental design task itself (object), background and incentives for the participants (actors), and the design process embodied in the experimental task.

The context of the two datasets is that of engineering system design where multiple designers are individually selecting properties of subsystem components and collectively working towards meeting the prespecified system goals. The setup of each experimental setting is such that one team member designs one subsystem component, but two components maybe interdependent. The design task involves parametric design whereby a designer chooses design parameter of the respective design component. The interdependence between a pair of subsystems is structured as the number of shared design parameters between the pair. The overall interdependence between all pairs of subsystems are represented using a design structure matrix in

Table 4.1. : A comparison of the model world and the reference world

Category	Dimension	NASA MDL (Reference)	Controlled experiment (Model)
Task	System	Conceptual spacecraft design	Parametric engine design
	Timescale	One week	One hour
	Disciplines	Typical aerospace disciplines	Engine components
	Team size	12 disciplines, 32 design variables	5 disciplines, 10 design variables
	Design interdependence	Theory-based mapping betwn. design variables and disciplines	Theory-based mapping betwn. design variables and disciplines
	Design objective	Satisfying requirements	Optimization of objectives
Subject	Design expertise	NASA engineers	Engineering students
Context	Communication channels	Face-to-face communication	Text-based communication and shared parameter database
	Cultural norms	Slightly different risks	No specific risk, flat distribution of risk
	Resource access	Access to discipline models, simulations, testbed	Access to component-level simulation models and catalogs
	Incentives	Career development, fixed salary employment	Performance-based monetary payment

which each cell stores the number of shared variables between the corresponding row and column subsystems, as shown in Figure 4.2.

4.1.1 Reference Setting: Spacecraft Design Task

The spacecraft design task involves conceptual design of spacecraft systems with close to 10 subsystems, e.g., advanced camera for survey (ACS), command and data handling (C&DH), communications, flight dynamics (FD), etc. Such task is conducted through 4-days long studies at the NASA mission design laboratory. For each study, co-located NASA engineers with specialized knowledge of one or more subsystems engage in the design of spacecraft subsystems. Additionally, a systems engineer facilitates the design process by integrating information from different subsystems, sharing design parameters, and resolving conflicts. One part of the design goals is to meet some requirements on the spacecraft dry mass quantifying the total mass of individual subsystem designs. Subsystem engineers periodically post their present subsystem mass to a common database accessible by others.

The subjects of each study interact face-to-face for meetings. The face-to-face meetings may include groups of two or more subjects. To maintain similarity with the pairwise interactions in the engine design dataset, the spacecraft design dataset converts a group talk into one-to-one undirected pairwise interactions between all those involved.

The dataset includes timestamped observations of pairwise interactions, the identities of subsystems involved in each interaction, and the system-level spacecraft dry mass from six different studies. The study subjects are full-time NASA engineers and engage in the design tasks as a part of their job. Most of the six studies have one engineer, mainly the same person, assigned to each subsystem with detailed prior knowledge of their respective spacecraft subsystems and relevant design interdependencies.

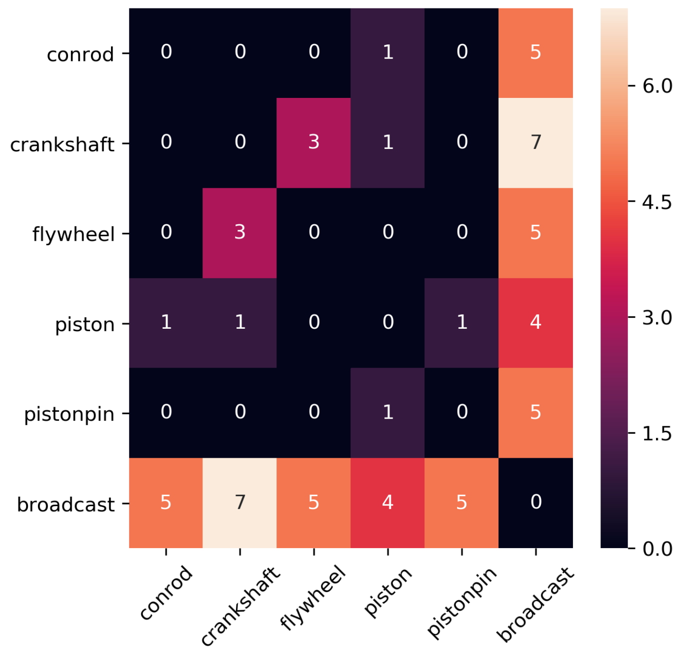
4.1.2 Engine Design Task

The engine design task operationalizes an engineering system design of an engine that consists of five subsystems components, viz., connecting rod, crankshaft, piston head, flywheel, and piston-pin. Design of each subsystem requires specification of two design parameters while anticipating the effects of other parameters shared with other subsystems. Figure 4.2 presents each subsystem's design variables and the design variables that it shares with others. The assumed design interdependence structure has subsystems with low interdependence such as piston pin, as well as subsystems with high interdependence such as piston and flyweel. The system-level design objectives in the experimental task are to minimize the total mass of the engine components and maximize the factor of safety against failure. The total mass is sum of individual masses, whereas the system-level factor of safety is minimum of individual factors of safety.

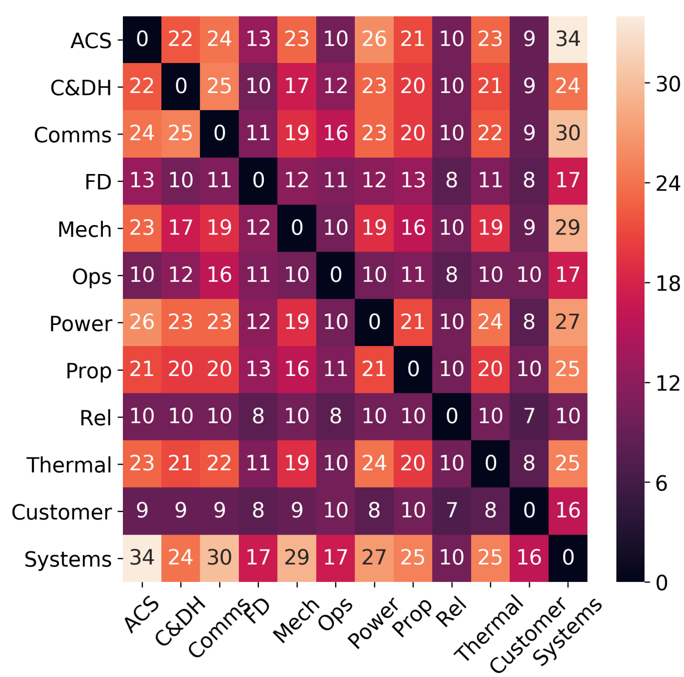
Members on the same team could communicate with one another through *one-to-one text messages*. Face-to-face communication as well as group text messages are restricted so that the interactions between different members on the same team can be tracked over time.

The participants in the experimental task were undergraduate students from an introductory machine design course. A total of 200 students participated in the experiment. Every five students were randomly grouped into a team and each student in a team was assigned a unique role. Another component of the team was a shared virtual screen, called "broadcast", that showed the current values of design objectives (see example in Figure 4) and design parameters in certain conditions. To ensure the experiment task was incentive compatible, the experiment paid monetary incentives proportional to the achieved design performance as a function of the system-level mass and the system-level factor of safety. The achieved design performance fell into one of the five levels. All subjects on a team received the same amount in prize money, between \$10 to \$20 at the end of the experiment. Figure 4.3 shows an example of how the payment was calculated from the system performance levels achieved.

Each team belong to either of the four experimental conditions that differed in two independent variables: i) design exploration using simulations on the continuous design space versus a catalog with pre-evaluated design points scattered across the de-



(a) Design structure matrix for the engine design problem



(b) Design structure matrix for the spacecraft design problem.

Figure 4.2. : Pairwise design interdependencies in the matrix cells represent the number of shared design variables

Table 4.2. : Design interdependence structure for design of an engine assembly.

Subsystem	Subsystem parameter 1	Subsystem parameter 2	Shared design parameters
Connecting rod	I-section thickness (t_I)	Ratio of rod length to crank length (r_1)	Piston bore diameter (D)
Crankshaft	Bearing offset (c)	Pin diameter (dc)	Flywheel diameter (d_s), Flywheel thickness (t_f), Piston bore diameter (D)
Flywheel	Flywheel diameter (d_s)	Flywheel thickness (t_f)	Bearing offset (c)
Piston head	Head thickness (t_H)	Bore diameter (D)	–
Piston pin	Length-diameter ratio (r_2)	Ratio of inner-diameter to outer diameter (r_3)	Bore diameter (D)

sign space, and ii) the global availability of shared design parameters on the broadcast screen versus no such availability. The factorial design of experiment generated four experiment treatments: i) design exploration using simulations but design parameters not visible (S-P_0), ii) design exploration using simulations and design parameters are visible (S-P_1), iii) design exploration using catalogs but design parameters are not visible (C-P_0), and iv) design exploration using catalogs but design parameters are visible (C-P_1). Figure 4.4 gives a snapshot of information provided during the two different types of deign exploration At the time of design exploration, the current values of key design parameters were either available on the broadcast screen

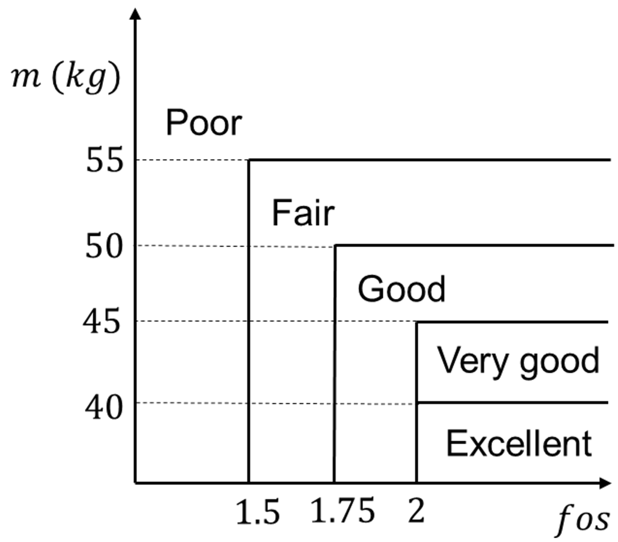
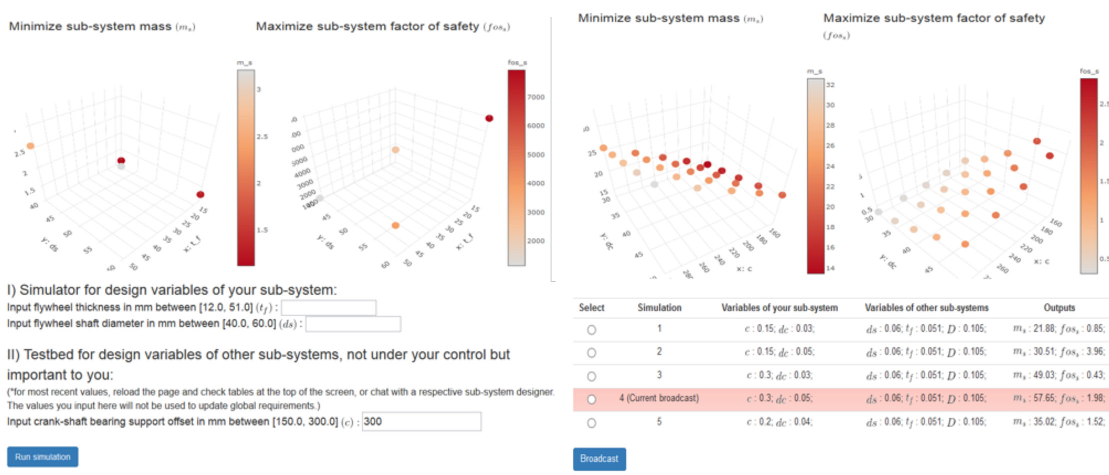


Figure 4.3. : Categorization of continuous design objectives into discrete levels.



(a) Simulation-based local design search

(b) Catalog-based local design search

Figure 4.4. : The user interface for two types of design exploration processes: a) simulation-based and b) catalog-based.

(conditions P_1), or this information was hidden (conditions P_0). Figure 4.5 provides a user interface showing the broadcast information. In both cases, the information about the current state of system-level objectives such as mass and factor of safety was present on the user interface.

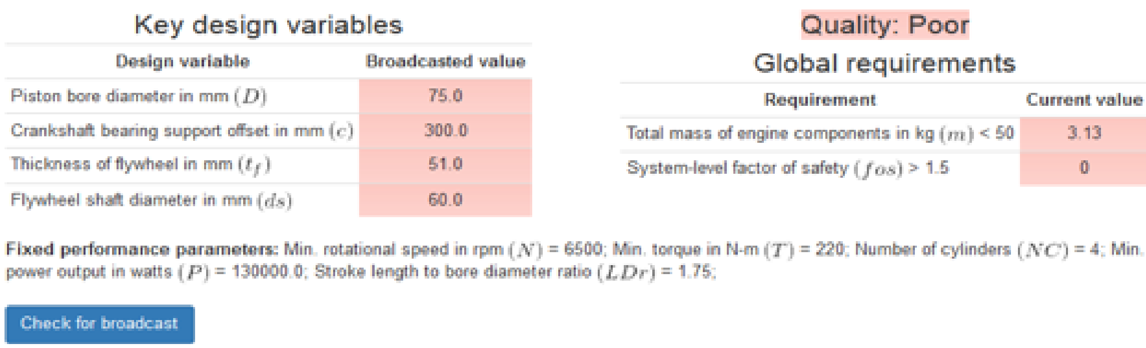


Figure 4.5. : An example of the global information shown on broadcast, which displays key design variables for condition P_1 only.

4.2 Descriptive Models of Communication-related Decisions

The stochastic models represent the probabilities of change in team communication and design performance for any given moment while considering their mutual dependence. Specifically, they models the probability that a designer interacts with fellow teammates given past communication and design performance at the time, and the probability that such communication has an impact on the systems performance.

Team communication spans a finite space $\mathcal{X}$ of directed networks between a fixed number of nodes. A sample directed network quantifies pairwise interactions (an edge property) in terms of an adjacency matrix. The observation times of communication are discrete, equidistant time steps $t_0, t_1, t_2, \ldots, t_M$ in a fixed time interval $\mathcal{T}$. A time series of team communication forms a discrete stochastic process $\{X(t_m) | \, t_m \in \mathcal{T}\}$ on the network space $\mathcal{X}$, where time step t_m assumes values in interval $\mathcal{T}$. A cell value, say $X_{ij}(t_m) \in \mathbb{Z}$, in the adjacency matrix at time t_m denotes the number of interactions between disciplines i and j during interval $(t_{m-1}, t_m]$. Additionally, a diagonal entry $X_{ii}(t_m)$ represents discipline i's interactions with itself such as the count of design searches. Therefore, note that the network represented by adjacency matrix $X(t_m)$ is a weighted network. The constant separation between two consecutive time steps $\delta = t_{m-1} - t_m$ is called the sampling window. The pairwise interactions within a sampling window are considered simultaneous and independent. Too large sampling window δ will misclassify interactions as independent. If δ is too small, interactions will be thinly scattered between the observation times. The adjacency matrix $X(t_m)$ is a directed graph if identification of the source and the recipient of each interaction is possible. Otherwise, an undirected graph converts to a directed graph by converting each edge in an undirected graph into two edges.

At any given moment, a K dimensional vector in the space $\mathcal{Z}$ denotes the design performance. The design performance over time is a discrete stochastic process $\{Z(t_m) | \, t_m \in \mathcal{T}\}$ on the space $\mathcal{Z}$ where time parameter t_m takes values in interval $\mathcal{T}$. Each design objective that constitutes a design performance vector is discrete but ordered (ordinal) variable whose higher level is preferred over a lower level. Using ordinal variables to represent design objectives implies that disciplines have a relative understanding of the design performance and do not know the design performance on a continuous interval.

The central hypothesis of the stochastic models is that the dynamic system consisting of team communication and design performance is an outcome of a Markov process. The Markov assumption means that the conditional distribution of future team communication $\{X(t_{m+1}) \mid \ t_{m+1} > t_m\}$ for any time $t_m \in \mathcal{T}$ given the past and the present depends only on the present team communication $X(t_m)$ and the present design performance $Z(t_m)$. Similarly, the distribution of future design performance $\{Z(t_{m+1}) | \ t_{m+1} > t_m\}$ is conditionally dependent on only the present states $X(t_m)$ and $Z(t_m)$.

4.2.1 Deciding the Rate of Interactions

In a system design team, a discipline acquires information from multiple channels to determine the opportunity for improvement in overall design performance and the performance of its own subsystems. Examples of such channels are pairwise communication with other design disciplines, communication with an integrator discipline (systems engineer), and exploration of the design space of its subsystem. The representation of team communication as a matrix encompasses these choices by including an integrator discipline as a part of the nodes in the adjacency matrix and including the design search counts as diagonal entries.

This model considers that the present state of design performance affects the frequency of selecting specific communication channel. If the past and present design performance is low, then the number of interactions would increase during immediately next future time step. Suppose that a discipline can access communication channels to initiate an interaction. Let vector $\boldsymbol{\beta}_c \in \mathbb{R}^K$ the effect of present design performance on the frequency of interactions through a particular channel c. Constant parameter $\gamma_c \in \mathbb{R}$ is an intercept parameter. Then, the expected number of interactions through channel c is

$$\lambda_c(t_m) = e^{\boldsymbol{\beta}_c \cdot Z(t_m) + \gamma_c}. \tag{4.1}$$

This time variant function $\lambda_c(t_m)$ is called the *rate function* of communication channel c. The probability distribution of the number of channel c interactions during the next time step is a Poisson distribution:

$$p(N_c(t_{m+1}) = y | \boldsymbol{\beta}_c, \gamma_c, Z(t_m)) = \frac{\lambda(t_m)^y}{y!} e^{\lambda(t_m)}. \tag{4.2}$$

The implicit assumption in the communication frequency model of Eq.(4.2) is that the interactions through different channels during any sampling window take place independently of each other. This assumption is acceptable when a discipline can simultaneously operate different channels without mutual interdependence.

4.2.2 Deciding Whom to Communicate With

The focus of modeling social selection is on understanding the specific reasons why a discipline chooses to interact with other disciplines. For this purpose, this model considers two types of factors affecting the social selection: social network factors such as acquaintances between discipline pairs, present popularity, etc., and technical factors such as design interdependencies. Social network factors are time-varying and instantaneous, whereas the design interdependecies are fixed for a given design problem. Therefore, modeling social selection would allow investigation of whether instantaneous social factors has an equal effect as the constant technical factors on team communication.

Suppose that discipline i's communicates with other disciplines to maximize its utility which is a function of some dyadic statistics such as pairwise reciprocity, nodal popularity, and design interdependence. Discipline i's prefers to maximize L different dyadic network statistics that depend on the present adjacency matrix $X(t_m) = \mathbf{x}$.

These network statistics are denoted by $\mathbf{f}_i(\mathbf{x}) = \{f_{i1}(\mathbf{x}), f_{i2}(\mathbf{x}), \ldots, f_{iL}(\mathbf{x})\}$. Parameter i in the subscripts implies that the values of network statistics are different for different disciplines. Then, from a future interaction with discipline j, discipline i can change its dyadic network statistics by amount $\mathbf{f}(\Delta_{ij}\mathbf{x}) - \mathbf{f}(\mathbf{x})$, where operator Δ_{ij} reflects the single change in the present adjacency matrix due to an interaction added to cell value x_{ij}. Similarity, by communicating with discipline j, discipline i increases total amount of design interdependence with the connected nodes by amount S_{ij}. Then, the overall utility function of discipline i for communicating with discipline j is:

$$U_{ij}(\mathbf{x}) = \boldsymbol{\beta} \cdot (\mathbf{f}_i(\Delta_{ij}\mathbf{x}) - \mathbf{f}_i(\mathbf{x})) + \gamma S_{ij} + \epsilon_{ij}, \tag{4.3}$$

where random parameter ϵ_{ij} changes between discipline pairs and changes with time (which assumed implicit to make notation simpler). Parameters $\boldsymbol{\beta} \in \mathbb{R}^L$ represents the relative preferences for different network statistics and γ is the preference for maximizing interactions with highly-interdependent disciplines. The utility function in Eq.(4.3) is called discipline i's *objective function.*

Consider two examples of dyadic network statistics assumed to drive social selection: nodal popularity and pairwise reciprocity. First, nodal popularity statistic measures the total incoming degree of the disciplines that discipline i is connected to. It is mathematically given by $f_{i,1}(\mathbf{x}) = \sum_{j \neq i} \mathbf{1}_{x_{ij}>0} \sum_{h \neq j} x_{hj}$, where indicator function $\mathbf{1}_{x_{ij}>0}$ is 1 if x_{ij} is a positive integer and 0 otherwise. Then, a positive preference parameter of nodal popularity statistic in the objective function would imply that nodes with high incoming interactions in the present are likely to be chosen for more interactions in the future. Second, the pairwise reciprocity statistic measures the number of mutually matched incoming and outgoing links with the connected disciplines. Mathematically, the reciprocity statistic is calculated as the minimum of incoming and outgoing links, $f_{i,2}(\mathbf{x}) = \sum_{j \neq i} \min\{x_{ij}, x_{ji}\}$. A positive preference parameter associated with the reciprocity statistic would imply that future interactions are responses to present interactions from others.

Next, the choice probability for discipline i to select a discipline for communication in the next time step is proportional to the objective function in Eq (4.3). If we consider that ϵ_{ij} has a standard Logistic distribution, then the probability that discipline i selects discipline j, given that discipline i is communicating, is:

$$p_{ij}(\mathbf{x}, S_{ij}) = \frac{e^{\boldsymbol{\beta}\cdot\mathbf{f}_i(\Delta_{ij}\mathbf{x})+\gamma S_{ij}}}{\sum_{h\neq i} e^{\boldsymbol{\beta}\cdot\mathbf{f}_i(\Delta_{ih}\mathbf{x})+\gamma S_{ih}}}. \tag{4.4}$$

This choice probability function has a multi-logit form similar to multinomial logistic regression and dependent on present communication state $\mathbf{x}$ and pairwise design interdependence S_{ij} Note that the present network statistics $\mathbf{f}_i(\mathbf{x})$ from the numerator and denominator cancel each other out. Additionally, because the interactions during a sampling window are to be independent of each another, the choices of discipline i during the next sampling interval $(t_m, t_{m+1}]$ has a multinomial distribution with independent probabilities for different discipline partners given by Eq. (4.4).

4.2.3 Effects of Team Communication on Design Performance

There are multiple reasons why team communication may affect the design performance in short term and long term. One reason is that designers discover inconsistencies between design parameters of different subsystems through communication. In short term, the identification inconsistencies may result in reduction in design performance. Through continued communication among the disciplines and partial or substantial rework, the design inconsistencies may get resolved in a long term. Another reason for reduction in design performance over long term may be misplaced communication leading to lost opportunity to discover inconsistencies and improve upon the existing design.

In the dynamic model of team communication and design performance, the predictors of design performance change are different amounts of communication corresponding to different interaction channels. Let vector $N(\mathbf{x})$ denote the numbers of interactions through different communication channels. To model these amounts' effect on the design performance, we consider a random utility function. The change from the present design performance to future performance should be in accordance with the maximization of this utility function. The utility maximization is a matter of conscious choice on part of a designer, rather the change occurs by the virtue of designers' design search decisions, subsystem models and integration of performance at the system level. The random utility of selecting level z for the k^{th}-dimension of the performance vector in the next time step $Z(t_{m+1})$ is:

$$U_k(z) = \boldsymbol{\beta}_k^z \cdot N(\mathbf{x})z + \gamma_k^z z + \zeta_k. \tag{4.5}$$

In Eq. 4.5, assuming positive values of effect parameters $\boldsymbol{\beta}_{\boldsymbol{k}}^{\boldsymbol{z}}$, the correlation term $N(\mathbf{x})z$ means the larger numbers of interactions increase the design performance. Parameter γ_k^z represents the basic shape of the utility function. Positive γ_k^z would signify that the dynamic process results in large values of performance levels in general, where as negative γ_k^z would signify the opposite. Random variable ζ_k is specific to the present time step and the dimension of design performance being considered.

From the possible levels, the dynamic process selects a particular level z for k^{th} performance dimension with the following probability:

$$p_k(z, \mathbf{x}) = \frac{\boldsymbol{\beta}_k^z \cdot N(\mathbf{x})z + \gamma_k^z z}{\sum_{q \in \mathcal{Z}_k} \boldsymbol{\beta}_k^z \cdot N(\mathbf{x})q + \gamma_k^z q}. \tag{4.6}$$

This choice probability function depends on the present communication state $\mathbf{x}$. The assumption implicit in the derivation of Eq.(4.6) is that the random variable ζ_k has a standard logistic distribution.

4.3 Results and Discussion

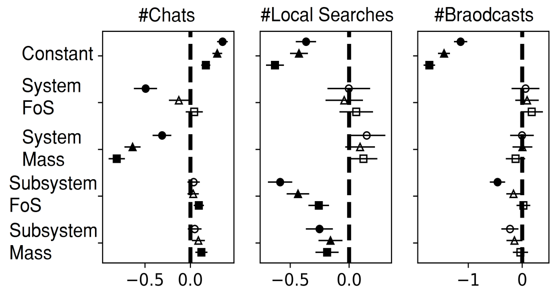
The results are divided into four sections: (i) posterior parameter estimates for preferences on when a subject runs local search, performs text-based interaction, or broadcasts own parameters, (ii) posterior estimates of how a subject selects whom to communicate with from the fellow designers working on other subsystems, (iii) posterior parameter estimates for how the rate of interactions affect the changes in design performance, and (iv) model checking of the descriptive decision models on the observed communication data.

Since the time duration of the tasks in the two settings is different, we define unit interval time for each setting so that we can compare communication patterns on the same timeline. The unit interval for the engine design setting is one minute and it is one hour for the NASA MDL setting. These particular unit intervals are selected because (i) they partition the respective task duration into a similar number of time steps (≈ 35) in two settings, and (ii) the average number of interactions by each discipline (average network degree) within a unit interval is approximately same (between 3 to 5) in two settings. The results include three sets of smoothing windows for predictor variables, no smoothing ($\tau = 1$), moving sum over past five unit time steps ($\tau = 5$), and moving sum over past ten unit time steps ($\tau = 10$).

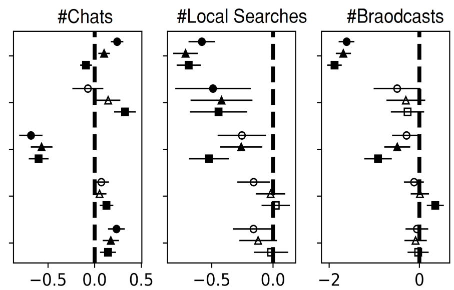
4.3.1 Estimated Strategies for Selecting the Rate of Interactions

In the engine design experiment, the first result across multiple experimental conditions is that the present system performance levels or present subsystem performance levels has negative effects on the future number of local design searches. This negative effect is measured by the negative parameter estimates in Figure 4.6. This effect may be explained using two mechanisms: i) the student subjects increase local design searches if they encounter low performance at any given time step, or ii) the student subjects reduces the future local searches once they reach a sufficient performance levels. In these two possibilities, the first possibility is more likely because the student teams do not reach high performance levels early in the duration of the design process. This effect is statistically significant in conditions S-P_0, S-P_1 and C-P_0, but not in condition C-P_1. In condition C-P_1, the design space search is efficient than other conditions due to the availability of catalogs and the global availability of design parameters. The availability of such global information may preclude the need for repetitive exploration of the design space. Further, the student subjects' attention differs from the subsystem performance levels in condition S-P_0 to the system performance levels in condition S-P_1 and C-P_0. This difference is justifiable because the student subjects in the most information-restrictive condition S-P_0 sequentially search their own subsystem design spaces and likely understand the state of the overall system performance from their subsystem performance. The second result observed across the experimental conditions is that the system mass levels has negative effect on the rate of text-based interactions. This implies that the low performance levels of the system mass in the present drive more one-to-one textual interactions in the future. Note the corresponding effect parameters for the system factor of safety levels are statistically insignificant, hinting that the system mass levels might have been bigger hindrance for achieving the highest possible system objectives. Indeed, a separate analysis confirms that the total Sobal sensitivity of all design parameters to the system mass is more than the corresponding the total Sobal sensitivity to the system factor of safety. In condition C-P_0, the observed negative effect of system mass performance levels on text-based interactions is only significant between consecutive time steps (τ = 1 minute) and is insignificant when system mass performance levels are smoothed over longer past (τ = 5 and τ = 10 minutes). The reasons behind this anomaly is not entirely clear.

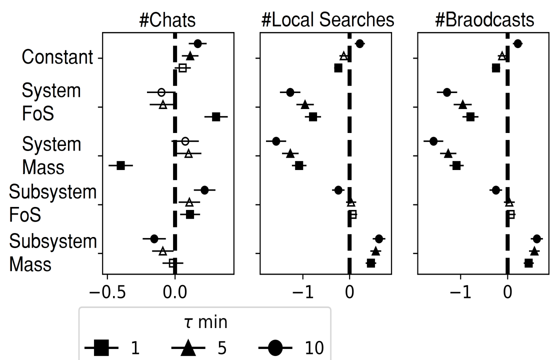
In the NASA MDL teams, the present spacecraft mass performance levels has a negative effect on the face-to-face communication between disciplines over short term but the spacecraft mass performance levels from a longer term in the past has a positive effect. This suggests that any reduction in spacecraft dry mass level (i.e. increase in actual mass) brings the disciplines together for communication in immediate future (τ = 1 hour). Also, good performance levels over time still fuel more face-to-face interactions in longer future (τ = 5 and τ = 10 hours), unlike in the student teams where the correlation between present performance levels and future text-based interactions is largely negative. The NASA engineers' willingness continue engagement in team communication even with large performance levels may highlight their incentives to focus beyond just reducing the spacecraft mass and on maintaining the



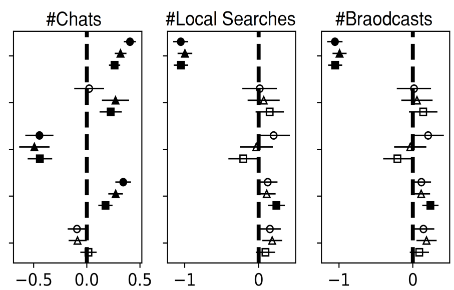
(a) Condition S-P_0



(b) Condition S-P_1



(c) Condition C-P_0



(d) Condition C-P_1

Figure 4.6. : Predicting the rate of interactions through different channels

overall integrity of the spacecraft design. Whereas, the student subjects' incentives are aligned with maximizing only the performance levels of the system-level mass and system-level factor of safety.

4.3.2 Estimated Strategies for Selecting Communication Partner

Across the experimental conditions in the engine design experiment, the design interdependence and the pairwise reciprocity are statistically significant predictors of how a discipline chooses a fellow discipline for text-based interactions. The positive effect of design interdependence highlights that the number of shared design variables is a key driver of inter-discipline communication. This positive effect of shared design variables is stronger in the conditions S-P_0 and S-P_1 than in conditions C-P_0 and

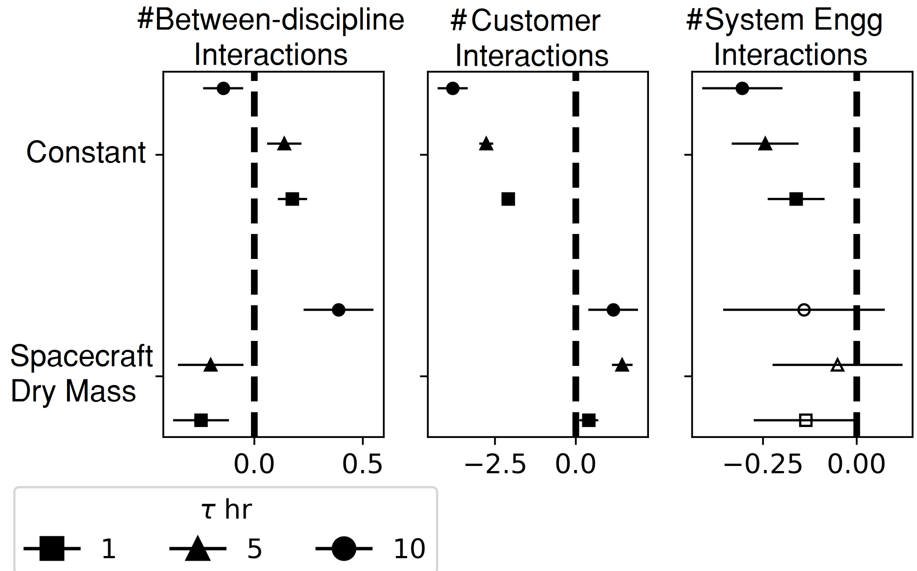
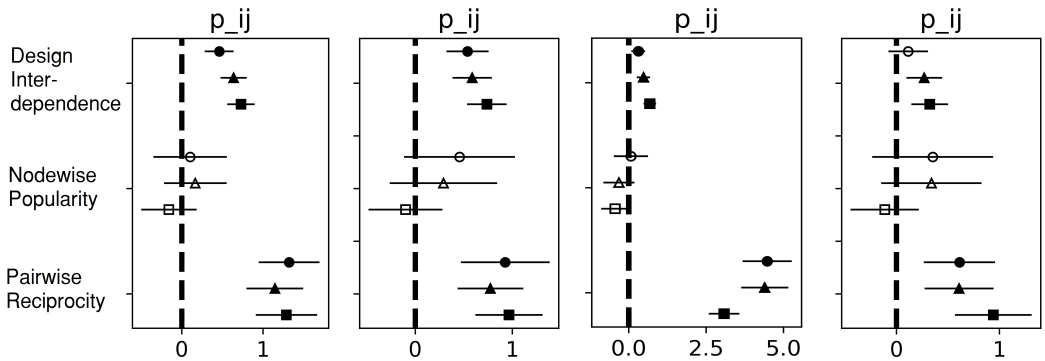


Figure 4.7. : Predicting the rate of interactions through different channels

C-P_1, which suggests there is more emphasize on text-based communication as a primary channel for exchange of design parameter values while sequentially simulating the subsystem design spaces than when using catalogs. Additionally, the strong influence of the pairwise reciprocity, i.e., the inclination to interact with disciplines with history of interactions, signifies the role of social network factors in driving interdisciplinary communication. The other social network effect due to nodewise popularity of disciplines is small and statistically insignificant, likely because of the difficulty in forming a power-law degree distribution in small, fully-connected communication networks such as the six-discipline engine design teams.

For the NASA MDL teams, the design interdependence, nodewise popularity, and pairwise reciprocity all have statistically significant effects on how the spacecraft disciplines select whom to interact with. These effects are evidence of the key roles that technical factors as well as social network factors play in driving the interdisciplinary communication in engineering design teams. Note that the social selection analysis models interactions among only the designing, core disciplines excluding integrative disciplines such as customers and systems engineering.



(a) S-P_0 (b) S-P_1 (c) C-P_0 (d) C-P_1

Figure 4.8. : Predicting the interdisciplinary social selection

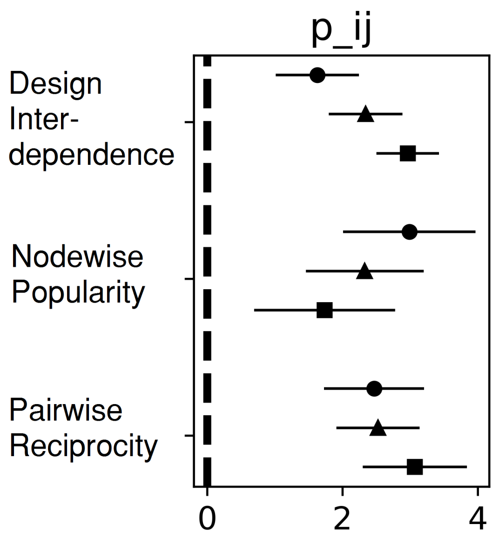


Figure 4.9. : Predicting the interdisciplinary social selection

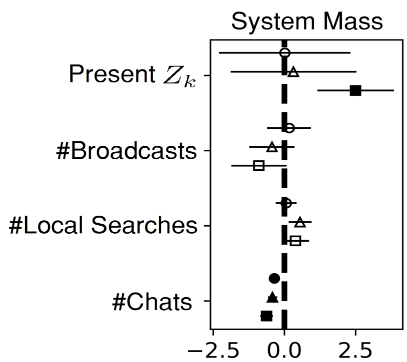
4.3.3 Estimated Effects of Team Communication on Design Performance

For the student teams designing an engine, the effects of amount of communication on the design performance change are small across the experimental conditions and statistically significant only in selective conditions. In condition C-P_0, the number of text-based interactions has a small but significant positive impact both on the system mass performance and the system factor of safety indicating benefits of communication for overall improvement in the design performance. In some cases, the number of interactions has negative effects, e.g. the effect of text-based interactions on the system mass performance levels in condition S-P_0 and the effect of the number of local design searches on the system factor of safety in condition S-P_1. These negative effects and the overall lack of strong positive effects may point to the downside of too much communication in engineering design teams, especially when the communication is not directed towards resolving inconsistencies. Additionally, the problem difficulty and the ruggedness of design space might also play roles in weak effects of communication on the design performance. For instance, the design performance trajectories over time in the engine design task has large variance indicating rapid changes in design performance without consistent positive improvement.

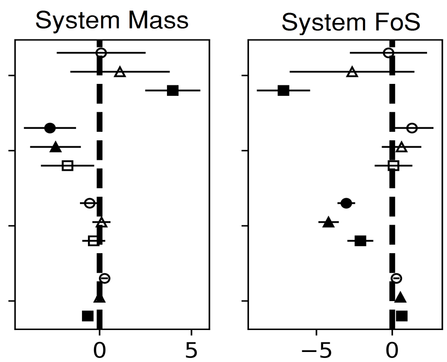
On the other hand, for the NASA engineers designing a spacecraft system, the number of inter-discipline interactions by core disciplines has a strong and statistically significant effect on the performance of spacecraft dry mass over long period of time (τ = 10 hours). The amount of communication does not appear to result in immediate improvement in the performance levels of spacecraft dry mass (τ = 1 or (τ = 5 hours) likely because a complex system such as a spacecraft may require longer discussions to achieve meaningful improvement in the design performance. In a separate effect, the number of interactions by systems engineer has a negative effect on the performance levels of spacecraft mass. It is unclear why, nevertheless, this negative effect is smaller that the the positive effect of the number of interactions by core disciplines.

4.3.4 Checking Model Accuracy

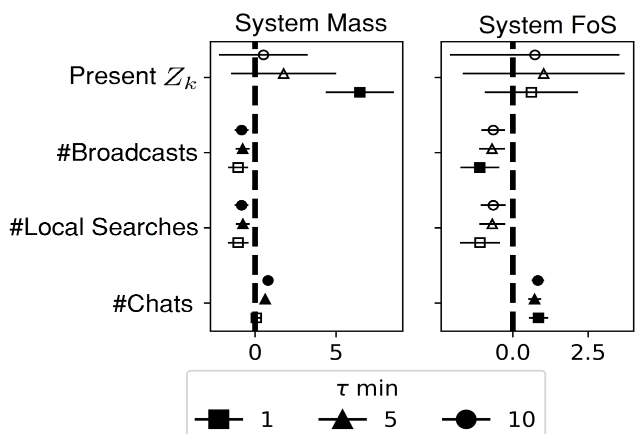
In order to evaluate the model performance, we randomly select one team in every condition to form testing data, except for condition S-P_0 whose testing data involves two randomly selected teams because of large number of teams available. Then, the stochastic models of team communication and design performance dynamics are



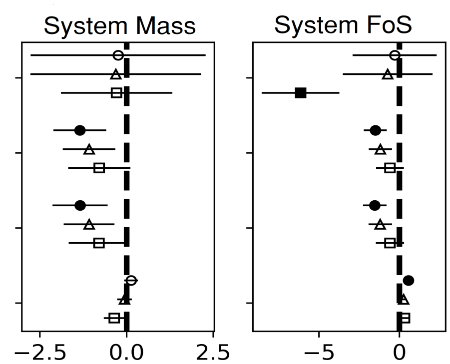
(a) Condition S-P$_0$



(b) Condition S-P$_1$



(c) Condition C-P$_0$



(d) Condition C-P$_1$

Figure 4.10. : Predicting the effect of amount of communication on design performance

trained on the training data and their performance is evaluated on the testing data, for each setting separately.

To characterize model performance, a Bayesian measure, Wantanabe Alaike Information Criterion (WAIC), is calculated on testing data, which is unseen by the model during the training process. The metric WAIC hedges the log pointwise predictive density of the testing data through adjustment terms for the number of model parameters, thereby penalizing a model with large number of model parameters. The WAIC is proportional to the number of data points, and lower values of WAIC are

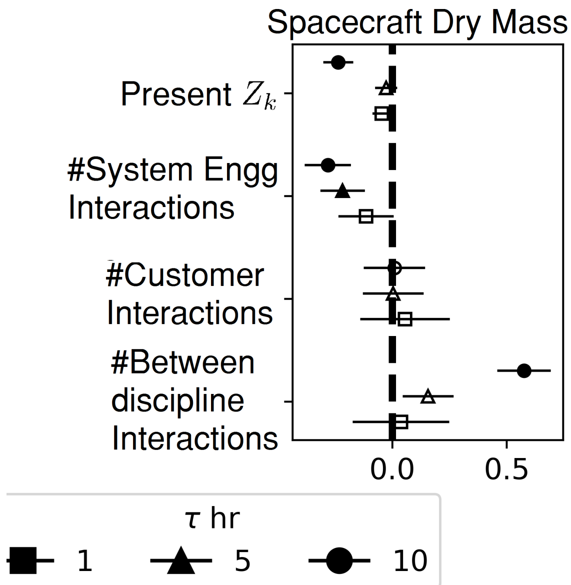


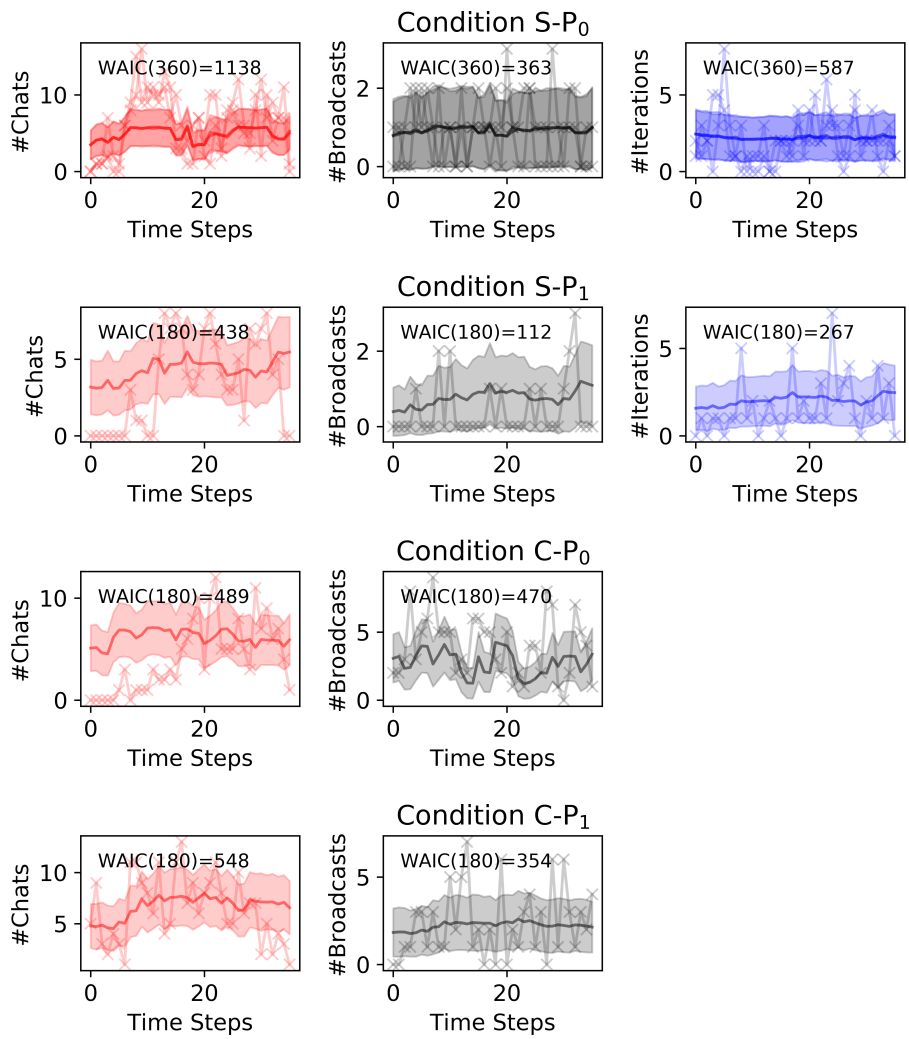
Figure 4.11. : Predicting the effect of amount of communication on design performance

preferred over lower values. Another approach for explaining the model performance is visual posterior predictive checking in which we visually compare model predictions with observed testing data.

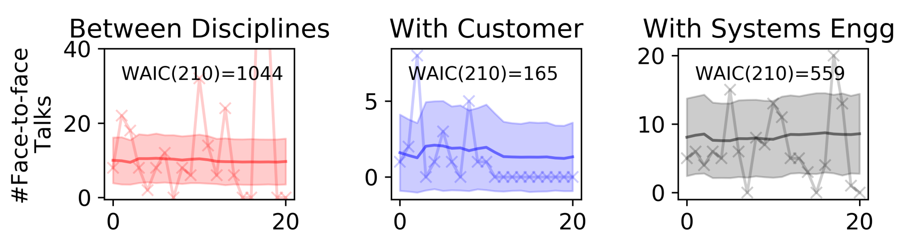
Figure 4.12 presents the posterior predictive checks for the models of rates of interactions in testing data. In the engine design dataset, these models predict the amount of text messages, the mount of local searches and the amount of broadcasts based on design performance states from past one time steps (τ = 1). In the spacecraft design dataset, they predict the amount of interactions between disciplines, by a customer, and by a systems engineer. The models approximately capture the average number of interactions over time for both the engine design and the spacecraft design datasets. They also capture the variance in testing data at any time step. However, the models in some cases do not capture increasing and decreasing trends over short time duration. For example, the observed number of text messages in testing data for conditions S-P_1 and C-P_0 increase significantly over time, but the respective model predictions show consistently high rate of text messages. In other cases such as conditions S-P_0 and C-P_1, the models appears to capture increasing and decreasing trends over time. Also, the number of face-to-face talks with the customer are large initially and reduce to 0 over time. These results highlight that there may be changes in the rates of communication which are not entirely described by the design performance and which occur explicitly due to the factor of time.

Figure 4.13 compares predictions from the multinomial model of social selection (Eq. 4.4) to the observed testing data at every time step. The Chi-squared error at given time step t_m, $\chi^2(t_m) = \sum_{i=1}^{N}\sum_{j=1;j\neq i}^{N}\frac{(x_{ij}(t_m)-\tilde{X}_{ij}(t_m))^2}{\tilde{X}_{ij}(t_m)}$, quantify the difference in observed interactions between two disciplines $x_{ij}(t_m)$ and the expected interactions $\tilde{X}_{ij}(t_m)$ averaged from 2000 posterior model predictions. From the social selection model in Eq. 4.4 and estimated parameters in Figure 4.8, we observe that the design interdependence and reciprocity are significant predictor variables of interdisciplinary communication in the engine design dataset, and the popularity effect is an additional predictor for the spacecraft design dataset. Figure 4.13, additionally, shows that the predictor variables are drivers of interdisciplinary selection mainly during the early design process, especially in conditions S-P_1, C-P_0, and the spacecraft design study. The large Chi-squared error suggests more dispersed communication between all disciplines. In the engine design dataset, large Chi-squared errors may be attributed to increased interactions by all disciplines. The drivers of increased interactions late in the process may be the changes in system design performance that is tied to the student subjects' payment. However, the exact reasons are unknown. In the spacecraft design dataset, a large Chi-squared error towards the end of the design process may be attributed to a tag-up event where all disciplines reconvene to verify their designs.

Figure 4.14 presents the posterior predictive checks for the predictions of design performance as a function of the amounts of interactions from last time step ($\tau = 1$ minute). In general, the amounts of communication such as text messages do not accurately predict the design performance in testing data over short duration. The prediction of average design performance is close to the actual levels in conditions

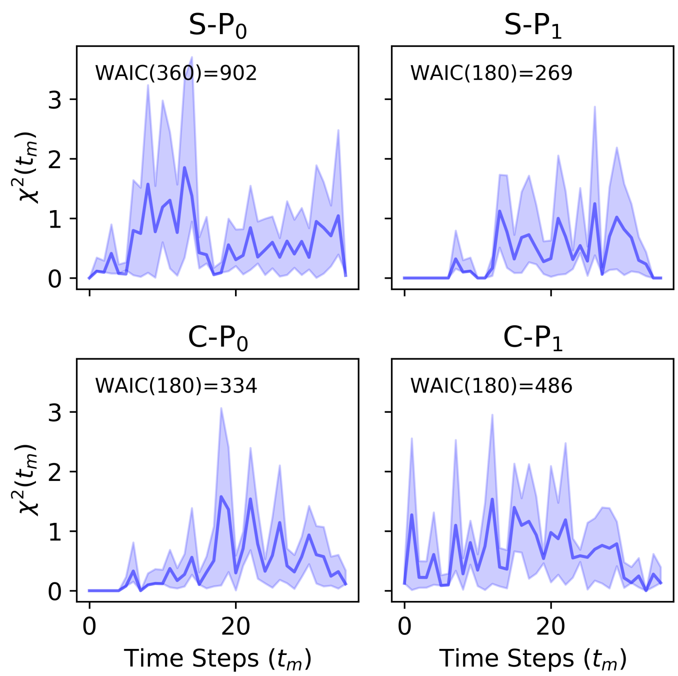


(a) Engine Design Dataset

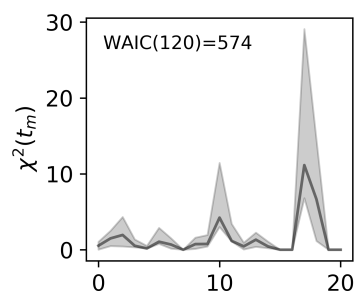


(b) Spacecraft Design Dataset

Figure 4.12. : Predictions of the rate of communication (with mean and 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile bounds) for the testing data.



(a) Engine Design Dataset



(b) Spacecraft Design Dataset

Figure 4.13. : Time-wise Chi-squared error in the multinomial test comparing predictions of social selection to the observed social selection from the testing data.

C-P_0 and C-P_1. A possible reason for poor model performance may be an insufficient number of data points (approximately 27 per each team) to train the models. Also, the present number of text-based interactions are significant predictors of design performance only in the condition C-P_0. These results hint at considering the content of communication in addition to the amount of communication as a predictor of design performance.

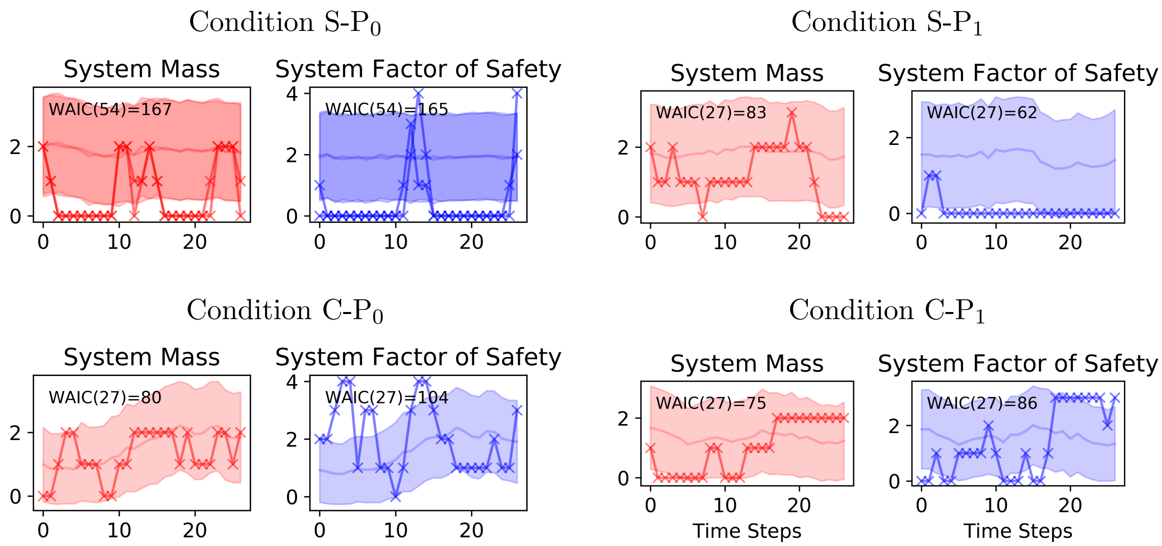


Figure 4.14. : Predicting the changes in design performance for the testing data in the engine design dataset.

5. EVALUATING REPRESENTATIVENESS OF THE ENGINE DESIGN EXPERIMENT

After estimating the decision strategies for communication-related decisions in collaborative teams in Chapter 4, the next step is to evaluate the representativeness of the controlled experimental setting (engine design experiment) with respect to the NASA Mission Design Laboratory (MDL). In simple terms, representativeness means the degree to which a controlled experimental setting mimicks desired decision making behaviors observed in a reference setting. If criteria for representativeness are satisfied, the insights from a controlled setting can be used to understand the behaviors in a reference setting. The research objective of this chapter are:

1. to evaluate the representativeness of the engine design experimental setting for the NASA MDL reference setting with regard to team communication, and
2. to synthesize lesson learned for building representative controlled experiments.

The hypothesis is that the representativeness is meant to be judged based on subject-task-context interrelationships rather than singular dimensions of subject, task, or context. To test this hypothesis, we compare the communication patterns between the engine design experimental setting and NASA MDL setting.

We begin by specifying the quantities used for measuring team communication patterns. Since these quantities depend on the task and the context, it is necessary to state some basic parameters of the two settings. Such parameters may be different between the settings, but note that First, the instances of communication between different disciplines are counted as pairwise interactions. A pairwise interaction in the NASA MDL setting represents one instance of face-to-face talk, whereas a pairwise interaction in the engine design setting may represent either one text message, one update or one check of the broadcast element. These measures are not exactly the same, since an MDL conversation likely involves the equivalent of several text messages. However, we do not aim to compare these numbers directly but rather to compare trends and relative changes in the number of interactions, such as changes over time or relative differences in communication across subsystems. Second, discipline pairs are dichotomized into tightly-coupled and loosely-coupled pairs using a threshold number of shared variables. The threshold divides the discipline pairs into two similar-sized groups in either setting. The discipline pairs with the the number of shared variables greater than the threshold are tightly-coupled pairs, and the rest of pairs are loosely-paired. The threshold for the engine design task is 0.5 and it is 15.5 for the spacecraft design task. Third, since the time duration of the tasks in the two settings is different, we define unit interval time for each setting so that we can compare communication patterns on the same timeline. The unit interval for the engine design setting is one minute and it is one hour for the NASA MDL setting. These particular unit intervals are selected because i) they partition the respective task duration into a similar number of time steps (≈ 35) in two settings, and ii) the average number of interactions by each discipline (average network degree) within a unit interval is approximately same (between 3 to 5) in two settings. Based on these assumptions, the following quantities are used for comparing communication patterns.

1. *Technical-Communication Mirroring*: This is defined as the correlation between the number of shared variables and the number of interactions between discipline pairs.
2. *Amount of Communication*: This is the total number of interactions among all disciplines.
3. *Technical-Communication Mirroring over Time*: To observe changes in mirroring over time, we define *fractional mirrored interactions*, which is the ratio of number of interactions between tightly-coupled discipline pairs and the total number of interactions among all disciplines. Mathematically, if $N \times N$ matrix $S = \{s_{ij}\}$ is a design structure matrix for N disciplines, δ is a threshold for dichotomizing the degree of coupling, and $N \times N$ matrix $X = \{x_{ij}\}$ is the matrix

of pairwise interactions, then the fractional mirrored interactions are defined as,

$$\text{Fractional Mirrored Interactions} = \frac{\sum_{ij} \mathbf{1}_{>\delta}(s_{ij})x_{ij}}{\sum_{ij} x_{ij}}. \tag{5.1}$$

The indicator function $\mathbf{1}_{>\delta}(s_{ij})$ is 1 when s_{ij} is greater than δ and 0 otherwise. We calculate this metric over unit time intervals and take its moving average using a three-interval window.

4. *Discipline-wise Centrality Indices:* Two indices are used to quantify network centrality of disciplines with respect to incoming and outgoing interactions. *Hub index* estimates a discipline's centrality based on outgoing interactions, whereas *authority index* which estimates a discipline's centrality based on incoming links. The estimation procedure for the two indices is given in Ref. [103] and its implementation is available in Python library networkx [104].

Section 5.1 describes the specific differences and similarities along the aforementioned quantities. This works has been submitted for publication at American Society of Mechanical Engineers (ASME) Journal of Mechanical Design [105].

5.1 Comparison of Communication Patterns

In this subsection, we present the results from quantitative comparison of communication patterns in the reference setting and the model world. Based on the results, we make a qualitative evaluation of which key features of the reference setting are preserved in the model world in Table 5.1.

5.1.1 Communication Among Subsystem Designers

A very basic expectation was that the design process would create a need for communication among the subsystem designers. Due to the smaller size of the engine design problem and design team, we expected a smaller amount of communication in the engine case. However, surprisingly, the results suggest that the total number of interactions is similar in the NASA MDL setting and engine experiment. This can be seen from Figure 5.1 which plots the empirical estimates for the probability distribution of total pairwise interactions per team. The surprising similarity in the number of interactions is due to the different ways in which interactions are counted in the two settings: at NASA, a whole conversation counts as one interaction, whereas in the engine experiment, several interactions (one-way messages) might make up a conversation. For our purposes, the total number is not important, because we only compare relative trends in the number of interactions. At this point, we simply note that, indeed, the design problem did drive communication among designers.

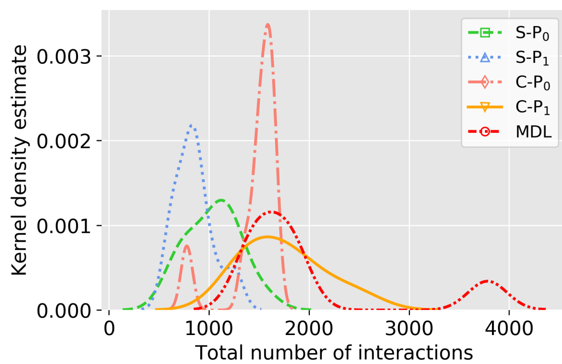


Figure 5.1. : Estimated distributions of total pairwise interactions per team. Kernel density estimates are found using a statistical data visualization library, seaborn [106].

5.1.2 Communication Mirroring Technical Dependencies

We expected to see greater communication among strongly coupled disciplines than among weakly coupled disciplines; in other words, that the strength of coupling would drive the amount of communication between pairs of disciplines. We therefore examine the correlation between the number of shared variables and the number of interactions between discipline pairs, in the engine and NASA MDL settings.

Figure 5.2 shows the results. As expected, the correlation between the number of shared variables and interactions is large and significant for both the NASA MDL and the engine case, particularly when a catalog is present. Specifically, for the NASA

MDL teams, the estimated slope of the linear fit between the number of interactions and the number of shared variables is 1.39 ($R^2 = 0.30$) with a two-sided p-value is less than 0.0001, using Wald's test whose null hypothesis is that the slope is 0 and the degrees of freedom (DoF) is 124. For the engine setting when catalogs are present, the C-P_0 and C-P_1 conditions, the slope coefficients are 3.97 ($R^2 = 0.30, \mathrm{DoF} = 52, p$-value < 0.0001) and 4.14 ($R^2 = 0.33, \mathrm{DoF} = 46, p$-value < 0.0001) respectively. When a catalog is not present, the S-P_0 and S-P_1 conditions, the correlation is lower (the slope coefficient is smaller). The slope coefficients are 1.91 (p-value < 0.001, DoF=82) and 0.42 (p-value=0.1, DoF=52) for the S-P_0 and S-P_1 conditions, respectively. The implications of these results, especially for condition S-P_1, are explored in Section 5.2.

The implication is that, indeed, the engine design task resulted in some degree of the mirroring seen in the reference NASA setting: communication patterns were correlated with technical dependencies. When catalogs were not present, this effect was diminished, which suggests that catalogs were important in enabling representativeness. This makes sense because catalogs were intended to help engine designers spend less time searching for subsystem-optimal solutions, perhaps enabling a larger focus on finding good solutions for the entire system by communicating with the other subsystem designers.

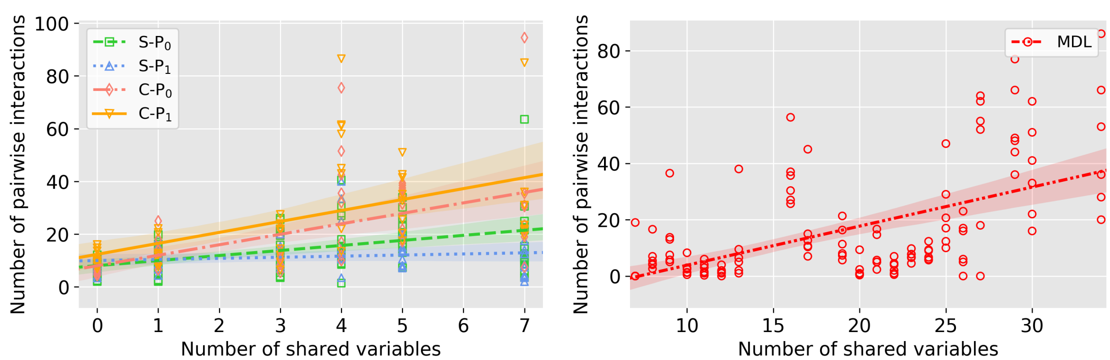


Figure 5.2. : Mirroring between technical dependence and the amount of communication in the engine design experiment (left) and in the NASA concurrent design facility (right).

5.1.3 Expertise in Resolving Interdependencies

We expected that the students' communication patterns should become more like those of the experts in the NASA setting over time, after they have learned about the problem – and in particular, learned which of the technical dependencies are most important in driving the performance of the final system. The NASA experts should know from the start which trades are most important and spend their communication effort on resolving those trades, while students would need to learn this first before their communication patterns would exhibit this characteristic.

To explore this in the data, we examine the 'fractional mirrored interactions' (defined earlier). Intuitively, this metric indicates what percentage of the interactions were between tightly-coupled disciplines. A value near 1 indicates that nearly all interactions were between tightly-coupled pairs and almost no interactions were between less-coupled pairs.

Figure 5.3 shows a moving average of this metric over time for the engine study (left) and the NASA reference setting (right). From Figure 5.3 (left), we observe that the fractional mirrored interactions in the engine setting remains close to 1 for first 4 to 5 time steps, but it reduces to about 0.7 or 0.8 after 15 time steps and then remains in this range. On the other hand, the NASA MDL teams start the spacecraft design task in the range of 0.7 to 0.8, with some slight increases up to about 0.9 towards the end of the time horizon.

The possible reason for this behavior is that the information about the design interdependencies that the students receive through the problem statement may drive their communication at an early stage. Indeed, they seem to communicate *only* about the most important interdependencies, to the exclusion of all else. However, partway through the task, this fraction reduces as low interdependence pairs communicate more, perhaps to explore trades that may help improve the system objectives. On the other hand, the NASA engineers know important interdependencies before the start of the task, but also focus throughout the task on surfacing issues that may require reconvening at the system level or coordinating with less-coupled subsystems. Towards the end, their focus may have narrowed to resolving one or two key trades among tightly-coupled subsystems.

The results suggest that, indeed, the students' communication patterns became more like those of the NASA experts after an initial learning period (the first 10-15 time steps). After this period, the fractional mirrored interactions were in the same range of 0.7 to 0.8 for most of the study. (The students did not exhibit the same slight rise in fractional mirrored interaction toward the end of the study.)

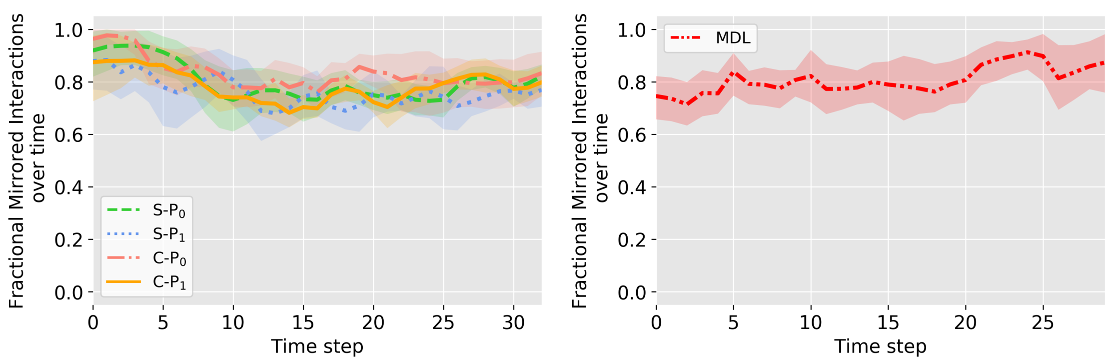


Figure 5.3. : Fractional mirrored interactions over time with 5^{th} and 95^{th} percentiles calculated for all discipline pairs.

However, a closer look at the data suggest that a large portion of the students' mirrored communication was through the broadcast function – which shows the state of the global objectives and, in the P_1 experimental setting, also shows shared design parameters. Figure 5.4 shows the fraction of mirrored interactions over time *excluding* integrative disciplines such as broadcast (in the engine experiment) and the systems engineer and customer (in the NASA MDL). While the students' proportion of interdependency-driven interactions went down significantly to around 0.5, the NASA experts' proportion remained roughly equal to its value when integrators were included (Figure 5.3), around 0.7 to 0.8.

These results suggest that the students came to rely upon the broadcast function for managing many of their trades, whereas the experts utilized more pairwise communication. It is possible that this difference in behavior is driven by the smaller size of the problem and the smaller number of pairwise interdependencies in the engine setting, and/or by the lack of a designer assigned to an integrative role in the engine setting. Better representativeness might require a more similar problem size and interdependency structure, and/or a specific integrative discipline such as systems engineering. Further study is necessary for testing this hypothesis.

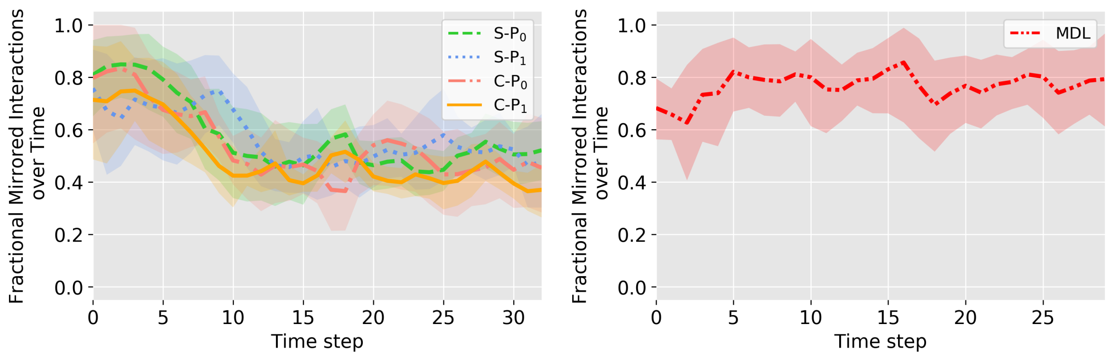


Figure 5.4. : Fractional mirrored interactions over time with 5^{th} and 95^{th} percentiles calculated after excluding integrative disciplines such as systems engineer, customer and broadcast screen.

5.1.4 Discipline-specific Communication Patterns

To further understand how the lack of a designated systems engineer might influence communication patterns in the engine experiment, we examine each discipline's hub and authority indices [103]. Hub and authority indices quantify how 'central' a discipline is in the communication network based on the number outgoing interactions and the number of incoming interactions, respectively. These indices use pairwise interactions from the entire duration of the tasks.

For NASA, Figure 5.5 suggests that the systems engineer is central to driving communication. In the engine experiment, on the other hand, there is no systems engineer. The broadcast function was intended to substitute for some functions of the systems engineer – computing the global objective value and, in the P_1 condition, sharing key design parameters. The results in Figure 5.6 show that the broadcast function is particularly important for outgoing links (see hub indices), which represent disciplines pulling updates about shared global objectives and design parameters. However, the broadcast has the lowest importance in the communication network as a receiver of information (indicated by small authority indices). Because our NASA data did not provide the direction of communication, it is not possible to distinguish the systems engineer's role in outgoing versus incoming communication. However, based on our qualitative understanding from extensive conversations with MDL systems engineers, their role is not the same as that apparently performed by the broadcast function – simply to update designers on objective and parameter values. Therefore, the results suggest that the broadcast function did not fulfill the same integrator role as the systems engineer, and that a more representative model world might require a designated systems engineer.

Figures 5.5 and 5.6 also provide insights into the communication patterns of the core disciplines *excluding* systems engineering and broadcast. The plots suggest that the differences in discipline-wise authority and hub indices are statistically insignificant in the engine experiment because of large variance in their values. Whereas at NASA MDL, these differences, albeit small, are statistically significant. The results suggest that the disciplines in the engine experiment contribute equally to the communication over the entire time horizon, and the NASA MDL disciplines are selective about with whom and how much they interact, likely due to their greater knowledge and experience in solving similar problems. Additionally, we do not observe significant differences in the centrality indices because of conditions C, S, P_0, and P_1. This suggests that the core disciplines' relative contributions to the overall communication likely remain unchanged despite changing how they search the design space and whether they can access global status of the design.

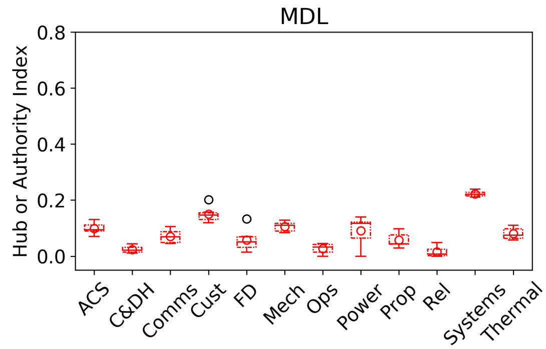


Figure 5.5. : Network centrality for the NASA MDL disciplines. Hub and authority indices [103] are the same for NASA MDL teams because pairwise communication is undirected.

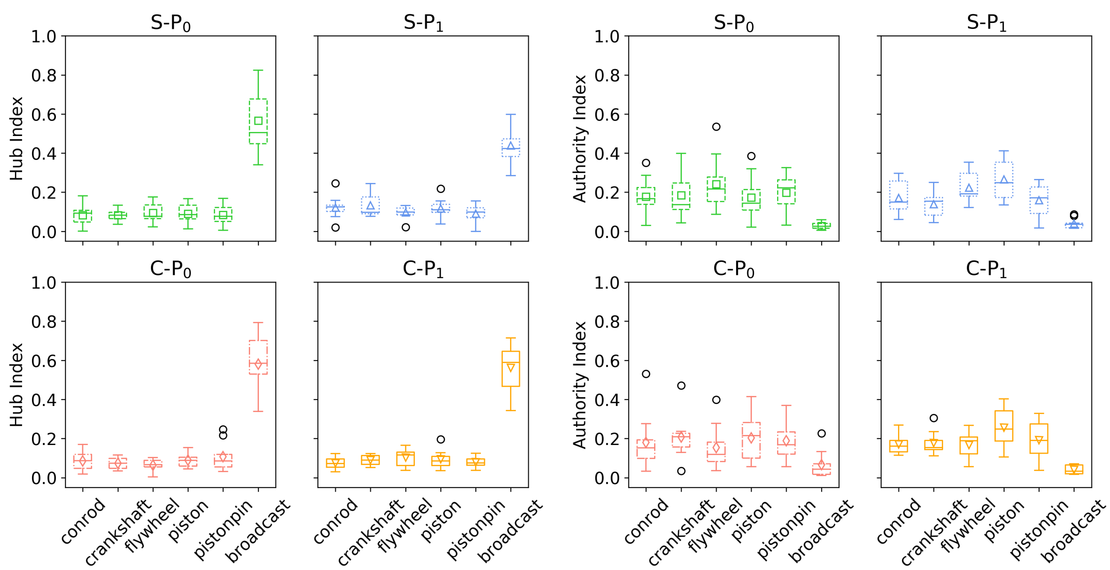


Figure 5.6. : Hub indices and authority indices [103] representing the disciplines' network centrality based on outgoing and incoming links, respectively. These network indices are computed using a network analysis library, networkx [104].

5.2 Generalizability of Engine Experimental Results

Next, we examine the results of the engine experimental treatments – the counterfactuals that could not be tested in the NASA setting. The results provide insights

Table 5.1. : The evaluation of whether key communication-related features are preserved in the model world

Category	**Evaluation**	**Evidence**
Communication among subsystem designers	Preserved	The design process creates a need for communication in both settings.
Communication mirroring technical dependencies	Preserved	The correlation between the number of shared variables and aggregate interactions is large and significant.
Expertise resolving interdependencies	Partially Preserved	In the engine setting, communication mirroring technical dependencies changes over time and the subjects rely overly on "broadcast" instead of textual communication.
Discipline-specific communication patterns	Not Preserved	The two settings used integrative functions differently, and in the engine setting, unlike the NASA setting, there are no significant differences in communication by the remaining individual subsystems (neither across time nor instantaneously).

about the effects of the design search method (catalogs vs. simulations) and global information availability (presence or absence of a parameter database) on the communication metrics. Subsequently, we consider the extent to which these results may be generalized to the NASA setting based on what was learned in Section 5.1 about which aspects of the engine experiment communication patterns matched those of the NASA reference setting – i.e., the engine setting's representativeness.

5.2.1 Effects of Using Catalogs Versus Simulations

When catalogs are used, interdisciplinary communication is more frequent (both in total across the task duration and at any given moment) as seen from Figure 5.1. The total number of interactions in the catalog treatments is greater on average than the total number of interactions based on an equal-variance t-test ($t(38) = 6.4, p = 0.001$). Also, the correlation between the number of interactions and the number of shared variables is also larger in the catalog treatments than the simulation treatments (see Figure 5.2). The implications of these results are presented in Section **??**.

5.2.2 Effects of Using Shared Parameter Databases

Within the engine design study, the global availability of design parameters does not appear to affect the total number of interactions (Figure 5.1). On the other hand, the mirroring results in Figure 5.2 show that, among all four experimental conditions, only the S-P_1 fails to exhibit mirroring (correlation between the number of interactions and the number of shared variables is insignificant). It is not entirely clear why this is the case. A possible reason is that more than half of the interactions there occur through text-based channels, as seen from Figure 5.7. And, as we observed before in Figures 5.3 and 5.4, the students' text-based interactions are in general less aligned with technical dependencies, while their interactions through broadcast are more aligned with technical dependencies (which also include many high-interdependence pairings). The condition S-P_1 means that students have easy access to their colleagues' design parameters but do not have easy access to a set of pre-computed solutions throughout their solution space. Perhaps the latter makes them less able to iterate on their designs and the former makes them less likely to communicate because they already have access to the information required. Further investigation is needed to confirm and to understand this result.

Across both experimental treatments, there is more and better-aligned communication with a catalog than with search within a continuous design space. This is a

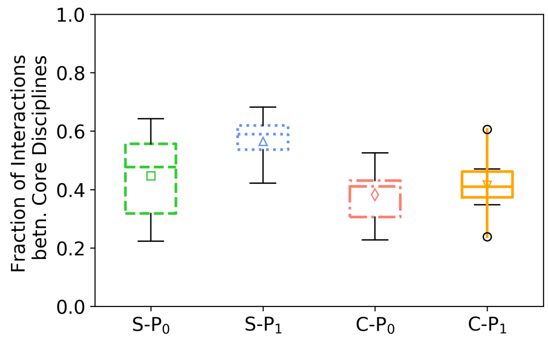


Figure 5.7. : The fraction of total interactions between the core disciplines, equivalently text-based interactions, for the experimental conditions.

consequence of having ex-ante evaluations of subsystem outputs available in a catalog. Quicker design exploration when using catalogs may facilitate a better understanding of input-output relationships, of how disciplines behave in response to the changes from shared variables, and of which designers need to coordinate with one another. Also, a catalog provides flexibility in broadcasting any design point from a large design space, thus encouraging more interactions with the broadcast and with other designers.

If we believe these findings are representative of NASA, they imply that design organizations such as NASA should consider maintaining or enabling catalog-based search, since it appears to lead to more and better-aligned communication. It remains to be seen, however, whether this translates into superior performance. On the other hand, global availability of design parameters does not, surprisingly, appear to improve communication patterns, suggesting that investment in IT backbones may need to focus on information exchange beyond just sharing design parameters, particularly in concurrent design settings with low barriers to interpersonal communication.

The key question, then, is whether we believe these findings are representative of NASA. The major differences found were around the key role of the systems engineer and the extent to which mirroring occurred through the broadcast function in the engine setting. It is possible that a skilled systems engineer could, in a continuous-search environment, compensate for the lower and less-aligned communication by prompting the right people to talk to one another, but the apparent advantages of catalog-based search would still be relevant, and therefore we consider it likely that catalogs would still prompt better communication patterns in the NASA setting. On the other hand, it is less clear why a shared parameter database made little difference in communication, so we recommend further investigation before applying these latter findings in practice.

5.3 Lessons for Designing Representative Experiments

This section presents suggestions for designing representative experiments with respect to the key features of system design process based on the "lessons learned" from our analysis and on knowledge from extant literature.

5.3.1 Align Design Expertise and Task Complexity

The design expertise of the subjects and the task complexity should be aligned to ensure that the subjects can achieve the given task objectives and that it prompts the desired types of behavior. For example, expert designers deploy different search strategies (depth-first more than breadth-first) depending on whether the design requirements are complex [107]. Engineering students in controlled studies take longer to complete coupled system design tasks as the number of design variables increases [108, 109]. Our study has shown that, in the context of studying interdisciplinary communication for parametric design, the engine design task possesses appropriate complexity for engineering students to successfully achieve the given objectives, as discussed in Section 4.1. Moreover, this subject-task interaction shows strong technical-communication mirroring because the students can understand the interdependencies of the problem and decide whom to interact with (see Figure 5.2). One particularly important result of this subject-task interaction is the design search behavior it prompts. This appeared crucial to driving representative communication patterns, since the teams with access to a catalog had more representative mirroring patterns (see Figure 5.2).

Thus, our results suggest that when designer expertise and task complexity are appropriately calibrated, novices can behave much more like experts. The engine problem is simpler than the spacecraft problem, but it was designed to be approximately as hard for the students as the spacecraft problem is for NASA engineers, and Section 5.1 suggests at least partial success in meeting this goal. Therefore, in designing a model world, it is critical to focus on matching subjects' behavior, which results from the interaction of the task complexity, time, information, & resources available, and subjects' expertise, rather than on matching task complexity or 'absolute' expertise level. To that end, Dorst [67]'s expertise framework is useful for thinking about the representativeness of subject-task interactions. Dorst categorizes such interactions into seven levels, for example, i) a naive designer makes a one-off choice from available options, ii) a novice designer follows strict rules or a formal process to meet fixed requirements given by exerts, iii) an advanced designer adapts a formal process for considering situational aspects, etc. If the subject-task interactions in different settings fall under the same level of Dorst's framework, then we can be more confident that behavior will be representative across the settings.

5.3.2 Allow a Burn-in Period

Compare behaviors after subjects are attuned to making decisions in the given controlled experimental setting. There are several reasons for this. First, subject behaviors from a transient period before becoming fully aware of the setting can be a result of framing effects or human biases such as anchoring bias and availability heuristic [110]. For example in Figure 5.3, we observe a high mirroring at the beginning of the engine design task because its problem framing indicated possible dependencies with other subsystems. Cash et. al [111] observed similar phenomenon where their student teams spent a lot of time initially finding information within a source such as a website for a product, however the equivalent time spent by advanced designers on information seeking initially was much lower (see Figure 6 in [111]). Yu et al. [112] compared student behaviors and practitioner behaviors for the parametric design of desalination systems. They observed that the initial jump from a given design point towards the desired design space was significantly bigger and faster for practitioners with high knowledge levels than for students with no knowledge.

A second reason why it is important to wait for burn-in is because it might be necessary in order to build subject expertise in the problem, in order to appropriately match the expertise to task complexity as advised in Section 5.3.1. If the problem cannot be simplified to match subjects' expertise (in our case because it was necessary to include a number of coupled design variables – see Section 5.3.4), then it appears possible to build task-specific expertise within the model world, by allowing this burn-in period. In the engine experiment, while novices did not behave like experts initially, after a relatively short period of exploration, they learned which interactions drove performance and settled into a pattern that better replicated expert behavior. This is akin to a burn-in period in reliability testing.

Other reasons why burning the first few decision steps may be necessary are: i) when subjects need to get familiar with the user interface [113]; ii) if two consecutive conditions require conflicting skills (order effects), e.g., using two different coding languages; initial decisions in the second condition may be tainted, because using one coding language may inhibit the skills required for using the other coding language; iii) if subjects receive benefits during the experiment, then the endowment effect will influence decisions of the remaining experiment [72]. For tips on building an environment for economic experiments, the reader may refer to Refs. [69, 70].

Therefore, our results and the literature suggest that students behave more like experts (but not exactly like experts) after they have had time to learn about the problem setting.

5.3.3 Consider Organizational Structure and Incentives

The organizational structure and incentives must be considered for their roles in constraining and motivating behavior. There are several aspects to this issue that arose in our model world.

First, imposed roles or organizational structure can influence behavior. One particularly important aspect of organizational structure in this case was the formal integrator role – the "systems engineer" in the NASA MDL setting. Having an integrative discipline or designer assigned in the engine setting could have improved representativeness with respect to the nature of system design process and interactions between disciplines. For instance, Figures 5.2 and 5.4 suggest that incorporating the broadcast discipline into the engine design teams enabled stronger technical-communication mirroring, comparable to the NASA MDL setting. However, the lack of a designated integrator role also led to differences in some communication patterns, which might have been better matched if such a role had been included in the engine setting. It is well known that engineers spend 40% to 60% of their time communicating with peers working on the same projects [114]. Among such peers, there are individuals, called "gatekeepers" (equivalently integrative disciplines), whom others heavily depend upon for internal as well as external sources of information. The "gatekeepers" of engineering teams provide efficient means to disseminate outside and internal information [115]. Computer-based tools can also be supportive of systems engineering for dispersed teams if appropriate sufficient means of data exchange and communication are incorporated [116]. Designers of future model worlds should consider the potential importance of an integrator role or tool in prompting representative behavior.

Second, more generally, other aspects of our organizational structure and incentives were more successful in constraining and motivating representative behavior. The similarities in amount and mirroring of communication between NASA and the engine experiment suggest that our incentives and organizational structure successfully motivated students to work together to find a good solution and constrained their communication pathways to represent those of the experts. Had we enabled the small 5-person teams to discuss things all together, rather than restricting them to pairwise communication, the patterns would likely have been very different and less representative of the MDL patterns. The literature bears out the importance of appropriate incentives and their interaction with design communication. Success in cross-functional teams requires setting appropriate project goals for effective communication [117]. Teams can address design inconsistencies through design dialogue if common goals are identified early in the process [15]. Moreover, formal and informal organizational channels for communication are known to influence how communication happens [115, 116], so it is important to set those up in a representative manner.

It is important to realize that "representative" does not always mean "same". In the MDL, the incentives for good performance revolve around career goals and employment. In the engine setting, performance was incentivized with a small monetary reward for better-performing designs. These incentives are a relatively poor match to the real world reference system, but it is very difficult to replicate long-term career incentives in a classroom setting. Our results suggest that classroom incentives motivated designers to design good systems and that was sufficient for representativeness.

5.3.4 Select Appropriate Degree of Design Coupling

For studying system design processes, the disciplines in a controlled setting should be selected so that they possess between-disciplines coupling that is comparable to the reference setting. It is well understood that a larger degree of coupling increases individual effort and total completion time while decreasing solution quality [40, 42]. In this study, to reduce effects of overly coupled problems, we used a mix of tightly coupled and loosely coupled disciplines in the controlled experiment to match the NASA MDL settings – see the distribution of shared design variables in Figure 4.2 – and the results suggest that it was sufficient to generate relatively similar communication patterns. Based on our results, it is not clear what degree of match in coupling among design variables is "enough" to generate similar communication patterns, but roughly matching the coupling worked well in this case.

Broadly, we can compare the overall degree of coupling between two settings using eigenvalue analysis of their design structure matrices (DSMs). Since a DSM quantifies degrees of coupling between discipline pairs, eigenvectors represent directions to different discipline groups in a latent space where nearby disciplines have similar coupling. The eigenvalues represent the degree of separation between different disciplines in the latent space [118]. The larger the eigenvalue the larger is the separation of its discipline from other disciplines. If we normalize a DSM with the maximum possible coupling, the eigenvalues greater than 1 represent dominant disciplines that are highly coupled. For instance, the largest eigenvalue of the spacecraft DSM in Figure 5.8 corresponds to systems engineering which is the most coupled discipline. For the engine DSM, two largest eigenvalues correspond to broadcast and crankshaft. Interestingly, when broadcast is omitted from the engine DSM, the relative differences in eigenvalues reduce. This implies that the disciplines are less separable in the latent space without broadcast and that broadcast enhances the degree of coupling for the crankshaft discipline.

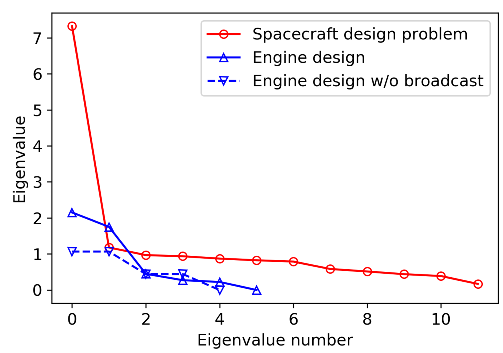


Figure 5.8. : Eigenvalues of the normalized design structure matrices.

5.3.5 Select Team Size Appropriate for Research Question

The size of a design team may be secondary to choosing an appropriate coupling, because successful collaboration is possible in small teams as well as in teams with large numbers of people [119, 120]. One can use various statistical analyses to determine the team size that provides the same strength of technical-communication mirroring between a model world and a reference setting. Suppose we are interested in understanding the correlation between two given properties of discipline pairs, say X and Y. The question is how many discipline pairs we need for such comparison. This question can be answered based on some rule-of-thumb or classical statistical tables [121]. If tables are inaccessible, alternative methods for approximating the sample size are available [122]. For multiple regression with m predictors and the estimated correlation of R^2 (equivalently effect size $f^2 = \frac{R^2}{1-R^2}$), the suggested number of samples is $N \geq \frac{6.4+1.65m-0.05m^2}{f^2}$. If the goal is to find the partial correlation of each individual predictor, then the number of samples should be $N \geq 8/f^2 + (m-1)$. For example, consider the correlation between the number of shared variables (X) and the number of interactions (Y) for which $m = 1$. If the observed correlation from the NASA MDL teams is $R^2 = 0.30$ ($f^2 = 0.43$), then the number of samples (i.e. the number of discipline pairs) for observing similarly significant correlation from the controlled setting is $N \geq 18.6$. Further, since the number of discipline pairs $N = (n^2-n)/2$ is a function of the number of teams n, we can estimate the number of disciplines as $n \geq 6.6$. Then, in hindsight, the choice of number of disciplines (recall, total number of components including broadcast is 6) in the engine experiment is roughly correct.

5.3.6 Gather Many Samples and Compare Aggregate Behaviors

Large number of observations of team behaviors are always preferable because repetition is helpful for identification of noise and accurate comparison of behaviors at the aggregate level. For instance, the results from Figure 5.1 and Figure 5.3 show noisy communication patterns for the student teams as well as the NASA MDL teams which are different when mean statistics are compared. However, large sample may not be always necessary. It is possible to get estimates of the sample size using classical statistical techniques. Suppose we are comparing team property X between a reference and a model world and interested in accessing the validity of null hypothesis that X has the same average in both the settings. The number of teams required per setting then can be determined from tables or using the normal cumulative distribution function if we assume that observations of X have normal distribution [123]. For example, if we assume that both X's have the same known variance σ but different means, then, to reject the null hypothesis against an alternative hypothesis that the mean difference is Δu, we require $N \geq \left(\frac{z_\alpha + \Phi^{-1}(1-\beta)}{\Delta u/\sigma}\right)$ number of teams [123], where Φ^{-1} is the inverse cumulative distribution function and $z_\alpha = \Phi^{-1}(\alpha)$ is the quantile of probability α. This method requires a priori specification of the probability of Type I error α and the probability of Type II error β.

6. CONCLUSIONS

This dissertation investigated how designers make information acquisition decisions and identified models that provide the best description of designers' decision strategies. The approach combined computational modeling and behavioral experiments with statistical Bayesian inference to quantify designers' decision strategies in design exploration and team communication, which are regarded as information sources. The two research studies found that the subjects relied upon simple heuristics shared via graphical interfaces and communication pathways for making the information acquisition decisions. The identified simple heuristics are fast and frugal in that they consume less time and involve relatively small cognitive effort. This chapter summarizes the case-specific decision strategies identified in two research studies. In addition, the chapter provides key contributions in terms of new case-specific descriptive models and managerial insights to improve the engineering design process. Finally, we discuss future research directions necessary for wider applicability of the proposed approach in engineering design.

6.1 Summary of Key Results

1. **Information acquisition decisions of engineering students**:
   The research study in Chapter 3 finds that the student subjects rely upon simple cues accessible via graphical interfaces for making the most of the **design exploration decisions**. This reliance on simple cues for making design decisions may be attributed to the relatively smaller cognitive effort involved in using simple cues. For example, the subjects select design points close to the highest upper confidence bound (UCB & 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 engine design experiment in Chapter 4 provides estimates of student subjects' **communication decisions**. The student subjects choose whether to performs local search or to communicate to other disciplines on the design team based on the present state of system and subsystem performance. If the present system performance levels or present subsystem performance levels are low (i.e. poor performance), then the future number of local design searches increases. The number of local searches reduces if the performance levels are high. Similarly, if the present system mass performance is large, then the rate of text-based interactions reduces. Finally, the student subjects select disciplines for text-based interactions with whom they have high the design interdependence and the pairwise reciprocity.

2. **Information acquisition decisions of expert engineers**:

   Chapter 4 provides insights about how expert designers **decide whom to communicate with** during the design process. In the NASA MDL teams, low spacecraft mass performance levels in the present results in larger the face-to-face communication between disciplines, but only in immediate future and not over a long term. The spacecraft mass performance levels from a longer term in the past has a positive effect on face-to-face communication. This suggests that any reduction in spacecraft dry mass level (i.e. increase in actual mass) brings the disciplines together for communication in immediate future ($\tau = 1$ hour). But good performance levels over time still fuel more face-to-face interactions in longer future (τ = 5 and τ = 10 hours), unlike in the student teams where the correlation between present performance levels and future text-based interactions is largely negative. Further, expert engineers' choices of whom to communicate with suggest statistically significant effects of the pairwise design interdependence, nodewise popularity, and pairwise reciprocity.

3. **Effects of incentives and restricted budget on students' design exploration decisions**

   From the sequential information acquisition experiment of Chapter 3, we observe that the subjects' decisions are affected by the amount of fixed budget and incentives to save budget. Exploration of design space increases with the increase in fixed budget. 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. The probability of stopping early at high values of remaining budget increases with increase in fixed budget and with the incentive-to-save budget.

4. **Effects of design exploration process and effects of the global availability of information on communication decisions**: Within the engine design study in Chapter 4, when catalogs are used, interdisciplinary communication is more frequent (both in total across the task duration and at any given moment). The global availability of design parameters does not appear to affect the total number of interactions.

5. **Effects of communication on design performance**

   For the student teams designing an engine, the effects of amount of communication on the design performance change are small across the experimental conditions and statistically significant only in selective conditions. On the other hand, for the NASA engineers designing a spacecraft system, the number of inter-discipline interactions by core disciplines has a strong and statistically significant effect on the performance of spacecraft dry mass over long period of time (τ = 10 hours). The amount of communication does not appear to result in immediate improvement in the performance levels of spacecraft dry mass (τ = 1 or τ = 5 hours) likely because a complex system such as a spacecraft may require longer discussions to achieve meaningful improvement in the design performance.

6.2 Contributions

1. The primary contribution of this paper is **an approach that combines computational models of decision making and behavioral experiments** to understand human decision making behavior in design under uncertainty. The methodology used for eliciting decisions and estimating decision strategies is particularly suited for the embodiment phase of the design process. The descriptive decision models and Bayesian inference methodology are general and can be applied to more design decision-making problems.
2. The paper points to **specific descriptive decision models** that describe the subjects' information acquisition decisions better than the counterpart expected utility-based models. With the models that incorporate simple heuristics, researchers can quantify design performance in terms of designers' decision strategies. The applications of this include design crowdsourcing where game-theoretic models lack design process models [51–53], and the agent-based models of engineering systems design where characterization of quality as a function of designer effort is difficult to achieve [54]. Further, 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 towards expected utility (EU)-based strategies, which are efficient for maximizing net payoff (design performance minus cost of evaluation).
3. Another contribution is the **insights for system engineers and managers** about nudging towards 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 exploration decisions. Further, if we believe the findings from engine design experiment are representative of NASA, they imply that design organizations such as NASA should consider maintaining or enabling catalog-based search, since it appears to lead to more and better-aligned communication. It remains to be seen, however, whether this translates into superior performance. On the other hand, global availability of design parameters does not, surprisingly, appear to improve communication patterns, suggesting that investment in IT backbones may need to focus on information exchange beyond just sharing design parameters, particularly in concurrent design settings with low barriers to interpersonal communication.

6.3 Future Research Directions

There is still a long way to go before achieving wider applicability of descriptive decision models and Bayesian methodology for the engineering design process. Future descriptive modeling efforts need to account for context-dependent design situations, where decision strategies depend upon the availability of problem-specific information or the lack thereof [124], an acceptable quantification of predictive uncertainty is absent [125], and the mapping between resources expended and the value of prototypes created varies across disciplines and knowledge domains [126]. Such design situations should include multiple objectives and/or multidimensional design parameters [127]. Also, many factors influencing team dynamics are yet to be studied such as gender, ethnicity, years of experience, and technical discipline [128]. Prior knowledge of designers plays a role in design decisions such as evaluating similarity [129] and responses to theory-related questions [130]. More importantly, the future research needs to address the following issues.

1. **Descriptive decision analysis workflow**: The guidelines for designing descriptive models and experiment design in Chapter 2 constitute a first step in

formalizing the descriptive decision analysis approach. Further improvements in the guidelines will benefit wide engineering design community. Future research on this direction can investigate key questions such as: (i) which components of the design process are suitable for descriptive decision analysis and how do we extend the approach to more components?, (ii) how can machine learning and artificial intelligence tools aid the automation of descriptive decision analysis?, and (iii) how can descriptive decision analysis support innovation (e.g. better design performance, faster design convergence, better coherence in team communication, etc.) in the engineering design process?

2. **Aggregating case-specific insights for theory building**: The key question that the future research should address is: How to integrate case-specific insights from different descriptive decision analyses for building theoretical foundations of engineering design? Given the situated nature of engineering design process, the descriptive decision analysis approach is useful for generating case-specific insights. Also, an important benefit of the descriptive decision analysis is its quantitative nature. However, every controlled experiment can ideally study one or two research questions. Studying the design of complex systems will require studying multiple research questions in many different aspects of engineering design. Studying the effects of every factors such as prior knowledge, problem specific factors such as complexity, incentives and communication pathways available, etc., will require designing and executing different controlled experiments. One way forward may be to develop a mathematical framework connecting descriptive decision models from different design conditions using hierarchical modeling. Such framework will be represented by a network graph with different decision making scenarios as its nodes.

REFERENCES

REFERENCES

[1] David L Marples. The decisions of engineering design. *IRE Transactions on Engineering Management*, (2):55–71, 1961.

[2] Jack Mostow. Toward better models of the design process. *AI magazine*, 6(1):44, 1985.

[3] Farrokh Mistree, Douglas Muster, Jon A Shupe, and Janet K Allen. A decision-based perspective for the design of methods for systems design. 1989.

[4] G. A. Hazelrigg. A framework for decision-based engineering design. *Journal of Mechanical Design*, 120(4):653–658, 1998.

[5] Wei Chen, Christopher Hoyle, and Henk Jan Wassenaar. *Decision-based design: Integrating consumer preferences into engineering design*. Springer Science & Business Media, 2012.

[6] Vijitashwa Pandey. *Decision based design*. CRC Press, 2013.

[7] George A Hazelrigg. An axiomatic framework for engineering design. *Journal of Mechanical Design*, 121(3):342–347, 1999.

[8] Roxanne A Moore, David A Romero, and Christiaan JJ Paredis. Value-based global optimization. *Journal of Mechanical Design*, 136(4):041003, 2014.

[9] Ying Xiong, Wei Chen, and Kwok-Leung Tsui. A new variable-fidelity optimization framework based on model fusion and objective-oriented sequential sampling. *Journal of Mechanical Design*, 130(11):111401–111401–9, 2008.

[10] G Gary Wang and Songqing Shan. Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical design*, 129(4):370–380, 2007.

[11] Haitao Liu, Yew-Soon Ong, and Jianfei Cai. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. *Structural and Multidisciplinary Optimization*, 57(1):393–416, 2018.

[12] Piyush Pandita, Ilias Bilionis, and Jitesh Panchal. Extending expected improvement for high-dimensional stochastic optimization of expensive black-box functions. *Journal of Mechanical Design*, 138(11):111412–111412–8, 2016.

[13] Piyush Pandita, Ilias Bilionis, Jitesh Panchal, BP Gautham, Amol Joshi, and Pramod Zagade. Stochastic multiobjective optimization on a budget: Application to multipass wire drawing with quantified uncertainties. *International Journal for Uncertainty Quantification*, 8(3), 2018.

[14] R Tebay, J Atherton, and SH Wearne. Mechanical engineering design decisions: instances of practice compared with theory. *Proceedings of the Institution of Mechanical Engineers, Part B: Management and engineering manufacture*, 198(2):87–96, 1984.

[15] Louis L Bucciarelli and Louis L Bucciarelli. *Designing engineers*. MIT press, 1994.

[16] Patricia M King and Karen Strohm Kitchener. Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational psychologist*, 39(1):5–18, 2004.

[17] Joshua D. Summers and Jami J. Shah. Mechanical engineering design complexity metrics: Size, coupling, and solvability. *Journal of Mechanical Design*, 132(2):021004–021004–11, 2010.

[18] Steven D Eppinger and Karl T Ulrich. *Product design and development*. New York, NY, 1995.

[19] Roger E Bohn. Noise and learning in semiconductor manufacturing. *Management Science*, 41(1):31–42, 1995.

[20] David H Jonassen. Instructional design as design problem solving: An iterative process. *Educational Technology*, pages 21–26, 2008.

[21] Robin S Adams, Jennifer Turns, and Cynthia J Atman. Educating effective engineering designers: The role of reflective practice. *Design studies*, 24(3):275–294, 2003.

[22] Jitesh H Panchal, Zhenghui Sha, and Karthik N Kannan. Understanding design decisions under competition using games with information acquisition and a behavioral experiment. *Journal of Mechanical Design*, 139(9):091402–091402–12, 2017.

[23] Murtuza Shergadwala, Ilias Bilionis, Karthik Kannan, and Jitesh H Panchal. Quantifying the impact of domain knowledge and problem framing on sequential decisions in engineering design. *Journal of Mechanical Design*, 140(10):101402–101402–13, 2018.

[24] Peter C. Fishburn. Expected utility: An anniversary and a new era. *Journal of Risk and Uncertainty*, 1(3):267–283, 1988.

[25] George A Hazelrigg. Validation of engineering design alternative selection methods. *Engineering Optimization*, 35(2):103–120, 2003.

[26] Henk Jan Wassenaar and Wei Chen. An approach to decision-based design with discrete choice analysis for demand modeling. *Journal of Mechanical Design*, 125(3):490–497, 2003.

[27] Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

[28] Donald R Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of global optimization*, 21(4):345–383, 2001.

[29] Luc Huyse, Sharon L Padula, R Michael Lewis, and Wu Li. Probabilistic approach to free-form airfoil shape optimization under uncertainty. *AIAA journal*, 40(9):1764–1772, 2002.

[30] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

[31] Susan Finger and John R Dixon. A review of research in mechanical engineering design. part i: Descriptive, prescriptive, and computer-based models of design processes. *Research in engineering design*, 1(1):51–67, 1989.

[32] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

[33] Gerd Gigerenzer and Wolfgang Gaissmaier. Heuristic decision making. *Annual review of psychology*, 62:451–482, 2011.

[34] Herbert A. Simon. A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99, February 1955.

[35] Cleotilde Gonzalez, Javier F Lerch, and Christian Lebiere. Instance-based learning in dynamic decision making. *Cognitive Science*, 27(4):591–635, 2003.

[36] Christopher McComb, Jonathan Cagan, and Kenneth Kotovsky. Capturing human sequence-learning abilities in configuration design tasks through markov chains. *Journal of Mechanical Design*, 139(9), 2017.

[37] Christopher McComb, Jonathan Cagan, and Kenneth Kotovsky. Mining process heuristics from designer action data via hidden markov models. *Journal of Mechanical Design*, 139(11), 2017.

[38] Zhenghui Sha, Karthik N Kannan, and Jitesh H Panchal. Behavioral experimentation and game theory in engineering systems design. *Journal of Mechanical Design*, 137(5):051405–051405–10, 2015.

[39] Thurston Sexton and Max Yi Ren. Learning an optimization algorithm through human design iterations. *Journal of Mechanical Design*, 139(10):101404–101404–10, 2017.

[40] Forest Flager, David Jason Gerber, and Ben Kallman. Measuring the impact of scale and coupling on solution quality for building design problems. *Design Studies*, 35(2):180–199, 2014.

[41] Paul Egan and Jonathan Cagan. Human and computational approaches for design problem-solving. In *Experimental Design Research*, pages 187–205. Springer, 2016.

[42] Paul T Grogan and Olivier L de Weck. Collaboration and complexity: an experiment on the effect of multi-actor coupled design. *Research in Engineering Design*, 27(3):221–235, 2016.

[43] Andy Dong, Andrew W Hill, and Alice M Agogino. A document analysis method for characterizing design team performance. *J. Mech. Des.*, 126(3):378–385, 2004.

[44] Christopher McComb, Jonathan Cagan, and Kenneth Kotovsky. Lifting the veil: Drawing insights about design teams from a cognitively-inspired computational model. *Design Studies*, 40:119–142, 2015.

[45] Richard Fujimoto, Conrad Bock, Wei Chen, Ernest Page, and Jitesh H Panchal. *Research challenges in modeling and simulation for engineering complex systems*. Springer, 2017.

[46] Daniel J Power. A brief history of decision support systems. *DSSResources.com*, 3, 2007.

[47] Olivier Casse. *SysML in Action with Cameo Systems Modeler*. Elsevier, 2017.

[48] Andrei Borshchev, Sally Brailsford, Leonid Churilov, and Brian Dangerfield. Multi-method modelling: Anylogic. *Discrete-event simulation and system dynamics for management decision making*, pages 248–279, 2014.

[49] Hans-Peter Hoffmann. Deploying model-based systems engineering with ibm® rational® solutions for systems and software engineering. In *2012 IEEE/AIAA 31st Digital Avionics Systems Conference (DASC)*, pages 1–8. IEEE, 2012.

[50] Zhenghui Sha, Qize Le, and Jitesh H Panchal. Using sysml for conceptual representation of agent-based models. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 54792, pages 39–50, 2011.

[51] Ashish M Chaudhari. *Crowdsourcing for engineering design: theoretical and experimental studies*. PhD thesis, Purdue University, 2017.

[52] Ashish M Chaudhari, Zhenghui Sha, and Jitesh H Panchal. Analyzing participant behaviors in design crowdsourcing contests using causal inference on field data. *Journal of Mechanical Design*, 140(9):091401, 2018.

[53] Zhenghui Sha, Ashish M Chaudhari, and Jitesh H Panchal. Modeling participation behaviors in design crowdsourcing using a bipartite network-based approach. *Journal of Computing and Information Sciese in Engineering*, 19(3):031010, 2019.

[54] Salar Safarkhani, Ilias Bilionis, and Jitesh H. Panchal. Understanding the effect of task complexity and problem-solving skills on the design performance of agents in systems engineering. volume 2A: 44th Design Automation Conference, 08 2018. V02AT03A060.

[55] Jitesh H Panchal and Zoe Szajnfarber. Experiments in systems engineering and design research. *Systems Engineering*, 20(6):529–541, 2017.

[56] Henry Montgomery. Decision rules and the search for a dominance structure: Towards a process model of decision making. *Advances in psychology*, 14:343–369, 1983.

[57] Glenn J. Browne, Mitzi G. Pitts, and James C. Wetherbe. Cognitive stopping rules for terminating information search in online tasks. *MIS Quarterly*, 31(1):89–104, 2007.

[58] Gerd Gigerenzer and Peter M Todd. *Simple heuristics that make us smart*. New York : Oxford University Press, 1999.

[59] Glenn J Browne and Mitzi G Pitts. Stopping rule use during information search in design problems. *Organizational Behavior and Human Decision Processes*, 95(2):208–224, 2004.

[60] Jerome R Busemeyer and Amnon Rapoport. Psychological models of deferred decision making. *Journal of Mathematical Psychology*, 32(2):91–134, 1988.

[61] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer School on Machine Learning*, pages 63–71. Springer, 2003.

[62] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[63] Konstantinos V Katsikopoulos. The less-is-more effect: Predictions and tests. *Judgment and Decision Making*, 5(4):244–257, 2010.

[64] Ali Borji and Laurent Itti. Bayesian optimization explains human active search. In *Advances in neural information processing systems, Lake Tahoe, Nevada*, pages 55–63, 2013.

[65] Darryl A Seale and Amnon Rapoport. Sequential decision making with relative ranks: An experimental investigation of the" secretary problem". *Organizational behavior and human decision processes*, 69(3):221–236, 1997.

[66] Sequential decision making: When to stop? `https://engineering.purdue.edu/DELP/sites/default/files/ME597/Lecture_15.pdf`. Accessed: 2020-6-30.

[67] Kees Dorst. Design research: a revolution-waiting-to-happen. *Design studies*, 29(1):4–11, 2008.

[68] Stephanie C Thompson and Christiaan JJ Paredis. An investigation into the decision analysis of design process decisions. *Journal of Mechanical Design*, 132(12):121009, 2010.

[69] Steven D Levitt and John A List. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives*, 21(2):153–174, 2007.

[70] John A List. Why economists should conduct field experiments and 14 tips for pulling one off. *Journal of Economic perspectives*, 25(3):3–16, 2011.

[71] William R Shadish, Thomas D Cook, Donald Thomas Campbell, et al. *Experimental and quasi-experimental designs for generalized causal inference/William R. Shedish, Thomas D. Cook, Donald T. Campbell.* Boston: Houghton Mifflin,, 2002.

[72] Keith M Marzilli Ericson and Andreas Fuster. The endowment effect. *Annual Review of Economics*, 6(1):555–579, 2014.

[73] John Salvatier, Thomas V Wiecki, and Christopher Fonnesbeck. Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2:e55, 2016.

[74] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[75] Model checking and diagnostics. `https://pymc-devs.github.io/pymc/modelchecking.html`. Accessed: 2020-6-30.

[76] Adrian E Raftery and Steven M Lewis. The number of iterations, convergence diagnostics and generic metropolis algorithms. *Practical Markov Chain Monte Carlo*, 7(98):763–773, 1995.

[77] Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. Visualization in bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):389–402, 2019.

[78] Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pages 199–213. Springer, 1998.

[79] David J Spiegelhalter, Nicola G Best, Bradley P Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639, 2002.

[80] Angelika Van Der Linde. Dic in variable selection. *Statistica Neerlandica*, 59(1):45–56, 2005.

[81] Sumio Watanabe. A widely applicable bayesian information criterion. *Journal of Machine Learning Research*, 14(Mar):867–897, 2013.

[82] Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for bayesian models. *Statistics and computing*, 24(6):997–1016, 2014.

[83] John Salvatier, Thomas V Wiecki, and Christopher Fonnesbeck. Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2:e55, 2016.

[84] Ashish M Chaudhari, Ilias Bilionis, and Jitesh H Panchal. Descriptive models of sequential decisions in engineering design: An experimental study. *Journal of Mechanical Design*, 142(8), 2020.

[85] HJM Freriks, WPMH Heemels, GJ Muller, and JH Sandee. On the systematic use of budget-based design. In *INCOSE International Symposium, Orlando, FL*, volume 16, pages 788–803. Wiley Online Library, 2006.

[86] Petru-Daniel Moroşan, Romain Bourdais, Didier Dumur, and Jean Buisson. Building temperature regulation using a distributed model predictive control. *Energy and Buildings*, 42(9):1445–1452, 2010.

[87] Yew San Hor, Anthony J Williams, Joseph G Checkelsky, Pedram Roushan, Jungpil Seo, Qiang Xu, Henny W Zandbergen, A Yazdani, NP Ong, and Robert Joseph Cava. Superconductivity in $cu_x bi_2 se_3$ and its implications for pairing in the undoped topological insulator. *Physical review letters*, 104(5):057001, 2010.

[88] Jon C Helton. Treatment of uncertainty in performance assessments for complex systems. *Risk analysis*, 14(4):483–511, 1994.

[89] Efstratios Nikolaidis, Dan M Ghiocel, and Suren Singhal. Engineering design reliability handbook. CRC Press, Boca Raton, FL, 2004.

[90] Christoph H Loch, Christian Terwiesch, and Stefan Thomke. Parallel and sequential testing of design alternatives. *Management Science*, 47(5):663–678, 2001.

[91] E. Dahan and H. Mendelson. Optimal sequential and parallel prototyping policies. In *Innovation in Technology Management. The Key to Global Leadership. PICMET, Portland, OR.* IEEE, July 1997.

[92] Timothy W Simpson, Mary Frecker, Russell R Barton, and Ling Rothrock. Graphical and text-based design interfaces for parameter design of an i-beam, desk lamp, aircraft wing, and job shop manufacturing system. *Engineering with Computers*, 23(2):93, 2007.

[93] WJ Chen and KM Wallace. Manufacturing information for designers. In *Proceedings of the International Conference on Engineering Design (ICED'93), The Hague*, volume 2, pages 1111–1114, 1993.

[94] Daniel L Chen, Martin Schonger, and Chris Wickens. otree-an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016.

[95] Fritz Strack. "order effects" in survey research: Activation and information functions of preceding questions. In *Context effects in social and psychological research*, pages 23–34. Springer, 1992.

[96] Ashish M. Chaudhari, Ilias Bilionis, and Jitesh H. Panchal. How Do Designers Choose Among Multiple Noisy Information Sources in Engineering Design Optimization? An Experimental Study. volume 2A: 44th Design Automation Conference, Quebec City, Canada. V02AT03A021.

[97] Robert B Gramacy and Herbert K H Lee. Optimization under unknown constraints. In J Bernardo, MJ Bayarri, JO Berger, AP Dawid, D Heckerman, AFM Smith, and M West, editors, *Bayesian Statistics 9*, pages 229–256. Oxford University Press, Oxford, 2011.

[98] Eldar Shafir and Amos Tversky. Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive psychology*, 24(4):449–474, 1992.

[99] Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for bayesian models. *Statistics and computing*, 24(6):997–1016, 2014.

[100] Alp Kucukelbir, Dustin Tran, Rajesh Ranganath, Andrew Gelman, and David M Blei. Automatic differentiation variational inference. *The Journal of Machine Learning Research*, 18(1):430–474, 2017.

[101] Connor Forsythe, Nikolai Joseph, Zoe Szajnfarber, and Erica Gralla. Visualizing communication patterns in design teams over time. In *Systems Engineering in Context*, pages 449–461. Springer, 2019.

[102] Zoe Szajnfarber and Erica Gralla. A framework for measuring the "fit" between product and organizational architectures. In *Disciplinary Convergence in Systems Engineering Research*, pages 483–499. Springer, 2018.

[103] Amy N Langville and Carl D Meyer. A survey of eigenvector methods for web information retrieval. *SIAM review*, 47(1):135–161, 2005.

[104] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[105] Ashish M Chaudhari, Erica Gralla, Zoe Szajnfarber, Paul T. Grogan, and Jitesh H Panchal. Designing representative model worlds to study socio-technical phenomena: A case study of communication patterns in engineering systems design. *Journal of Mechanical Design*, 2020.

[106] Micheal Waskom. Statistical data visualization.

[107] Linden J Ball, Balder Onarheim, and Bo T Christensen. Design requirements, epistemic uncertainty and solution development strategies in software design. *Design Studies*, 31(6):567–589, 2010.

[108] Paul T. Grogan and Olivier L. de Weck. Collaboration and complexity: an experiment on the effect of multi-actor coupled design. *Research in Engineering Design*, 27(3):221–235, 2016.

[109] N Hirschi and D Frey. Cognition and complexity: an experiment on the effect of coupling in parameter design. *Research in Engineering Design*, 13(3):123–131, 2002.

[110] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.

[111] Philip J. Cash, Ben J. Hicks, and Steve J. Culley. A comparison of designer activity using core design situations in the laboratory and practice. *Design Studies*, 34(5):575–611, September 2013.

[112] Bo Yang Yu, Tomonori Honda, Mostafa Sharqawy, and Maria Yang. Human behavior and domain knowledge in parameter design of complex systems. *Design Studies*, 45:242–267, 2016.

[113] Lester Diamond and F Javier Lerch. Fading frames: Data presentation and framing effects. *Decision Sciences*, 23(5):1050–1071, 1992.

[114] Carol Tenopir and Donald W King. *Communication patterns of engineers*. John Wiley & Sons, 2004.

[115] Thomas J Allen. Communication networks in r & d laboratories. *R&D Management*, 1(1):14–21, 1970.

[116] David Harris. Supporting human communication in network-based systems engineering. *Systems engineering*, 4(3):213–221, 2001.

[117] Edward F McDonough III. Investigation of factors contributing to the success of cross-functional teams. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, 17(3):221–235, 2000.

[118] Hamdi A Bashir, Khalid AlZebdeh, and Jamil Abdo. An eigenvalue based approach for assessing the decomposability of interdependent design project tasks. *Concurrent Engineering*, 17(1):35–42, 2009.

[119] Fay Sudweeks and Sheizaf Rafaeli. How do you get a hundred strangers to agree? *Computer Networking and Scholarly Communication in the Twenty-first Century, New York: State University of New York*, pages 115–136, 1996.

[120] Bob Abarbanel and William McNeely. Fly thru the boeing 777. In *ACM SIGGRAPH 96 Visual Proceedings: The art and interdisciplinary programs of SIGGRAPH'96*, page 124. ACM, 1996.

[121] Jacob Cohen. Statistical power analysis. *Current directions in psychological science*, 1(3):98–101, 1992.

[122] Samuel B Green. How many subjects does it take to do a regression analysis. *Multivariate behavioral research*, 26(3):499–510, 1991.

[123] Helena Chmura Kraemer and Christine Blasey. *How many subjects?: Statistical power analysis in research.* Sage Publications, 2015.

[124] William R Binder and Christiaan JJ Paredis. Optimization under uncertainty versus algebraic heuristics: A research method for comparing computational design methods. volume 2B: 43rd Design Automation Conference, Cleveland, Ohio, page V02BT03A057, 08 2017.

[125] Timothé M. Sissoko, Marija Jankovic, Christiaan J. J. Paredis, and Eric Landel. An empirical study of a decision-making process supported by simulation in the automotive industry. volume Volume 7: 30th International Conference on Design Theory and Methodology, Quebec City, Canada, 08 2018. V007T06A018.

[126] Edward Tiong, Olivia Seow, Bradley Camburn, Kenneth Teo, Arlindo Silva, Kristin L. Wood, Daniel D. Jensen, and Maria C. Yang. The economies and dimensionality of design prototyping: Value, time, cost and fidelity. *Journal of Mechanical Design*, 143(3):031105, 2018.

[127] Ashish M Chaudhari and Jitesh H Panchal. An experimental study of human decisions in sequential information acquisition in design: Impact of cost and task complexity. volume Research into Design for a Connected World, Bangalore, India, pages 321–332. January.

[128] Andrew Gerard Carrillo and LJ Leifer. *Engineering design team performance: Quantitative evidence that membership diversity effects are time dependent.* PhD thesis, Stanford University, 2002.

[129] Ashish M Chaudhari, Ilias Bilionis, and Jitesh H Panchal. Similarity in engineering design: A knowledge-based approach. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 59278, page V007T06A045. American Society of Mechanical Engineers, 2019.

[130] Atharva Hans, Ashish M Chaudhari, Ilias Bilionis, and Jitesh H Panchal. Quantifying individuals'theory-based knowledge using probabilistic causal graphs: A bayesian hierarchical approach.