Designing Representative Model Worlds to Study Socio-Technical Phenomena: A Case Study of Communication Patterns in Engineering Systems Design

Ashish M. Chaudhari
Graduate Research Assistant
School of Mechanical Engineering
Purdue University
West Lafayette, Indiana 47907

Erica L. Gralla
Associate Professor
Engineering Management & Systems Engineering
The George Washington University
Washington, District of Columbia 20052

Zoe Szajnfarber
Professor and Department Chair
Engineering Management & Systems Engineering
The George Washington University
Washington, District of Columbia 20052

Paul T. Grogan
Assistant Professor
School of Systems & Enterprises
Stevens Institute of Technology
Hoboken, New Jersey 07030

Jitesh H. Panchal*
Associate Professor
School of Mechanical Engineering
Purdue University
West Lafayette, Indiana 47907

September 13, 2020

*Corresponding author. Email: panchal@purdue.edu

Abstract

The engineering of complex systems, such as aircraft and spacecraft, involves large number of individuals within multiple organizations spanning multiple years. Since it is challenging to perform empirical studies directly on real organizations at scale, some researchers in systems engineering and design have begun relying on abstracted model worlds that aim to be representative of the reference socio-technical system, but only preserve some aspects of it. However, there is a lack of corresponding knowledge on how to design representative model worlds for socio-technical research. Our objective is to create such knowledge through a reflective case study of the development of a model world. This "inner" study examines how two factors influence interdisciplinary communication during a concurrent design process. The reference real world system is a mission design laboratory (MDL) at NASA, and the model world is a simplified engine design problem in an undergraduate classroom environment. Our analysis focuses on the thought process followed, the key model world design decisions made, and a critical assessment of the extent to which communication phenomena in the model world (engine experiment) are representative of the real world (NASA's MDL). We find that the engine experiment preserves some but not all of the communication patterns of interest, and we present case-specific lessons learned for achieving and increasing representativeness in this type of study. More generally, we find that representativeness depends not on matching subjects, tasks, and context separately, but rather on the behavior that emerges from the interplay of these three dimensions.

Keywords: Socio-technical systems, representativeness, model worlds, communication patterns

1 Introduction

Design and development of large-scale complex systems, such as aircraft and spacecraft, takes multiple years and involves large numbers of individuals who may be spread across different organizations, each geographically and culturally distinct. Scientific understanding of such design and development projects requires an integrated view of not only the technical aspects of the system being designed but also the social phenomena within and across organizations and the interdependencies between social and technical aspects. However, studying these phenomena in real organizations is limited by both empirical and practical challenges such as difficulties in gaining access to observe or collect data, presence of large numbers of geographically distributed individuals, and projects spanning multiple years or decades.

Therefore, researchers in systems engineering and design have started to rely on abstracted *model worlds* that aim to be representative of the reference socio-technical system being studied, but only preserve some aspects of it [1]. Such model worlds enable more tractable observation and data collection as well as, potentially, experimental manipulation. Model worlds may involve theoretical models [2] or simulated environments in laboratory settings [3, 4, 5] or field settings [6, 7, 8]. The realism, fidelity, and level of abstraction differ significantly across these model worlds, but they all aim to preserve particular aspects of the reference systems they represent – the aspects relevant to answering the research question of interest. In other words, they aim to be *representative* of the reference system with respect to the research question.

While there is increasing acceptance of the importance of carefully selecting or designing representative model worlds, so far, there is a dearth of specific guidance on rigorous implementation of the concept. Creating a model world involves a series of research design choices about which features of the real world to include, and which to abstract. For example, a popular airplane manufacturing simulation uses Lego blocks instead of airplane subsystems to drastically simplify the process while maintaining a tactile interaction that is known to ground behavior [9]. However, while each of these choices can have a strong impact on one's ability to generalize outcomes, there is no wide agreement on standards for assessing whether an abstracted model world is sufficiently representative of the reference real-world setting to answer a given research question.

To that end, our motivating question for this paper is: *How should we design representative* *model worlds for socio-technical research in systems engineering and design?* The specific objective of this paper is to begin to answer this question by illustrating relevant issues through a “case study” of model world design: we design a representative model world for a research purpose and compare it to a real-world reference system, in order to (1) elaborate on and illustrate the large number of research design choices that go into generating a representative model and (2) synthesize lessons learned about the consequences of these choices. In our case study, the model world was designed to answer a separate, “inner” research question: to study how two factors influence interdisciplinary communication in the design process. This “inner” research question is described further in Section 3.1, since it is only a means to the end of exploring how to design representative model worlds. We see this paper as a step on the path toward more formal approaches to adopting and/or evaluating representativeness, such as dimensionless scaling relationships that relate real-world problems to representative model worlds.

This paper is organized as follows. Section 2 discusses the concepts of model worlds and representativeness, then reviews related work in the engineering design literature. Next, Section 3 describes the case study in which we design a model world to answer a research question about a real-world reference system. We introduce the research question, describe the real-world reference setting and the features we wish to preserve in the model world, then discuss the thought process in designing the model world experimental setting. Section 4 utilizes data collected in *both* the model world and real-world settings to compare the phenomena we attempted to preserve, discusses the generalizability of model world results to the real-world setting, and collects insights and “lessons learned” for designing this and future model worlds. Finally, Section 5 concludes the paper.

2 Literature Review

This section first introduces the concept of a model world in engineering research. Next, a discussion of model world representativeness compares similar concepts such as generalizability and validity in other fields. Finally, a review of related work in the engineering design literature focuses on design task similarity and subject expertise to assess model world representativeness.

2.1 Model Worlds in Engineering Research

Engineering research traditionally relies on physical models as a simplified setting with reduced spatial or temporal scales where engineers and researchers perform experiments. For example, architectural scale models visualize aesthetics of new building concepts, model engines help understand thermodynamic cycles in a classroom, and wind tunnels analyze aerodynamic lift and drag at reduced scales based on equivalence in Reynold's number. Advances in information technology enable computational model worlds where mathematical expressions of material properties, system state, and dynamics permit numerical simulation and analysis of system attributes and behaviors. Unbounded by the physical domain, modeling and simulation applications seek to represent real-world behavior as computable expressions.

Research in engineering design and systems engineering goes beyond engineering artifacts to also consider the social and organizational elements influencing their design. In these settings, more details about the human participants, their responsibilities, and the surrounding environment are critical elements. Broadening the concept of a model to include these contextual factors leads to our characterization of a "model world." The term groups all research settings that aim to proxy a real world setting, including both traditional models and lab experiments.

Drawing from an earlier work [1] and from community discussions, we define a model world as an *abstraction* of the real world that *represents* ***subject(s)*** *performing* ***task(s)*** *in* ***context*** for the *purpose* of advancing research objectives. This definition emphasizes several common elements across all model worlds. First, a model world must have a defined purpose to advance a specific research objective. Second, a model world abstracts the real world in some way to improve access to information, for example, through reduced cost, size, time, or risk. Finally, a model world arises from the interplay among three elements:

1. the subject is the individual, organization, or artifact under study,
2. the task is the research-relevant function or behavior performed by the subject, and
3. the context consists of the environment in which subjects interact with tasks; it may include access to information, incentive structure, environmental conditions, etc.

While these elements are often considered separately, we demonstrate in this paper how the interplay among them is more critical in generating useful (representative) model worlds. This interdependency can describe any number of physical relationships (between product components), social interactions, or causal relationships in a model. Later, we use communication patterns to compare two settings in systems design.

While this definition is most closely tied to design research where the subjects are human designers, the task is design work, and the context is the design environment, it can also be applied to other topics such as aerodynamics (subject: airfoil, task: provide lift, context: fluid flow) and optimization (subject: algorithm, task: minimize mass, context: input parameters).

2.2 Model World Representativeness

When choosing to adopt or design a model world, a researcher must assess whether it is representative for their purposes. Representativeness is not a feature of the model world, but rather a relative notion that only has meaning with respect to a specified research objective. Scientific communities discuss issues of representativeness in different ways based on the defining characteristics of their model worlds. For example:

- *Statistical representativeness* describes whether inferences made about characteristics of a sample can be transferred to the population from which it was drawn [10].
- *Analytical generalizability* describes whether theory developed by studying an empirical instance can be applied to other similar instances [11].
- *Ecological validity* describes whether a subject's response to stimuli in an artificial setting such as an experimental laboratory corresponds to their behavior in the associated real world setting (i.e., their ecology) [12].

These examples all consider the relationship between observations in a limiting environment (e.g., statistical sample, case study, or experiment) and some other, target environment (e.g., real world).

Enabled by the emergence of computational models, the field of modeling and simulation formalizes representativeness as model validity, defined as "substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" [13]. Establishing relationships between different system representations (e.g. between a source system and a model world) relies on a morphism mathematical relationship [14].

The simplest form of replicative validity assesses the input/output correspondence between source system and generative model. Stronger forms of predictive and structural validity add additional requirements of state and component state correspondence to predict and explain previously-unseen system behavior. Model worlds with human actors (described as games or gaming simulations) demand a different type of validation compared to computational model worlds. Raser distills game validity into four criteria: psychological realism for players, structural isomorphism, process isomorphism, and outcome correspondence, with weightings dependent on each application [15].

Whether the model world is adopted or designed, it is defined by the interplay between subject and task in a context where at least one aspect is abstracted from the real world. Evaluation of model world representativeness must consider how selection of the subject(s), task(s), and context influence outcomes with respect to the research objective. Some research objectives may permit substantial abstraction (e.g. using computational model worlds) while others may not.

We define model world *representativeness* as a measure of *whether, and the extent to which, observations made in a model world* <u>*can*</u> *be used to answer a specific research question about the real world.* This definition encompasses many of the other perspectives on representativeness but focuses only on the model world itself and not on other elements of research design such as sampling, replication, and randomized controls. Additionally, this definition does not seek absolute correspondence between the model world and the real world—selected abstractions specific to the research question may allow comparisons only at a theoretical level.

2.3 Representativeness in Engineering Design Literature

Assessing the representativeness of research settings is fundamental to produce generalizable insights and a necessary precursor to build on each other's research. Within the engineering design literature, a number of efforts evaluate how well the research studies real design scenarios. Specifically, the studies can be classified into two types: those that focus on the representativeness and similarity of design problems/tasks, and those that evaluate the differences between experts and novices. We discuss the design literature in these two areas in the following subsections.

2.3.1 Studies Focused on Design Problem/Task

It is well accepted in the design community that the nature of the problem plays a significant role in various design phenomena. One way to characterize the space of design problems is to identify dimensions of similarity. Anandan and coauthors [16] argue that the similarity of design problems depends on their (i) product design specifications, (ii) constraints (cost, time, weight, safety), (iii) function structures, (iv) concepts, (v) CAD models, (vi) shapes, and (vii) manufacturing process plans. Durand and co-authors [17] identify the following *structural features* of design problems that should be considered while designing problems for design research: problem size, functional coupling, participants' familiarity with the design problem/solutions and underlying principles, nature of solution space: size and constraints, effort required to solve the problem, domain of design problem and degree to which analogous solutions can be retrieved.

Kumar and coauthors [18] compare 55 design problems from the literature based on their representation using protocol analysis and latent semantic analysis. They characterize the problems based on five structural elements: goals of the problem, functional requirements, non-functional requirements, information about end user, and reference to an existing product. Sosa [19] characterize design tasks for experimental creativity research (idea generation and conceptual design) using the metrics of semantic score, lexical ambiguity, precedent analysis, and readability. Some studies have used experiments to compare different design problems. Levy and coauthors [20] experimentally compare four design problems (peanut shelling, corn husking, coconut harvesting, and personal alarm clock) for equivalence in design ideation tasks. Problem comparison uses the ideation metrics of quantity, quality, novelty, variety, and completeness.

2.3.2 Studies Focused on Expertise

One of the key aspects of representativeness of model worlds relates to ensuring that the behaviors of the individuals (engineers) in the real world match the behaviors of the subjects (typically, students) in the model world. It is commonly argued that students behave differently than experts. There is extensive literature on expertise in general, and design expertise in particular. Cross [21] presents a detailed review of the literature on the differences between experts and novices. Experts and novices differ in their problem solving strategies. Novices adopt a 'depth-first' approach to problem solving, whereas experts adopt a top-down and breadth-first approach. Experts are able to mentally to stand back from the specifics of the problem and draw upon abstract knowledge in their domain of expertise. Experts can recognise underlying principles, rather than focusing on the surface features of problems. Based on study of two expert designers, Cross and Cross [22] conclude that design experts take a systemic view of the design situation, choose to frame their view of the problem in a challenging way, and draw upon first principles to guide both their overall concept and detailed design.

Through a comparison of expert and novice designers, Ahmed and coauthors [23] conclude that novice designers tend to reason backwards and to use a deductive approach, whereas experienced designers tend to reason forward and, when solving more complex problems, alternate between forward and backward reasoning. Novice designers rely on trial and error while experienced designers use particular design strategies. The differences between experts and novices are also studied in the literature on design education that compares students with experts. Crismond and coauthors [24] highlight that students differ from informed designers (experts) in how they: understand the design task, build knowledge around the problem, generate ideas, represent ideas, weigh options and make decisions, conduct experiments, troubleshoot problems, revise/iterate, and reflect on the design process. Atman et al. [25] concluded that experts spend more time than novices (students) in product realization activities, which encompasses decision making and communication.

Cash et al. [26] compare the behaviors of subjects in laboratory experiments with design practitioners for three core design activities: information seeking, ideation, and design review. They find that many of the observed behaviors are common across design contexts (experiments vs. practice) and populations. They found differences in the frequency and time spent on debating and communication. As in the study by Atman et al. [25], the practitioners had a greater frequency of communication, and spent more time per instance than in the laboratory settings. They attributed the differences to the embedded nature of practice within a pre-existing design process and the associated importance of communication.

While much of the existing literature is focused on independent analysis of the design problem or the expertise, few studies analyze the differences between model worlds and the real design settings that they are intended to represent. There is a lack of studies that systematically analyze the process of designing model worlds for specific phenomena under investigation. There is a need to consolidate insights from these previous studies and to investigate additional (potential) dimensions of representativeness, and the interplay among these dimensions, through empirical comparisons between model worlds and real worlds. This paper takes a step on that path.

3 Model World Design Case Study

To explore the various possible dimensions of representativeness, this paper uses a case study in which a model world was designed to represent a real world setting to advance a study objective. Data were collected in *both* settings to explore model world representativeness. Section 3.1 describes the "inner" research question that motivates the design of the model world. Section 3.2 discusses the reference system - a NASA concurrent design facility. A model world created to represent the reference system is presented in Section 3.3, along with the thought process followed in creating the model world.

3.1 The "Inner" Research Question

Recall from Section 1 that the objective of this paper is to learn from a "case study" in which a model world was designed and compared to a real-world reference system, in order to (1) elaborate on the research design choices involved in generating a representative model world and (2) to synthesize lessons learned. Within this case study, the model world was designed to answer a separate, "inner" research question: How do two factors influence interdisciplinary 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. Thus, this paper is a "study-within-a-study": a study of communication in design is used to explore how to design representative model worlds.

The motivation for this "inner" research question is to support design organizations (such as NASA, in this case) to determine the appropriate information technology infrastructure and communication pathways to use in their design processes, e.g., whether to develop catalogs or continuous search tools for designers, and how frequently designers should receive design variable updates. Communication is an important aspect of systems engineering and design [27]. Good communication in an organization is essential to address technical dependencies, ensure good design outcomes and promote efficient design and systems engineering processes. Poor communication is a major cause of failure in design processes. Communication is particularly important in large scale complex systems design such as automobiles, aircraft and spacecraft.

We would ideally perform studies within the real organization. However, there are many barriers to conducting such studies, including the size of the organization, the long duration of design projects, and the costs involved. Collecting data during real projects is also restricted. Since the real world setting often does not allow us to control the environment and collect fine-grained data, we need to rely on other settings such as model worlds. Establishing representativeness of the model world with respect to certain communication aspects would help advance research because the model world affords more granular data collection over time including repeated studies in quasi-experimental research designs [28, 29].

3.2 Real World Reference System: NASA's Concurrent Design Facility

The real world reference setting is a conceptual spacecraft design process at the NASA Goddard Space Flight Center's concurrent design facility, the Mission Design Lab (MDL). The MDL develops high-level conceptual designs for spacecraft, relying on a team of disciplinary experts co-located in a single room who focus on this task for one entire week. Figure 1, below, shows the layout of the room and where each disciplinary expert sits. Using the 'classic' categories of subject, task, and context, the setting is described in detail in the following paragraphs. These descriptions focus on those aspects of the subjects, task, and context whose interplay may influence communication during the design process, such as the subjects' expertise, the nature of the task, the formal and informal organizational processes and structure, the available communication channels, the incentives, and the tools that facilitate communication and design.

3.2.1 Subjects

The MDL is staffed by NASA engineers who are experts in their disciplines. They occasionally participate in MDL studies in addition to their regular jobs at Goddard. MDL engineers also have experience in the MDL concurrent engineering process. Typically, when an engineer starts

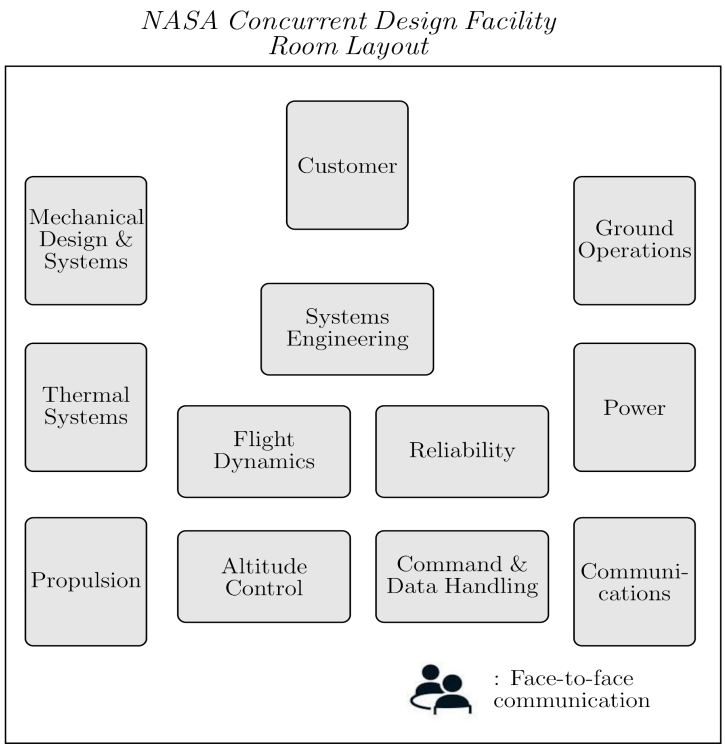


Figure 1: The seating chart of disciplines in NASA's mission design lab

working with the MDL, there is an 'apprenticeship' period in which they train with an experienced MDL team member from their discipline. As a result, they understand how the MDL team works together, what is expected to happen when, and how to use the tools provided in MDL.

3.2.2 Task

The task is to produce a conceptual design for a spacecraft, captured in a report consisting of detailed slides, an oral presentation, and other documentation. A conceptual design consists of a feasible "point design" for a spacecraft with most of the key engineering parameters determined and costed (Concept Maturity Level 4, according to [30]). The objective, not explicitly stated mathematically, typically consists of finding a feasible design that meets the engineering and science requirements and a cost cap; implicitly, mass is minimized as a proxy for cost.

The design task is accomplished by a team of 12–20 experts in typical aerospace disciplines such as mechanical systems, thermal systems, power, propulsion, avionics, attitude control, command and data handling, and systems engineering. A disciplinary expert is responsible for the design of a corresponding subsystem on the spacecraft (and some disciplinary experts have integrative roles).

The design of each subsystem is dependent on that of others through shared design variables. Some subsystems are tightly coupled and others are loosely or not at all coupled. The MDL completes the task in four work days of approximately eight hours per day. Some decisions must be made relatively quickly to complete a feasible design in available time. Some engineers experience time pressure to solve problems rapidly, but there is typically enough time to finish a design.

3.2.3 Context

There is limited organizational structure in the MDL. A team leader runs the full-team meetings and sets expectations. Two systems engineers help to ensure needed side meetings happen, important problems and trades are surfaced and solved, and the design closes on time. There is, however, a clear process. Days begin with a briefing on progress and key problems. A similar briefing is held just after lunch. There are clear expectations about what kind of work will happen when during the study and how communication will occur.

Communication is typically face-to-face. All disciplinary experts may easily communicate in person, since they are in the same room and typically seated near others with whom they usually need to speak (see Figure 1). Because they are co-located, the barrier to communication is low. In addition to face-to-face communication, a spreadsheet maintains key design parameters, but it is not typically used to communicate design parameters. Experts may also phone or email others in their disciplines outside the room for assistance. Engineers have shared expectations of many goals and constraints that are not explicitly described, such as acceptable amounts of risk in the design, what constitutes a reasonable mass and how it translates into cost.

The incentives for each individual are varied, and have not been specifically measured. Success in design closure is indirectly relevant to career advancement. Engineers are motivated by the work itself and by learning from each other. They are also, of course, paid a fixed salary for their jobs, of which the MDL is a part. Engineers use a combination of catalogs and simulations to inform design decisions. Some disciplines have specific tools that search and optimize over the solution space, but many also pick from a set of existing and proven designs.

3.2.4 Data Collection

The collected data includes time stamped observational records of interdisciplinary communication over the course of the week-long study: who communicated with whom and for how long. The NASA engineers sometimes talk in groups of three or more, but, for comparison with the model world data, we consider that a group discussion consists of one-on-one interactions among those involved. Such data are collected from seven different studies, all of which took place in the NASA MDL.

Part of the motivation for designing a model world to represent this real-world context is that there are many kinds of data that cannot be collected in this setting. Because MDL studies are proprietary, there are restrictions on sharing the final designs. It is not feasible to capture all the potential design solutions explored by designers (such as what they thought about, did back-of-envelope calculations for, or tested in models) nor the exact content of all of their communications. Finally, it is not possible to 'experiment' with alternative tools and processes. Model worlds, in which such data can be collected and experiments run, can potentially support deeper analyses of certain types of questions, if they can be made representative of the real-world setting.

3.2.5 Real World Features to Preserve

It is important to reflect on which aspects of the real world setting need to be preserved in a representative model world to answer the research question on interdisciplinary communication in design. There are many factors that affect communication in design teams, including technical dependencies among subsystems (which drives what design information needs to be passed among team members), the formal organizational structure (which creates the explicit channels and processes for communicating among individual roles), and the informal cultural norms (which drive expected behavior, such as preferences among which sources of information to trust or what constitutes a good enough solution) [31, 32]. In reflecting on the way these concepts are embodied at the MDL, the following key areas emerged as crucial to preserve in the model world because they are expected to drive communication.

1. Technical interdependencies in the design problem: Greater technical dependence among subsystems requires greater coordination between the teams working on those subsystems [33]. Therefore, strongly coupled systems are more likely to have greater communication links, compared to weakly coupled subsystems. In the MDL, the patterns of communication appear to be strongly driven by the technical dependencies in the physical system (i.e., the spacecraft), and we expect limited communication among disciplines that do not share any technical parameters. Moreover, some interdependencies are particularly challenging to resolve; these drive a significant portion of the communication. Therefore, the first important feature of the model world is *a design task for a technical system with complex interdependencies, where some are more important than others.*

2. Organization structure, process, and knowledge and expertise: Generally, the organization structure and processes, both formal and informal, drive communication across disciplines. In the MDL, however, since all subject matter experts (SMEs) are collocated in the same physical space and explicitly encouraged to work collaboratively and quickly, there is limited formal organizational structure. Instead, the process is driven by (1) clearly defined disciplinary roles and (2) an informal but well understood expectation of the design process. Regarding (1), for example, there is a defined facilitator role (similar to a project manager) who ensures that the values of key design parameters are communicated to the other disciplines who need them by walking around the room and stimulating pairwise exchanges. Regarding (2), there is a shared expectation for how the design process *should* go, based on thousands of past MDL studies. While each new spacecraft has unique features, experts report that there is a “rhythm to the room” that is common across studies. By this they mean that most of the design decisions are consistent, even though the specific values are study-specific. For example, the communication engineer always needs to make a link budget which depends on data needs of the science instrument and opportunities for downlink coverage with ground stations. Both of these are driven by orbit choices and by mission operations timelines. All of this necessitates design trades and associated discussion, but those discussions usually happen around the same time in the study and among the same people. While some studies involve unique features (e.g., multiple components flying in formation), these new discussions represent a small fraction of the overall interdisciplinary communication in the study.

In short, rather than formal organizational structures and processes, communication in the MDL is driven more by the knowledge and expertise of its members. Teams of experts coordinate their behavior through shared mental models [34, 35]. As a team, they are aware of each others’ role on the team, knowledge, skills, and goals. Experts have a better understanding about which other subsystems they should communicate with to improve performance or to resolve conflicts, as compared to novices. In the MDL in particular, since all the MDL participants are SMEs, we would not necessarily expect to see equal communication among all disciplines that share design parameters. The experts are expected to know which technical dependencies are more important, and hence, focus their communication on the corresponding dependencies. Part of why the MDL is efficient is that there is a strong expectation for which design trades are most salient and will drive future decisions; therefore, we expect the SMEs to spend most of their communicating time on the subset of technical dependencies that are most important to the final system. To embody these characteristics, *the subjects in the model world should have insight into their discipline-specific design spaces, and should know which trades with other disciplines are likely to be most important.*

In addition, the *subjects must be incentivized toward a collaborative goal, so that they are motivated to identify and analyze these tradeoffs with one another.*

3. Nature of the Design Process: A third aspect of the MDL that should be preserved is the design search behavior of the SMEs. Given the short timeline, the MDL culture tends to focus on 'good enough' solutions. Their primary goal is to assess whether the design constraints are feasible, not to come up with a final design for the spacecraft. As a result, most of the disciplines work from catalogues from trusted manufacturers, rather than doing de novo design at the component level. For example, if the propulsion engineer is asked by a trajectory analyst whether a particular burn profile is feasible, (s)he will not start designing from first principles; instead, (s)he might answer with the operating limits from a known manufacturer's design.

Although the MDL has a long history of efficiently delivering spacecraft designs in a week, its success has led to requests for more varied kinds of spacecraft, and several have questioned the scope of applicability of the process. It is efficient precisely because it does not search the whole design space, but how much efficiency is good? What if valuable designs are missed because the experts in the room are not talking as they should. Part of the motivation for a representative model world is to be able to probe these what-ifs, outside of the professional setting. The model world *should enable us to explore the "counterfactual" of continuous/optimizing rather than discrete/good-enough search of the design space, which we cannot observe nor experiment with* directly in the professional setting.

In summary, the model world for studying communication in concurrent design settings should consist of a reasonably complex task with dependencies among disciplines, subjects with relevant knowledge of respective disciplines and shared mental models, a collaborative context where subjects can interact and exchange information, and incentives to collaborate and make design tradeoffs. Finally, it should enable exploration of "counterfactual" scenarios that cannot be observed nor experimented with directly in the professional setting, such as continuous design search rather than choosing among off-the-shelf solutions.

3.3 Designing the Model World: An Abstract Engine Design Setting

3.3.1 Key Decisions in Designing the Model World

In designing this model world, our first decision was selecting the subject pool. This decision was driven in part by the opportunity to recruit students from a homogeneous subject pool from a large-enrollment design course. One of the authors was teaching a junior-level machine design course with an enrollment of over 200 students across multiple sections. Recruiting the students from the course allowed a reasonable sample size across different treatments.

Once the subject pool was fixed, the next step was to design the task. The factors that were considered included the subjects' knowledge of the domain, the need for multiple subsystems/components and technical dependence among them, and a reasonable completion time. We also aimed to replicate key features of the real world reference system in the design problem, including multiple objectives, time pressure, discipline interdependence, and the need for collaboration.

Considering desired factors and the context of the machine design course, we designed an abstract engine design problem with five components - connecting rod, crankshaft, piston, piston pin, and flywheel. The students were familiar with these components through the course and the scientific knowledge (e.g., stress analysis, fatigue analysis, etc.) required to design these components. The task was framed as a parametric design problem of the five engine components with the objective to maximize the overall factor of safety and to minimize the overall mass of the engine. At a higher level, they had a shared understanding of the role of the components in the engine, and some understanding of how their design decisions affect, and are affected by, other components.

Real engine design is a complex design activity. However, we decided to keep the engine design problem significantly simpler than the spacecraft design problem in MDL, with a smaller number of subsystems and only a few variables in each subsystem. This decision was mainly driven by the need to complete the activity in a shorter time frame of around 30 minutes instead of multiple days. (The short timeframe fit the requirements of the design course in which the study was run, and also offered us the opportunity to evaluate whether a much-shorter task could still be representative of the real world.) The component-level simulations were also simpler than the sub-system design activities in the MDL study, which allowed us to conduct multiple replications of the experiment. Despite the simplicity, the engine design preserved the characteristics of dependencies between subsystems which required the subjects to communicate with each other to complete the task.

In contrast to domain experts at MDL, however, the students do not have experience with the specific design problem. Therefore, the shared mental model of the design team is expected to be different from the shared mental model of the experts at MDL, resulting in differences in communication patterns between MDL and the model world. Studies on team effectiveness show a strong reciprocal linkage between team cognition and behaviors such as communication. According to Kozlowski and Ilgen [36], "repeated interactions among individuals that constitute processes tend to regularize, such that shared structures and emergent states crystallize and then serve to guide subsequent process interactions." Hence, the lack of shared mental model is expected to diminish as the students start interacting with each other within the experiment. For example, at the start of the experiment, the students know which design variables from other components influence their design, but do not know the extent of influence and relative importance of the variables. This knowledge is gained over time as they perform the design task. Therefore, we expect students' behavior to better represent experts in the latter part of the experiment.

To represent discipline-specific knowledge, each student was seated at a different workstation and given different information as seen from the room layout in Figure 2. The extent of their knowledge about their own subsystem and that of the others depended on the information and tools provided, how they used those tools to explore the design space, and how they used the available communication channels to explore the design space collaboratively. (The details of these communication channels and search tools are provided later.) Finally, there was an incentive for the team to perform well on this task (also detailed below). This setup was intended to give the students discipline-specific expertise and to motivate them to communicate with teammates to

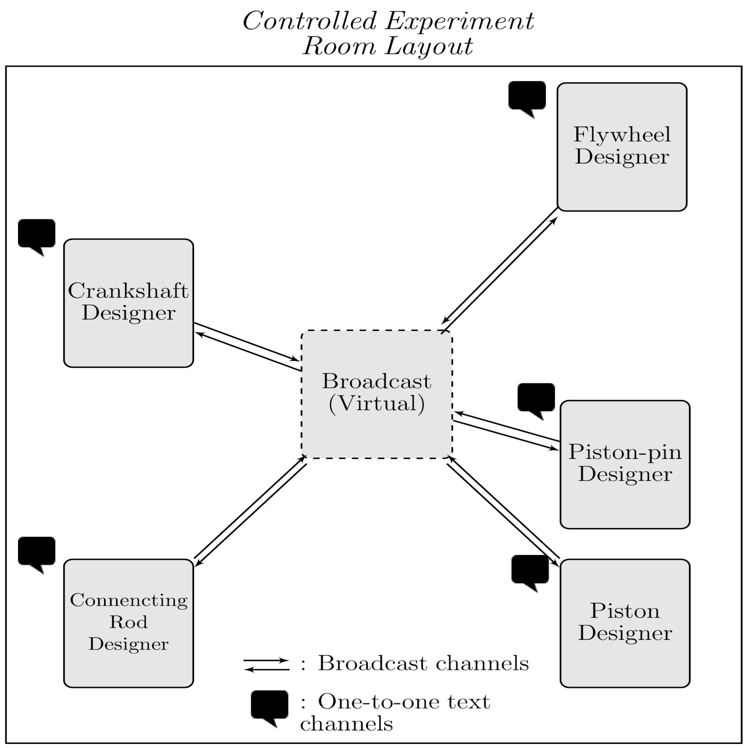


Figure 2: The disciplines in the engine experiment were randomly spread across the experiment room.

discover and make important trade-offs.

Not all aspects of the experiment were designed to represent behavior in the MDL. We randomly assigned individuals to components and teams, instead of having individuals self-select components and form their own teams. The rationale for this decision was to control for any individual-specific effects, any effects of prior collaborations, personality effects, and acquaintances on their collaboration. While communication in real world design organizations (including the MDL) is affected by these factors, our intent was neither to study these factors, nor was it feasible to identify and replicate such influences within the experiment.

Although MDL engineers can communicate in smaller and larger groups, we decided to limit the communications in the model world to pairwise (one-to-one) and a limited "broadcast" where subjects can push design information and other disciplines can pull the current information. This was partially because of the small team size in the model world (5 in engine design vs 18–20 in MDL) and partially because allowing multi-person chat conversations may lead teams to create one single team-wide chat, which would not be able to distinguish who was really talking to whom, as in the MDL data set. On a related note, although in the MDL the SMEs can communicate verbally and use various visual aids, the students were set up at different screens and could only communicate through the user interface (either one-on-one chat or the broadcast function).

The experiment did not include a formal coordinating role. The MDL has a facilitator and one or two systems engineers, whose roles involve identifying key trades among disciplines and ensuring that the respective SMEs communicate and resolve those trades. Given the small size of the engine design task and teams, we believed such a role would be unnecessary to enable coordination.

Finally, to address the research objective presented in Section 3.1, we designed a factorial experiment with two variables, i) the nature of the design solution space and ii) the availability of global information. We also designed the user interface to implement different experimental treatments. Since the experiment was intended to enable exploration of the "counterfactual" cases of continuous/optimizing rather than discrete catalog-based search of the design space, we allowed the first factor to vary along two levels: local search of design space using just simulation (label S) and local search of design space using just catalogs (label C). The experiment was also intended to explore the importance of a central data-sharing repository such as the parameter database. Hence, we allowed the second factor to vary along two levels: where the parameter database is absent (label P_0) and where it is present (label P_1)

Further details of the subjects, design task, and the organizational context are discussed in the following sections. These details are also summarized in Table 1.

3.3.2 Subjects

The subjects in the model world are undergraduate mechanical engineering students from a required junior-level machine design class. The student subjects have the knowledge of relevant fundamentals of mechanisms, machine design and failure analysis. They do not have experience in working on real engine design problems, but there is familiarity with component-level parametric design problems similar to the abstract engine design problem used in the model world.

3.3.3 Task

The engine design task is a parametric design problem which requires specification of the geometry parameters of an internal combustion (IC) engine to optimize the mass and factor of safety. At the

Table 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

discipline level (i.e., the component level), the objective is to minimize the subsystem component's mass and maximize its factor of safety. The system-level objectives are to maximize the overall factor-of-safety (which is the minimum of the subsystem-level factors of safety) and minimize the total mass of all engine components. The students remain seated during the entire duration to avoid disturbing other teams in the same room. Each team consists of five disciplines corresponding to the five engine components: piston, crankshaft, flywheel, connecting rod, and piston pin. There is no central discipline (e.g., system engineering) to manage global requirements and team interactions. Instead, a broadcast screen displays current values of the system objectives (total mass and factor-of-safety). In the P_1 conditions, it also displays the key design variables.

Different disciplines are coupled through the shared variables similar to the reference setting. Therefore, we define the design interdependence between a discipline pair as the number of variables shared. There are discipline pairs with strong coupling (e.g. flywheel-crankshaft) as well as discipline pairs with weak coupling (e.g. piston pin-connecting rod). Each discipline's subsystem outputs, such as subsystem mass and subsystem factor-of-safety, may depend on the design parameters of other subsystems. For example, the crankshaft discipline requires the piston bore diameter, which is controlled by the piston discipline, to determine the maximum gas piston force on the crankshaft pin and estimate its factor-of-safety. The shared variables between any discipline and "broadcast", the broadcast screen that displays system objectives and key design variables, include the system-level mass, system factor-of-safety, discipline-specific design parameters, and design parameters shared by the discipline with other disciplines. All pairwise design interdependencies are summarized in the design structure matrix (DSM) shown in Figure 3, right, and compared with that of the real world reference system, left. The "broadcast" element, a counterpart of "systems engineering" and "customer" from the NASA MDL, is supposed to be an integrative agent that facilitate information exchange [37] and, therefore, it is included in the engine DSM. While the two matrices are not identical, they both show variability in the strength of coupling across various disciplines pairs.

An exact match between the problems is not critical for representativeness; rather, what matters is whether the interplay among this design task, the students' expertise, and the design problem context generates representative patterns of communication. As in the real world reference system, subjects need to search the design space and work in a team to learn the technical dependencies in the engine design problem. Design collaboration is essential because the subjects need to work together to maximize the global objectives. Whether this achieves representativeness is evaluated later, in Section 4.

3.3.4 Context

With the task and subject pool determined, the context was designed to balance the constraints of a classroom setting with the aim of replicating important contextual features of the real world reference system. Unlike in the real world reference system, the organizational structure is entirely decentralized. There is no systems engineer – only a broadcast that displays current system objective values and, in condition P_1, design variables. The broadcast feature enabled the team to keep in mind the global goals and infer tradeoffs among disciplines, as a systems engineer might do.

The following are specifics of how we implemented *simulation modules* and *catalogs* in the experiment. When using simulations for evaluations of the mass and factor of safety, the subjects feed design parameters to the discipline-specific models. Simulations do not have an associated cost, which implies the subjects can run as many simulations as they like. Evaluations from past simulations are available on the screen for reference, as shown in Figure 4 (left). On the other hand, when using catalogs for design exploration, the pre-evaluated performance is known at discrete

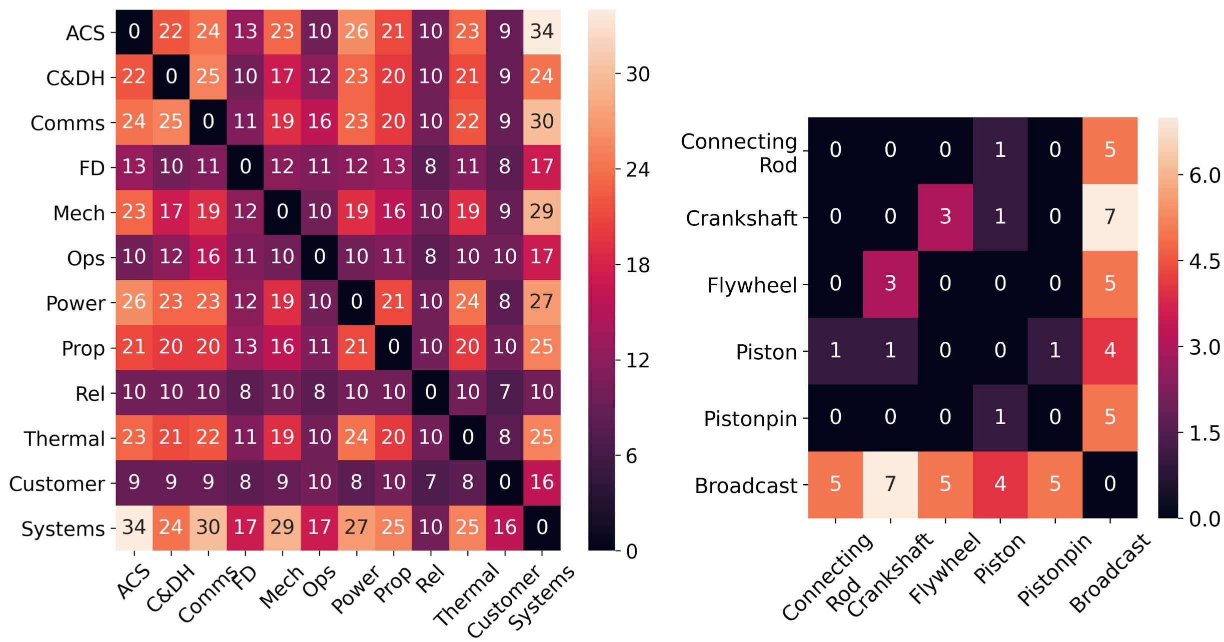
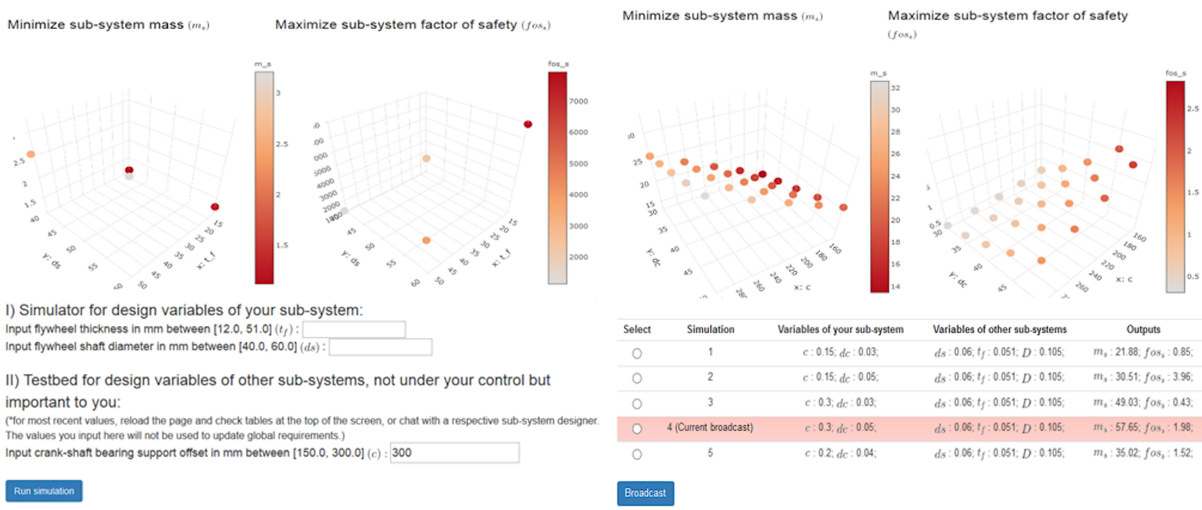


Figure 3: Design structure matrix for the spacecraft design problem (left) and the engine design problem (right). Pairwise design interdependencies, i.e., the number of shared design variables, are shown in the matrix cells.

points in the design space. The disciplines also have access to a testbed where they can view different catalogs by changing values of the design variables (shared variables as well as those under their own control). Figure 4 (right) provides a screenshot of the user interface available while using catalogs. Further, Figure 5 is a screenshot of broadcast showing design parameter values (only in the P_1 conditions) and global objectives information.

To motivate the students to achieve the system-level objectives, the engine design task included a monetary incentive proportional to the system performance, i.e., total mass and the overall factor-of-safety. Based on the preference over their values, the system outputs were categorized into five quality levels: *poor*, *fair*, *good*, *very good*, and *excellent*. In general, total engine assembly mass and factor-of-safety exhibit negative correlation. The system level quality was the minimum of the two quality levels. The poor quality level payed \$10 per discipline, whereas the excellent quality level payed \$20 per discipline. The payments per discipline for fair, good, and very good levels, respectively, were \$12, \$15, and \$17. We chose equal reward for all subjects on a team because group-level incentives are important for initiating effective collaboration [38], and, also, this ensures incentive-compatibility by aligning subject behaviors with the task objective.



(a) Simulation-based local design search (b) Catalog-based local design search

Figure 4: The user interface for two types of design exploration processes: a) simulation-based and b) catalog-based.

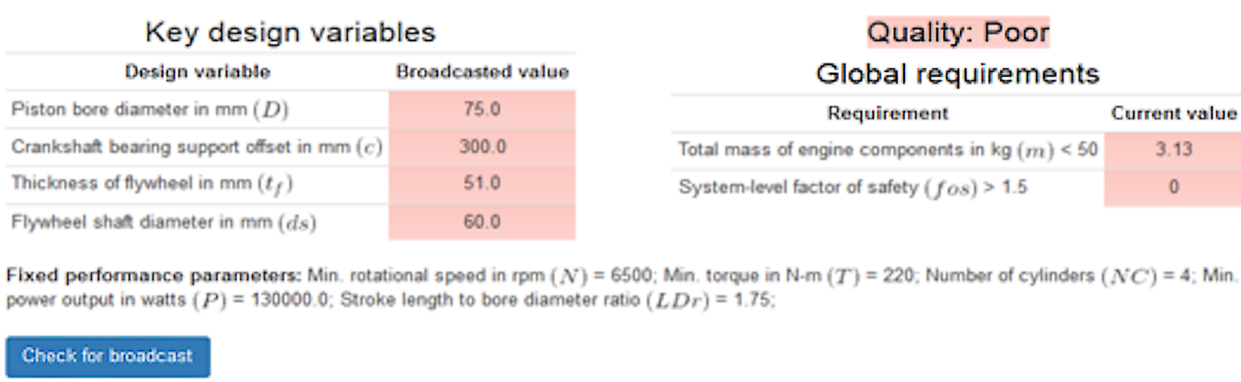


Figure 5: An example of the global information shown on broadcast, which displays key design variables for condition P_1 only.

3.3.5 Data Collection

In the model world, the collect data include timestamped logs of i) pairwise text messages, and ii) discipline-wise updates to and from the "broadcast". Such pairwise interactions are used to represent the amount of communication.

4 Evaluating Representativeness of the Engine Design Model World

In this section, we evaluate the representativeness of the engine design setting with respect to the NASA MDL by comparing observed team communication patterns. 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 research settings. Such parameters may be different between the settings, but note that the representativeness should be judged based on multidimensional subject-task-context interrelationships and not based on singular dimensions along subject, task, or context. 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,

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

   The indicator function $\mathbf{1}_{(\delta,\infty)}(s_{ij})$ 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. [39] and its implementation is available in Python library networkx [40].

Section 4.1 describes the specific differences and similarities along the aforementioned quantities.

4.1 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 2.

4.1.1 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 6 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.

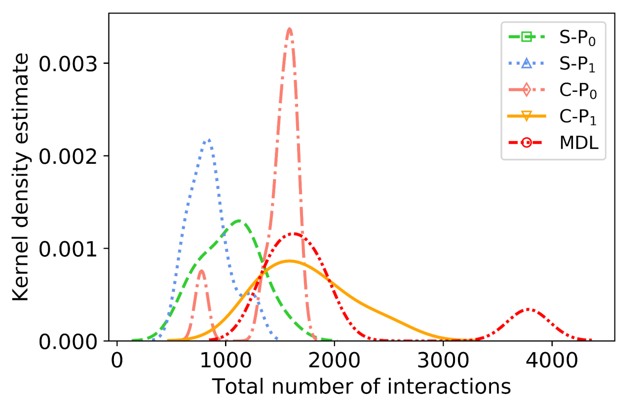


Figure 6: Estimated distributions of total pairwise interactions per team. Kernel density estimates are found using a statistical data visualization library, seaborn [41].

4.1.2 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 7 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 ($R^2 = 0.30, \mathrm{DoF} = 52, p\text{-value} < 0.0001$) and 4.14 ($R^2 = 0.33, \mathrm{DoF} = 46, p\text{-value} < 0.0001$) 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 4.2.

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.

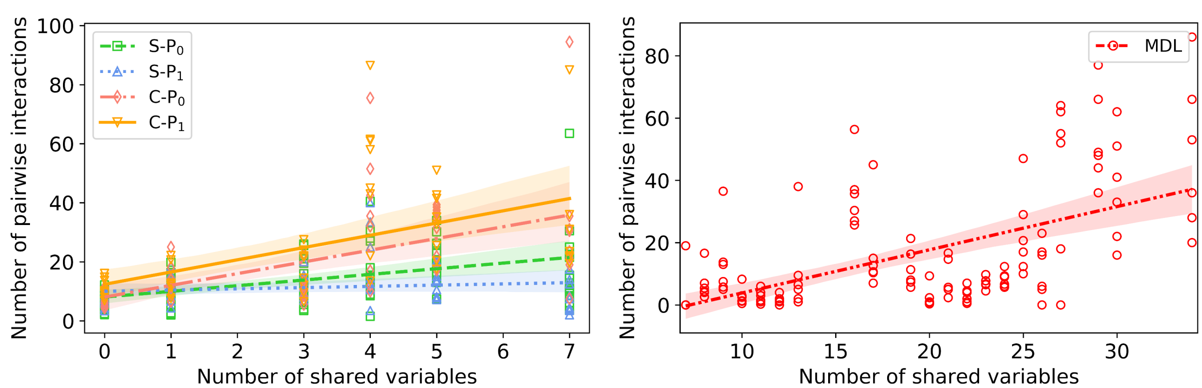


Figure 7: Mirroring between technical dependence and the amount of communication in the engine design experiment (left) and in the NASA concurrent design facility (right).

4.1.3 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 8 shows a moving average of this metric over time for the engine study (left) and the NASA reference setting (right). From Figure 8 (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.)

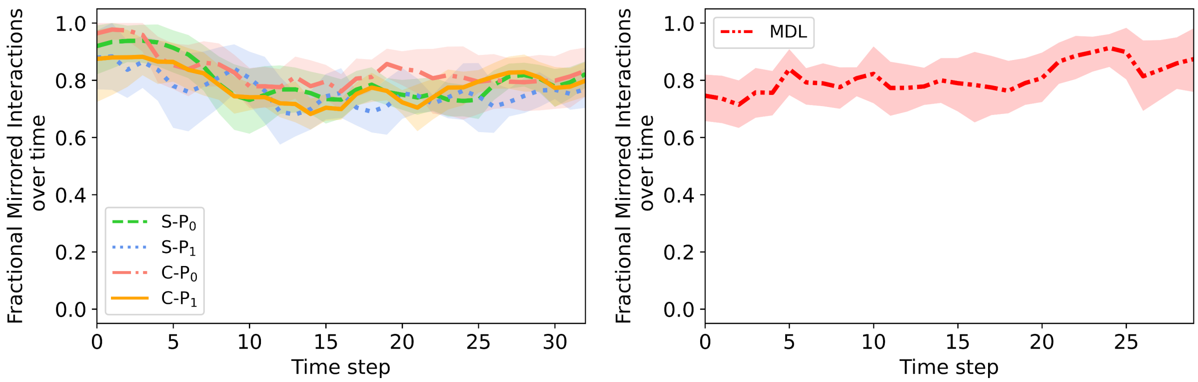


Figure 8: Fractional mirrored interactions over time with 5th and 95th percentiles calculated for all discipline pairs.

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 9 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 8), 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.

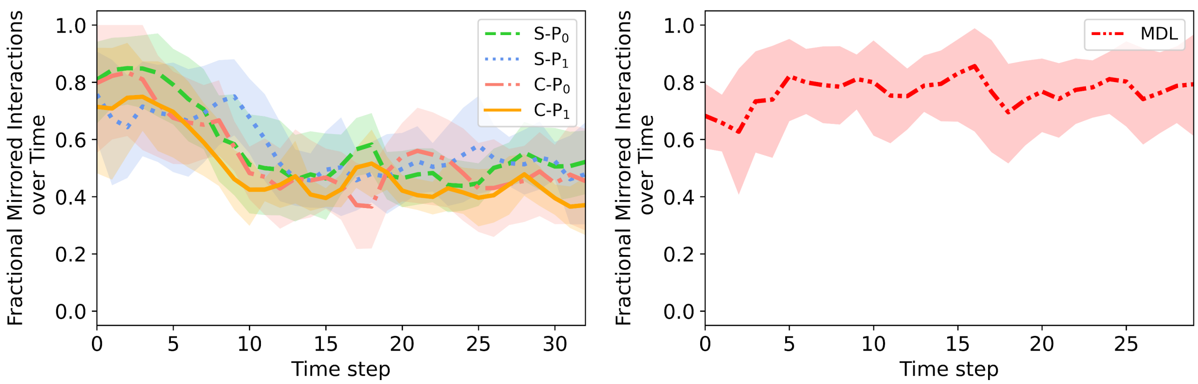


Figure 9: Fractional mirrored interactions over time with 5^{th} and 95^{th} percentiles calculated after excluding integrative disciplines such as systems engineer, customer and broadcast screen.

4.1.4 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 [39]. 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 10 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 11 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 10 and 11 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.

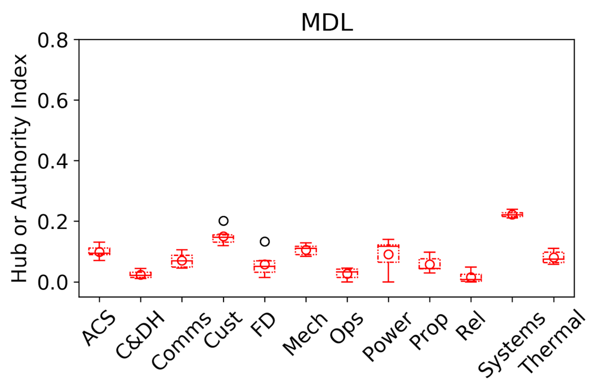


Figure 10: Network centrality for the NASA MDL disciplines. Hub and authority indices [39] are the same for NASA MDL teams because pairwise communication is undirected.

4.2 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 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 4.1 about which aspects of the engine experiment communication patterns matched those of the

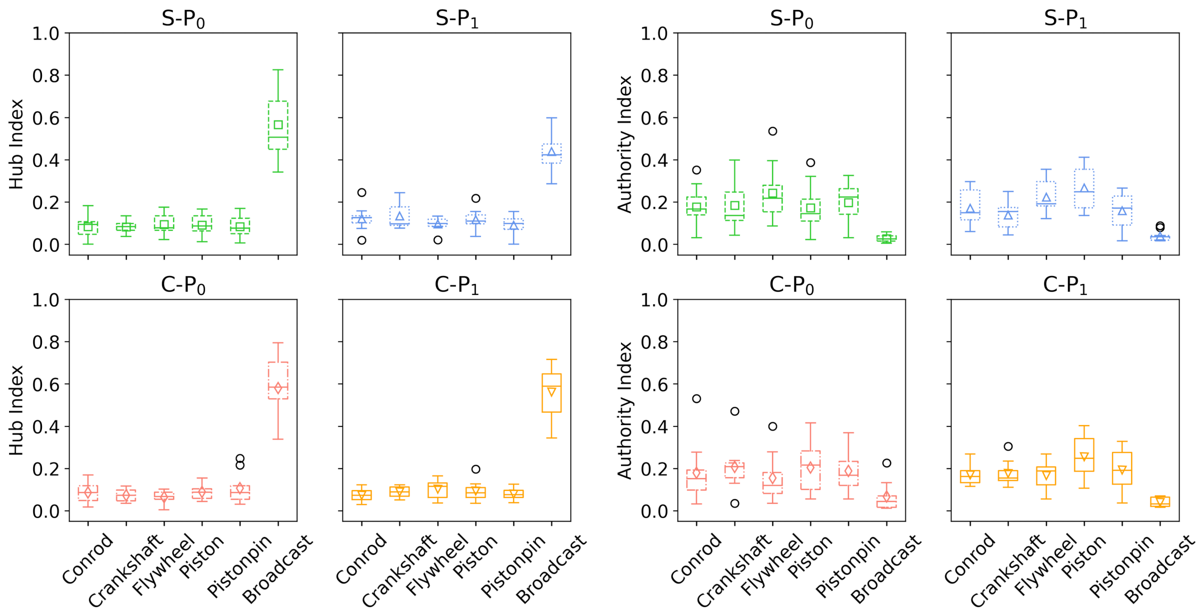


Figure 11: Hub indices and authority indices [39] representing the disciplines' network centrality based on outgoing and incoming links, respectively. These network indices are computed using a network analysis library, networkx [40].

NASA reference setting – i.e., the engine setting's representativeness.

4.2.1 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 6. 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 7). The implications of these results are presented in Section 4.2.3.

4.2.2 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 6). On the other hand, the mirroring results in Figure 7 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

Table 2: 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).

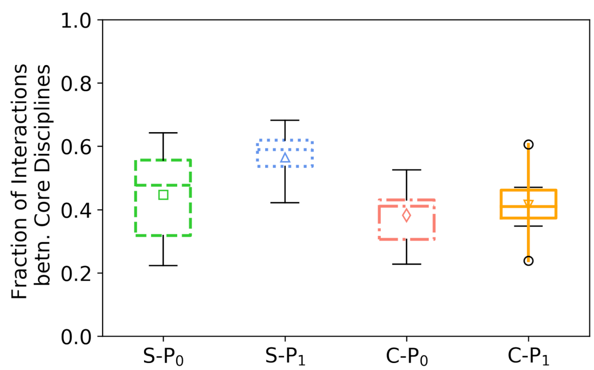


Figure 12: The fraction of total interactions between the core disciplines, equivalently text-based interactions, for the experimental conditions.

interactions there occur through text-based channels, as seen from Figure 12. And, as we observed before in Figures 8 and 9, 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.

4.2.3 Discussion and Implications

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 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 in the analysis in Section 4.1 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.

The results presented in this paper are limited by the measurement choices made in the two settings. Specifically, we chose a 1-minute time-step in the engine design experiment and an hour in the NASA MDL setting. The unit of communication in the engine study is a chat message or a broadcast message, whereas in the NASA MDL setting, the unit of communication is a face-to-face conversation. While these measures allow us to compare the trends in communication patterns, they do not capture similar amount of information exchanged across the two studies. The content of information exchanged in one chat message is clearly much lower than the content of information exchanged in one face-to-face discussion. Similarly, the information exchanged in one minute in the engine experiment may not be equivalent to the information exchanged in one hour at NASA MDL. Further investigation is necessary to establish equivalent measures of information content across the two studies, which are likely to depend on the complexity of the problem, the mode of communication, and the prior knowledge and expertise of the subjects.

4.3 Lessons Learned for Designing Representative Model Worlds

This section presents suggestions for designing representative model worlds with respect to the key features of system design process described in Section 3.2.5, based on the "lessons learned" from our analysis and on knowledge from extant literature.

4.3.1 Align Design Expertise and Task Complexity for Subject-Task Interactions

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 [42]. Engineering students in controlled studies take longer to complete coupled system design tasks as the number of design variables increases [3, 43]. 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 3.3.3. Moreover, this interplay between the subject and task shows strong technical-communication mirroring because the students can understand the interdependencies of the problem and decide whom to interact with (see Figure 7). 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 7).

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 4.1 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 interplay among the task complexity, time, information, & resources available, and subjects' expertise, rather than on matching task complexity or 'absolute' expertise level. To that end, Dorst's expertise framework [44] 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.

4.3.2 Allow a Burn-in Period Before Observations

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 [45]. For example in Figure 8, 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 [26] 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 [26]). Yu et al. [46] 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 4.3.1. 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 4.3.4), 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 technical dependencies 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 [47]; 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 [48]. For tips on building an environment for economic experiments, the reader may refer to Refs. [49, 50].

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.

4.3.3 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 communication among disciplines. For instance, Figures 7 and 9 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 [51]. 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 [52]. Computer-based tools can also be supportive of systems engineering for dispersed teams if appropriate sufficient means of data exchange and communication are incorporated [53]. 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 interplay with design communication. Success in cross-functional teams requires setting appropriate project goals for effective communication [54]. Teams can address design inconsistencies through design dialogue if common goals are identified early in the process [32]. Moreover, formal and informal organizational channels for communication are known to influence how communication happens [52, 53], 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.

4.3.4 Select Appropriate Degree of Coupling for Tasks

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 [55, 56]. 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 3 – 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 [57]. 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 13 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.

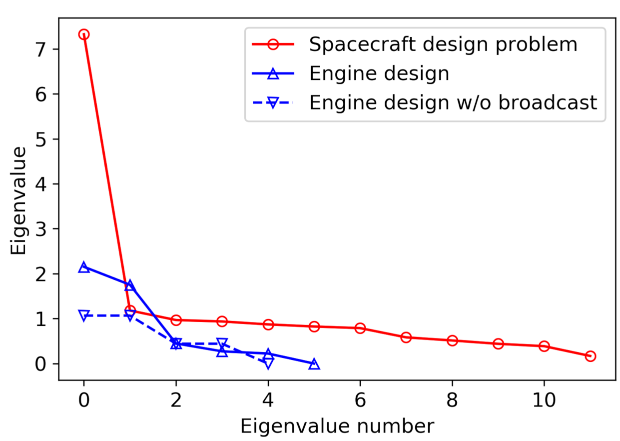


Figure 13: Eigenvalues of the normalized design structure matrices.

4.3.5 Select Team Size Appropriate for Technical-Communication Mirroring

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 [58, 59]. 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 [60]. If tables are inaccessible, alternative methods for approximating the sample size are available [61]. For multiple regression with m predictors and the estimated correlation of R^2 (equivalently effect size $f^2 = \dfrac{R^2}{1-R^2}$), the suggested number of samples is $N \geq \dfrac{6.4 + 1.65m - 0.05m^2}{f^2}$. 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.

4.3.6 Gather Independent 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 6 and Figure 8 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 [62]. 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 [62], 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.

5 Conclusions

This paper set out to illustrate the importance and nuances of designing representative model worlds. By collecting data on both the reference setting and the model world, we were able to compare the results and evaluate the extent to which our model world design succeeded in generating representative behavior. This enabled us to contribute lessons in what to worry about when designing model worlds more generally.

Specifically, our analysis generated several lessons learned about how to design model worlds for a purpose similar to that in our case study (detailed in Section 4.3): (1) choose the subjects and task such that the interplay between subject expertise and task complexity prompts the desired types of behavior; (2) if expertise in the particular task and context are important, allow a burn-in period so that subjects can learn the setting; (3) consider two aspects of context, organizational structure and incentives, because they constrain and motivate behavior; (4) in studying system design processes, ensure there is appropriate coupling between disciplines; (5) select an appropriate team size to support this need for coupling; (6) gather independent samples and compare aggregate behaviors.

It is worth highlighting that all of the above insights relate to the interplay among subject, task and context, a key feature of our model world definition. This is a critical takeaway because, to the extent that past studies have considered representativeness, it has been in terms of single dimensions (e.g., student subjects vs. experienced engineers). While these considerations are important too, our study reveals that it is not enough to consider these dimensions on their own. Moreover, it appears that it might be feasible in some cases to obtain overall representativeness of the model world, without high fidelity on any one dimension (e.g., lesson learned (1)).

The overarching lesson is that *model world designers should assess whether they have achieved a representative model world in terms of whether the behaviors match, rather than matching the dimensions of subject, task, and context individually.* Achieving matching of behavior usually involves representative interplay among subjects, tasks, and context. For example, the combination of the students' (limited) expertise and the simpler engine design task can generate communication behaviors observed at NASA MDL which consists of expert engineers working on a complex spacecraft design task. Similarly, what matters is not whether the same incentives are used for students and professionals but choosing (potentially different) incentives that are motivating for each group.

In other words, *representativeness is different from similarity.* Increasing similarity in individual dimensions may not necessarily increase representativeness – in fact it may even reduce representativeness if the effect is not balanced across the model world. What can be done if a model world does not preserve the intended behaviors and features needed to achieve representativeness? Our case study points to two ways forward. First, assess whether these non-preserved behaviors are crucial to answering the research question. Model worlds need not match all aspects of a real world in order to be representative for a particular research question. Second, if these behaviors are crucial, the model world must be modified to increase representativeness. These modifications will be specific to the setting and the research question; we discussed several such modifications for our particular setting in Section 4.3.

Future work should focus on refining and expanding the lessons learned from the use of the model world (Section 4.3). For instance, future work may investigate whether larger problem size, better-aligned interdependency structure, and the availability of an integrative discipline improves the model world representativeness in the collaborative design context. Future work should also explore model worlds for studying phenomena other than communication – such as creativity or search – to determine what is critical for representativeness in those cases. To the extent possible, using theories to identify the impact of choices in the model world on behaviors can help in reducing the number of iterations for developing a good model world.

As highlighted earlier, designing good model worlds of socio-technical systems is challenging, and there is a lack of shared understanding and standards for developing model worlds to support engineering systems design research. There is a need for further research in this direction. In the long term, we hope the community can move towards a set of "scaling relationships" for capturing key dynamics of real-world problems in representative model worlds, but this will be a challenging task since we are dealing with socio-technical systems.

In the meantime, in an effort towards achieving shared understanding within the engineering systems design community, we recommend that publications that use model worlds should clearly specify the objective of the model world, the target real world and the phenomena they are trying to represent. The publications should also clearly specify the thought process used in developing model worlds.

Acknowledgements

The authors gratefully acknowledge financial support from the National Science Foundation through NSF CMMI EDSE collaborative Grant Nos. 1841192 [ZS, EG], 1841062 [JP], and 1841109 [PG] and also NSF CMMI 1563408 and 1662230. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. The authors also wish to acknowledge Nikolai Joseph, Connor Forsythe, and Victoria Nilsen for the NASA MDL data collection, and Vikranth Reddy for design of the engine design problem.

References

[1] Szajnfarber, Z., Grogan, P. T., Panchal, J. H., and Gralla, E., 2020. "A call for consensus on the use of representative model worlds in systems engineering and design". *Systems Engineering.* In review.

[2] Panchal, J. H., and Szajnfarber, Z., 2017. "Experiments in systems engineering and design research". *Systems Engineering,* **20**(6), pp. 529–541.

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

[4] Grogan, P. T., and de Weck, O. L., 2016. "Collaborative design in the sustainable infrastructure planning game". In 2016 Spring Simulation Multi-conference Annual Simulation Symposium, SCS.

[5] Chaudhari, A. M., Bilionis, I., and Panchal, J. H., 2020. "Descriptive models of sequential decisions in engineering design: An experimental study". *Journal of Mechanical Design,* **142**(8).

[6] Gralla, E., Goentzel, J., and Fine, C., 2016. "Problem formulation and solution mechanisms: A behavioral study of humanitarian transportation planning". *Production and Operations Management,* **25**(1), pp. 22–35.

[7] Azhar, A., Gralla, E. L., Tobias, C., and Herrmann, J. W., 2016. "Identification of subproblems in complex design problems: A study of facility design". In ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, ASME.

[8] Chaudhari, A. M., Sha, Z., and Panchal, J. H., 2018. "Analyzing participant behaviors in design crowdsourcing contests using causal inference on field data". *Journal of Mechanical Design,* **140**(9), p. 091401.

[9] Gralla, E., and Szajnfarber, Z., 2018. "Games and Exercises for Teaching and Research: Exploring How Learning Varies Based on Fidelity and Participant Experience". In Volume 2B: 44th Design Automation Conference, American Society of Mechanical Engineers, p. V02BT03A026.

[10] Kruskal, W., and Mosteller, F., 1979. "Representative sampling, III: The current statistical literature". *International Statistical Review,* **47**(3), Dec., p. 245.

[11] Yin, R. K., 2009. *Case Study Research: Design and Methods*. Applied social research methods series. Sage Publications, Beverly Hills, CA.

[12] Brunswik, E., 1949. "Systematic and representative design of psychological experiments. with results in physical and social perception". In Proceedings of the [First] Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, pp. 143–202.

[13] Sargent, R. G., 2013. "Verification and validation of simulation models". *Journal of Simulation,* **7**(1), pp. 12–24.

[14] Zeigler, B. P., Praehofer, H., and Kim, T., 2000. *Theory of Modeling and Simulation*, second ed. Academic Press, New York.

[15] Raser, J. R., 1969. *Simulation and Society: An exploration of scientific gaming.* Allyn and Bacon, Boston.

[16] Anandan, S., Teegavarapu, S., and Summers, J. D., 2006. "Issues of similarity in engineering design". In Volume 1: 32nd Design Automation Conference, Parts A and B, ASMEDC.

[17] Durand, F., Helms, M. E., Tsenn, J., McAdams, D. A., and Linsey, J. S., 2015. "In search of effective design problems for design research". In Volume 7: 27th International Conference on Design Theory and Methodology, American Society of Mechanical Engineers.

[18] Kumar, V., and Mocko, G., 2016. "Similarity of engineering design problems to enable reuse in design research experiments". In Volume 7: 28th International Conference on Design Theory and Methodology, American Society of Mechanical Engineers.

[19] Sosa, R., 2018. "Metrics to select design tasks in experimental creativity research". *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science,* **233**(2), June, pp. 440–450.

[20] Levy, B., Hilton, E., Tomko, M., and Linsey, J., 2017. "Investigating problem similarity through study of between-subject and within-subject experiments". In Volume 7: 29th International Conference on Design Theory and Methodology, American Society of Mechanical Engineers.

[21] Cross, N., 2004. "Expertise in design: an overview". *Design Studies,* **25**(5), Sept., pp. 427–441.

[22] Cross, N., and Cross, A. C., 1998. "Expertise in engineering design". *Research in Engineering Design,* **10**(3), Sept., pp. 141–149.

[23] Ahmed, S., Wallace, K. M., and Blessing, L. T., 2003. "Understanding the differences between how novice and experienced designers approach design tasks". *Research in Engineering Design,* **14**(1), Feb., pp. 1–11.

[24] Crismond, D. P., and Adams, R. S., 2012. "The informed design teaching and learning matrix". *Journal of Engineering Education,* **101**(4), Oct., pp. 738–797.

[25] Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., and Saleem, J., 2007. "Engineering design processes: A comparison of students and expert practitioners". *Journal of Engineering Education,* **96**(4), Oct., pp. 359–379.

[26] Cash, P. J., Hicks, B. J., and Culley, S. J., 2013. "A comparison of designer activity using core design situations in the laboratory and practice". *Design Studies,* **34**(5), Sept., pp. 575–611.

[27] Hales, C., 2000. "Ten critical factors in the design process". In Proceedings of ASM/ASME Conference on Failure Prevention Through Education–Getting to the Root Cause, Strongsville, pp. 49–55.

[28] Szajnfarber, Z., and Gralla, E., 2017. "Qualitative methods for engineering systems: Why we need them and how to use them". *Systems Engineering,* **20**(6), pp. 497–511.

[29] Broniatowski, D. A., and Tucker, C., 2017. "Assessing causal claims about complex engineered systems with quantitative data: internal, external, and construct validity". *Systems Engineering,* **20**(6), pp. 483–496.

[30] Sherwood, B., and McCleese, D., 2013. "JPL innovation foundry". *Acta Astronautica,* **89**, pp. 236–247.

[31] Eppinger, S. D., 2002. Patterns of product development interactions. ESD Working Papers ESD-WP-2003-01.05-ESD Internal Symposium.

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

[33] Sosa, M. E., Eppinger, S. D., and Rowles, C. M., 2004. "The misalignment of product architecture and organizational structure in complex product development". *Management science,* **50**(12), pp. 1674–1689.

[34] DeChurch, L. A., and Mesmer-Magnus, J. R., 2010. "The cognitive underpinnings of effective teamwork: A meta-analysis.". *Journal of Applied Psychology,* **95**(1), pp. 32–53.

[35] DeChurch, L. A., and Mesmer-Magnus, J. R., 2010. "Measuring shared team mental models: A meta-analysis.". *Group Dynamics: Theory, Research, and Practice,* **14**(1), pp. 1–14.

[36] Kozlowski, S. W., and Ilgen, D. R., 2006. "Enhancing the effectiveness of work groups and teams". *Psychological Science in the Public Interest,* **7**(3), Dec., pp. 77–124.

[37] Eppinger, S. D., 2001. "Innovation at the speed of information". *Harvard business review,* **79**(1), pp. 149–158.

[38] Sbea, G. P., and Guzzo, R. A., 1987. "Group effectiveness: What really matters?". *Sloan Management Review (1986-1998),* **28**(3), p. 25.

[39] Langville, A. N., and Meyer, C. D., 2005. "A survey of eigenvector methods for web information retrieval". *SIAM review,* **47**(1), pp. 135–161.

[40] Hagberg, A., Swart, P., and S Chult, D., 2008. Exploring network structure, dynamics, and function using networkx. Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

[41] Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., Augspurger, T., Halchenko, Y., Cole, J. B., Warmenhoven, J., de Ruiter, J., Pye, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Bachant, P., Martin, M., Meyer, K., Miles, A., Ram, Y., Yarkoni, T., Williams, M. L., Evans, C., Fitzgerald, C., Brian, Fonnesbeck, C., Lee, A., and Qalieh, A., 2017. mwaskom/seaborn: v0.8.1 (september 2017).

[42] Ball, L. J., Onarheim, B., and Christensen, B. T., 2010. "Design requirements, epistemic uncertainty and solution development strategies in software design". *Design Studies,* **31**(6), pp. 567–589.

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

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

[45] Tversky, A., and Kahneman, D., 1974. "Judgment under uncertainty: Heuristics and biases". *science,* **185**(4157), pp. 1124–1131.

[46] Yu, B. Y., Honda, T., Sharqawy, M., and Yang, M., 2016. "Human behavior and domain knowledge in parameter design of complex systems". *Design Studies,* **45**, pp. 242–267.

[47] Diamond, L., and Lerch, F. J., 1992. "Fading frames: Data presentation and framing effects". *Decision Sciences,* **23**(5), pp. 1050–1071.

[48] Marzilli Ericson, K. M., and Fuster, A., 2014. "The endowment effect". *Annual Review of Economics,* **6**(1), pp. 555–579.

[49] Levitt, S. D., and List, J. A., 2007. "What do laboratory experiments measuring social preferences reveal about the real world?". *Journal of Economic perspectives,* **21**(2), pp. 153–174.

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

[51] Tenopir, C., and King, D. W., 2004. *Communication patterns of engineers.* John Wiley & Sons.

[52] Allen, T. J., 1970. "Communication networks in r & d laboratories". *R&D Management,* **1**(1), pp. 14–21.

[53] Harris, D., 2001. "Supporting human communication in network-based systems engineering". *Systems engineering,* **4**(3), pp. 213–221.

[54] McDonough III, E. F., 2000. "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), pp. 221–235.

[55] Flager, F., Gerber, D. J., and Kallman, B., 2014. "Measuring the impact of scale and coupling on solution quality for building design problems". *Design Studies,* **35**(2), pp. 180–199.

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

[57] Bashir, H. A., AlZebdeh, K., and Abdo, J., 2009. "An eigenvalue based approach for assessing the decomposability of interdependent design project tasks". *Concurrent Engineering,* **17**(1), pp. 35–42.

[58] Sudweeks, F., and Rafaeli, S., 1996. "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*, pp. 115–136.

[59] Abarbanel, B., and McNeely, W., 1996. "Fly thru the boeing 777". In ACM SIGGRAPH 96 Visual Proceedings: The art and interdisciplinary programs of SIGGRAPH'96, ACM, p. 124.

[60] Cohen, J., 1992. "Statistical power analysis". *Current directions in psychological science,* **1**(3), pp. 98–101.

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

[62] Kraemer, H. C., and Blasey, C., 2015. *How many subjects?: Statistical power analysis in research.* Sage Publications.

List of Tables

List of Figures