Computationally Facilitating the Problem-Solving Design Process Via Real-Time Process Management

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy
in

Department of Mechanical Engineering

Joshua Thomas Gyory

B.S., Mechanical Engineering Department, University of Pittsburgh
M.S., Mechanical Engineering Department, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

August 2021

Acknowledgements

Wow, what a journey it has been. It is strange (and in some ways, perhaps a bit anticlimactic?) to think that five years of hard work and dedication culminated into this document. Going for and attaining a PhD is certainly not trivial, and I would not have made it through this journey if it was not for my support network and the love and friendship of so many people.

As I tell every prospective student that I meet, your advisor is THE most important part of your PhD experience. My experience is no different. Jon, I honestly do not think words can describe how much I have appreciated your mentorship and how you have served as my role model the past five years, academically, professionally, and personally. Whether it is your constant work ethic or your care and service to others, you have made me the person I am today – my leadership roles and my accomplishments. I wish I could dedicate those to you. Thank you for everything. And Ken, you too have also served as a source of inspiration to me these past five years with your wit and intellect. You never fail in asking the hard questions, pushing and expanding my intellectual limits. I will forever approach all my research and questions with your philosophical lens and curiosity. Thank you for your constant advice, support, wisdom, and humor.

To my other thesis committee members, Conrad Tucker and Andy Dong. Thank you both for invaluable feedback and for lending your time and expertise. The conversations I have had with both of you, at conferences and over zoom meetings, solidified in my mind that I wanted you to be part of this work. I look forward to the many future conversations and collaborations we may have as I begin my academic journey.

Chris and Kosa. As my academic [older] brothers, I have looked up to you more than I think you have realized. Through coursework, qualifying exams, research help, and even just joining the lab in the first place, I have always appreciated your advice on both a personal and

academic level. Thank you for always being there when I needed help and for serving as role models for me, especially in the early stages of my PhD experience.

To the iDIG lab members, both old and new. From a non-biased perspective, honestly, I think we are the best lab. The amount that we bond, both inside and outside of work, has really made my Carnegie Mellon experience so much better. From our daily lunches in the office, to our trips to Cedar Point, escape rooms, conferences, and dinners, I not only consider you my colleagues, but as amazing friends. Thank you for all the feedback on my presentations and research and for the unforgettable memories outside of the lab. I know we will all continue to say in touch and friends wherever our paths will take us.

Thank you to my Graduate Student Assembly (GSA) network and all those that have entered my life because of it. While I must admit it was a stressful experience, it was undoubtedly a large part of my life and PhD experience that I certainly do not regret. A special shoutout to the GSA 2019-2020 and 2020-2021 Executive committees! You are all incredible individuals and I so admire your determination, morals, and work ethics, as we all balanced our academic careers while still supporting the entire graduate student community. I have gained so much personally and professionally from each and every one of you, and I will take with me and learn from these, for the rest of my life.

Thank you to my FitFam family. You have always been there through the most stressful moments of my PhD. You were also there, and we continued our workouts virtually, throughout the entire pandemic. I thank you for those workouts, the distractions from school, and the focus on mental health, wellness, and work-life balance. Thank you for always being there and for the friendships that will continue for a lifetime.

Thank you to all my friends that I have made along the way at the University of Pittsburgh and at Carnegie Mellon. You have filled my life with such cherishable memories, food, and fun. Thank you also for the work-life balance and being my support system away from home. You are my life-long friends and I cherish all the memories we made and the many more to come.

Finally, thank you to my family, especially my parents and my brother. You all have been by my side, my support network, and shared your love throughout my entire journey. Words will never express the love you have for me and the love I have for you. This entire dissertation is dedicated to you. Love you with all my heart.

And last, but certainly not least, I want to acknowledge the funders of this research to make it possible in the first place. These include the Air Force Office of Scientific Research (AFOSR) under grant No. FA9550-16-1-0049 and grant No. FA9550-18-1- 0088, and the Defense Advanced Research Projects Agency (DARPA) through cooperative agreement No. N66001-17-1-4064.

Abstract

Teams are a major facet of engineering and are commonly thought to be necessary when solving dynamic and complex problems, such as engineering design tasks. Even though teams collectively bring a diversity of knowledge and perspectives to problem solving, previous work has demonstrated that in certain scenarios, such as in language-based and configuration design problems, the production by a team is inferior to that of a similar number of individuals solving independently (i.e., nominal teams). Aid in the form of design stimuli catalyze group creativity and help designers overcome impasses. However, methods for applying stimuli in the engineering design literature are largely static; they do not adapt to the dynamics of either the designer or the design process, both of which evolve throughout the problem-solving process. Thus, the overarching goal of this dissertation is to explore, better understand, and facilitate problem solving computationally, via adaptive, process management.

This dissertation first compares individual versus group problem solving within the domain of engineering design. Through a behavioral study, our results corroborate previous findings, exhibiting that individuals outperform teams in the overall quality of their design solutions, even in this more free-flowing and explorative setting of conceptual design. Exploiting this result, we consider and explore whether a human, process manager can lessen this underperformance of design teams compared to nominal teams, and help teams overcome potential deterrents that may be contributing to their inferior performance. The managerial interactions with the design teams are investigated and post-study interviews with the human process managers are conducted, in an attempt to uncover some of the cognitive rationale and strategies that may be beneficial throughout problem solving. Motivated from these post-study interviews, a topic-modeling approach then analyzes team cognition and the impact of these process manager interventions. The results from

this approach show that the impacts of these interventions can be computationally detected through team discourse. Overall, these studies provide a conceptual basis for the detection and facilitation of design interventions based on real-time, discourse data.

Next, two novel frameworks are studied, both of which take steps towards tracking features of design teams and utilizing that information to intervene. The first study analyzes the impact of modulating the distance of design stimuli from a designers' current state, in this case, their current design solution, within a broader design space. Utilizing semantic comparisons between their current solution and a broad database of related example solutions, designers receive computationally selected inspirational stimuli midway through a problem-solving session. Through a regression analysis, the results exhibit increased performance when capturing their design state and providing increased stimulus quality. The second framework creates an artificial intelligent process manager agent to manage the design process of engineering teams in real-time, tracking features of teams' actions and communications during a complex design and path-planning task with multidisciplinary team members. Teams are also placed under the guidance of human process managers for comparison. Across several dimensions, the overall results show that the AI manager agent introduced matches the capabilities of the human managers, showing potential in automating the management of a complex design process. Before and after analyses of the interventions indicate mixed adherence to the different types of interventions as induced in the intended process changes in the teams, and regression analyses show the impact of different interventions. Overall, this dissertation lays the groundwork for a computational development and deployment of adaptive process management, with the hope to make engineering designs as efficient as possible.

Table of Contents

List of Figures

List of Tables

Chapter 1 : Introduction

1.1 Motivation

Teamwork pervades the practice of engineering, both in academia and industry. Providing a diversity of knowledge and perspectives, team problem solving is particularly crucial when the task may be too challenging or complex for individual expertise and effort. To effectively problem solve, team members must be able to collaborate and communicate with one another efficiently [1,2]. Beginning in the early stages of their careers, engineers are trained via problem- and project-based experiences that provide them with these hands-on, team-oriented skills to prepare them for the rigors and demands of engineering design in practice [3–5]. Sharpening these collaborative techniques can improve upon the competencies of making decisions, exchanging technical ideas, and resolving conflicts to help abate any dissension that may arise within the team [6,7]. Effective communication strategies can also lead to a common, shared mental model of the problem among team members, fostering team synergy and improving the collective team performance [8,9].

Although there are potential benefits to teams, there also exist numerous, generally acknowledged deficiencies, highlighted and studied across different research fields [10]. Social loafing, for example, is a psychological phenomenon in which individuals tend to expend less effort when working in a group as opposed to working individually, even to the extent of withholding information from the rest of their group members [11,12]. This is due to the belief that an individual's work (or lack of) will be obscured relative to the efforts by the rest of the team. Furthermore, group think is a psychological drive for a consensus among a group in an attempt to avoid any potential confrontations or disagreements. This phenomenon may lead to a premature decision by the group, even one that some members may not agree with, and in turn, can hinder creative thinking [13,14]. A dominant individual can also emerge from the group, who pressures other team members and consequently has a greater share of impact on the final outcome of the team, often resulting in more polarized decisions [15]. This influence may not necessarily stem from a higher degree of knowledge or problem-solving ability, but from that individual simply being more persuasive and persistent. Thus, many of these deficiencies in teams have to do with the dynamics of the problem-solving process itself, as opposed to individual experience or knowledge. These deficiencies can even lead teams to underperform to individuals on the same task, including brainstorming studies, Remote Associate Test (RAT) puzzles, and configuration engineering design tasks. Perhaps, as Maier put it, there is a "*need for an integration function*" that could automatically mitigate some of these process shortcomings in groups by acting as their central nervous system [10]. The goal of such an integration function would be for it to intervene when a team veers off course to stimulate problem-solving and mitigate some of these process pitfalls.

In the field of engineering design, stimulus techniques have already been shown to be effective in impact the problem-solving process [16,17]. However, one of the critical limitations of these interventions lies in their static nature. Specifically, stimuli provided during cognitive studies are determined a priori to problem solving, and do not adapt to the dynamics of the designer(s), design space, or process, all of which change and evolve throughout problem-solving. Some work with analogical stimuli shows that providing near and/or far stimuli to the problem domain creates beneficial and different impacts on ideation outcomes. For example, providing far stimuli has been shown to increase the novelty of solution outcomes, but this result does not necessarily hold with near stimuli [18]. What if these interventions not only modulate to the problem domain itself, but also the state of the designer(s) and design progress? To fully realize this goal, these techniques need to adapt to the ongoing design state and be automatically implemented in real-time, perhaps via an artificial intelligent (AI) agent, which could serve as Maier's "integration function." This opens the question of how designers and the design state can be actively monitored during problem-solving by such an agent.

The field of AI and human collaboration is quickly evolving and achieves superior outcomes by taking advantage of the complementary strengths of each, combing the creativity of humans with the analytical power of AI. Research focuses on the effectiveness of human-AI collaboration in augmenting the ability of humans and/or AI in solving complex problems. For example, Hu and Taylor exhibit the benefits of computer-aided design intelligent tutors, assisting students apply their learned knowledge to solve novel problems by guiding their exploration [19]. Instructional design agents support novice designers in exploring complex design spaces through a design study with a solar farm design problem [20]. Song *et al.* report that AI agents can effectively improve the performance of hybrid teams in the configurational and operational design of drones. The effectiveness of human-AI hybrid teams has also been seen in other contexts, such as tutoring and education [21], job shop scheduling [22], and clinical imaging [23].

Moreover, many efforts work towards the understanding, design, and improvement of human-AI interaction to facilitate collaboration [24,25]. For example, according to the information flow and role distribution between humans and AI, a group of researchers classify human-AI collaboration into two forms, human intelligence in the loop of artificial intelligence (human-in-the-loop) and AI in the loop of human intelligence (AI-in-the-loop) to clarify the required characters of interactive AI [24]. In terms of AI design, the common language and the explainability and transparency of AI are considered essential to improve humans' understanding of and trust towards AI [25,26]. Van Den Bosch and Bronkhorst study human-AI collaboration in complex decision making, categorizing six levels according to the type of interaction and level of collaboration [27]. Yang, Steinfeld, and Rosé identify design challenges of human-AI interaction and propose strategies for addressing them [28]. In this dissertation, AI will instead be used to improve the behaviors and problem-solving processes of teams in the field of design, a field requiring unique problem-solving strategies. The advantage of an AI agent in this research application is twofold: its ability to be able to track multiple metrics simultaneously and over time, and its ability to ascertain underlying patterns within data, such as team communication, that may not be perceivable via direct human inspection.

Within teams, members collaborate and communicate. Team communication is a fundamental component of complex engineering design processes, especially those that engage multiple disciplines. It enables team members to integrate specialized knowledge, bridge gaps, and negotiate, playing a boundary spanning role that supports design knowledge sharing, exploration, collaboration, and coordination [29-31]. Prior studies recognize communication as a prime success factor of shared leadership in teams [32], impacting team creativity [33]. In design teams aiming at complex problem solving, the effectiveness of interpersonal communication influences team efficiency, performance, and progress toward their design. Inadequate or ineffective communication can hinder task and team achievements [34]. Moreover, studies on team communication facilitate our understanding of team cognitive processes. Accordingly, tracking team discourse and interactions could serve as a critical measure for an integrative AI agent to monitor and understand, in real-time, the evolving states of the designer(s) and design progress during problem-solving.

1.2 Thesis Statement

Traditional human studies and team research in the engineering design theory and methodology field do not fully address real-time tracking of team dynamics and/or consider the evolving state of problem-solving teams. In order to effectively aid and provide feedback to teams, methods need to capture these dynamics that organically arise throughout the problem-solving process. Natural language processing and design stimulus techniques are tools studied across many different fields, such as computer science, design engineering, and human-computer interaction, to name a few. Integrated together, these techniques can serve as a fundamental basis for tracking the state of teams and intervening in real-time to improve the problem-solving process of engineering teams. Based on this framework, the following thesis statement is proposed for this dissertation:

> *Real-time process management, via the monitoring of design cognition and discourse, can adapt to the state and dynamics of the designers and design progress, thereby facilitating the overall problem-solving design process.*

1.3 Dissertation Outline

This overarching thesis will be supported through two main phases of work in this dissertation. The first phase includes an exploration study in order to uncover some of the mechanisms and nuances of real-time, process management as well as the utilization of team discourse as a measure to track teams. To begin, **Chapter 2** presents a behavioral study with human process managers, who intervene with their design teams via a "manager bank" of prescribed stimulus techniques. However, unlike traditional stimulus methods, these stimuli are dynamically provided by the managers in real-time when deemed appropriate depending on the state of the teams' actions and discourse, thus adapting to their current state. Post-study interviews with the human managers reveal that a desire to invoke shifts in team discourse is a critical motivation for intervening. Driven by this result and the criticality of team communication in team problem-solving, **Chapter 3** analyses teams' discourse data via a topic modeling approach. This analysis uncovers whether team members' verbalizations can be leveraged to computationally detect the impact and effects of these managerial interventions and produce the desired topic shifts. These two chapters provide insights into the strategies of an effective AI manager and the feasibility of monitoring design discourse.

The second phase of this dissertation works towards the facilitation of real-time interventions *during* problem-solving. **Chapter 4** presents a framework that monitors the design state of designers midway through problem-solving and modulates the distance of design stimuli relative to that current state. Latent Semantic Analysis (LSA) indicates the distance of the design stimuli through semantic similarity, how semantically near or semantically far a stimulus lies. Computing semantic comparisons between design progress and a large database of potential design stimuli, these modulations are done, and the interventions are provided in real-time. Stimulus distance affects final design outcomes in different ways. **Chapter 5** integrates many aspects of the previous chapters and demonstrates the development of an artificial intelligent process manager. Trained on previous team problem-solving data, this process manager tracks several features of teams (team communication and team action) and intervenes at discrete points in time. The AI agent is implemented in the context of a highly complex drone design and operations path-planning problem, and completed fully online, so teams are geographically distributed and interact/communicate with each other through an online platform. **Chapter 6** then presents additional investigations on the process manager study from the preceding chapter. Before and after analyses are done to identify the impact of the interventions on the process behaviors of the teams and regression models are trained to explore the correlation between the interventions and overall team performance. Finally, **Chapter 7** provides a summary of the dissertation and major results, presents the main contributions to the research literature, as well as avenues for future work in these areas.

Chapter 2 : An Exploratory Study – The Effects of Real-Time Process Management[1]

2.1 Introduction & Background

While teams are a fundamental component of engineering, they may not always perform optimally or as expected [35]. Previous work in psychology and configuration engineering design has supported this assertion that teams may not always be maximally proficient [36-38]. Many of these studies utilize the term nominal team to refer to an equivalent team of participants who work individually, but whose efforts are pooled together with the best solution chosen to form the collective team performance. Nominal teams are necessary in these studies to be able to make comparisons between equivalent sized groups of individuals and interacting teams. In brainstorming studies, interacting groups are shown to generate fewer ideas than nominal brainstorming teams [39,40]. In language-based, Remote Associate Test (RAT) puzzles, interacting groups also perform worse than non-interacting nominal groups [41]. Remote Associate Tests are word-retrieval protocols which measure analytical and convergent thinking ability [42]. The test is done by having the participant form compound words/phrases through a given set of provided cue words by finding the common term that links with all three cues. It is inherently a language-based task because of this dependence on semantically related word

[1] The work presented in this chapter is published in:
Gyory, J. T., Cagan, J., and Kotovsky, K., 2019, "Are You Better off Alone? Mitigating the Underperformance of Engineering Teams during Conceptual Design through Adaptive Process Management," Res. Eng. Des., 30(1), pp. 85–102.

retrievals. Additionally, in configuration engineering design problems, McComb et al. behaviorally studied and computationally simulated human design teams. Configuration design refers to a subset of design problems that involve pre-defined set of components that can be combined based on interface constraints and requirements of the task [43]. In their work, they showed that under certain problem characteristics, such as objective alignment and global structure, zero interaction frequency among team members resulted in optimal performance [44]. This zero-interaction frequency finding indicates that individuals, i.e., nominal teams, would be the preferred structure under these circumstances.

The current chapter exploits previous findings of individual and team performance but on a different type of design problem, conceptual engineering design. Conceptual design is free flowing and exploratory in the early stages as with brainstorming but narrows down to seek high quality and practical solutions. It is also structured and constrained as with configuration design, but while conceptual design is open-ended, configuration design is limited to the combinatorics of a finite set of enumerated components. Thus, because configuration design problems utilize these predefined components to work with, there is normally a single best assembly/solution [45]. Conceptual design is initially more open-ended, in that there could be multiple preferred solutions that satisfy the problem, although a single concept must be selected by the end amongst these design alternatives. Moreover, unlike previous brainstorming studies, this chapter presents an engineering design problem that more accurately emulates one that would be encountered in practice, with dynamically changing design constraints [46]. Thus, this chapter seeks, in part, to understand how teams behave in this early, albeit critical, part of the engineering design process.

Engineering teams may often reach impasses during problem-solving [47], as do people solving RAT problems [48]. Additional aid in the form of analogies and solution examples has been shown to stimulate group creativity and improve solution quality, particularity during brainstorming tasks and the ideation stage of conceptual engineering design [49, 50]. However, such aid can also lead to fixation, which is a blind adherence to a set of ideas or concepts. Fixation can suppress team creativity and lead to inferior overall performance [51,52]. Work on creativity tasks has tried to lessen this fixation effect, for example, through providing expansive versus restrictive examples [53], effectively managing non-verbal devices such as pictures or sketches [54], as well as examining the specificity and level of abstraction of such [55]. Furthermore, theoretical models have been developed for leadership strategies in creativity tasks [56]. Perhaps some of the aforementioned deficiencies in team problem-solving can be ameliorated in conceptual design if a portion of the resources used for solving the problem are instead used to guide and control the team's design process. This chapter considers allocating these resources to a human manager, who acts as a team's central processing system, to gather feedback and provide adaptive management. Previous work on design creativity provides stimuli to problem-solvers that are static in nature, such as in constant modalities and/or at pre-defined intervals. The integrative role of the process manager in this current work provides real-time feedback to design teams that is fluid in both modality and timing, adapting to a team's current state and how they evolve throughout the problem-solving process.

Consequently, this chapter is motivated by two goals. The first seeks to determine whether the underperformance of team problem-solving, identified in language-based and configuration design, applies to the conceptual phase of design, by comparing individual and team performance in this domain. The second goal is to explore whether a portion of the resources used for problem-solving could instead be used to dynamically manage and control the design process of the remainder of the team and improve performance via a human process manager. The effect of the process manager on design teams is studied and the performance compared to that of both unmanaged and nominal teams. After this comparison, an amalgamative analysis, with both quantitative and qualitative features, is done to uncover some of the managerial strategies, and the rationale behind which, are most beneficial to design teams.

To form a conceptual framework, this chapter (Chapter 2.2) begins by introducing the experimental architecture and logistics of the behavioral study that is run to answer the aforementioned research goals. Here, three different types of team conditions will be defined, for consistency, and used throughout the remainder of the chapter: a managed team (problem-solvers who work together and are overseen by a more experienced process manager), an unmanaged team (problem-solvers who work together and are not overseen by a more experienced process manager), and a nominal team (individual problem-solvers who do not work together but are randomly placed together to form an artificial team). Also in this section is a description of how the process managers are chosen for this study and the methods by which they can intervene with their design teams. The next sections of the chapter provide an overview of the analysis metrics and techniques that are used to compare problem-solving performance (design quality, design novelty, and team cohesion), followed by the presentation of the results and the process manager interventions. The chapter concludes with a discussion of the strategies and cognitive rationale of the process managers, as well as a supplemental experiment studying the effect of verbalization. As the primary means of conveying ideas, teams are required to communicate and verbalize with each other, which is one of the main differences between individual and team problem solving. Also motivated by previous research studying this effect, whether verbalization acts as a cognitive deterrent during conceptual engineering design problem-solving is examined. Finally, limitations of the study and future work are acknowledged, followed by brief remarks on the implications of this chapters results for engineering design teams and design practice.

2.2 Methodology

2.2.1 Experimental Conditions

To address these two primary research goals, a behavioral study was run with freshman engineering students at Carnegie Mellon University in Pittsburgh, PA, USA. Participants for the study were recruited from the “Fundamentals of Mechanical Engineering” class, a freshman-level course in the Mechanical Engineering department. The intention was to recruit students with comparable and little to no prior exposure to techniques and theories in engineering design methods and conceptualization. In total, 95 freshman engineering students participated in the study. Because students were recruited through their engineering course and participated in the study in lieu of a scheduled lecture, they were not monetarily compensated for participating. However, during the period they were not participating in the study, they received an educational lecture on engineering ethics.

These students, or novice designers, were randomly assigned to one of three different team conditions: a managed team, an unmanaged team, or a nominal team (see Figure 2.1). A managed team was composed of four freshman engineering students collectively solving the problem, with one mechanical engineering graduate student as a manager overseeing their design process. An unmanaged team was comprised of five freshman engineering students and no graduate student manager. The additional problem-solver in the unmanaged teams was to keep the number of problem-solving resources equivalent across the two experimental conditions. Lastly, a nominal team was composed of five randomly chosen freshman engineering students (of the 23 participants) who solved the problem individually but did not interact with each other. Instead, the best solution was chosen from amongst the five individual solutions. In total, there were eight managed and eight unmanaged teams, and 23 individuals from which eight nominal teams were artificially created to compare with the other team conditions. The method for generating the nominal teams will be discussed later in the chapter.

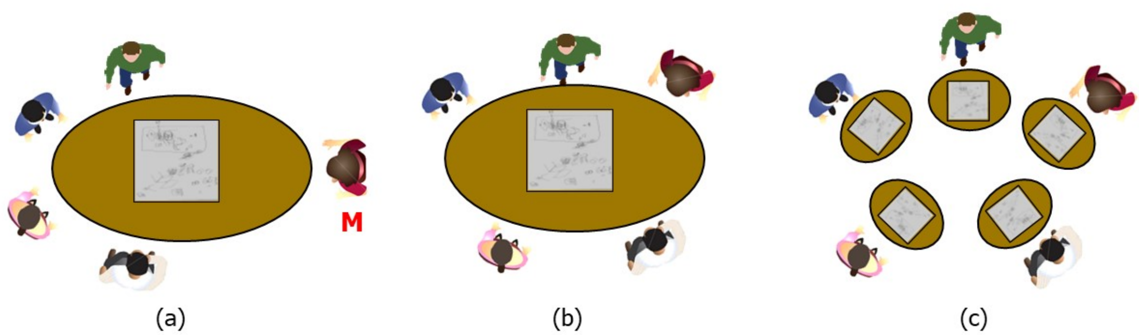


FIGURE 2.1: PARTICIPANTS WERE RANDOMLY ASSIGNED TO ONE OF THE THREE TEAM CONDITIONS: A (a) MANAGED TEAM, AN (b) UNMANAGED TEAM, OR A (c) NOMINAL TEAM

2.2.2 Process manager selection and manager bank creation

The managed teams were composed of four freshman novice designers and one graduate student process manager. These managers were able to intervene with their team to affect the solving process with different stimuli but could not directly contribute to the problem solution. The graduate students were selected for the study via a recruitment survey that was disseminated to a portion of the graduate student population in the Mechanical Engineering Department at Carnegie Mellon University. The response rate from the surveys was 68%, with 40 graduate students completing it.

The desired managerial characteristics were the possession of engineering design knowledge and prior experience in leading a team. Using a Likert scale assessment, questions on the survey queried the graduate students to self-assess their mechanical engineering design knowledge, as well as their leadership experience. To minimize survey bias due to over- (or under-) confidence in their own abilities, supplemental questions asked them for specifics such as undergraduate/graduate classes they had taken in engineering design, the area of their primary research, areas of interest outside of their primary research, and specific examples they had in leading a team. The eight graduate students with the highest level of design knowledge and leadership experience were selected as the process managers for the study.

Even though the managers were instructed to intervene when they felt it necessary, they were only allowed to interact with items from a bank of prescribed stimuli. Three distinct categories of interventions were chosen, theoretically grounded in previous literature for different modalities and stimuli for improving ideation and problem-solving effectiveness. Inspired by the approaches of design by analogy and metaphor, keywords were selected, a technique known to help designers divergently think about and reframe the problem in a different domain [57,18]. Cognitive priming with solution examples, shown to increase quality output [49], inspired the use of different design components. Moreover, because visual representations have been shown to be a more effective modality [58] these components were pictorially depicted. Heuristics for creative problem-solving and management [59,60] influenced the use of design strategies, where designers with more structured planning and approach perform better. Thus, previous literature motivated the following intervention types that were, respectively, chosen: keywords, design components, and design strategies.

The collection of permissible stimuli was created for this particular problem from questions in the manager recruitment survey. In the survey, each potential manager was asked to generate possible items in each of those categories that they might provide to a hypothetical engineering team to aid in solving the design problem used in the study. After the graduate students with the highest level of engineering design knowledge and leadership experience were chosen through the survey, their answers to these questions were compiled. For the three intervention types, the six answers with the highest frequency between the chosen managers were aggregated to form the manager bank, shown in Figure 2.2. During the ensuing study, the process managers were only allowed to select from these 18 specific examples, that comprised the three categories, to intervene with. This compilation was done to control for variability and to maintain some consistency in the types of interventions. The keywords and design components in the bank were printed on cards that were physically handed to the design teams, while the design strategies were verbally spoken by the managers, all done when deemed appropriate. All in all, the study was purposefully designed to minimize the impact of differences or limitations in managerial skill, by recruiting graduate students with both design knowledge and leadership experience, and creating the manager bank, from which they were restricted to apply.

Prior to the experiment, the graduate students were required to participate in a 30-min training session. During this session, which was led by one of the research investigators, the managers were trained on the experimental procedures. Other than reading the instructions, answering logistical questions, and intervening with a design strategy, the managers were not allowed to speak during the experiment. The managers were also told to keep notes on the exact times and types of interventions they used during the study with their teams. It was also emphasized to them that they were not to help their team in directly solving the design problem but were there only to manage their team's design process.

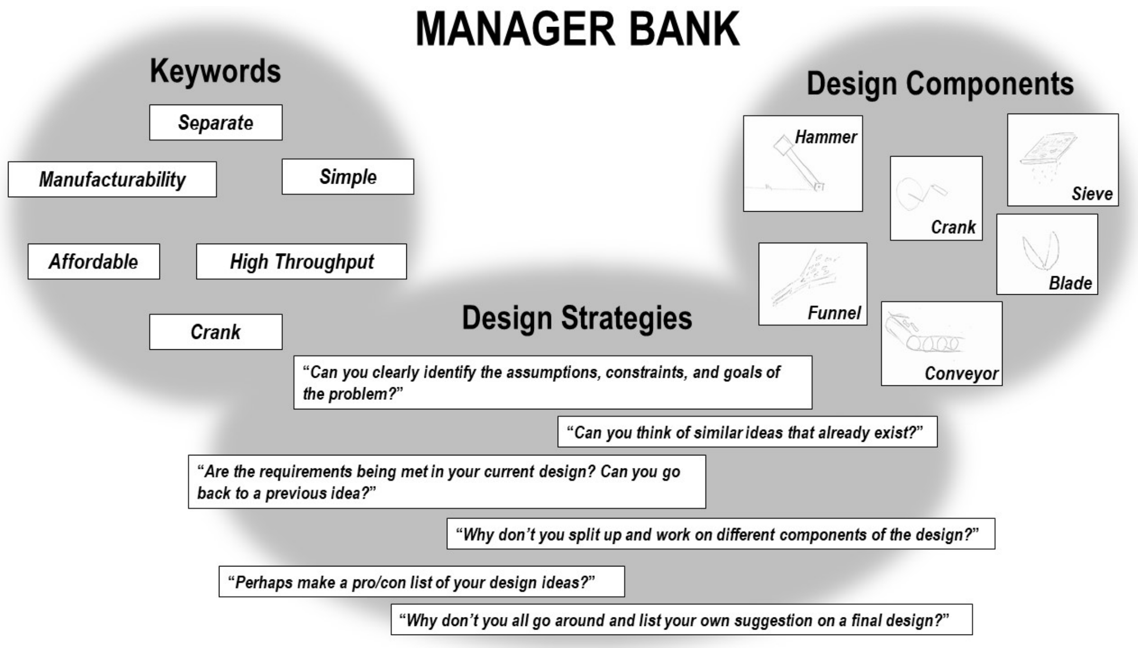


FIGURE 2.2: THE COLLECTION OF INTERVENTION STIMULI WHICH THE PROCESS MANAGERS WERE ABLE TO USE DURING THE EXPERIMENT

Within 2 days following the experiment, a post-study interview was also conducted with each manager. During these interviews, the research investigator went through each intervention and asked the managers: "What made you interact [with item x]", "Why did you interact with what you did", and "What was the effect of your interaction?" The primary goal of these interviews was to determine what prompted the managers to interact with their teams, to gain a deeper understanding into the rationale underlying these interventions. During these interviews, the managers were allowed to refer to the notes they had taken during the study, to facilitate in their recollection of events, if necessary.

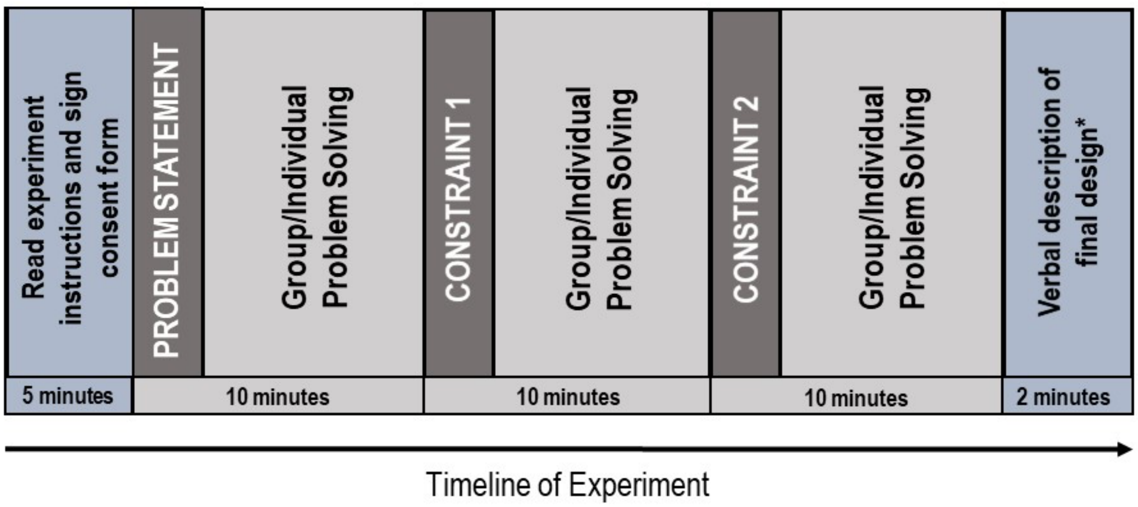
2.2.3 Experimental protocol and materials

The students' 1-h and 50-min class period was broken up into two 55-min intervals. Half of the class was assigned to group A while the other half of the class was assigned to group B. While group A participated in the experiment for the first 55-min period, the other group received a lecture on engineering ethics, an agreed-upon request between the researchers and course instructor for running the study during the class period. For the second half of the class, the two groups switched. Within a group, students were evenly and randomly distributed among the different experimental conditions. The same experimental materials were provided to all participants, regardless of the team condition they were in. These items include a pen for everyone on a team and 3 sheets of 11″×17″ white, multipurpose paper. In the managed and unmanaged team conditions, the three sheets of paper were shared among the entire team to promote collaboration, while in the nominal condition, everyone received three sheets. Prior to participating in the experiment, individuals were asked to read and sign a consent form.

During the experiment, participants were given 30 min to solve the following engineering design problem [49, 61, 62]:

Problem Statement:
Design a low-cost and easy to manufacture device that removes the outer shell from a peanut.
Constraint 1: *The device is meant to be utilized in developing countries where electricity may not necessarily be available as a power source.*
Constraint 2: *In addition to the previous constraint, the proposed design must be able to separate a large quantity of peanuts from their shells, while causing minimal damage to the inner peanut.*

To simulate a real-world, dynamically changing engineering problem, the two constraints were sequentially introduced 10 min and 20 min, respectively, into the study. These added constraints were meant to exacerbate the problem difficulty throughout problem-solving. An overview of the experimental timeline is shown in Figure 2.3.



*Audio was not collected from the individuals (i.e., nominal teams), so they did not provide a verbal description

FIGURE 2.3: THE EXPERIMENTAL TIMELINE SHOWING THE INTRODUCTION OF THE PROBLEM STATEMENT, CONSTRAINTS, AND BLOCKS OF PROBLEM SOLVING

Participants were allocated the entire 30 min to problem solve. To distinguish this task from a pure, creative brainstorming session, participants were told from the onset that they could initially (and were encouraged) to discuss and sketch about as many possible ideas for the problem, but by the end of the experiment, had to come up with a single design solution. This instruction was to encourage problem solvers to eventually bring closure to (i.e., select a concept), and potentially, iterate upon their final design. Each time a constraint was introduced during problem-solving, they were instructed to continue their sketches on a new sheet of paper that was provided to them. By the end of the experiment, each team had three sheets of sketches, with the last sheet containing their final solution. Using audio recorders, both managed and unmanaged teams' discourse was also collected throughout the experiment. At the conclusion of the experiment, one team member from these two conditions was also asked to provide a brief, verbal description of the group's final design. The individuals and unmanaged teams were under the supervision of passive experimenters, who only read the experiment instructions, monitored time, and provided the experimental materials (sheets of sketching paper and constraints to the problem). In the managed condition, these same procedural responsibilities were assigned to the process managers.

2.3. Data Analysis

2.3.1 Ideation Metrics

To compare the performance between the collaborative and nominal teams, their final designs from the end of the experiment are evaluated. According to Shah et al., accurate measures of ideation effectiveness, as well as how problem solvers explore within the design space, can be seen in the novelty, quantity, quality, and variety of their design output [63]. However, because the participants in this experiment are told to only come up with one final design solution, and both quantity and variety are a function of the number of ideas generated per individual, only the novelty and quality are used in this analysis. In practice, products need to work, but they do not need to be novel; thus, an effective solution is required while a novel solution is only desired. Thus, novelty, or uniqueness, of a solution is secondary to quality, given that low-frequency solutions are not necessarily good solutions, they are simply rare. Therefore, the teams' final designs are evaluated based on these two ideation metrics, with quality taking precedence.

2.3.2 Design Quality

The quality of a design refers to its technical feasibility and how well a particular solution satisfies the engineering specifications of the problem. Two mechanical engineering graduate students at

Carnegie Mellon University, with extensive experience in design theory and methodology, rated the quality of the final designs based on how well a solution satisfies the constraints of the design prompt. Ratings are coded into the three distinct categories shown in Table 2.1.

TABLE 2.1: QUALITY RATING CATEGORIES

Rating Category	Category Description
0	*The design violates a constraint/function of the design problem.*
1	*The design poorly satisfies the constraints/functions of the design problem.*
2	*The design effectively satisfies all constraints/functions of the design problem.*

This three-point rating scale was chosen to minimize the subjectivity inherent in larger rating scales, at the expense of losing resolution in the scores, while enabling a judgement of excellent, acceptable, and poor. The raters were provided with brief instructions, the problem statement and constraints used in the study, and the corresponding category descriptions from Table 1, but did not receive any further training on how to score the designs. Because each design is scored by both raters, each design has two associated quality values.

2.3.3 Design Novelty

The *novelty* of a design solution defines how unconventional or unusual an idea is compared to other designs within the set of designs generated within the experiment. This definition represents the breadth of search through the design space. The authors prefer the term *uniqueness* as these designs may not have value or be truly rare beyond this study; however, for consistency with the design literature and by assimilation of the Shah et al. metric, the term novelty is used in this chapter. A posteriori evaluation of novelty is computed, in which comparisons are made relative to the ideas generated between participants during the experiment. Therefore, the design space is populated with only those designs from the experiment. The novelty is calculated by looking at the different sub-functions of a design and identifying what mechanism is used to satisfy that sub-function (Equation 1 and Equation 2). To meet all the engineering requirements of the problem statement from the study, an adequate solution can be broken down into five distinct sub-functions that must be satisfied by the design (each with an associated weight f). These include: an energy conversion (human/natural to mechanical) mechanism (f_1), transportation of the peanuts through or along the device (f_2), crushing/de-shelling of peanut shells (f_3), sorting of the intact peanuts from their crushed shells (f_4), and the collection of the harvested peanuts (f_5). Shah et al., formulate the posteriori computation of the novelty, N, of a team's design as:

$$N = \sum_{j=1}^{n} f_j \sum_{i=1}^{m} \frac{T_{ji} - C_{ji}}{T_{ji}} \times 10 \times p_i, \quad (1)$$

where T_{ji} is the total number of ideas generated for sub-function j, C_{ji} is the count of the current solution for function j, f_j is the weight assigned to function j, signifying its importance, n is the total number of sub-functions (in this case, $n = 5$), m is the particular stage of the design process, and p_i is the weight associated with that stage. Because the focus was only on one phase of the design process, the ideation phase, the above equation reduces to:

$$N = \sum_{j=1}^{n} f_j \frac{T_{j1} - C_{j1}}{T_{j1}} \text{ x } 10. \quad (2)$$

The multiplication of the constant 10 is to normalize the novelty scores on a scale from 0 to 10. The weights, f_j, for each of the five sub-functions are chosen based on the experimenter's estimated importance of the sub-functions' contribution to the overall design problem. Accordingly, the chosen weights for the sub-functions to compute the overall novelty scores are $f_j = \{0.25, 0.10, 0.35, 0.20, 0.10\}$, where $1 \le j \le 5$.

2.3.4 Generation of nominal teams

To compare the performance of both the managed and unmanaged teams with the individual problem solvers, *nominal teams* are generated. Nominal teams are teams composed of individual problem solvers who did not interact during the study but are artificially placed together to form a team. Then, the cumulative best solution amongst the individuals' solutions is considered the product of the entire team. In this experiment, five individuals are placed together so that they can be compared with the other team conditions. The nominal teams are computationally formed with a random number generator, under the following constraints: (1) all 23 individual problem solvers need to be placed on a team, (2) all but six individual problem solvers needed to be assigned to two different teams (assuming eight total teams with five students on each), and (3) every team had to consist of five unique individual problem solvers. This algorithm is repeated until eight valid nominal teams are generated. This random assignment removes any bias that could be formed because all members were equally likely of being selected for a team. In this experiment, the solution with the highest quality was chosen as the team solution. The intuition behind this is that, in practice, when individuals on a team select between possible designs (or their supervisor selects amongst candidate solutions), the design with the best quality would likely be the one chosen. If two designs are comparable with respect to their quality, then the design with the higher novelty is selected as the preferred idea. With this method, it is possible that a significantly superior design (in terms of quality and novelty) could appear on multiple teams. However, at most, an individual could only be placed on two different teams, and therefore, that design could only be considered at most twice.

2.3.5 Team cohesion

In addition to measuring the solution output from the design teams, the teams' dynamics are investigated. The audio recordings from the experiment are used for this purpose to measure the similarity of a teams' discourse, i.e., the team cohesion, and how it evolves throughout problem solving. Latent semantic analysis (LSA) has been shown to quantify this level of semantic convergence in language-based communication between members in design teams [64]. The degree of a team's cohesion has been shown to be directly proportional to their cognitive representation of the design problem and is accepted as an accurate measure of a team's design performance [9]. Because the individual problem solvers that comprised the nominal teams were not audio recorded during the experiment, only the collaborative teams can be compared with one another with this measure. Therefore, this analysis will facilitate in addressing only the second research goal of this chapter: the impact of process management interventions on engineering teams.

As background, LSA uses singular value decomposition (SVD) to determine the underlying patterns within text (here speech) by projecting the co-occurrence of words across documents (here speakers) to a lower rank (dimensional) approximation of the semantic space [65]. The SVD of a matrix, X, is shown in Equation 3:

$$X = USV^T . \qquad (3)$$

For this experiment, matrix X is an [n x m] occurrence matrix with n number of words and m speakers, U is an [m x r] concept vector matrix with rank r, S is an [r x r] singular values matrix, and V is an [n x r] speaker matrix, containing speaker vectors. By mapping the co-occurrence of words into this new r-dimensional semantic space, the cosine similarity between speaker vectors can be computed to determine how closely related speakers' semantic coherence is. The overall semantic coherence of a design team is then computed by taking the average of all pairwise cosine similarities between members of a team.

The pipeline for post-processing the audio files and running LSA to compute the semantic coherence is depicted in Figure 2.4. The text corpus for LSA is generated from the audio files recorded during the study. First, each audio file is transcribed, via an outsourced vendor, into a single transcript for each design team (step 1). Each transcript is then checked and verified for proper speaker identification. The speakers are then segmented out of the full experiment transcripts to obtain a text document representing each speaker on a team (step 2). For the managed teams, the manager document is excluded from the analysis because they are not included when computing the semantic similarity of a design team. The co-occurrence matrix, *X*, is then generated for each team (step 3) and weighted using the global, log-entropy (Equation 4) across speaker documents of a team [56]:

$$W = 1 + \sum_{j} \frac{p_{ij} \log_2(p_{ij})}{\log_2(m)}. \quad (4)$$

P_{ij} is the ratio of the frequency of each term in a document to the frequency of each term over all documents. This global entropy weighting is used to dampen the effects of large differences in the frequency of words (i.e., gives less weight to terms that occur frequently or are commonly used, and more weight to terms that are less frequently used). After weighting the occurrence matrix, SVD is performed (step 4) to reduce the dimensionality of the matrix and speaker vectors are then extracted. The cosine-similarity is then computed between all pairs of speakers (step 5) and the average of those comparisons is taken as the team's semantic similarity (step 6).

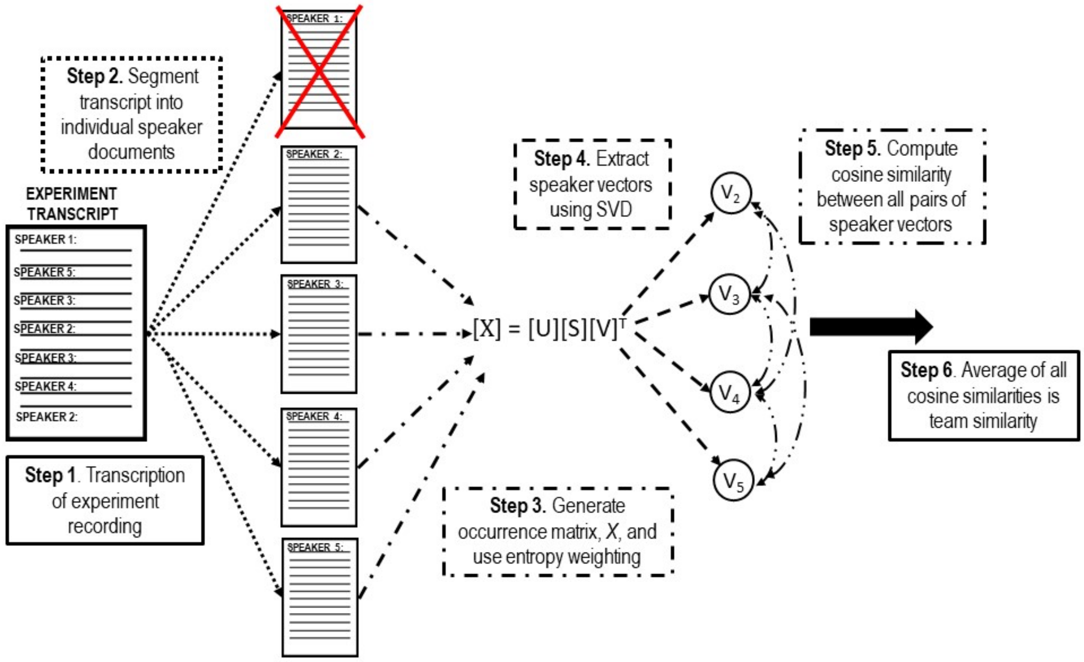


FIGURE 2.4: REPRESENTATION OF THE PIPELINE FOR RUNNING LATENT SEMANTIC ANALYSIS (LSA) ON THE AUDIO TRANSCRIPTIONS

2.3.6 Manager Interventions

The transcripts from the experiment are utilized, in conjunction with the post-study interviews, to designate the timing and type of managerial interventions. This exploratory analysis will provide insight into some of the managerial techniques and strategies utilized by the managers at different points in the design generating process. In order to do this, comparisons are made across three, equal 10-minute intervals of the experiment, which are delineated based off of when the constraints were added to the problem (i.e., before the first constraint was given, between the first and second constraints, and after the second constraint).

2.4. Results

At the conclusion of the study, participants were instructed to have their single, final design circled. For both the managed and unmanaged teams, this solution was collaboratively chosen by all members on a team and therefore is representative of an entire team's effort. These elected designs are extracted from the last sheet of sketches and utilized in the assessment of the previously discussed ideation metrics. A sample set of final design solutions from each team condition is shown in Figure 2.5.

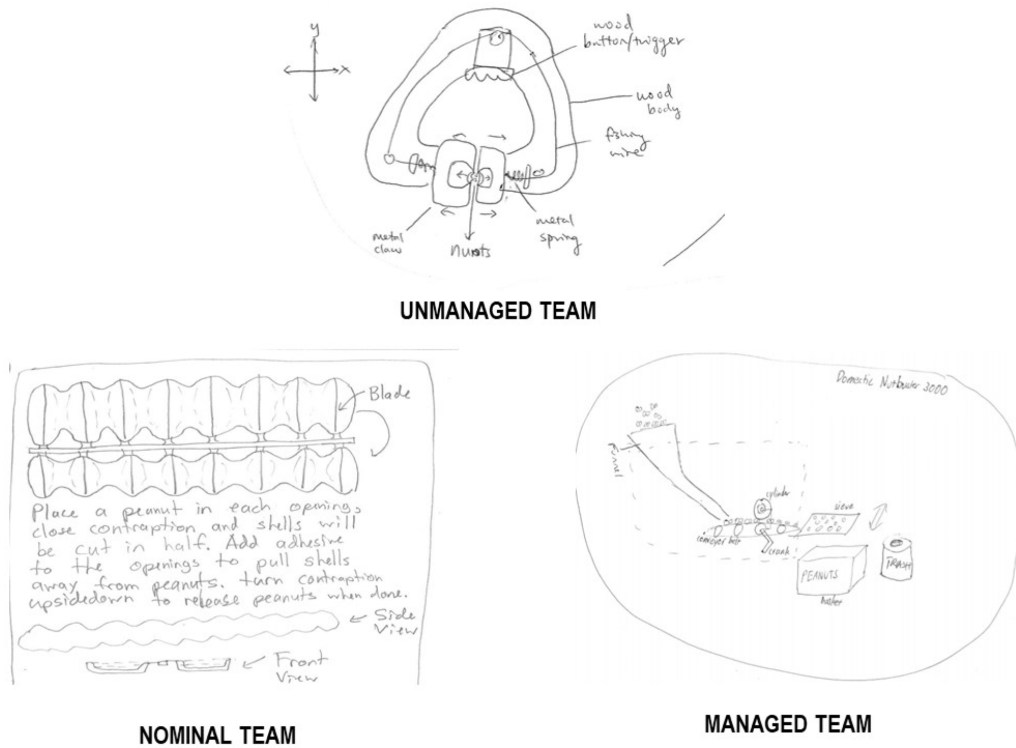


FIGURE 2.5: EXAMPLE FINAL DESIGN SOLUTIONS FROM AN UNMANAGED (TOP), A NOMINAL (BOTTOM-LEFT) AND A MANAGED (BOTTOM-RIGHT) ENGINEERING DESIGN TEAM

2.4.1 Design Quality

As discussed in the analysis section, each design is rated for quality by two mechanical engineering graduate students as either 0, 1, or 2, depending on how well the solution satisfies the problem statement and design constraints. Figure 2.6 shows the frequency of designs binned into each of the categories, with the x-axis representing the three quality categories and the y-axis being the frequency of designs in each respective bin. Due to some subjectivity in this assessment, each design is scored by both graduate student raters to maintain consistency and objectivity in the scoring. To gain a sense of the degree to which the raters are consistent with one another, the intraclass correlation coefficient (*ICC*) is computed across all designs (*ICC* = 0.58), which is an acceptable correlation [67].

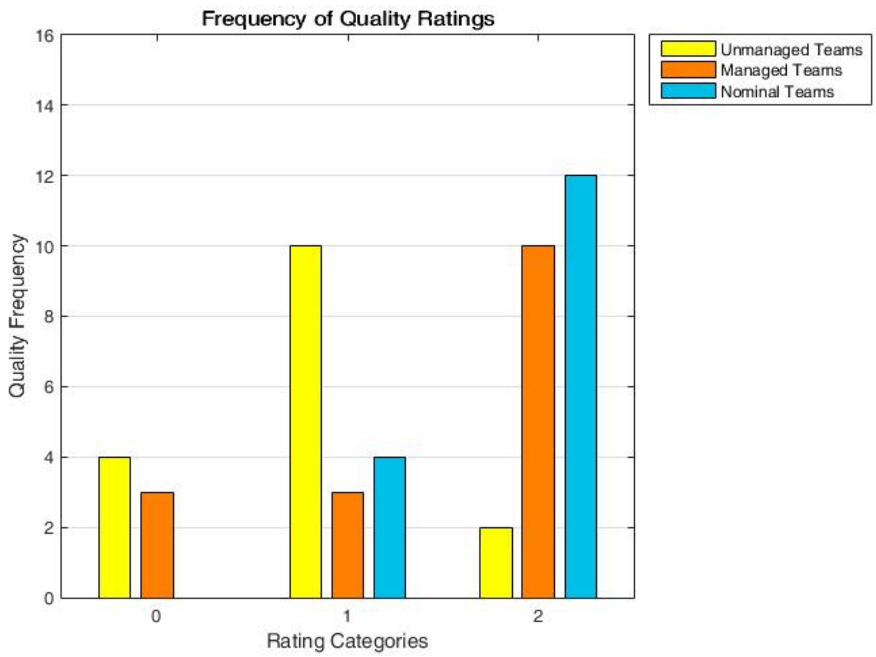


FIGURE 2.6: FREQUENCY OF QUALITY RATINGS FOR UNMANAGED TEAMS, MANAGED TEAMS, AND NOMINAL TEAMS

Because the data is ordinal and not normally distributed, a Mann-Whitney U-test, the non-parametric version of the t-test, is run. The test reveals that both the managed teams ($p < 0.014$, *effect* $r = 0.39$) and nominal teams ($p < 0.001$, *effect* $r = 0.64$) generate designs of significantly higher quality than the unmanaged teams. The nominal teams' designs tend to be of slightly higher quality than those of the managed teams ($p < 0.15$, *effect* $r = 0.18$). Thus, the first research goal of whether the performance of individuals surpasses teams in the domain of conceptual engineering design, can now be addressed. Indeed, the higher quality designs generated by the nominal teams, compared to unmanaged teams, are indicative of better overall performance.

2.4.2 Design Novelty

The final designs are evaluated using a posteriori measure of novelty (i.e., *uniqueness*), which describes how uniquely each of the five sub-functions are satisfied in a given solution, compared to the other solutions within the experimental design space. Table 2.2 provides a summary for each of the five sub-functions and the count (C_j) of each mechanism used to satisfy them. Summing these counts for each sub-function identifies how many teams in total, T_i, satisfy a particular sub-function, where i ranges from 0 to 24, for 24 total design teams (eight in each of the three conditions).

TABLE 2.2: THE COUNT OF DIFFERENT MECHANISMS USED IN DESIGN TEAMS' SOLUTIONS

Novelty scores, a posterior mechanism counting

Energy Conversion ($f_1 = 0.25$)	C_{11}	Transportation ($f_2 = 0.10$)	C_{21}	Crushing ($f_3 = 0.35$)	C_{31}	Sorting ($f_4 = 0.20$)	C_{41}	Collection ($f_5 = 0.10$)	C_{51}
Bike Pedal	3	Wind	1	Rollers	5	Sieve	8	Bucket/Basket/box	11
Crank	8	Conveyor	4	Lead Screw	1	Fan	2		
Lever	1	Inclined Slope	4	Hammer	1	Mesh	2		
Pivot	1	Gravity	3	Blade	1	Mold	1		
Spring	2	Funnel/Cylinder	4	Cylinder	2	Wedge	1		
Hinge	3	Rollers	2	Scissors	1	Spikes	1		
Human/Manual	4			Friction	4				
				Crusher	7				
				Claw	1				
				Human force	1				
T_{11} =	22	T_{21} =	18	T_{31} =	24	T_{41} =	15	T_{51} =	11

In reference to the total counts, T_j, the only sub-function that was satisfied by all 24 teams is the crushing function, f_3. The energy conversion sub-mechanism is the next highest, followed by the transportation function. This result is consistent with the chosen weights, f_j, that were discussed previously. For example, the sub-functions given the higher weights end up being fulfilled by more design teams, signifying their lack of ambiguity from the problem statement and importance to the overall design task. Substituting the counts (C_j) and totals *(T_j)* from Table 2.2 into Equation 2, the novelty scores can be calculated for each team. Figure 7 shows that the managed teams have a significantly higher measure of novelty than both the nominal ($p < 0.01$, $d = 1.35$) and unmanaged teams ($p < 0.007$, $d = 1.41$). There is no significant difference between novelty of unmanaged and nominal teams' designs.

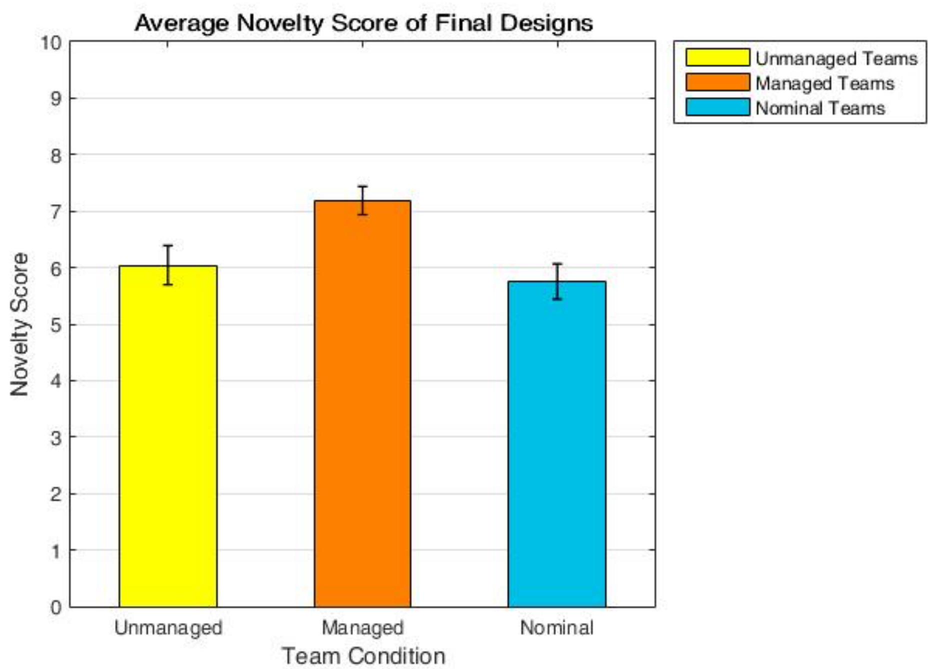


FIGURE 2.7: AVERAGE NOVELTY SCORES FOR UNMANAGED, MANAGED, AND NOMINAL TEAMS. (ERROR BARS SHOW ±1 S.E.)

To determine whether the experimenter's chosen values of the weights impact the results of the novelty, a sensitivity analysis is performed on the novelty formulation. The functions are first weighted equally ($f_j = 0.2$), for each sub-function (equivalent to no weighting at all). Re-computing the novelty, the results are identical to that originally found, with the managed teams exhibiting higher novelty than both unmanaged and nominal teams; the latter two having no significant difference. Furthermore, to determine the sensitivity of the novelty score to the value of the individual weights, each weight is successively perturbed ±10% from its original value and the novelty is re-determined. When one weight is perturbed, the four remaining weights are readjusted so that the total weight sum remains at 1 ($\sum_{j=1}^{5} f_j = 1$). After recalculating the novelty with each new combination of weights, in every case, the managed team's novelty score remains highest, and the unmanaged and nominal teams lower and similar. This analysis confirms that the scores are not sensitive to the originally chosen weights (which were, f_j = [0.25, 0.10, 0.35, 0.20, 0.10], for $1 \leq j \leq 5$), and this weighted set will be the one used for purposes of this chapter.

2.4.3 Team Cohesion

Now that nominal teams have been shown to produce higher quality solutions than unmanaged teams, the second research goal can be addressed, namely, whether a manager is able to mitigate the costs associated with collaborative engineering design teams. From Figure 2.6 and Figure 2.7, the teams under the guidance of a manager generated solutions that were both significantly more novel and of slightly higher quality than the unmanaged teams. Post-processing the audio transcriptions according to the pipeline outlined in Figure 2.4 and running LSA on these transcriptions (Figure 2.8), the managed teams' discourse exhibit higher semantic similarity over all three experimental intervals ($p = 0.016, p = 0.026, p = 0.015,$ respectively). This result further supports the claim that the process managers can mitigate some of the performance costs of engineering design teams. It is also interesting to note that both team conditions exhibit a similar and consistent trend, showing a decrease in cohesion between the first and second intervals, followed by a small increase in the third interval.

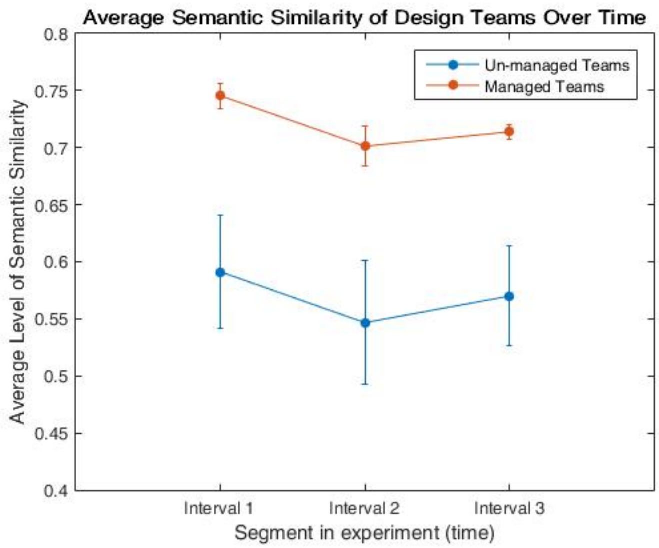


FIGURE 2.8: LATENT SEMATIC ANALYSIS ON AUDIO RECORDINGS OF DESIGN TEAMS. (ERROR BARS SHOW ±1 S.E.)

2.4.4 Process Management Interventions

Since process managers are shown to be beneficial, a preliminary analysis into some of these constructive managerial strategies can be done by examining the frequency and types of interventions throughout the experiment. A general summary of all the manager interventions is shown in Figure 2.9. As a recap, the allowable types of manager interventions are the *keywords*, *design components*, and *design strategies* that were depicted in the *manager bank* (Figure 2.2).

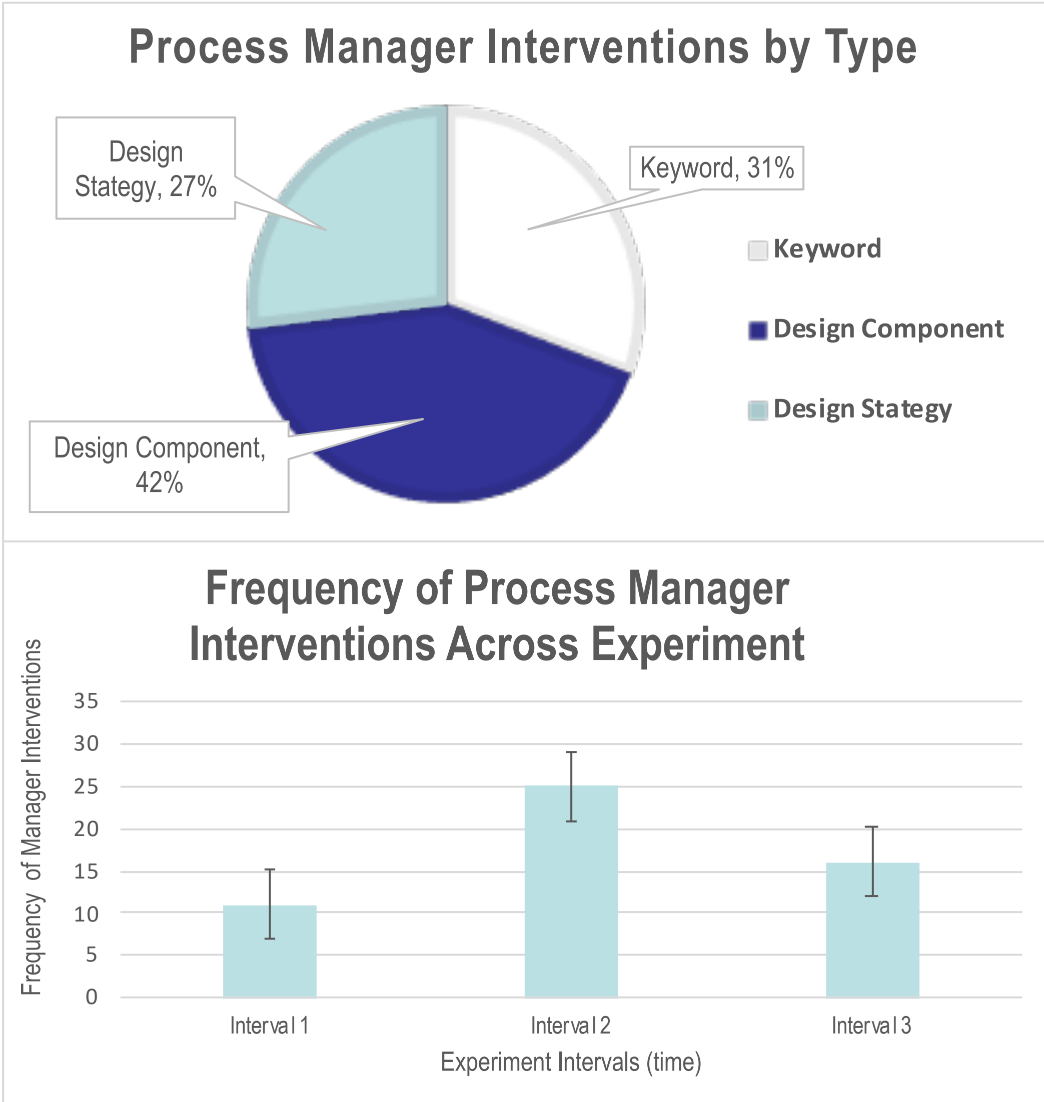


FIGURE 2.9: OVERVIEW OF MANAGER INTERVENTION USAGE OVER THE ENTIRE EXPERIMENT BY TYPE (TOP) AND FREQUENCY (BOTTOM).

The managers were able to interact with those prescribed items but could not otherwise speak with the design teams or help in directly solving the problem. In total, the managers intervene 52 times with 11 interventions in the first interval, 25 interventions in the second, and 16 during the final interval. Of the 52 interventions, 42% are design components, 31% are keywords, and 27% are design strategies. In order to understand the evolution of managerial strategy over the different problem-solving phases, Figure 2.10 depicts the temporal evolution of interventions. The percentages shown are relative to the number of interventions per interval. For example, 45% of the design strategy interventions in the first interval equate to 5 distinct design strategy interventions. Overall, the design strategies and keywords comprise the majority of interventions in the first segment of the experiment, with a more equal distribution among all three types in the middle. By the end of the experiment, the largest proportion of managerial interventions are design components as the final design ideas are instantiated.

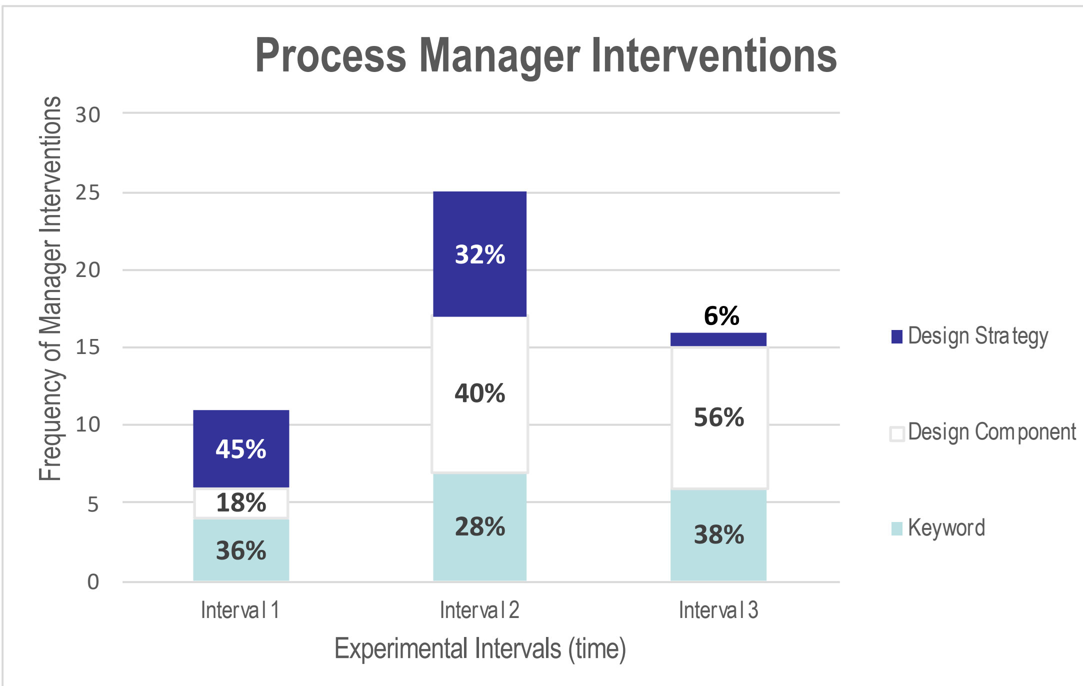


FIGURE 2.10: THE EVOLUTION OF MANAGERIAL INTERVENTIONS DURING EACH EXPERIMENTAL INTERVAL

The post-study interviews with the graduate student managers are also evaluated to uncover the underlying motivations. This is useful for determining the rationale, and consequently, the feedback the managers used in deciding the most opportune times of intervening. After analyzing the interviews and sorting the interventions into common themes, four salient motivations emerged: assist the team in generating new ideas, help the team promote their current thought, remind the team of the engineering design requirements, and improve the team dynamics. To get a better sense of how these motivations are binned, consider the ones shown in Table 2.3.

TABLE 2.3: EXAMPLE MOTIVATIONS FROM POST-STUDY INTERVIEWS WITH MANAGERS

Managerial Motivations
Help Generate New Ideas: "*There was no structure to their thought process, and there was no direction.*"
Promote Current Thought: "*They were thinking through a bunch of human interfaces and they hadn't really considered a crank, and I thought a crank would be a useful extension to the one's they had considered"*
Remind of Engineering Design Requirements: "*They had only focused on crushing the shell at that point and not thought about how to actually get the center of the peanut out of the shell*"
Improve Team Dynamics: *"There was really one person leading it and I wanted everyone to have something to do and have them take different tasks."*

For example, asking a manager why they intervened with the design strategy, "*Can you think of similar ideas that already exist,*" they responded with the first statement from Table 3: "*There was no structure to their thought process, and there was no direction*." This intervention is characterized as "help generate new ideas," because the engineering design team was not focused and having difficulty deciding how to approach the problem and brainstorm initial ideas. Each of the 52 manager interventions is analyzed and categorized in this way, by associating with it an underlying motivation.

Similar to the intervention types, these motivations are examined across the three experimental intervals. The progression of managerial motivations is shown in Figure 2.11. The percentages are, again, relative to the number of interventions per interval. For example, 18% of the "reminder of engineering design requirements," in the first interval equates to 2 distinct interventions. One central trend captured in Figure 2.11 is the steady increase of "reminder of engineering design requirements". This evolution suggests that assisting design teams in focusing on the constraints of the problem and attributes of their designs is an increasing objective of the process managers throughout the entire problem-solving process. This is true, particularly near the end of the experiment, when 75% of the interventions are categorized by this motive.

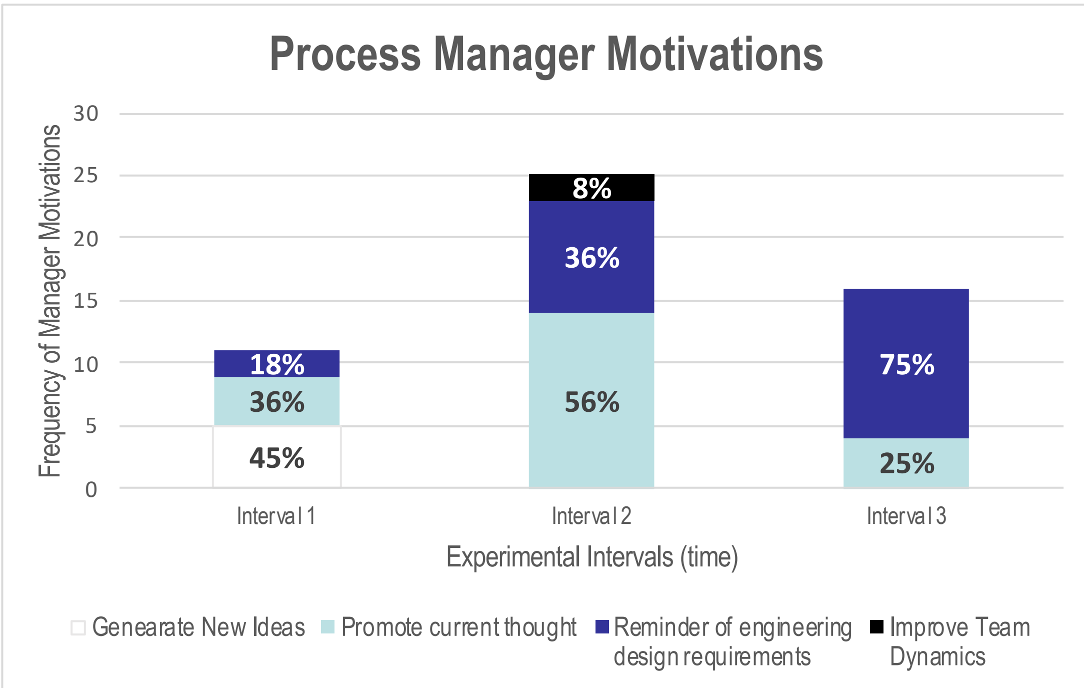


FIGURE 2.11: THE EVOLUTION OF MANAGERIAL MOTIVATIONS DURING EACH EXPERIMENTAL INTERVAL

2.4.5 A Possible Mechanism: Verbalization

Even though process management, in this experiment, is shown to be beneficial to design teams, individual problem solvers still perform marginally better in terms of their collective solution quality. One of the main differences between individual and collaborative team problem solving is the fact that teams need to verbalize to communicate ideas with one another. Sio et al. previously investigated the effect of this communicative process on RAT problems and found that thinking-aloud nominal groups were impaired in comparison to nominal groups who solved the problem quietly [41]. Perhaps this could be one of the main cognitive hindrances and costs of group problem solving, and one that the process management in this experiment is not able to mitigate.

Consequently, a supplemental condition is run to examine the effects of verbalization during problem solving and to see if individuals who verbalize perform worse than individuals who do not verbalize during problem solving. The experimental architecture and logistics are identical to the previous nominal team group, except that participants are also told to think aloud during problem solving so that their thoughts on conceptualization could be followed. The time was carefully monitored, and if an individual went 10 seconds or longer without expressing their thoughts, the experimenter reminded them to continue verbalizing. In total, 22 additional freshman engineers participated in this condition of the experiment, and nominal teams were generated following the same algorithm discussed earlier. The designs were evaluated in the same way, and the results are shown in Figure 2.12.

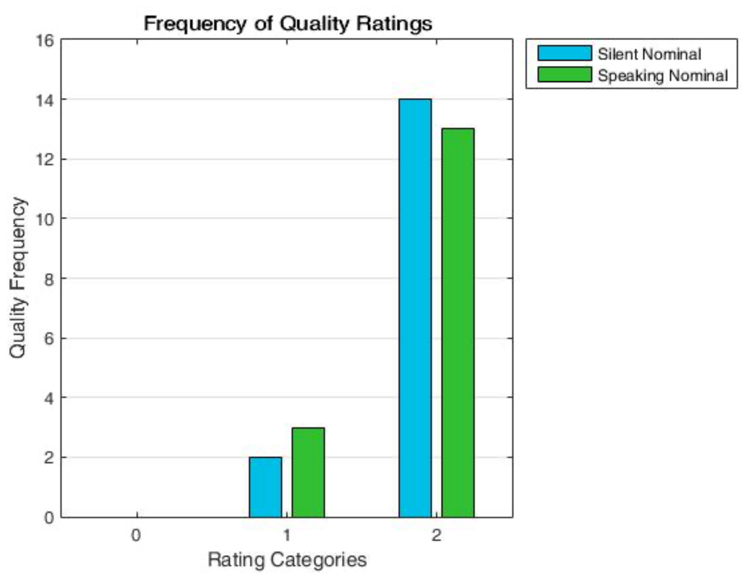


FIGURE 2.12: FREQUENCY OF QUALITY RATINGS BETWEEN SILENT NOMINAL TEAMS AND SPEAKING NOMINAL TEAMS

While the differences are in the right direction (silent outperforming speaking), there is no significant difference in the quality of design solutions between individuals who problem solve silently and those who concurrently verbalize. The inconsistency of this result with the results from the RAT problem task are likely due to a fundamental difference in the tasks: in particular the difference in the cognitive processing of spatial and verbal tasks. As shown by Brooks [68], tasks that are verbal in nature, such as the RAT problems, will be hindered by concurrently performing a task that is also verbal, such as thinking out loud, while spatial tasks, in his case visualizing action on a block letter, will not be so affected. Conceptual design is more of a spatial oriented problem [69]; thus, verbalization should not, and turns out does not, act as a cognitive barrier while concurrently solving the design problem, due to this difference in the form of processing. Thus, verbalization does not seem to be a direct cognitive barrier to this type of team problem solving, and additional work must be done to try to identify what features of collaborative teams put them in a disadvantage to individuals in conceptual engineering design.

2.5 Discussion

2.5.1 Effects and general strategies of process management

As introduced at the beginning of the chapter, this research presents two primary goals. The first is to determine whether the performance of individuals (i.e. *nominal teams*) is superior to that of collaborative team problem solving in conceptual engineering design, a domain which shares the free-flowing, creative aspects of brainstorming, but possess the goal-oriented, and structured characteristics of configuration design. The results from this study support this claim, showing that unmanaged engineering design teams are not as proficient as individual problem solvers. Nominal teams generate design solutions of significantly higher quality. Exploiting this result, the second aim of this chapter is to examine whether the underperformance of these design teams can be mitigated with resources allocated to the management of the design process via a human manager. The effects of a manager to the problem-solving process show that these teams do benefit, even when resources are taken away from directly solving the problem. The managed teams, with one less member and therefore fewer direct problem-solving resources, still perform better, with significantly higher quality design solutions than unmanaged teams. These teams also produce final designs that are of greater novelty (*uniqueness* across the set generated) - and are more cohesive throughout the experiment, as measured through the semantic similarity of their verbal discourse. Thus, the findings from this study show that real-time management of the design process closes the performance gap between individual problem solvers and teams. However, the nominal teams still produce designs of marginally higher quality than the teams that are managed, suggesting that there are still some deficiencies in design teams that managers are not able to mitigate.

Given that the process management of design teams is shown to be beneficial, analyses can now focus on uncovering some of the constructive intervention strategies. Generally, the managed teams satisfy a greater number of the five identified sub-functions of the design problem, and in more unconventional ways than teams in either of the other conditions, resulting in higher novelty of their final designs. Of the 12 mechanisms, in Table 2.2, that are used by only a single design team (i.e., the most unique mechanisms), half of those are from managed teams. The set of these mechanisms includes lever, pivot, wind, lead screw, hammer, and scissors. Also, as depicted in Figure 2.11, reminding teams of the engineering design requirements is the only motivation that considerably increases throughout the entire experiment. This suggests that the managers play a major role in getting their teams to think about all requirements of the design.

Another significant trend in the manager interactions is the usage of the design component stimuli, which increases from 18% of the manager interventions in the first interval to 56% by the final segment. Because the design components are all specific mechanical elements, this trend is also consistent with the managers reminding their teams to consider the functional aspects of the design. For example, providing a team with the conveyor stimuli could prompt a design team to focus on the transportation of the peanuts through the device. Similarly, providing a team with the sieve component helps teams to concentrate on sorting the peanuts from their shells. One could argue that the managed teams become fixated and directly use the components that are provided to them during the interventions. Because the managers are trained not to speak when intervening with the design components, some of the participants may have perceived these particular interventions as additional requirements to the design problem. Even so, this does not undermine the fact that, overall, the managed teams' designs are more novel. Also, as shown in Table 2.2, out of the most novel mechanisms (those with a count of 1), only the hammer design component is taken directly from the manager bank. As such, fixation on the specifics of manager suggestions is an unlikely implication from the interventions.

Collectively examining the types and frequency of interventions over all three experimental intervals (Figure 2.10), coupled with the compilation of motivations (Figure 2.11), yields valuable insights into the managerial strategy and how it evolves throughout problem solving. The design strategies comprise the largest proportion of interventions in the first interval of the experiment (46%). This result suggests that toward the beginning of ideation, the managers want their teams to follow a more encompassing and exploratory search of the design space. Specifically, "*Can you think of similar ideas that already exist,*" and "*Can you clearly identify the assumptions, constraints, and goals of the problem*" are the two significant design strategies suggested by managers in the early stages of the experiment. Similarly, in the first 10 minutes of the study, 46% of the interventions are motivated toward helping their design teams generate new ideas. This result agrees with the types of interventions, because in the early stages of brainstorming and ideation, effective exploration of the design space is important. As problem solving proceeded, the most frequent interventions focused on design components, with 56% of the interventions in the final experimental interval. The increasing implementation of design components, particularly near the end of the experiment, indicates that the managers try to get their teams to home in on a specific region, or subset of solutions, within the design space, to instantiate a final effective solution. Overall, it seems that the predominant tendency in managerial behavior is to push their teams to follow an exploratory-to-convergent search of the design space, which has been shown to be an effective strategy for concept generation and creativity in design [70,71], and thinking in design teams [72]. This exploratory-to-convergent funneling of design team efforts is also mirrored by the manager motivations, with 75% of the interventions in the last interval being prompted by reminding teams of the engineering design requirements. This rationale is critical in final design convergence and selection when improvements and iterations must be done to achieve all the engineering design specifications.

2.5.2 Limitations and future work

It should be acknowledged that the results from this chapter are prognostic as opposed to purely diagnostic. The underlying reason of why nominal team performance is superior is still an open question and requires future work to make any definitive conclusions. Perhaps some of the recognized deficient team characteristics from other literature, such as social loafing, may be at play here [73]. In observing the interaction of teams during the study, for example, it appeared that some members did not participate as much as others; thereby indicating that social loafing could be the cause. Also of note, both the teams and individuals evolve what they perceive to be their best design. However, in the team condition, biases in decision-making and other team characteristics may possibly have influenced and inhibited the selection of their actual best design [74,75]. By the current definition of a nominal team, the best design amongst the collection of individuals is automatically chosen for them. This discrepancy in selection could have affected the difference in collaborative and nominal team performance, though this, and other theories that might account for the inferior performance of the design teams, are left for future investigation. Nonetheless, supplying process management did mitigate the negative effects of working in teams.

Moreover, the team structure in the experiment is both static and free; the structure does not change throughout problem solving and all members are free to communicate with all other members without any limitations. This structure could also have negatively affected and led to the inferior performance of design teams. Additional work on modeling different team structures which have shown to be more effective [76,77], such as hybrid teams, where individuals initially work on the problem separately before coming together to collaborate [78], could be an interesting direction.

To gain a deeper understanding into managerial behavior, future work can focus on a more refined and in-depth analysis of the manager interventions to extract specific modalities and timing that are most beneficial to design teams. The evolution of each teams' designs may also be tracked through the experiment to see how designs are affected by these interventions and whether managers help teams overcome some of the stumbling blocks associated with problem solving. The process managers were also recruited with similar skillsets and constrained in the types of interventions that could be used; this was purposefully done to equalize management capabilities and reduce variability. Although fixation was not seen as directly impacting the teams (as the managed teams had more novel ideas), it would be interesting to see how enlarging the manager bank to create a more expansive example set, as well as changing the managers' expertise, impact the design process [53]. Different evaluation metrics could also be studied, such as creativity, to determine how feedback influences such [79].

Because this study utilizes freshman engineering students from an academic institution and limits the problem-solving process to a short time frame, there are still unanswered questions about the generalization of the results outside the lab setting. Thus, future work can also consider if the results and observations from this current chapter extend to a larger scale and apply to engineering design teams in practice.

2.6 Summary

This chapter presents an empirical study to investigate individual versus group performance in the domain of conceptual engineering design. Accordingly, a behavioral experiment is run, in which freshman engineering students solve a conceptual design problem individually or in a collaborative team setting. Corroborating previous findings, those who work on the problem individually, when the cumulative best solution is selected, end up generating solutions of much higher quality than those whom work in unmanaged teams. An attempt to mitigate some of the deficiencies associated with design teams is then made by introducing a third condition, where partial resources are taken away from problem solving and reallocated to the process management of the team. Teams that are guided by this process management perform nearly as well as individuals, suggesting that, perhaps under the proper direction, teams can become as efficient as individual problem solvers (i.e., *nominal teams*).

After demonstrating the beneficial effect of managing resources applied to engineering design teams, a preliminary analysis into this process management is done. This analysis involves tracking the evolution and motivation of these interventions throughout the experiment. The general pattern emerging is an exploratory-to-convergent managerial strategy. Overall, managers seem to promote a breadth of search within the design space early on in the ideation process, resulting in more novelty and uniqueness in the solutions. Near the end ideation, management is used to help teams think about the engineering specifications and requirements of the design problem, and to refine search, as closure is brought to the process. Furthermore, managers also help their teams maintain cohesion in their thought, as measured by the semantic similarity of the team's discourse.

There could be several different explanations of why the unmanaged teams did not perform better than individuals, and taken alone, these results are not sufficient to provide any complete explanation. However, one of the main differences between team collaboration and independent problem solving is the role of verbalization. Thus, in an attempt to begin answer this question of the possibility that verbalization, acting as a cognitive barrier to problem solving, is also studied. An appendage to the current study is run where participants individually solve the same conceptual design problem, but this time, while simultaneously verbalizing their ideas out loud. Results show that verbalization does not act as a direct deterrent to problem solving, as those individuals who thought aloud generate solutions of nearly equal quality compared to those individuals who problem solved silently. Future work can identify other potential obstacles to team problem solving, as, at least in conceptual design, verbalization does not seem to be a factor.

The empirical results from this chapter expand growing evidence that individuals are more effective than teams in a variety of problem-solving situations, including conceptual design [80].

Ultimately, the hope is to understand why teams are not always maximally proficient, in what types of circumstances they significantly underperform, and what methods are most effective in assisting them. The study presented in this chapter is a step towards uncovering approaches and methods that can build more focused and efficient engineering design teams, which has major implications for how design teams work together, and problem solve in practice.

Chapter 3 : Enabling Automation via a Topic Modeling Approach[2]

3.1 Introduction & Motivation

Design cognition studies investigate how designers think and strategize during design activities [81]. A variety of methodologies exist to analyze these processes of thought including, but not limited to, case studies, protocol analysis, and empirical performance tests [82,83]. Concurrent verbalization, or thinking aloud, also reveals certain characteristics of design cognition, such as the interactions between design problem and design solution [84]. By analyzing and codifying team discourse data, Stempfle, et al., propose a two-process-theory of thinking in design teams [72]. Their work theorizes that four basic cognitive processes encapsulate design thinking, including the operations of generation, exploration, comparison, and selection. More computational approaches have also been utilized to study design team communication, including Latent Semantic Analysis (LSA). LSA has been shown to be effective in modeling mental knowledge representations by analyzing design team communications such as emails, reports, and automating the annotation and tagging processes of team discourse to predict performance [9, 49, 64,85,86]. As the aforementioned works demonstrate, studying design team communication comprises a rich area. More critically, analyzing team interactions and communication amongst designers provides valuable insight into design processes and cognition, as well as predicting the state and effectiveness of problem-solving. Knowledge of this state/cognition of designers can

---

2. The work presented in this chapter is published in:
Gyory, J.T., Kotovsky, K. and Cagan, J., 2021. The Influence of Process Management: Uncovering the Impact of Real-Time Managerial Interventions via a Topic Modeling Approach. *Journal of Mechanical Design*, *143*(11), p.111401

provide feedback into identifying the types of facilitation and mediations teams may require during problem solving.

Apart from design communication, a large body of work in engineering design theory research studies methodologies that facilitate problem-solving effectiveness. These include providing near and far analogies, patents, example solutions, and/or functional structures as inspirational stimuli, among others [50, 53, 87-89]. Along this vein, more recent work furthers these findings by studying the effects of intervening with designs stimuli in near *real time*. Work by Goucher-Lambert, et al., modulates the distance of design stimuli from designers' current design state *during* concept generation [90]. Midway through problem solving, designers provide their current solution to a design problem. Then, a stimulus is computationally provided in real time to support ideation, either near to or far from the current state of the designer. In that work, Latent Semantic Analysis (LSA), a method utilizing the co-occurrence of words and singular value decomposition to represent the semantic space, computes the semantic similarity between current design states and a design space corpus of potential stimuli. Furthermore, work by Gyory, et al., explores the effects of adaptive process management on design problem-solving [80,91]. Process managers guide design teams through problem solving and intervene in *real time*, when deemed appropriate. That work shows that teams under the guidance of process managers significantly outperform teams that are not, in terms of their final design solutions. While both of these works move towards real-time, adaptive interventions during design problem-solving, neither utilize discourse information in an algorithmic way to either monitor design progress or measure stimulus effects on design cognition. This chapter builds upon these studies (in fact, using data from the latter), and utilizes topic modeling to computationally analyze team discourse and detect the effects of real-time, process management.

Topic modeling spans a wide variety of algorithms, applications, and research domains. At the highest level, topic modeling constitutes a text mining process of automatically extracting themes or *topics*, (which are sometimes latent) from corpuses of text [92]. By "latent," occasionally these themes/topics are not directly observable via direct inspection but are distinguishable through more nuanced and underlying similarities within the text. A specific type of method, used in this chapter, probabilistic topic modeling has been shown to be effective in identifying themes in unstructured text corpuses [93]. In all cases, the number of topics, specified by the researcher as an input to the algorithm, are assumed to be distributed across the entire corpus. In this way, the text corpuses can be modeled purely by their distributions over these topics. As an example, Rosen-Zvi, et al., create an author-topic model from a collection of 1,700 conference proceedings and 160,000 abstracts, and illustrate its predictive power by revealing relationships between authors, documents, topics, and words [94]. Similar analyses have been done on other large collections of documents, such as in articles from Science [95].

Topic modeling has emerged as a valuable tool for researchers in the fields of engineering and engineering design theory. In transportation analysis, models are trained on corpuses of journal articles to identify sub-fields and provide a more holistic perspective on the current research landscape. In design research, Ball, et al., use it to model the expertise of members within multi-disciplinary design teams to predict their success and performance as teams in mass collaboration efforts [96]. Further examples in engineering design utilize topics and sentiment analysis to study design spaces [97]. Specifically, a Bisociative Information Network, composed of conceptual similarities between design topics, represents inspiration in idea generation as a conceptual bridge between domain topics [98]. Additionally, in requirements engineering, Bhowmik, et al., leverage topic modeling, augmented with part of speech tagging, to automatically generate requirements from stakeholders' comments to support combinatorial creativity in requirements engineering. Kim's group analyzes product designs and attributions through customers' perspectives through online reviews by using Latent Dirichlet Allocation and other semantic methods [99-101]. While topic modeling has emerged in the field of engineering, thus far, it has not seen much utilization to dynamically study design team cognition under the impact of process manager interventions, which offers promising opportunities to research engineering design teams, as explored in this chapter.

Consequently, this research utilizes a topic modeling framework to model discourse between members of a design team to study the impact of process manager interventions during problem-solving. Given that discourse provides insights into designer cognition and the state of the team over the course of problem solving, the specific goal is to explore and computationally detect the impact of these interventions during the problem-solving process. To this end, Section 3.2.1 of this chapter presents the cognitive design study from which the discourse data was collected in prior work [91], as well as a brief introduction to the field of topic modeling and the topic modeling framework for processing the verbalization data. Section 3.3 follows with results on the overall difference in topic structures between managed and unmanaged design teams, including a dynamic look at these topic structures over time. The second main analysis studies the direct impact of the interventions on team cognition, by focusing on and comparing the team members' discourse immediately prior to and following an intervention. Section 3.4 and 3.5 conclude with a discussion of the results, particularly regarding process management and team discourse for engineering design teams, closing with opportunities for future research regarding the extension to real-time monitoring and intervention for design teams.

3.2 Background

This section provides background for two relevant areas of this chapter: Section 3.2.1 provides an overview of the cognitive study from which the transcript data originates, including the different experimental team conditions, the design prompt, types of design interventions, and data collection. The subsequent sub-section (Section 3.2.2) introduces the topic modeling framework for intervention assessment, discussing the core text analysis method selected as an illustration in this chapter, including the bag-of-words model and Latent Dirichlet Allocation (LDA).

3.2.1 Initial Motivating Study – Problem Solving under Process Management

The verbalization data presented in this chapter was collected from a behavioral study run with undergraduate, engineering students at Carnegie Mellon University. While only a general overview of the experiment methodology will be presented here, a more comprehensive outline is discussed in Gyory, et al., 2019 [91]. Student designers were randomly placed into teams and allowed 30 minutes to solve the following engineering design problem:

Problem Statement:
Design a low-cost and easy to manufacture device that removes the outer shell from a peanut.
Constraint 1: *The device is meant to be utilized in developing countries where electricity may not necessarily be available as a power source.*
Constraint 2: *In addition to the previous constraint, the proposed design must be able to separate a large quantity of peanuts from their shells, while causing minimal damage to the inner peanut.*

The two constraints were dynamically added during problem solving (10 minutes and 20 minutes into the experiment, respectively), to both exacerbate the difficulty of the problem and better emulate a dynamically evolving design task.

The experiment included three distinct team conditions, two of which are relevant to this study: managed teams and unmanaged teams. The managed teams comprised four student designers, all actively engaged in the problem-solving process and under the guidance of a human process manager (experienced, mechanical engineering graduate students). In this condition, the process managers intervened with their design teams, when they deemed it appropriate, in order to facilitate problem solving. The interventions, described briefly next, were standardized across all managers. On the other hand, the unmanaged teams consisted of five student designers, all actively engaged in the problem-solving process and under the direction of a passive experimenter. These passive experimenters could only read instructions, provide the design prompt and constraints, and answer questions prior to the start of the experiment; otherwise, no communication between these facilitators and participants was permitted. The reduction in the number of student designers in the managed condition was meant to equalize the problem-solving resources between the managed (four active participants + manager) and unmanaged (five active participants) design teams.

The process managers in the managed team condition intervened with their design teams to affect the problem-solving process. While free to intervene when they deemed it necessary, the managers could only select interventions from a pre-determined *manager bank*. Their interventions were limited to this set. The process managers were also not allowed to speak with their design teams other than to answer questions related to the experiment. This manager bank contained six design keywords, six design components, and six design strategies, all motivated from pre-existing strategies and techniques for increasing ideation/problem-solving effectiveness [57,59,60,102]. The design keywords and design components were related to the specific design prompt (e.g., "*sieve*" as a design component and "*high throughput*" as a design keyword), while the design strategies were tactics appropriate and generalizable to any design scenario (e.g., "*Are the requirements being met in your current design? Can you go back to a previous idea?*").

An audio recorder collected the design team discourse throughout the experiment. Sixteen transcripts were collected in total, with eight teams in the managed condition and eight teams in the unmanaged condition. The transcripts were transcribed via an outsourced vendor and manually checked for proper speaker identification. In addition to the transcripts, other data collected from the study included design teams' final designs (both a sketched drawing and a two-minute verbal description) and a complete recounting of manager interventions, including the timing and type. These were actively noted by the managers during the experiment.

In the immediate days following the study, the researchers conducted a post-study interview with each of the process managers. These post-study interviews queried the managers on the motivations for and the perceived effects of their interventions ("*What made you intervene with intervention* [X]?", "*What was the effect of your interaction?*"). The most common motivations noted include trying to get all team members to equally contribute to the process, reminding the team of the constraints, requirements, and goals of the problem, pushing teams to focus on a functional topic which they were either far away from our close to and needed an extra push, or to get the team back on track because they strayed completely away from the task [103]. These latter two motivations serve as the overarching inspiration for utilizing topic modeling as the computational approach in this chapter. If the managers felt that teams strayed from appropriate or conducive topics for effective problem solving, can these topic shifts be computationally detected? Thus, this research addresses this question via a topic modeling framework to compare topic structures in the teams' discourse.

3.2.2 Text Analysis – Latent Dirichlet Allocation

While a variety of methods for semantic text analysis and topic modeling exist and can be used within the framework introduced in this chapter, the algorithm chosen for use in this chapter is Latent Dirichlet Allocation (LDA) with a unigram, bag-of-words assumption [104]. As both a generative and probabilistic method, LDA models a corpus of documents as a collection of underlying topics. No knowledge of these topics necessarily exists a priori, as the training procedure tests across varying values. LDA then generates topics from the distribution of words and documents in the corpus and probabilistically determines the fit with each number of topics. The researcher determines the number of topics and topic model with the best fit to the data and research goals, as presented in this chapter in Section 3.1. Other methods for topic modeling include Latent Semantic Indexing (LSI), probabilistic latent semantic indexing (pLSI), non-negative matrix factorization (NMF), and Term Frequency – Inverse Document Frequency (TF-IDF) [105-107]. However, LDA overcomes some of the shortcomings of these precursor methods. Furthermore, LDA has been widely used on corpuses of many sizes, from micro-tweets, tweets, and micro-blogs, up to complete journal repositories [108-110], and results in more descriptive output, namely the actual topics that can be interpretable, and more differentiation to allow for potential explainability into the impact of the manager interventions. Bayesian probabilistic models, similar to LDA, have also been applied to discourse analysis [111,112]. Consequently, LDA was chosen for utilization in this chapter.

LDA assumes a bag-of-words model representation of the text corpus, one of the most commonly utilized methods in text analysis, information retrieval, and topic modeling [113]. Bag-of-words captures the frequency of terms or words in each document and across the entire corpus of documents. The model in this chapter treats each individual word as its own feature (i.e., unigram model), though features can be represented by a sequence of n words through an n-gram model which retains information about word ordering [114]. The bag-of-words model assumes that the order of which these features appear in the text does not matter. In text analysis, a *word* represents the most basic unit of vocabulary, and quantified by a basis unit vector: $\{1, \ldots., V\}$. In this case, a word signifies a unique utterance by a team member. It is from the co-occurrence of these words that underlying semantic similarities and topics can be inferred. *Topics* are defined as probability distributions over the set of words. *Documents* represent sequences of the words in the vocabulary, as denoted by $\mathbf{w} = (w_1, w_2,, w_3, \ldots, w_N,)$ for N words in each document, and the documents are generated by random mixtures over these topics, with topic mixtures, $(\theta_1, \ldots, \theta_M)$. The entire *corpus*, D, defines the entire collection of these M documents, $D = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots.. \mathbf{w}_M\}$.

For LDA's generative process, a Dirichlet distribution is used to represent the topic-word distributions across the documents, given by Equation 5:

$$p(\theta|\alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^{K} \theta_i^{\alpha_i - 1}, \qquad (5)$$

with K topics and concentration parameter, α. The multivariate Beta function, B, normalizes this probability distribution. Then for each of the N words in each document, a topic is chosen, z_n, from a multinominal distribution with parameter, θ, and a word, w_n, is chosen from a multinominal probability conditioned on the topic z_n, $p(w_n \,|z_n\,, \beta)$. Finally, the joint distribution of a topic mixture, given the above parameters, is:

$$p(\theta, \mathbf{z}, \mathbf{w}\,|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{K} p(z_n|\theta) p(w_n\,|z_n, \beta) \qquad (6)$$

for a set of K topics, $\boldsymbol{z}$.

3.3 Methodology

A topic model first needs to be trained on the corpus of transcript data from the previously mentioned process management study. This section of the chapter describes this training framework and sets up subsequent analyses on the output of the topic model. All the natural language processing and text analysis algorithms throughout this chapter utilizes MATLAB's implementation of LDA and supporting solvers.

3.3.1 Topic Modeling Framework

Figure 3.1 depicts the framework for training the topic model. The discourse of the design teams is identified by speaker, noting that the managed and unmanaged teams differ in the number of speakers (the unmanaged teams consist of five problem solvers while the managed teams consist of four problem solvers). The manager and experimenter dialogue are also removed from the discourse.

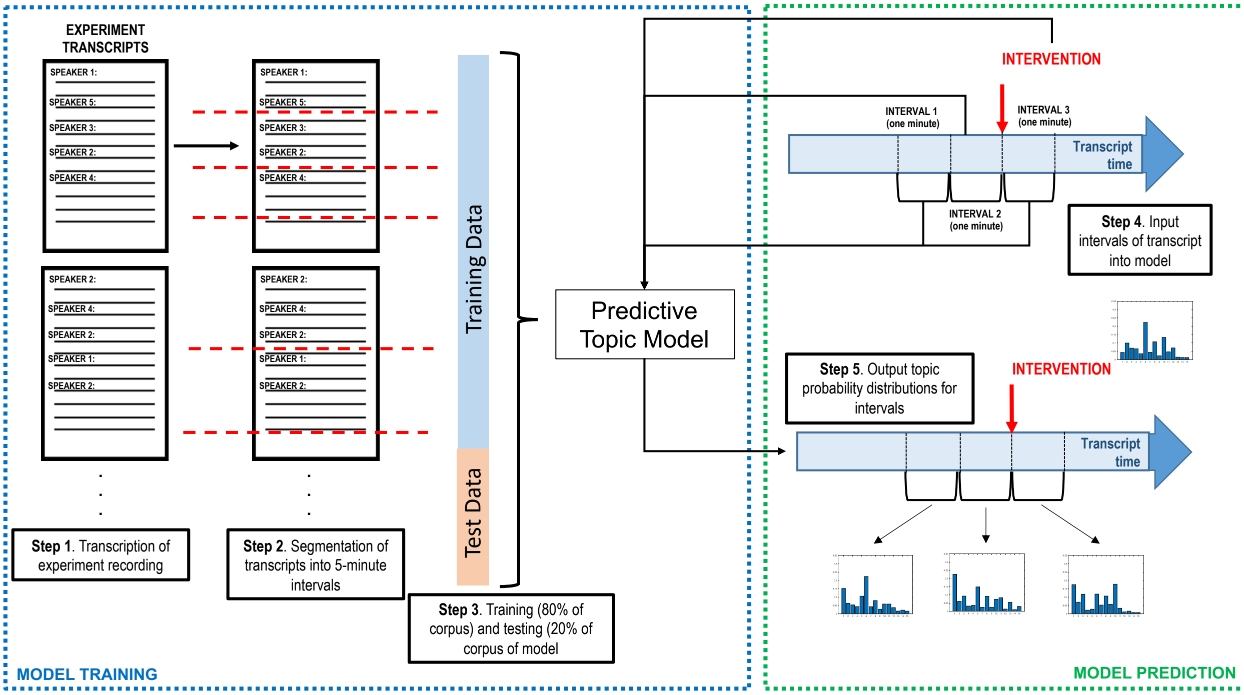


FIGURE 3.1: TOPIC MODELING FRAMEWORK WITH MODEL TRAINING AND MODEL PREDICTION

The transcripts are then further segmented into five-minute intervals to increase the corpus size per document, as needed to train sufficiently stable LDA models. In total, with eight managed and eight unmanaged teams, the entire corpus D consists of $M = 96$, distinct documents and $N = 741$ unique words. The average number of words per each five-minute interval document is 162 words, with a standard deviation of 25 words, and the average number of tokens per document being 452 tokens with a standard deviation of 102 tokens.

Prior to training the model, the transcripts undergo several pre-processing steps. The documents are first tokenized so that vectors of words represent each of the documents. All punctuation are removed, as well as the stop words identified in the Natural Language Toolkit (NLTK) [115]. Additional pre-processing steps includes the removal of infrequent words (those with frequencies less than 2), the removal of short (those less than two characters) words to eliminate noise with articles and non-words, as well as the stemming and lemmatization of words.

These two processes remove prefixes and suffixes of the vocabulary, so that all words remain in the same tense and return to their dictionary forms.

During the topic model training, the corpus D is randomly split into a training set and a test set. The training set incorporates 80% of the corpus (77 documents), while the test set contains the remaining 20% of the corpus (19 documents). All five-minute interval documents are considered during the randomization of splitting between these sets. While the two added constraints during the experiment (recall the first was added 10 minutes in with the second added 20 minutes) could add some noise during training, since the model is trained and testing on multiple randomizations of these two sets, and significant confounding effects from these constraints should be mitigated. Because the number of topics is not determined a priori for LDA, the topic model needs to be trained across a varying number of topics. Then, the number of topics with a better fit to the test data can be chosen. In this chapter, validation perplexity is used as one of the measures testing the fit to the data. Perplexity, a metric describing the goodness-of-fit, indicates how well the model, including the number of extracted topics, represents the documents; the lower the perplexity indicates a better fit [116–118]. The perplexity is mathematically defined in Equation 7 as:

$$perplexity(D_{test}) = exp\left\{ - \frac{\sum_{d=1}^{M} logp(\mathbf{w}_d)}{\sum_{d=1}^{M} N_d} \right\}, \qquad (7)$$

with N_d being the total number of words and M being the number of documents.

3.3.2 Comparing Topic Mixture Outputs

Following the training of the topic model, the model can be leveraged to transform topic mixtures of smaller discourse segments within the transcripts. The topic model dimensionally reduces documents into a probability distribution over the topic space (i.e., topic mixtures). As mentioned from the prior study, process managers intervened with a bank of prescribed stimuli which included functional design components (hammer, conveyor, etc.) and design keywords (high throughput, sieve, etc.). As noted by the managers, these interventions were injected to direct team discourse. Therefore, comparing the discourse intervals immediately prior to and following an intervention should enable detection of the impact of that intervention. As shown in Figure 3.1 three different intervals are used to analyze the impact of the interventions, where I_t designates the time of the intervention. Interval 1 (I_{t-2}) includes the discourse between two minutes and one minute prior to a process manager intervention, Interval 2 (I_{t-1}) includes team communication between one minute prior to and immediately up to the point of the intervention, and Interval 3 (I_{t+1}) includes the time immediately following the intervention and up to one minute after. An assumption of this chapter defines an "effective" topic shift as whether an intervention causes a topic shift *towards* the topic mixture of the intervention itself, indicating converging discourse on the intervention topic. As an illustrative example, an allowable design component intervention includes a sieve. The manager may intervene with the sieve component to get their team to start focusing on sorting of the peanuts from the crushed shells. The goal is to computationally uncover whether the team starts discussing not necessarily sieve specifically, but the concept, design, and/or function of sorting in general. Again, this idea of shifts in topic originates from the post-study interviews conducted following the behavioral study, in which topic-related motivations emerged as a common theme throughout the manager interventions. The immediate goal is to determine whether those motivations are realized in producing effective behavior change. This overall notion motivated the inspection of Interval 2 and Interval 3. Interval 1 is included in the analysis as a control, controlling for the idea that teams were already moving closer to the intervention topic. These analyses are described in more detail later in the chapter.

To fully carry out the above analysis, a topic mixture for the intervention itself must be defined. Recall that the design keywords and design function interventions consist of a single word and/or image (e.g., the text and image of a *sieve* for a design component, and the text of *high throughput* for the design keyword). Consequently, in order to define the intervention topic mixtures, the full dictionary definitions of the words are used. For example, Merriam-Webster dictionary defines *sieve* as "*a device with meshes or perforations through which finer particles of a mixture (as of ashes, flour, or sand) of various sizes may be passed to separate them from coarser ones, through which the liquid may be drained from liquid-containing material, or through which soft materials may be forced for reduction to fine particles*" [119]. On the assumption that the words in the definition are relevant to the term, this text document for the intervention, when mapped into the topic space using the trained model, can now be used to define the topic mixtures of the interventions. In this above example, an intervention is perceived as effective not only if the team starts talking specifically about a sieve, but if they also start discussing meshes, the concept/function of separating, etc. Thus, including the dictionary definition as the topic mixture for the interventions provides this additional level of detail for detection. The intervention documents are not included in the training of the topic model itself.

After extracting the topic probability distributions from the topic model, the distributions from the different intervals and the intervention need to be compared with each other. To measure the similarity between the topic probability distributions, the Kullback-Leibler (KL) divergence is used, also known as information divergence or relative entropy [120,121]. The KL divergence computes how different one probability distribution is from another probability distribution. It is more precisely defined in Equation 8, for a discrete probability distribution, as:

$$D(P||Q) = \sum_{\mathbb{R}^K} p(x) \log \frac{p(x)}{q(x)} dx \, , \qquad (8)$$

where P and Q represent two discrete, probability distributions, over the same variable, x (which in this case are the topics). As P and Q approach one another in similarity, the KL divergence approaches zero. It should be noted that the order for KL divergence matters, as the metric is not symmetric and does not follow the triangle inequality [122,123]. In other words, the divergence from P to Q does not necessarily equal the divergence from Q to P. Thus, in order to make relative comparisons, all analyses in this chapter follow the same temporal ordering (relative to the timeline of the experiment) for the distributions.

3.4 Results

Section 3.4.1 first discusses the selection of the topic model with the specific number of topics. This includes results from different optimization solvers, perplexity, pointwise mutual information, as well as a parametric analysis on the concentration priors that describe the prior word and topic distributions. Next, analysis on the differences in design cognition, modeled as topic mixtures, between the managed and unmanaged teams is presented both statically (the overall transcripts themselves) and over time (Section 3.4.2). Then, a before and after analysis of specific manager interventions on the discourse detects the impact of the process managers on design cognition via directed topic shifts (Section 3.4.3).

3.4.1 Topic Model – Validation & Selection

As discussed in the previous section, the number of topics for topics models is not determined a prior, so the model must be trained over a varying number of topics. The model is trained across a range of one to thirty topics, and for each distinct number of topics, trained for 100 iterations. Each iteration randomly selects the training and test sets from the corpus. Recall that the training set consists of 80% of the corpus, or approximately 77 documents, while the test set consists of the remaining 20% percent of the corpus, or 19 documents. Figure 3.2 shows the average validation perplexity of the 100 iterations over the range of topics: from one to thirty. Different optimization solvers are used in the fitting of the model to test performance. These include stochastic approximate variational Bayes [124,125], collapsed Gibbs sampling [126], approximate variational Bayes [127], and zeroth order, collapsed variational Bayes [127,128]. While more in-depth comparisons between the four solvers lies outside the scope of this chapter, here the different solvers are compared by their fit to the data via validation perplexity on the held-out test set.

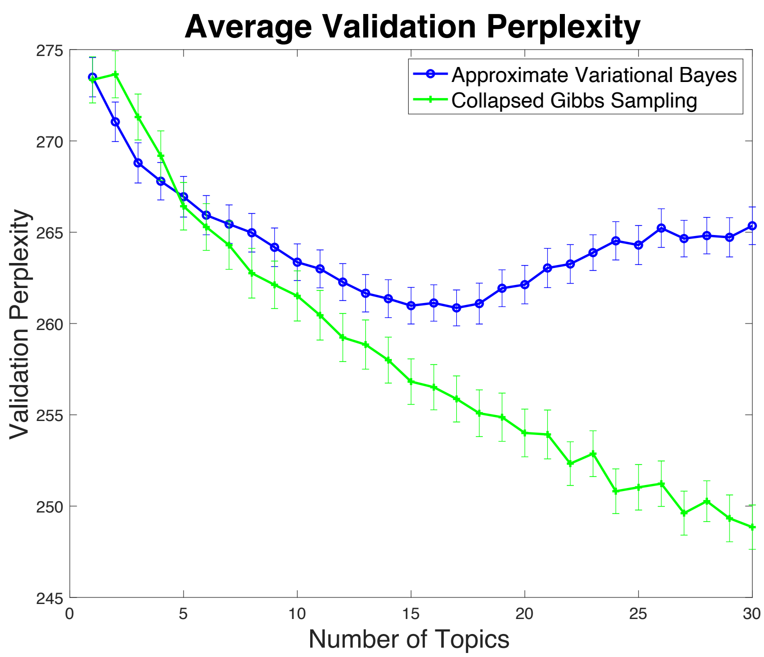


FIGURE 3.2: AVERAGE VALIDATION PERPLEXITY FOR TWO LDA ALGORITHMS OVER A VARYING NUMBER OF TOPICS (ERROR BARS SHOW ±1 S.E.)

Figure 3.2 shows validation perplexity for the two better performing optimization solvers: approximate variational Bayes and collapsed Gibbs sampling. The lower the perplexity indicates a better fit. The other two solvers result in significantly worse performance, with stochastic variational Bayes showing no increase in performance with increasing number of topics, and thus omitted from the figure. The two solvers shown behave a bit differently. While the validation perplexity decreases with an increasing number of topics for both, collapsed Gibbs sampling continues to decrease while approximate variational Bayes reaches a plateau before starting to increase in validation perplexity. While approximate variational Bayes plateaus at around 15 topics, collapsed variational Bayes starts to become noisy around 20 topics and greater. Due to this difference in behavior, both are considered in choosing the number of topics as well as an additional measure – pointwise mutual information.

Pointwise mutual information (PMI), or topic coherence, is measured across the same range of topics as the perplexity. PMI is occasionally used in lieu of, or in conjunction with, perplexity to characterize topic models, as it has been shown that perplexity can negatively correlate with human perception of the topics [129]. While human judges can be used, previous work in this area has demonstrated that it is possible to automatically measure topic coherence with near-human accuracy using topic coherence [130,131]. Specifically, pointwise mutual information scores the probability of pairs of terms taken from topics and their appearance across topics. In other words, PMI identifies overlap of information contained in topics. The same two better performing algorithms (in terms of validation perplexity) are graphed in Figure 3.3, which shows the average normalized PMI, calculated on the corpus, across the topics and models. As shown in the figure, both algorithms see a similar trend in pointwise mutual information gains. Both experience significant increases early on with smaller number of topics, both starting to settle between 12 and 18 topics. These trends are mutually considered with the previous trends from the validation perplexity to characterize and choose a topic model.

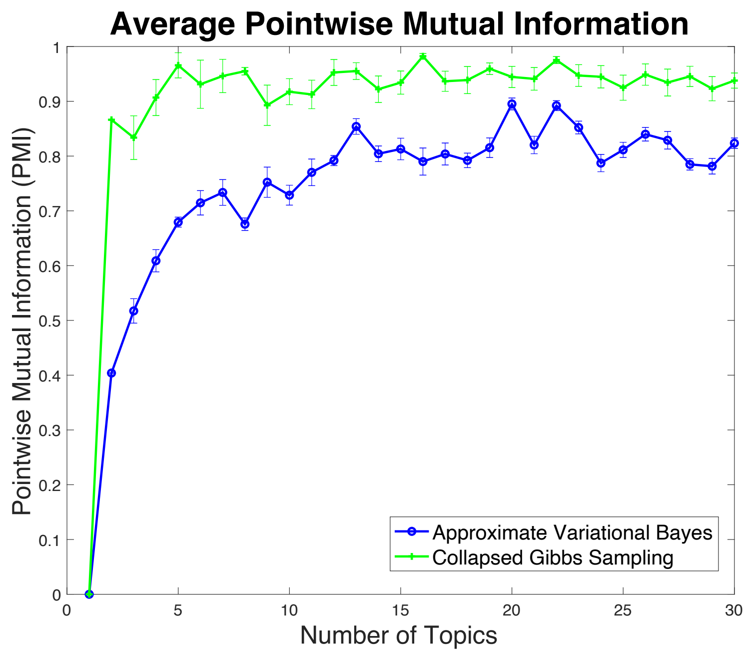


FIGURE 3.3: AVERAGE POINTWISE MUTUAL INFORMATION FOR TWO LDA ALGORITHMS OVER A VARYING NUMBER OF TOPICS (ERROR BARS SHOW ±1 S.E.)

Taken together, both the validation perplexity and pointwise mutual information measurements support a similar range of topics. Since there are no significant differences in validation perplexity over the range from 15 to 20 topics (the two solvers are within a range of 4.0), and both solvers' PMI level around a similar range, 15 topics (the plateau in perplexity) is chosen throughout the remainder of this chapter and analysis. Now regarding the specific model, recall that 100 different topic models are averaged for each number of topics. For those models using 15 topics, the model with the lowest validation perplexity, across both solvers, is selected. The model chosen possesses a validation perplexity on the test set of, $p(D_{test}) = 219.53$. As a final level of validation, the cosine similarity is computed between all pairs of resulting topics. Measuring the intra-topic similarity between topics provides an additional measure of overlap between topics – the greater the similarity and overlap, the less distinct the topics are. Ideally, topics should not significantly overlap. Excluding self-similarities, the average pairwise cosine similarity between topics for the selected model is $\mu_{c.s.} = 0.06$.

After selecting a specific topic model, a parametric analysis is performed to identify the sensitivity of this model with varying values of the two hyperparameters. The two hyperparameters for Latent Dirichlet Allocation, α and β, describe the prior distributions on the topic and word concentrations, respectively, when fitting the model. Larger values of α define documents as being composed of a wider variety of topics while larger values of β define topics as being composed of a wider variety of words. Figure 3.4 shows the resulting surface from the parametric analysis on the hyperparameter priors. For the analysis, values of the priors vary from, $\alpha, \beta \in [0.01 - 1]$ in 100 equal intervals. As shown, with increasing values of both α and β, the validation perplexity value decreases. Only with low values of the hyperparameters, specifically at approximately $\alpha \leq 0.4$, does the result become sensitive, with a significant and steep increase in validation perplexity. Accordingly, the chosen model uses values of these hyperparameters with lower and less sensitive perplexity.

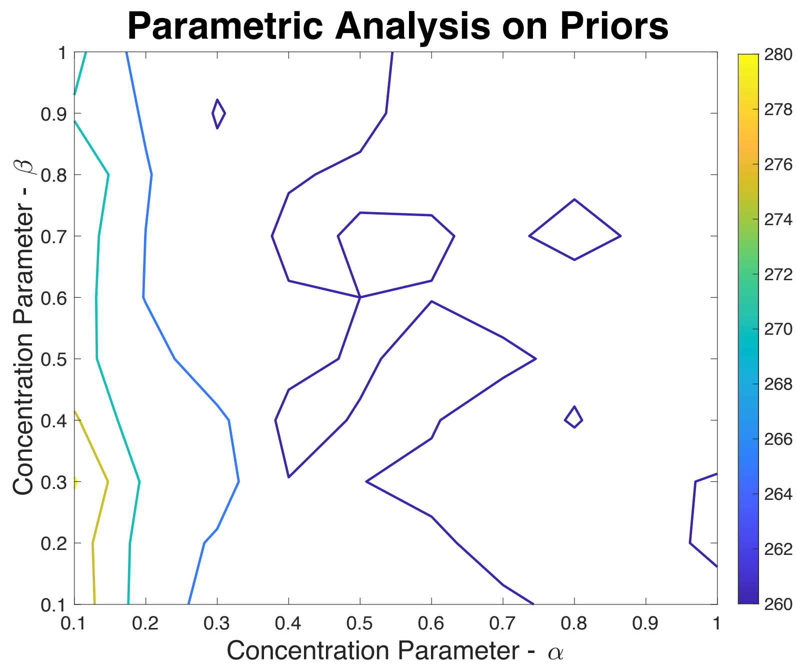


FIGURE 3.4: PARAMETRIC ANALYSIS ON VALIDATION PERPLEXITY, VARYING PRIOR PARAMETERS α AND β

As defined earlier, the topics are represented by a vector, equal in length to the number of words in the entire corpus. The entire document corpus D, from the 16 transcripts, contains $N = 741$ distinct words after the aforementioned pre-processing steps, and thus represents the size of each topic vector. Each word in these topic vectors is assigned a probability value that appears in that particular topic. Therefore, these topics can be "visualized" by viewing the most probable words in the respective topics. Table 3.1 shows the ten most probable words for each of the 15 resulting topics from the chosen topic model. Additionally, to illustrate a sample distribution within a topic, Table 3.2 shows the ten most probable words for Topic 11, along with their associated probabilities of occurrence.

TABLE 3.1: THE 15 EXTRACTED TOPICS WITH THE TEN MOST PROBABLE WORDS IN EACH TOPIC

TOPICS	TOP 10 WORDS
Topic 1 *(nPMI = 0.664)*	side, good, wait, top, bottom, draw, large, middle, view, kinda
Topic 2 *(nPMI = 0.862)*	peanut, easy, manufacture, cost, remove, low, guess, metal, electricity, plastic
Topic 3 *(nPMI = 0.752)*	peanut, hole, sort, big, hard, guess, bit, split, talk, apply
Topic 4 *(nPMI = 0.783)*	design, kind, open, idea, cut, thing, work, nutcracker, slide, start
Topic 5 *(nPMI = 0.742)*	mmhmm, affirmative, time, push, crush, inside, constraint, work, fall, claw
Topic 6 *(nPMI = 0.811)*	peanut, crack, kind, device, blade, size, roll, half, move, mechanism
Topic 7 *(nPMI = 0.749)*	draw, circle, guy, minute, sheet, long, thing, mumble, handle, suppose
Topic 8 *(nPMI = 0.850)*	crush, small, piece, thing, separate, break, basically, pressure, force, move
Topic 9 *(nPMI = 0.740)*	peanut, thing, hand, pull, press, clamp, spring, speaker, fall, hold
Topic 10 *(nPMI = 0.686)*	sieve, nut, funnel, high, add, shake, good, simple, amount, person
Topic 11 *(nPMI = 0.742)*	crank, conveyor, peanut, belt, fall, roller, final, turn, attach, leave
Topic 12 *(nPMI = 0.778)*	feel, wood, pretty, fine, flat, cheap, happen, edge, stuff, easily
Topic 13 *(nPMI = 0.819)*	basically, wheel, hmm, connect, power, process, spin, human, large, edge
Topic 14 *(nPMI = 0.799)*	laughs, gear, lot, laugh, wire, pretty, box, gap, foot, version
Topic 15 *(nPMI = 0.689)*	replaces, nut, true, rotate, draw, screw, wall, sense, cylinder, call

TABLE 3.2: SAMPLE TOPIC SHOWING THE TEN MOST PROBABLE WORDS AND THEIR ASSOCIATED PROBABILITIES

Word	Probability
crank	0.122
conveyor	0.063
peanut	0.062
belt	0.0621
fall	0.0606
roller	0.0484
final	0.0469
turn	0.0378
attach	0.0363
leave	0.032

3.4.2 Comparing Topic Mixtures Between Managed and Unmanaged Teams

After the topic model has been sufficiently trained, the model can be used to transform different documents into the topic space. This dimensionality reduction of discourse using the trained model produces a topic mixture profile for the input document. The topic mixture shows the probability for all fifteen topics appearing in that specific document. This initial analysis shows the topic mixtures of the entire transcripts themselves. That is, each of the eight managed and unmanaged teams' transcripts are projected into the topic space, and the results shown in Figure 3.5. Each bar represents the average probability of the topic for both the unmanaged and managed teams. For example, in terms of Topic 1, the average probability that this topic appears across the managed teams' discourse is 0.053, or 5.3%, while the average probability across the unmanaged teams' discourse is 0.12, or 12%.

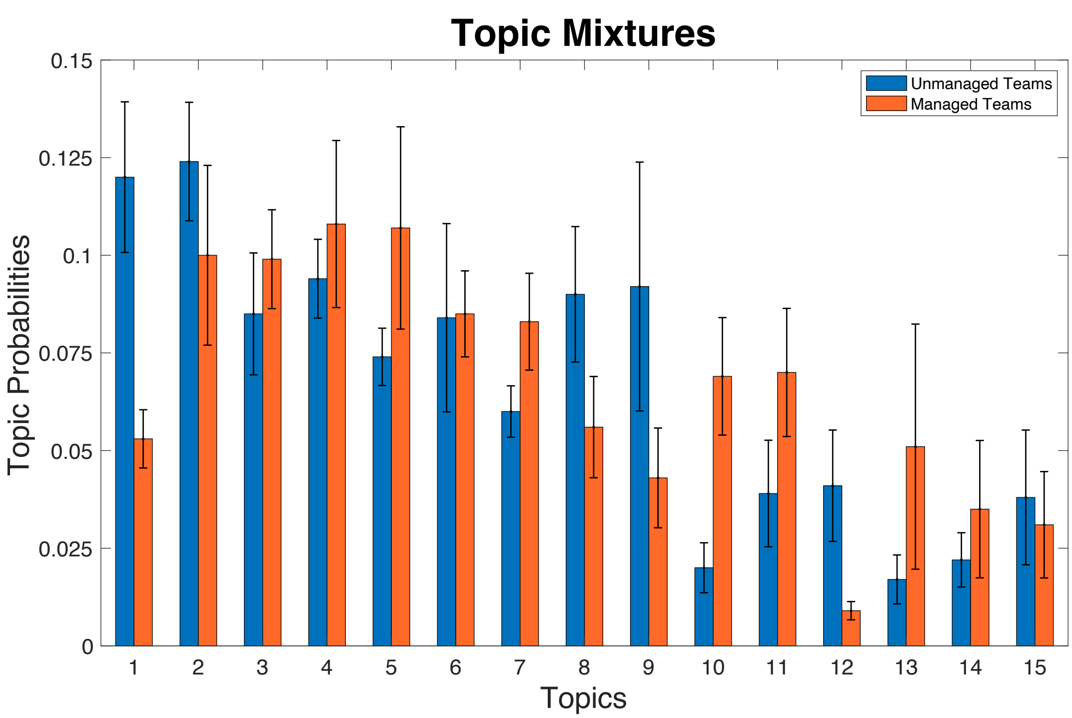


FIGURE 3.5: TOPIC MIXTURES (TOPIC PROBABILITY DISTRICTIONS) FOR MANAGED AND UNMANAGED TEAMS' TRANSCRIPTS (ERROR BARS SHOW ±1 S.E.)

Across the 15 different topics, using a two-tailed, non-parametric Mann Whitney U-test, four topics exhibit significant differences between the two team conditions: Topic 1 ($p < 0.0047$), Topic 10 ($p < 0.028$), Topic 11 ($p < 0.05$), and Topic 12 ($p < 0.027$). Topics 10 and 11 highlight an important finding from this analysis, which exhibit significantly higher probabilities of appearing in the managed teams' discourse. These topics, visualized by their ten most probable terms (Table 3.1), contain more design components, for example: sieve, funnel, belt, roller, etc., with some of these functional components coming directly from the manager interventions. Thus, the managed teams are significantly more likely to discuss these functional components and are directly influenced by the manager interventions, which contain such functional concepts.

In addition to mapping the entirety of the transcripts into the topic space, five-minute discourse intervals can also be mapped (i.e., $t \in \{0, 5\}$, $t \in \{5, 10\}$, $t \in \{10, 15\}$, $t \in \{15, 20\}$, $t \in \{20, 25\}$, $t \in \{25, 30\}$, where $t \in \{x, y\}$ defines the interval from x minutes to y minutes in the experiment). Transforming to these smaller intervals provides a dynamic look at the topic distributions over the course of problem-solving, as well as the differences between the managed and unmanaged teams over these time periods. Figure 3.6 shows the difference in probability distributions over the topic space for those six specific time intervals. A positive difference indicates that the topic exhibits higher probability in the managed teams while a negative difference indicates higher probability in the unmanaged teams.

Overall, computing the sum of the squared differences (SSD), the second half of the experiment shows greater disparities between the discourse of the managed and unmanaged teams, with the final interval exhibiting the largest, as shown in Table 3.3.

TABLE 3.3: SQUARED DIFFERENCES BETWEEN MANAGED AND UNMANAGED TEAMS' DISCOURSE VIA THE TOPIC SPACE

Time (t – minutes)	$t \in \{0, 5\}$	$t \in \{5, 10\}$	$t \in \{10, 15\}$	$t \in \{15, 20\}$	$t \in \{20, 25\}$	$t \in \{25, 30\}$
Squared Difference (SSD)	***SSD =*** 0.015	***SSD =*** 0.010	***SSD =*** 0.011	***SSD =*** 0.026	***SSD =*** 0.021	***SSD =*** 0.033

This result shows that the process managers create a larger impact during the second half of the experiment, as these intervals contain the largest overall differences in team discourse. Again, utilizing a two-tailed, non-parametric Mann Whitney U-test, several significant differences in each of the interval's individual topics emerge between managed and unmanaged teams: five minutes (Topic 7 ($p < 0.01$) and Topic 8 ($p < 0.01$)), ten minutes (Topic 9 ($p < 0.04$)), fifteen minutes (Topic 3 ($p < 0.01$)), twenty minutes (Topic 1 ($p < 0.05$), Topic 10 ($p < 0.01$), and Topic 13 ($p < 0.01$)), twenty-five minutes (Topic 1 ($p < 0.01$), Topic 8 ($p < 0.05$), Topic 10 ($p < 0.02$), and Topic 12 ($p < 0.04$)), and thirty minutes (Topic 1 ($p < 0.001$) and Topic 12 ($p < 0.02$)). Topics 10 and 11, which both contain functional concepts, become increasingly more integral to the managed team members' discourse as problem solving progresses, with Topic 10 becoming significantly more integral during the 20-minute and 25-minute intervals. On the other hand, Topic 1 becomes increasingly more integral to the unmanaged team members' discourse, and this trend continues throughout the entirety of the experiment, becoming most significant in the final interval. Visualizing Topic 1 from Table 3.2 (the most probable words in the document), the topic focuses more on the abstract design process and structure rather than on concrete design functions, such as "*draw,*" "*top,*" "*bottom,*" "*middle*," "*view*." Linking this to the semantics of the design process, the discourse in this topic seems to be based on visualization and orientation, and contradicts the process manager strategies, whose focus near the end was to home their teams in on functional concepts and ideas. This significant focus, particularly at the end of the experiment when designs need to be past the abstract/conceptual stage, could have been one of the factors harming the unmanaged teams, ultimately leading them to their inferior performance. However, further analyses to isolate this effect would be needed to determine the extent to which this focus detrimentally impacted the unmanaged teams.

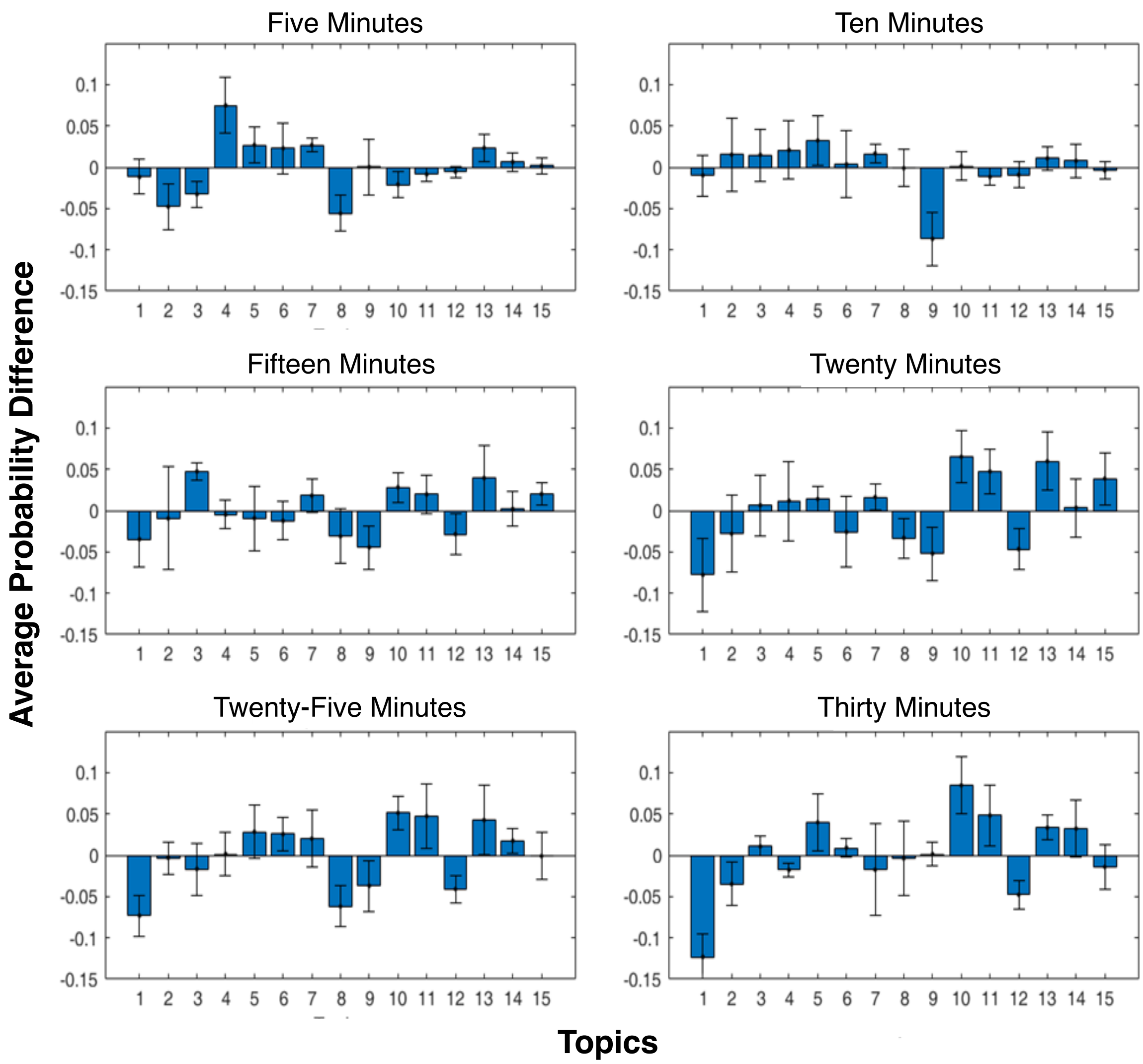


FIGURE 3.6: DIFFERENCE IN TOPIC MIXTURES (TOPIC PROBABILITY DISTRIBTIONS) FOR MANAGED AND UNMANAGED TEAMS OVER TIME (ERROR BARS SHOW ±1 S.E.)

3.4.3 Detecting Effects of Manager Interventions

The next analysis focuses on the effects of individual manager interventions on team discourse. As depicted in Figure 3.1, the topic mixture of the interval leading up to an intervention (I_{t-1}) is compared to the topic mixture of the interval immediately following an intervention (I_{t+1}). The assumption of this analysis relies on the idea that the team members' discourse should be more aligned with the intervention immediately following the manager intervention compared to before. For this analysis, 20 distinct interventions are studied, because: 1) the intervention is either a *design keyword* or a *design component,* or 2) the intervention is the specific design strategy intervention of, "*Can you identify the assumptions, constrains, and goals of the problem*?", and 3) no other interruptions (i.e., constraints or manager interventions) occurred within the one minute prior to and following the intervention. The first two requirements ensure more topic-focused, concrete interventions. The remaining design strategies are more process related and thus unsuitable for study via the topic modeling framework. Accordingly, future work can consider how to computational detect these remaining design strategy interventions. The third criteria (that no other interruptions occur within the one-minute intervals), controls for other confounding variables in the analysis that could potentially cause additional topic shifts. As mentioned previously, these 20 interventions came from a broader set of 52 total process manager interventions.

For these 20 distinct interventions, the KL divergence computes the similarity between the topic probability distributions of the one-minute interval prior to the intervention (I_{t-1}) and the one-minute interval after the intervention (I_{t+1}), both against the topic probability distribution of the intervention itself. Utilizing a two-tailed, non-parametric Mann Whitney U-test, results show that the topic mixtures in the minute following the intervention are *significantly more similar* to the intervention topic mixture than prior to the intervention (I_{t-1}= 0.69, I_{t+1}= 0.46, $p < 0.005$, *effect* $r = 0.44$, $U_{crit} = 127.5$). This result indicates that the interventions generate a significant impact on the topic structure and design cognition of the design teams, leading them to direct the focus of their discourse on the provided topics of interventions.

To further corroborate this finding, an LSA model is also trained on the data. Similar in spirit to LDA, LSA instead utilizes singular value decomposition for dimension reduction as opposed to a probabilistic approach as LDA does. The semantic distances (via cosine distance, D_c) between the one-minute discourse intervals, both before and after the interventions, are computed with the intervention documents. Across the same 20 interventions, the average cosine distance prior to an intervention is $D_c = 0.67$ with the average cosine distance following the interventions is $D_c = 0.45$, with the average change (using a temporal change with respective to experimental time, i.e., before intervention minus after intervention) in distance being ΔD_c = 0.124. Again, the results indicate that the discourse immediately after an injected intervention is more similar (smaller D_c) than prior to the intervention.

A follow-up analysis provides further evidence that this similarity in topic mixtures is a direct effect of the interventions. In order to rule out the possibility that the design teams incrementally move closer and closer to a given topic over time on their own, the KL divergence between all three intervals' topic mixtures with the intervention topic mixtures are computed. Thus, for the above notion to hold true, the divergence between the first interval's topic mixture and the intervention topic mixture should be the largest, and then decrease through I_{t-1} and I_{t+1}. After computing the divergences, the changes between the before and after intervals of the intervention can be computed, as in Equation 9 and Equation 10:

$$\Delta_{12} = D(\theta_{I_{t-2}}||\theta_{Inter}) - D(\theta_{I_{t-1}}||\theta_{Inter}), \quad (9)$$

$$\Delta_{23} = D(\theta_{I_{t-1}}||\theta_{Inter}) - D(\theta_{I_{t+1}}||\theta_{Inter}), \quad (10)$$

where $\theta_{I_{t-2}}$ is the topic mixture for I_{t-2}, $\theta_{I_{t-1}}$ is the topic mixture for I_{t-1}, $\theta_{I_{t+1}}$ is the topic mixture for I_{t+1}, θ_{Inter} is the topic mixture for the intervention, and D is the KL divergence operator. For example, the first term in Equation 9 ($D(\theta_{I_{t-2}}||\theta_{Inter})$) computes the KL divergence from I_{t-2}'s topic mixture to the intervention's topic mixture.

Conceptually, Equation 9 shows whether the teams members' discourse becomes more similar or dissimilar to the intervention *prior to* the intervention, while Equation 10 provides the same information, but *over* the intervention. Again, utilizing a two-tailed, non-parametric Mann Whitney U-test, results indicate that the changes between the intervals' topic mixtures from Equations 9 and 10 are significantly different ($\bar{\Delta}_{12}$= -0.14, $\bar{\Delta}_{23}$ = 0.23, p < 0.003, *effect* r = 0.48, U_{crit} = 113.4). Not only are the average divergences significantly different, but they are opposite in sign. Consequently, in the two, one-minute intervals prior to the intervention, the team members' discourse drifts away from that of the intervention, (i.e., becomes more *dissimilar* to the intervention), while during the one-minute intervals before and after the intervention, the team members' discourse converges back to the intervention topic (i.e., becomes more *similar* to the intervention). Thus, the possibility that the design teams incrementally move closer and closer to a given topic over time is disproven, further corroborating that the topic shifts are caused directly by these manager interventions.

3.5 Discussion

The utilization of topic modeling for the framework in this chapter, to detect the effect of intervention on design team process, is motivated by several factors. The first, as mentioned previously, involves the post-study interviews conducted with the process managers. After querying the human managers on their rationale for intervening with their design teams, many consisted of topic-related rationales. For example, one manager mentioned, "*Uhh, they* [the team] *were getting really close to the idea and they had been really close to the idea of a blade for a long time, but it was just to push them a little bit farther in that direction. They had the drawing and were really close, but they started adding more complicated things that I did not really think would be helpful.*" Additionally, another manager indicated, "*They* [the team] *started talking, they were very close to the idea, wanting to have a sieve, but couldn't come up with the idea themselves. They were trying to think of much more complicated solutions for that*." These quotes from the managers, representing just two of many, highlight the concept of topic pushes or topic shifts. Topic modeling provides an algorithmic way to computationally detect these changes and shifts in topic. Accordingly, the overarching question this research answers is whether these topics shifts can be computationally detected and reveal the plausible mechanism underlying the effectiveness of the managerial interventions. An additional motivation for utilizing topic modeling to answer the aforementioned research question lies in the algorithm output. The output of LDA, the illustrative technique used in this chapter, shows a probability distribution over a range of topics. Accordingly, this distribution provides a more holistic representation of the discourse data, as it maps team interactions along a spectrum of topics. This contrasts with other work in this area using manual/automatic coding, which maps lines of discourse to a single coding scheme.

The topic model goes through an exhaustive training procedure, varying over a number of topics, optimization solvers, and training and test sets. For each number of topics and optimization solver, the model runs for 100 iterations (each data point in Figure 3.2 and Figure 3.3 averages across those 100 runs). The results for the two better performing optimization solvers, collapsed Gibbs sampling and approximate variational Bayes, show that the perplexity behaves a bit differently. While the validation perplexity decreases with an increasing number of topics for both, collapsed Gibbs sampling continues to decrease while approximate variational Bayes reaches a plateau at around 15 topics. Collapsed variational Bayes starts to become noisy at 20 topics and greater. In addition to perplexity, pointwise mutual information (PMI), or topic coherence, is measured across the same range of topics. The two better fitting solvers in terms of validation perplexity are graphed in Figure 3.2, which shows the average PMI across the topics for a model. Both experience significant increases early on with smaller number of topics and start to settle between 12 and 18 topics. For the selection of the number of topics, this two-pronged approach acknowledges the limitations of perplexity when it comes to the perceptibility of topics via direct human inspection. While perhaps not the absolute optimal model, taking all these factors into consideration results in a pragmatically sufficient topic model for further analyses. A parametric analysis also tests the sensitivity of the hyperparameters describing the prior distribution, and a cosine similarity metric tests the overlap in the resulting topics.

The goal of this chapter is not to compare the efficiencies or performance of different types of topic models and/or algorithms. Rather, the goal is to study the effects of managerial interventions via design discourse. Since there does not exist a ground truth for the discourse data (a prior knowledge of the exact topics to extract), it is difficult to formalize a precise measure of optimality or the "absolute best" model. The rigor of the model selection process for this chapter explores the space of models with LDA and selects the one that provides acceptable performance on the chosen metrics. Used consistently across conditions, the chosen model can then be used to analyze differences between teams and segments of discourse. In addition, a Latent Semantic Analysis (LSA) model further supports the overall findings in the convergence of discourse as an impact of the process manager interventions. Future work can explore how different modeling algorithms compare to LDA (besides LSA which already corroborates results in this chapter) and perform on the discourse data. An additional opportunity can also explore training on larger corpuses such Wikipedia and Google News.

With the model trained, two different analyses test whether the interventions can be detected within the design team discourse. The first (Section 3.4.2) maps entire transcripts, comparing the managed and unmanaged teams, along the topic space by outputting their associated topic mixtures (i.e., the topic probability distributions). Two interesting findings emerge. The first involves the topics most relevant to the design interventions, specifically Topic 10 and Topic 11. These two topics are significantly more probable in the managed teams, including more functional concepts and are more representative of the design component interventions. This result is validated by both 1) visualizing these two topics with their 10 most probable words, and 2) mapping the interventions themselves into the topic space (where topics 10 and 11 become more prominent). The second interesting finding involves Topic 2 – Topic 4. All three of these topics are nearly equal between the managed and unmanaged teams, and, apart from topic one, contain the highest probabilities across the space. These topics most pertain to the constraints and goals of the problem, so it is not surprising (and further validates the framework) that these topics appear highly and equally probable across both team conditions.

A dynamic look at the topic mixtures across the experiment is also performed. Two interesting trends emerge from comparing the topic mixture space between the managed and unmanaged team conditions. Topics 10 and 11, which both contain more functional concepts, become increasingly more integral to the managed team members' discourse throughout the experiment, particularly near the end. Furthermore, Topic 1 becomes increasingly more integral to the unmanaged team members' discourse, reaching significance from the managed teams in the last two intervals of the experiment. The visualization of this topic shows more abstract, design process related activity. Taken together, these two results point to managers guiding the solving process toward completion and refinement of the final designs, while the unmanaged teams tend to be more focused on design visualization. Future work can consider smaller intervals of the transcript, to gain more resolution in the evolution of the topic structures throughout problem solving.

In order to ascertain the impact of the interventions, the second analysis (Section 3.4.3) performs a before and after investigation on 20 distinct manager interventions. Consequently, as shown in Figure 3.1, the one-minute intervals prior to an intervention and immediately following an intervention are mapped into the topic space. Using KL divergence to compare the similarity of probability distributions, the topic mixtures of these transcript intervals are compared to the topic mixture of the interventions themselves. Results reveal that, on average, the discourse becomes significantly more similar to the intervention immediately after the intervention. Taken together, these findings validate the detection of the topic shifts in discourse, which many of the managers claimed as their motivation for intervening. Of these 20 interventions studied, in four instances, the topic mixtures of the interval following the intervention actually become more *dissimilar* to the topic mixture of the intervention. These cases deserve a more thorough investigation. One of these cases has a near-zero change, while one of these four cases has a significantly larger difference than the others. In this particular instance, the manager directly perceives the intervention as ineffective, saying that "*It really did nothing, not at all*." In this case, it is interesting to note consistency in how the manager perceives the intervention with the detection within the discourse, as the team members' topic mixtures become more dissimilar immediately after the intervention. In general, deeper dives into these outlying cases can provide additional insight when considering the implementation and effectiveness of a real-time, intervention framework.

Overall, this chapter shows promise in a more automated approach to track design team discourse in real time. While Latent Dirichlet Allocation has been applied on smaller discourse such as tweets and micro tweets, recent developments and work in word embeddings, hierarchical topic models, and dynamic topic models have emerged for these purposes [132-134]. LDA was chosen for this specific chapter for its well-developed and wide utilization on texts of many sizes, accessibility, and previous application on discourse data. The comparison of LDA with these additional topic modeling algorithms lies outside the direct scope of this chapter (the main goal is not to find the most efficient topic modeling approach), but future work can consider the sensitivity of the corpus size and resulting topics with these other methods. Additionally, the domain, context, and process manager interventions for this problem are quite specific. The problem statement asks participants to design a peanut sheller, and over half of the process manager interventions are tailored towards this specific goal with the design components and design keywords. To fully understand the generalizability of this framework, analysis of different types of problem types and domains can be studied, as well as expanding beyond conceptual design problem solving. Finally, while this chapter identifies intended topics shifts in the behavior of the managed teams, additional experimental conditions need to be run to completely isolate this effect from all potential confounding variable, directly link this to overall more effective team performance, and explore other modes of process manager strategies.

3.6 Summary

This chapter utilizes a topic modeling approach to study the effects of process manager interventions via analysis of design team discourse. The transcripts, collected from the prior research study covered in Chapter 2, contain discourse of design teams solving a design problem under either the guidance or absence of a human process manager. The inspiration of this topic modeling perspective derives from post-study interviews conducted with the process managers. The goals of imbuing functional concepts into the discourse, and shifting to more relevant topics, serve as some of the primary motivations of the managers for intervening with the design teams.

The topic modeling framework, in this instance Latent Dirichlet Allocation, can be leveraged to predict topic mixtures of the team discourse at different segments during problem solving. Training over several topics, optimization algorithms, and training and test sets, 15 topics emerge as the number of topics based on validation perplexity and pointwise mutual information metrics. After this exhaustive training procedure, and corroborative analyses with LSA, the topic model can now be used to transform different intervals of the transcripts into the topic space. The model outputs a probability distribution of the segments of transcripts over the 15 topics. This output allows a more holistic perspective on team discourse, as opposed to mapping to a single topic or coding scheme.

In order to uncover the influence of the process managers, this chapter presents two analyses to detect the impacts of their interventions. First, the topic model framework is leveraged to predict the overall topic structures between managed and unmanaged teams. This includes both a holistic perspective, through the transformation of the entire transcripts themselves, as well as a more dynamic perspective, through the transformation of smaller, five-minute intervals over time. An additional analysis studies the direct impacts of the interventions by predicting the topic mixtures immediately prior to and immediately following the interventions. All these analyses corroborate similar findings, and show convergent effects on team discourse, and thus direct impacts of these inventions on design team cognition.

Design team interactions and discourse provide valuable insight into the state and cognition of designers, and effectively analyzing them can facilitate the design process. This research provides a computational perspective on not only studying design team verbalizations, but also leveraging communication to detect the effects of design interventions via shifts in topic mixtures. Overall, this chapter contributes towards the goal of an automated approach to track design team discourse in real time. Particularly as the collaboration of human and artificial intelligence designers to solve problems becomes more prevalent in practice, being able to computationally track the design state will be critically important to understand what types of interventions may be needed to maximize performance.

Chapter 4 : Modulating Distance of Design Stimuli to Design Progress[3]

4.1 Introduction

A wide variety of literature demonstrates the impactful nature of inspirational stimuli on design ideation, such as their ability to assist designers in developing solutions with improved characteristics (e.g., increased solution uniqueness and/or feasibility) [17, 135-137]. Additionally, the distance of the inspirational stimulus from the problem domain modulates its impact [88]. Typically, the "distance" of an inspirational stimulus refers to some measure of a stimulus' proximity to the problem or design space currently occupied by the designer. When viewed on a continuum, the measure of distance of a stimulus is quantifiable using a variety of techniques, such as semantic similarity comparisons or the similarity between functional representations of designs. One can think of a "near" or "close" inspirational stimulus as one that comes from the same or a closely related domain as the problem. Conversely, a "far" stimulus comes from a distant domain. It has also been noted that near stimuli share significant surface level (object) features with the target, while far stimuli share little or no surface features [50]. However, a critical, and currently overlooked consideration impacting these findings is that the relative position of a designer solution within the design space is not static; it dynamically changes throughout problem solving.

Since the distance of an inspirational stimulus is relative, a design solution evolving during ideation therefore directly determines what constitutes a stimulus as being either near or far. In

[3] The work presented in this chapter is published in:
Goucher-Lambert, K., Gyory, J. T., Kotovsky, K., and Cagan, J., 2020, "Adaptive Inspirational Design Stimuli: Using Design Output to Computationally Search for Stimuli That Impact Concept Generation.," ASME J. Mech. Des., 142(9), pp. 1–10.

other words, a far stimulus at the onset of design ideation is not necessarily the same distance as a far stimulus at the midpoint of design ideation. A goal of our research is to develop an individualized tool that enables designers to leverage the full power of inspirational stimuli during design ideation and problem solving. For this to be the case, such a tool should adapt to the current state of the designer in order to provide a stimulus that reflects the designers' solution within the design space. However, most current approaches to selecting design stimuli are not responsive or adaptive to the dynamic state of designers. Typically, stimuli presented during cognitive studies are determined *a priori*.

The work presented in this chapter contributes to and advances this ongoing research area by computationally selecting inspirational stimuli based upon a measure of the real-time status of designers' completed activity. First, our approach determines the location of a designer's solution within a larger design space halfway through an ideation session (referred to as their current "state"). To accomplish this, we employ a method of semantic similarity comparisons, computationally comparing the textual work of the designer to a pre-existing database of design concepts collected as part of a prior research study [135]. Using this information, an adaptive intervention is provided to the designer via a stimulus that is either near or far based upon the semantic similarity between all designs within the database and their current design. Thus, the overall goals of this chapter include: 1) determining whether or not the chosen method of design state detection and adaptive intervention is feasible (i.e., whether a design state can be measured in real-time), quantifiable (i.e., whether textual similarity can be used to provide near and far adaptive stimuli that are significantly different), and perceivable (i.e., can participants distinguish the differences between these categorizations) to designers, and 2) understanding the impact of adaptive stimuli on measurable design outcomes including the novelty, feasibility, and usefulness of design concepts.

4.1.1 Analogical Reasoning in Engineering Design

Prior work on the role of inspirational stimuli has predominately focused on "analogical reasoning" applied to design. However, there is currently debate as to whether this term is always appropriate for the design contexts in which it is used [138]. Formally, analogical reasoning is the process of retrieval and mapping of relations or information from a source to a target [139–141]. In this chapter, the term "inspirational stimulus" is utilized to more broadly encompass other types of stimuli intended to support design ideation that may not satisfy both of these conditions (retrieval and mapping). For example, a prior solution provided to a designer may enhance the likelihood of retrieving a useful concept from memory but does not guarantee a direct mapping incorporating aspects from the stimulus in a new solution for the problem.

The relationship between the distance of inspirational stimuli and solution outcomes is also well studied. One intriguing result is the notion of a "sweet spot" (between near and far) of distance from a stimulus to the problem domain in which a stimulus is most impactful [88]. Because defining a sweet spot for a given research problem is an open area of research itself, most research investigations rely on the comparison between near and far stimuli. Recently, Goucher-Lambert and Cagan analyzed the impact of stimuli distance on the novelty (i.e., uniqueness), feasibility, and usefulness of solution concepts across a wide variety of conceptual design problems from literature [135]. That work revealed that near stimuli improve the usefulness and feasibility of design solutions compared to a control, whereas far stimuli improved the novelty of solutions. Additionally, separate work by Goucher-Lambert et al., utilized functional magnetic resonance imaging (fMRI) to study neural activation patterns underpinning generating design concepts with and without inspirational stimuli of varying distances [8]. In that work, inspirational stimuli defined as close to the problem space activated a unique set of brain regions supporting memory retrieval and solving problems via insight (rather than by analysis) [18]. Across these two studies, closer stimuli were found to more reliably associate with positive ideation outcomes based on both behavioral and neuroimaging data. Other researchers have also found supporting evidence that conceptually near stimuli may in fact lead to better design outcomes than far stimuli [142]. As an additional contribution, the work presented in the current chapter explores the impact of inspirational stimulus distance on design outcomes, building on the aforementioned findings.

4.1.2 Finding and Applying Design Interventions

When should inspirational stimuli be provided to aid designers? Previous research has demonstrated that interventions are best introduced to problem solvers when there exists an open goal (i.e., when the solver has an understanding of the goal(s) they are trying to accomplish, but have not yet become fixated on a specific solution) [143–146]. Based on this, it would appear that inspirational stimuli should be presented at some point during problem solving before the designer has become fixated or has reached an impasse. However, the difficultly lies in determining the time frame that someone has reached such an impasse.

Instead of trying to provide a designer with an inspirational stimulus at the correct moment, a different approach allows designers to search for stimuli on their own using structured inputs. One such example are ontology-based frameworks where designers can search for text or image-based stimuli by specifying the object (e.g., chair) and function (e.g., to sit) of their ideas [147–149]. Recently in the Human and Computer Interaction (HCI) community, computational tools have been developed that allow for semi-directed analogy mining. Past approaches at solving this problem included word-embedding models such as GloVe [150], and an analogical search engine by Gilon et al., which looks for distant analogies for specific aspects of a product or design [151]. Additionally, Chan et al. developed an approach termed SOLVENT, which draws on pre-annotations by humans regarding different features of possible stimuli (e.g., purpose, mechanisms, findings) and makes connections based upon semantic representations [152]. While this work is promising, future work in this area is necessary to reduce the burden on designers using these tools to search for relevant analogous examples. The current approach differs from these past contributions by trying to determine design stimuli based on unstructured rather than structured inputs.

Another approach is to recruit the resources of an expert to help guide a designer towards unexplored areas of the design space. One initial step towards real-time management was an empirical study by Gyory et al., that investigated the characteristics of process management which are most effective for design teams [91]. In that exploratory work, a human process manager oversaw the problem-solving process of a collaborative design team solving a conceptual design problem. The managers tracked the state of the designers within the team, and freely intervened with prescribed stimuli (e.g., design components, select keywords, and/or design strategies) to affect the solving process when deemed necessary. These interventions adapt to a teams' state, since the managers provided stimuli they felt were necessary in reaction to the design teams' activity. Teams that were under the guidance of these process managers significantly outperformed teams that were not in terms of the quality of their solution output. The work by Gyory et al. exemplifies the benefits of real-time management and intervention in design teams. In the current chapter, we build on this idea further by computationally providing real-time adaptive stimuli through semantic similarity comparisons.

4.1.3 An Approach to Compare Design Content: Latent Semantic Analysis

In order to conduct the semantic similarity comparisons that determine designers' current state, as well as select the specific adaptive stimulus to provide them, Latent Semantic Analysis (LSA) was used. LSA computes the semantic similarity between text-based corpuses and has been shown to be well suited to a variety of semantic comparisons relevant to design. For example, LSA has been used to quantify the level of semantic convergence in language-based communication between members in design teams [9, 64], uncovering patterns in design repositories such as the US patent database [89], and visualizing the similarity between existing design concepts within a pre-defined design space in a network model [153].

LSA uses singular value decomposition (SVD) for dimension reduction [65]. Within this reduced space, semantic patterns can be uncovered between text-based documents by tracking the co-occurrence of words (represented as vectors). The cosine similarity between document vectors, which analytically computes semantic similarities, varies between zero (if the vectors are completely orthogonal and exhibit no similarity) and one (if the documents are identical). The current chapter leverages this analytical power of LSA to select design artifacts (inspirational stimuli) semantically near and semantically far from the designer's current concept. The design concept is input as an unstructured description of what the designer believes is currently their best design solution.

4.1.4 Approaches for Measuring and Evaluating Conceptual Designs

In order to study the impact of the introduced adaptive design intervention, this chapter relies on evaluations performed by trained expert raters. One such evaluation performed by expert raters is to assess the overall design *quality* of each conceptual design. Design quality is a prominent measure throughout the design literature, with the most common definition of design quality being what Shah, et al., term "a measure of the feasibility of an idea and how close it comes to meet the design specification" [63]. In their popular paper on metrics for ideation effectiveness, Shah, et al., represent quality as both tangible, physical characteristics of a design, as well as the functional, performance metrics, describing the nature of designs. Ahmed et al., provide a similar, but more precise definition for the utility of a design as: “a measure of the designs’ performance and can depend on multiple domain dependent factors like functionality, feasibility, usefulness, impact, investment potential, scalability, etc.” [154]. Some works have used less specific derivatives. For example, Hu and Reid attribute quality to be characteristics of "the physical property, user adoption, and cost-benefit ratio" [155]. Although this list is not exhaustive, it is clear that most, if not all, definitions realize quality as a multi-dimensional construct; some definitions focus on function, some focus on form, and others an amalgamation of the two.

How then, are raters supposed to accurately assess such a metric when attempting to take into consideration, or even deduce, its various sub-dimensions? The subjectivity of measuring quality may very well stem from its dimensional and semantic uncertainty [156]. Furthermore, without a more discrete hierarchy, it is possible for raters to internally weigh the underlying sub-dimensions differently during assessments, leading to yet another source of subjectivity. Motivated by this concern, the current chapter explores the use of a new measure to represent the overall innovative potential of conceptual designs. Unlike quality, this new measure consists of three distinct sub-dimensions (feasibility, usefulness, and novelty) and directly describes how they should be combined.

While various forms for design quality exist, they undeniably have certain dimensional commonalities. Even from the few definitions mentioned above, both Shah, et al., and Ahmed, et al., consider *feasibility*, or the level at which an idea is physically realizable [157]. Additionally, any design artefact must be able to satisfy its intended goal and meet all the design specifications and engineering constraints. Otherwise, the concept would not be *useful* in any form or function. Evidently, these two sub-dimensions (feasibility and usefulness) are commonly considered factors in the various definitions of design quality, even if not explicitly termed by researchers as such.

Novelty is a less common design metric to associate with quality. Still, many researchers consider novelty in terms of ideation effectiveness and divergent thinking [63], with a common definition being the uniqueness of a design within a pre-defined set of concepts [158]. Important in ideation applications, novelty can lead to a higher probability of producing higher quality solutions via expanding the design space. In terms of product deployment, the novelty of a design can set products apart from each other, especially when products are similarly effective in their function. Therefore, novelty is a facet of innovation in an increasingly competitive marketplace, and a dimension considered moving forward [159,160].

A previous study correlated quality with each of these dimensions (feasibility, usefulness, and novelty), on a corpus of design concepts originating from a cognitive experiment with 1106 designs [135]. The concepts represented solutions to 4 distinct design problems (electricity: n = 254, phone: n = 290, joint: n = 276, surface: n = 286). External evaluators, all Mechanical Engineering graduate students, rated the designs on the metrics of feasibility, usefulness, novelty, and quality, each on a range from zero to two. The interclass correlation coefficient was calculated on a subset of the designs for each design metric separately *(feasibility: ICC = 0.77, usefulness: ICC = 0.65, novelty: ICC = 0.71, quality: ICC = 0.50). All resulted in good or excellent consistency among raters, except for quality, which exhibited only a fair consistency. Correlations between dimensions, e.g., quality/feasibility (r = 0.43) and quality/usefulness (r = 0.73)* were moderate to strong. However, there was no correlation between novelty and quality *(r = 0.04)*. The latter result is not surprising, as the authors do not expect novelty, when considered by itself, to represent quality (i.e., novel designs may be poor designs). *But* as mentioned previously, novelty is still an important dimension for innovation. Thus, a focus of this chapter is on defining a new, and aptly named, measure for design potential as ***design innovation***, ***I*****,** which considers the feasibility, usefulness, and novelty of a design concept.

4.2 Methodology

To test the feasibility and impact of utilizing LSA to determine inspirational stimuli in response to the current design state of the designer, a human cognitive experiment was developed. The experiment explored the effects of two different LSA-determined distances of inspirational stimuli (near vs. far), as well as a control condition where participants were not provided with any stimulus. Their intermediate and final designs were evaluated across several outcome measures of interest, including feasibility, usefulness, and novelty.

4.2.1 Participants

Sixty-six participants (17 male, 49 female) were recruited for the cognitive study using a call for participation at Carnegie Mellon University and offered $10 compensation for their participation. All participants read, agreed to, and signed a consent form. Demographics consisted of both university undergraduate and graduate students from a variety of majors and research interests including Engineering, Fine Arts, Computer Science, and Social Sciences. Data from six participants became corrupted during data collection, and thus excluded from the analysis.

4.2.2 Experiment Overview

Participants recruited for the cognitive study completed the 30-minute experiment, outlined in Figure 4.1. For the entirety of the experiment, participants interacted with a graphical-user interface (GUI), coded in MATLAB, which displayed the experiment instructions, the problem statement, and a countdown timer during the problem-solving blocks. After reading through the experiment instructions, participants received the problem statement, which asked them to think of solutions to "*minimize accidents from people walking and texting on a cell phone*" (abbreviated). This design problem, adopted from work by Miller et al. [162], has been previously utilized by a portion of the current research team in similar concept generation and design ideation tasks [18, 135, 153,161].

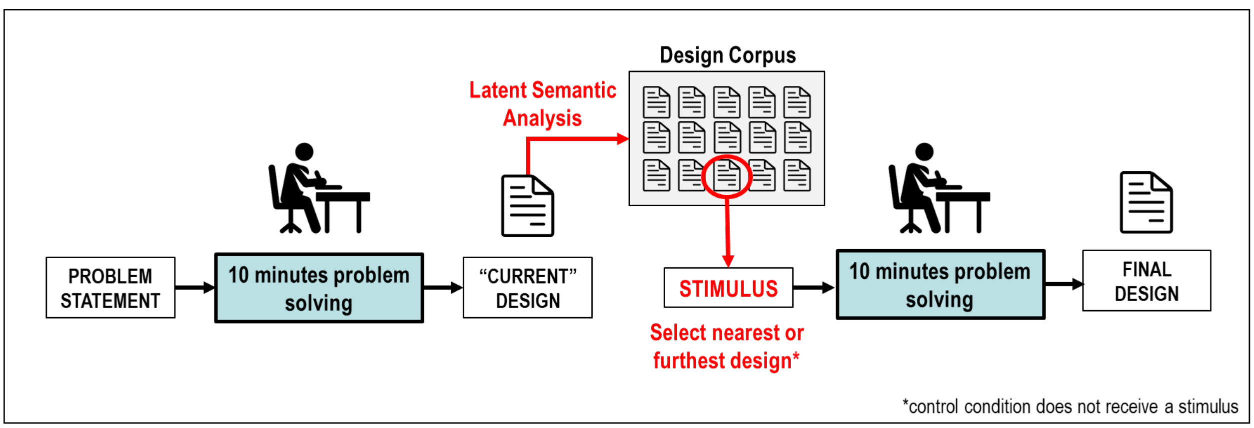


FIGURE 4.1: OUTLINE OF COGNITIVE STUDY

After reading through the experiment instructions and problem statement at their own pace, participants began generating solution concepts using paper and a digital pen (Neo Lab M1). The digital pen operated in the same way as a traditional pen but tracked pen strokes using a built-in camera (not analyzed in this study). Participants had 10 minutes to ideate and were encouraged to generate as many concepts as they wanted using any combination of textual and/or pictorial representations. At the end of the of the first 10-minute problem-solving block, the GUI instructed participants to type a 75-word textual description of one design solution in response to the following prompt: "*Please provide what you consider to currently be your best solution*". Using this textual description of each participant's current solution, LSA was run to make semantic comparisons between their solution text document and each of the 115 existing stimuli text documents within the design database. The resulting [116 x 116] output matrix from the SVD algorithm was unique to each participant, as it was comprised of both the 115 design stimuli (determined a priori), as well as a participant's newly developed design.

A balanced experimental design separated participants into one of three experiment conditions: near or far inspirational stimuli, and control, with 20 participants in each condition. Participants only saw one experimental condition during the experiment. The participants in either of the inspirational stimulus conditions (near or far), were immediately, in real time, provided with an inspirational stimulus for review (under 3 seconds of computational time). These stimuli were modulated based on their current design state. For the near inspirational stimulus condition, the stimulus provided was the *closest* stimulus within the design database (115 possible stimuli) to where they were at that point, based on the largest cosine similarity from LSA. In the far inspirational stimulus condition, the *furthest* stimulus within the database was given (lowest cosine similarity). Participants in the control condition immediately transitioned back to ideating after completing the write-up of their best design from the first ideation session. Finally, after the second 10-minute ideation period, all participants completed a write-up of their final "best" design solution. LSA was again performed between this final design write-up and the 115 set of design stimuli for data analysis purposes. The two different LSA comparisons between a participant's midpoint or final design and the 115 design stimuli allowed for a way to computationally measure the impact of the design stimuli on problem-solving behavior. By computing the semantic distance between participants' designs and the fixed stimulus space, a sense of the relative movement of a designer within this design space was extracted.

4.2.3 Design Database

The design database contained 115 possible inspirational stimuli adapted from prior work by Goucher-Lambert and Cagan [135]. During the prior research study, individuals generated 386 solutions for the same design problem employed in the current chapter. All of the 386 hand-written solutions contained a mixture of text annotations, text descriptions, and drawings. Three mechanical engineering Ph.D. students, previously trained to evaluate outcome measures (e.g., novelty, feasibility, and usefulness) of the same designs, transcribed descriptions of the content for a random subset of 115 of the 386 design solutions. Each transcription contained a minimum of 75 words. Initial pilot testing identified this word count threshold for each document as being necessary in order to obtain meaningful differences in LSA comparisons. Each of these 115 documents (text descriptions) became one of the potential inspirational stimuli. In the prior work, all 115 inspirational stimuli were evaluated for their novelty, feasibility, and usefulness [135]. Consequently, this study utilized these same rating criteria to make assessments regarding the influence of stimuli on design solution outcomes.

4.2.4 Analysis of Design Solutions Generated During the Cognitive Study

External raters evaluated both the intermediate designs (D_1, after the first 10-minute ideation period) and the final designs (D_2, after the second 10-minute ideation period) on the following outcome measures to understand the impact of the computationally selected inspirational stimuli:

1. **Feasibility**: rated on an anchored scale from 0 (the technology does not exist to create the solution) to 2 (the solution can be implemented in the manner suggested).
2. **Novelty**: rated on an anchored scale from 0 (the concept is copied from a common and/or pre-existing solution) to 2 (the solution is new and unique). Of note: "novelty" is considered as the uniqueness of the solution with respect to the entire solution set.
3. **Usefulness**: rated on an anchored scale from 0 (the solution does not address the prompt and/or consider implicit problem constraints) to 2 (the solution is helpful beyond status quo).
4. **Quality**: rated subjectively by each rater on a scale from 0 (low) to 2 (high).

Two trained mechanical engineering Ph.D. candidates, both specializing in design methodology, performed all ratings for solution characteristics. The intraclass correlation coefficient (*ICC*) assessed the consistency between the two design raters using a 25% subsample of the entire dataset. The *ICC* values for novelty (0.78), feasibility (0.65), and usefulness (0.79) were all good or excellent [67]. The *ICC* value for quality was 0.51 (moderate) and therefore excluded from further analysis for being markedly lower than the other measures.

In addition to the metrics noted, participants also provided self-ratings regarding the perceived usefulness and relevance of the provided inspirational stimuli. This information was collected at the end of the experiment, after participants had already written the description of their final design. Participants provided a self-rating for each metric ranging from 1 (low) to 5 (high).

The goal in collecting these ratings was to investigate whether or not the computationally determined levels for the inspirational stimuli (near vs. far) aligned with participants' perceptual notion of these categories. Participant self-ratings were not compared to expert evaluations, and therefore a separate scale with a wider range was utilized.

4.2.5 Design Innovation Measure: A Measure to Assess the Overall Potential of a Design Concept

In addition to assessing the underlying sub-dimensions of the designs (e.g., feasibility, usefulness, novelty), it is important to holistically determine an idea's overall potential. To explore the overall "goodness" of the designs in the current research study, the research team adopted the design innovation measure, *I*, for conceptual design assessment. Design innovation, *I*, is an encapsulating measure for the overall goodness of a concept, leveraging more well-defined design attributes. These include the feasibility *(F),* usefulness *(U),* and novelty *(N)* of a design, all of which are important for innovation. Accordingly, this measure was defined as follows:

$$\boldsymbol{I} = (\boldsymbol{F} \cdot \boldsymbol{U}) + \boldsymbol{N}\,. \quad (11)$$

To determine the accuracy and robustness of these underlying sub-dimensions to the overall goodness of concepts, additional formulations of the innovation measure, *I*, were explored in relation to quality (Table 4.1). Correlations were run between the design innovation measure and quality, not to necessarily equate the two measures, but to uncover if they follow similar trends.

The variables *F*, *U*, and *N* represent the same sub-dimensions as in Equation 11, and the weights, w_1, w_2, and w_3 are determined from a Principal Component Analysis (PCA) run on the rating data. The first formulation placed a greater penalty on feasibility and usefulness as opposed to novelty (i.e., if either *F* or *U* scores a 0, the entire (*F*-*U*) part of that formulation becomes 0). This formulation was motivated by the correlations discussed in Section 3.1.4 (i.e., feasibility and usefulness being more significant and robust than novelty). Variants two (Equation 12) and three (Equation 13) both assumed equal weighting for the three sub-dimensions, and, consequently, equal importance to innovative potential. However, the second variation allowed for more resolution in the score range and thus also when comparing designs. The multiplicative nature of Equation 13 yielded a larger penalty for scoring zero on any one of the sub-dimensions and introduces a non-linearity for comparable increases to Equation 12. Equation 14 presented a linear combination of the dimensions, weighted by the importance of each in a reduced dimensional space obtained by performing PCA. Justification for the chosen formulation (Equation 11) is presented later in the results section.

TABLE 4.1: FORMULATIONS OF DESIGN INNOVATION MEASURE, I

$\boldsymbol{I = (F \cdot U) + N}$ **(11)**
$\boldsymbol{I = F + U + N}$ **(12)**
$\boldsymbol{I = F \cdot U \cdot N}$ **(13)**
$\boldsymbol{I = w_1 F + w_2 U + w_3 N}$ **(14)**

4.3 Results

The resulting data from the methods outlined in Chapter 4.2 were analyzed to determine the impact of the computationally adaptive stimuli on design solution output. Specifically, the research objectives include: 1) to determine whether the computational method of design state detection and adaptive interventions via LSA was feasible (i.e., whether the design state can be measured in real-time), quantifiable (i.e., whether textual similarity can be used to provide near and far adaptive stimuli that are significantly different), and perceivable to designers (i.e., can participants distinguish the differences between these categorizations), and 2) to understand the impact of these adaptive stimuli on overall design outcomes (e.g., based on the design innovation measure score), and across final design sub-dimensions, including the novelty, feasibility, and usefulness of solutions.

4.3.1 Near vs. Far Inspirational Stimuli

The first objective involves determining the utility and validity in using Latent Semantic Analysis to monitor a designer's state. One way to verify the effectiveness of LSA is through examining whether or not the computational approach produced distinct categorizations of the inspirational stimuli provided to the designers. This categorization is determined using the two separate approaches described next. Figure 4.2 illustrates examples of a participant's midpoint design description and the stimulus they were provided with, in both the near and far condition.

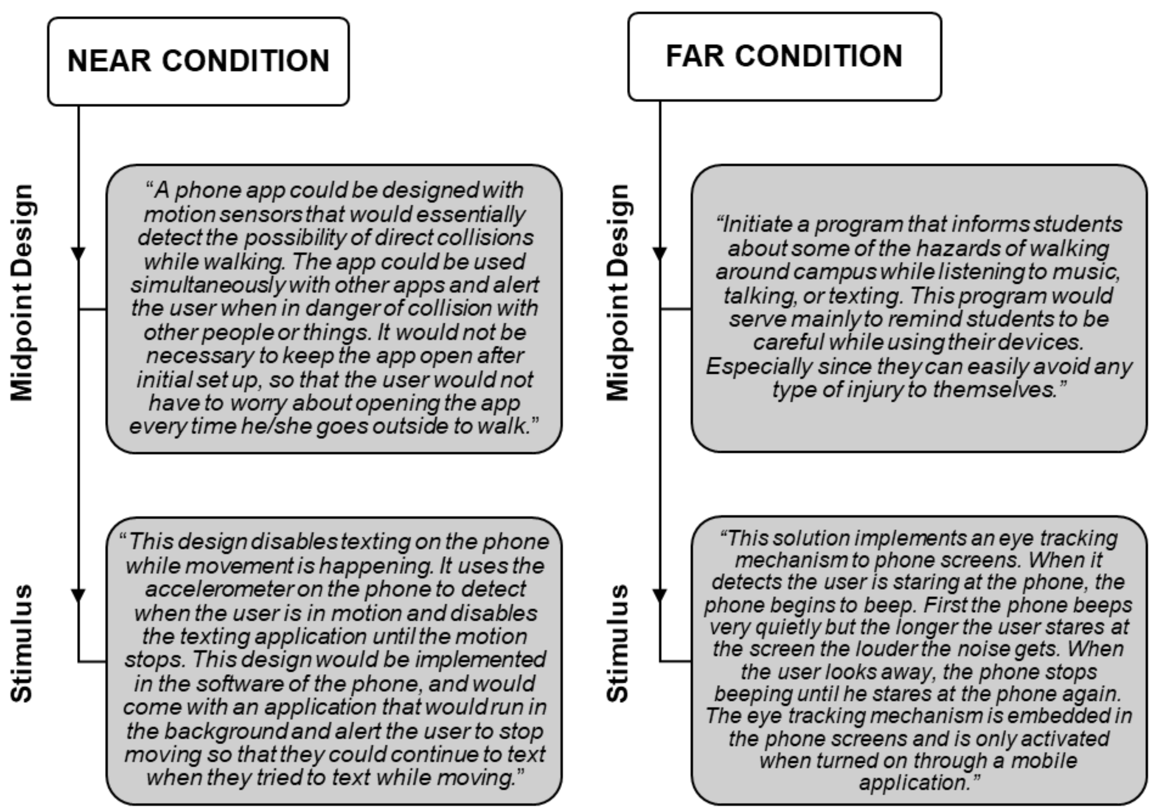


FIGURE 4.2: EXAMPLE MIDPOINT DESIGNS AND RESPECTIVE PROVIDED STIMULI FOR BOTH THE NEAR AND FAR CONDITION

The first approach to verify the effectiveness of LSA compares the average semantic similarity between the stimuli provided in each experimental condition with the state (midpoint design) of that participant. From this analysis, a clear separation between the near and far inspirational stimuli emerges (Figure 4.3). Near inspirational stimuli have an average cosine similarity of 0.54, whereas the far stimuli have a much lower similarity of 0.28 ($p << 0.01$). In other words, near inspirational stimuli, as intended, are *much* closer to the (real-time) calculated state of the designer than the far stimuli. This verified that the quantitative method for determining stimulus distance worked appropriately and created substantially difference near vs. far categories.

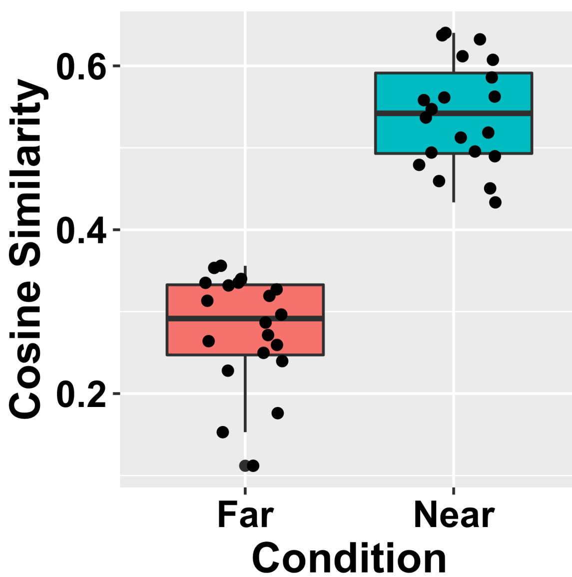


FIGURE 4.3: THE MEANING OF NEAR AND FAR STIMULI: TWO EXPERIMENTAL CONDITIONS WERE DEFINED USING LATENT SEMANTIC ANALYSIS AND THE CONDITIONS WERE SIGNIFICANTLY DISTINCT FROM ONE ANOTHER (BOXES SHOW UPPER AND LOWER QUARTILES)

The second approach leverages ratings-based data collected at the end of the experiment from participants. Each participant was asked how: 1) relevant their provided inspirational stimulus was to their current design on a 1 (not relevant) to 5 (very relevant) Likert scale and 2) helpful their provided stimulus was in developing a solution in response to the problem, again on a scale from 1 (not helpful) to 5 (very helpful). Results indicate that participants perceive near inspirational stimuli as significantly more relevant to their intermediate design solutions than the far stimuli (near: $\mu = 4.225 \pm 0.16$ S.E., far: $\mu = 3.35 \pm 0.36$ S.E., $p < 0.02$, $d = 0.70$). However, participants only found near field stimuli to be marginally more helpful during problem solving compared to the far stimuli (near: $\mu = 3.5 \pm 0.27$ S.E., far: $\mu = 3 \pm 0.35$ S.E., $p < 0.13$, $d = 0.36$). These results validate the computational approach to identify significantly different categorizations of near and far stimuli in response to the current state of designers. Furthermore, these categorizations match the perceived distances of the designers. However, designers perceive both conditions of stimuli as equally helpful to problem solving.

4.3.2 The Impact of Near vs. Far Inspirational Stimuli on Design Problem Solving

The most important goal of this chapter involves understanding how these computationally derived stimuli affect design ideation. The overall impact of each stimulus on a participant's design output can be measured in a variety of ways. In this chapter, the two methods employed are: 1) the amount of convergence on the stimulus by the designer and 2) the designer's relative movement within the design space.

The amount of convergence refers to the semantic similarity between the final design and the provided stimulus for both the near and far conditions based upon the LSA cosine similarity value (Figure 4.4.A). From this analysis, results indicate that participants provided with semantically near stimuli converged significantly closer to those stimuli by the end of the experiment ($p \ll 0.01$, mean cosine similarity values: near: $\mu = 0.33$, far: $\mu = 0.13$). However, while participants' designs remain more similar to near stimuli at the end of the experiment, far stimuli may have had a larger impact in the amount of participants' "movement" within the design space. A relative measure of the overall distance was determined by calculating both the semantic similarities between participants' first design and the stimulus, as well as the final design and the stimulus, and taking the difference between them (Figure 4.4.B). The distances were calculated relative to the design stimuli themselves, because in order to compare this distance across participants, there needed to be a common reference point across participants. The design stimuli served as these common points of reference within the design space. From this analysis, there is a significant difference between the two conditions ($p < 0.016$). Participants provided with far inspirational stimuli move a greater distance in the design space from the beginning to the end of the design ideation period (near: $\mu = 0.07$, far: $\mu = 0.12$).

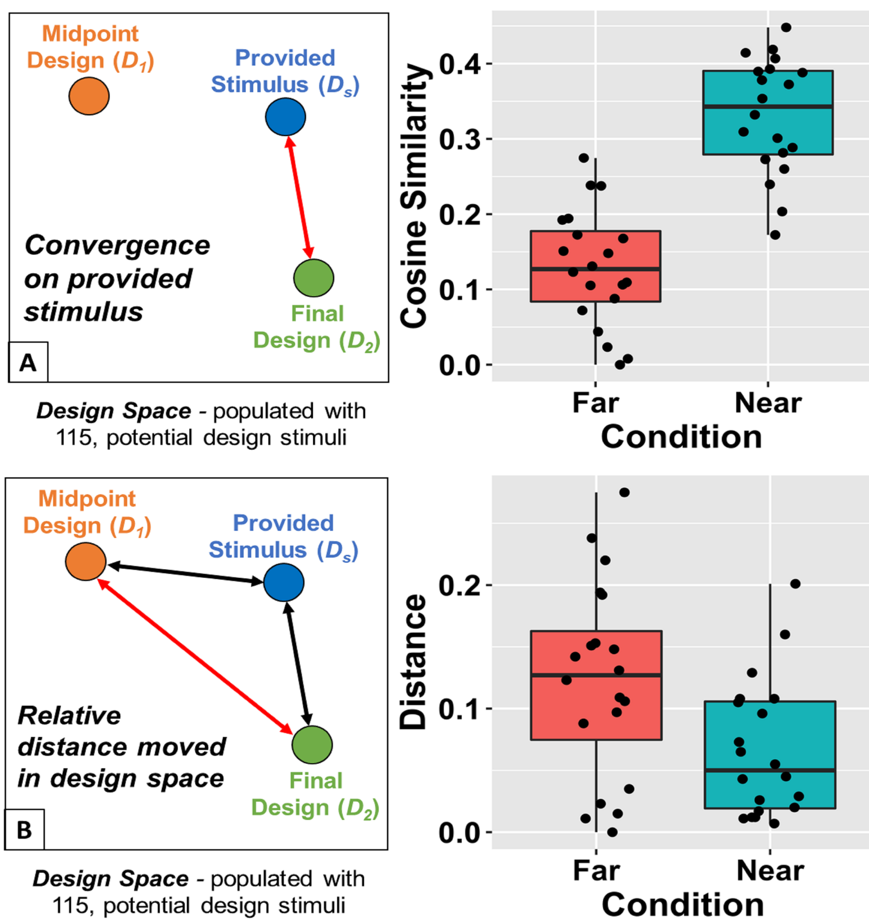


FIGURE 4.4: THE IMPACT OF INSPIRATIONAL STIMULI ON FINAL DESIGNS. THE LEFT SIDE OF FIGURES 4(A) AND 4(B) CONCEPTUALLY ILLUSTRATES THE DISTANCE MEASURED IN THE ACCOMPANYING PLOTS. A) PARTICIPANTS' FINAL DESIGN CONCEPTS ARE MORE SIMILAR TO THE STIMULUS WHEN PROVIDED A NEAR STIMULUS; (B) FAR INSPIRATIONAL STIMULI LED TO MORE RELATIVE MOVEMENT (I.E., DISTANCE) WITHIN THE DESIGN SPACE.

4.3.3 The Impact of Near vs. Far Inspirational Stimuli on Sub-Dimensions of Final Designs

In order to understand the impact of different types of computationally derived inspirational stimuli on sub-dimensions of final designs (novelty, feasibility, and usefulness), the expert ratings of these design metrics are used. As discussed previously, two trained experts evaluated both the midpoint (D_1) and final design solution (D_2) concepts produced by each participant during the cognitive study. To completely understand the impact of the stimuli on performance, one needs to consider where a designer ended up (i.e., their D_2) in reference to where they started prior to an intervention (i.e., their D_1). By analyzing performance in this manner, one can see if providing a near or far stimulus is beneficial or detrimental to problem solving. Using these ratings, the difference between the final design and the prior design is calculated separately for each of the sub-dimension metrics (novelty, feasibility, usefulness), and each experimental condition (Figure 4.5). Results indicate that there is no significant difference between the conditions in terms of novelty. In other words, providing a participant with a near, far, or no stimulus did not significantly increase or decrease the rarity of their designs from D_1 to D_2. However, intervening with semantically near inspirational stimuli significantly increases the feasibility of designs compared to providing no stimulus ($p = 0.05$, $d = 0.5$). Additionally, providing semantically far stimuli significantly decreases the usefulness of designs ($p < 0.01$, $d = 1.1$).

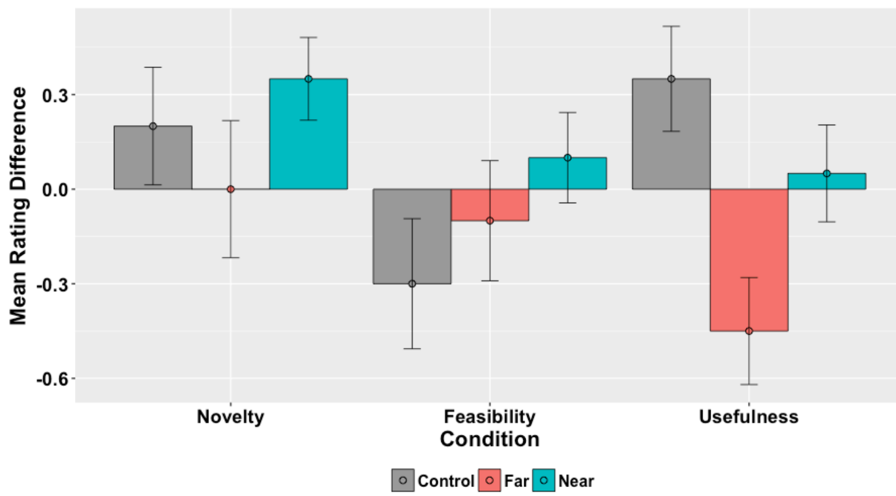


FIGURE 4.5: MEAN DIFFERENCE BETWEEN FINAL (D_2) AND INTERMEDIATE (D_1) DESIGNS FOR THREE METRICS, SEPARATED BY EXPERIMENTAL CONDITION

4.3.4 Exploring Multiple Formulations of the Design Innovation Measure

Correlations between design quality and design innovation for the different formulations of the design innovation measure (Table 4.1) are presented in Table 4.2. The results in Table 4.2 are separated based upon the 4 distinct design problems explored in prior work [135]. Specifically, these include: "*A device to remove the shell from a peanut in areas with no electricity*" (electricity), "*A way to minimize accidents from people walking and texting on a cell phone*" (phone), "*A device to immobilize a human joint*" (joint), and "*A device that disperses a light coating of a powdered substance over a surface*" (surface) [136,162-164]. The average sub-dimensional weights found through Principal Component Analysis (which was run separately on the three different design problems) was: $w_1 = 0.77$, $w_2 = 0.54$, and $w_3 = 0.37$. Overall, the chosen formulation of Equation 11 shows the highest correlation between innovation and quality on all 4 design problems. Along with its simplicity, Equation 11 is the chosen design innovation measure:

$$\boldsymbol{I} = (\boldsymbol{F} \cdot \boldsymbol{U}) + \boldsymbol{N}. \qquad (11)$$

However, it should be noted that all formulations exhibit fairly strong correlations, which supports both the accuracy and robustness of the three chosen sub-dimensions. Furthermore, when taken together, they are representative of the overall goodness of designs and possess merit as a useful design measure, without the burden of the less consistent quality evaluation.

TABLE 4.2: CORRELATIONS BETWEEN QUALITY AND FOUR FORMULATIONS OF THE DESIGN INNOVATION MEASURE

Problem \ Equation	(1)	(2)	(3)	(4)
Electricity	0.60	0.56	0.46	0.57
Phone	0.80	0.74	0.71	0.73
Joint	0.72	0.58	0.044	0.71
Surface	0.64	0.60	0.40	0.62

4.3.5 The Impact of Inspirational Stimuli on Overall Final Design Innovation

An alternate method to examine the impact of the inspirational stimuli is to examine whether a specific intervention led to an *overall* better final design (instead of focusing on sub-dimensions as described previously). Recall that one of the design metrics originally rated by the external evaluators was the quality of designs. However, the ICC value for quality was much lower than the other design metrics *(feasibility: ICC = 0.65, novelty: ICC = 0.78 usefulness: ICC = 0.79, quality: ICC =0.51).* Consequently, due to this inconsistency between raters, quality cannot serve as a consistent measure of impact of the inspirational stimuli. Instead, the design innovation measure I is used to holistically encapsulate the goodness of design concepts.

Utilizing the design innovation measure, the overall innovative potential of the design stimuli, $I(D_S)$, and both participants' intermediate and final designs ($I(D_1)$ and $I(D_2)$, respectively) are calculated (Equation 15). Similar to the analysis presented previously (Figure 4.5), performance is assessed by examining the difference in innovation scores between D_1 and D_2. Again, the results support a similar finding, with no significant effect in the change of innovation scores (either increase or decrease) in relation to the stimulus condition.

However, it is not completely accurate to only consider whether a stimulus is near or far when measuring its impact on problem solving. One limitation to the above analysis is that it does not consider the quality of the provided stimuli. For example, if a participant received a poor-quality stimulus, regardless of its distance, one should not necessarily expect their final design to improve. Because quality is hard to consistently assess, the design innovation measure is again used to represent the overall goodness of stimuli. Here, the different stimuli that designers received during the experiment varied significantly regarding this measure. Consequently, a follow-up analysis is performed, which considers the innovation score of the stimuli:

$$Corr\ ([I(D_2) - I(D_1)], [\,I(D_s) - I(D_1)]). \qquad (15)$$

Equation 15 measures the correlation between a participant's final design innovation score ($I(D_2)$) and the innovation score of the received stimulus ($I(D_S)$), both in reference to their intermediate design solution ($I(D_1)$). From this analysis (Figure 4.6), it can be seen that an inspirational stimulus with a higher innovation score, relative to the intermediate design, is significantly correlated with a better final design (i.e., an increase in I from D_1 to D_2; $r(38) = 0.67, p < 0.001$). To ensure this correlation was independent of the bias introduced via the transformation in Equation 15, an additional analysis was performed. First, 1000 random samples for each of $I(D_1)$, $I(D_2)$, and $I(D_S)$ were drawn from a uniform distribution and fed through Equation 15 (r_{bias}). Next, 1000-tuple samples (each tuple set containing an $I(D_1)$, $I(D_2)$, and $I(D_S)$) were taken from the experimental data (with replacement) and fed through Equation 15 (r_{exp}). A Fisher z-transformation showed that the empirically derived correlation value (r_{exp}) was stronger and significantly different from r_{bias} ($p << 0.001$), revealing the independence of the result from any introduced bias. Using the aforementioned distribution to determine r_{exp}, the following correlation value and 95% confidence interval were obtained: $r_{exp} = 0.67$, 95% CI [0.63, 0.70].

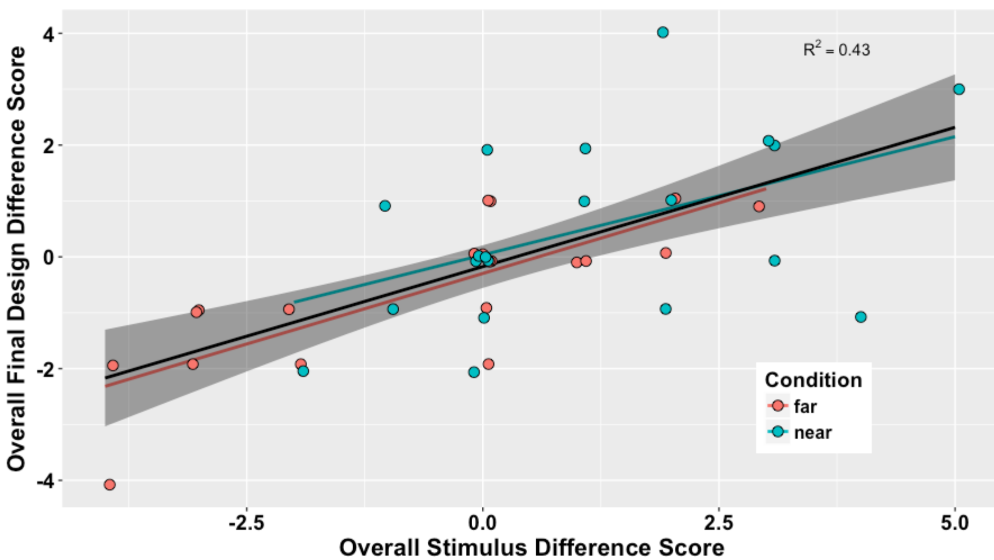


FIGURE 4.6: THE OVERALL INNOVATIVENESS OF THE PROVIDED INSPIRATIONAL STIMULI IS PREDICTIVE OF THE FINAL DESIGN OUTCOME

The positive correlation between the innovation score of the stimulus and final design is true regardless of whether a participant received a near or far stimulus. Conversely, a participant that received a less innovative stimulus is more likely to produce a less innovative final design. These results provide an interesting, tangential perspective. When intervening during problem solving via inspirational stimuli, adapting stimuli based on relative quality (as measured here by design innovation, I) highly correlates with final design outcomes.

4.4 Discussion

Overall, this chapter provides an initial step toward real-time intervention during engineering design problem solving. Unlike other work that provides inspirational stimuli for designers a priori, the methods introduced here respond dynamically to an evolving state of participants' design output. These stimuli are related design solutions sourced from a pre-existing database; they are intended to increase designers' ability to recall useful ideas from memory that may aid in their ability to generate solutions with increased positive characteristics (e.g., feasibility and novelty; see discussion in Chapter 4.1). In this chapter, interventions are provided midway through problem solving and adapted to designers' current solution output. The adapted stimuli, determined using Latent Semantic Analysis (LSA), represented solutions either semantically near or semantically far from designers' present solutions. Thus, these stimuli occupy positions in the semantic design space either close to or far from the designers' relative location.

The results from this study demonstrate the applicability of semantic similarity measures, such as LSA, for identifying stimuli based on the current state of a designer. When the semantically near and far stimuli are extracted from the design space, two distinct and significantly different (in terms of semantic similarity) clusters emerged. This supports the notion that computationally defined near (or far) stimuli are, indeed, near and far. Furthermore, it provides evidence that the design space of stimuli contained designs distinct enough from one another. If the design space did not contain sufficiently distinct designs, it would not have been appropriate to categorize the designs as near and far. Additionally, results from the qualitative analyses showed that participants in each condition perceived the stimulus provided to them as equally helpful. Participants self-rated near stimuli as significantly more relevant to the design problem compared to far stimuli, but not significantly more helpful. Therefore, participants perceived the inspirational stimuli as equally helpful, but not significantly different in terms of their relevance.

While this chapter applied LSA to adaptively select inspirational stimuli, there are other possible approaches. LSA bases its comparisons on semantic similarity; as such, this method can only handle and compare *textual* outputs of designers. Other topic modelling methods to handle text-based comparisons are Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet Allocation (LDA) [104,106]. The work presented in this chapter demonstrates LSA's ability to quickly find relevant stimuli for use by designers. However, LSA could be compared to other approaches for adaptively finding textual stimuli for designers. Furthermore, theoretically, there is nothing preventing similar vector-based similarity comparisons from being made between images or other modalities of stimuli. Future work should also consider additional modalities of inspirational stimuli and ways to logically compare similarities between two or more different modalities other than text (i.e., mapping conceptually near and far between sets of images).

After demonstrating the distinctiveness of both near and far inspirational stimuli, this study then explored their impact on design outcomes. In this chapter, evaluations of midpoint and final design solutions were used to assess the performance of the designers. Specifically, external raters evaluated the feasibility, usefulness, and novelty of each design. Results showed that participants provided with no stimuli had significantly less feasible design solutions, while those provided with far stimuli had significantly less useful design solutions. From this perspective, participants provided with near stimuli benefitted more from the intervention than those provided with the far stimuli. This corroborates previous findings from the authors which suggests near inspirational stimuli may be more helpful than far inspirational stimuli [135]. In contrast to the far and no stimulus conditions, the designs in the near condition were not negatively affected in their feasibility or usefulness. Nonetheless, an important piece to the puzzle is still missing: the overall goodness of the stimuli themselves.

This chapter utilizes a measure to capture the overall "goodness" of solutions to assess design concepts. The motivation for this measure stems from the ambiguity in a common design metric prevalent throughout the engineering design field: quality. Many works utilize this metric when assessing design artefacts, yet its holistic terminology and the variations in its definition, sometimes cloud the dimensions underlying its true meaning. For this reason, this chapter explores the use of a different measure. Literature review supports the concepts of feasibility and usefulness when considering the overall goodness of a design. Both are included in our newly defined ***design innovation measure***. Novelty, while not as common of a sub-dimension for overall goodness, adds the element of uniqueness to the measure.

In this chapter, different forms of the innovation measure are presented, motivated, and analyzed. Correlations with quality corroborate the underlying dimensions and the specific formulation proposed in this chapter (i.e., Equation 11) [90,153]. These four specific formulations were considered to explore a general subset of variations on the sub-dimensions and understand the sensitivity of the formulation to these variations. Future work should further investigate the stability and robustness of the proposed formulation with additional datasets.

Participants' midpoint and final solutions, as well as the provided inspirational stimulus, were analyzed using the newly defined design innovation measure. A similar analysis was carried out as described previously, but this time looking at the change in innovation scores between final designs and midpoint designs. Again, no significant change existed between conditions (near, far, or no stimulus). This suggests that providing an adaptive stimulus, either semantically near or semantically far, does not improve the innovation potential of solutions. In contrast, the innovation score of the provided stimulus did impact designers' outcomes: when provided with an inspirational stimulus with a relatively higher innovation score, designers were more likely to produce a more innovative final design, regardless of other inspirational stimulus conditions.

Previous research has proposed a "sweet spot" for analogical stimuli, representing an analogy that lies somewhere in between the near and far fields and yields the most benefit for positive design outcomes [88]. Yet, the stimuli in this experiment occupy the far ends of the spectrum, as opposed to this sweet spot. With the analytic nature of LSA, perhaps a more precise designation of this sweet spot can be identified in order to understand where between the near and far fields this sweet spot truly lies. Based on the present study, it may be important to not only consider the distance of the provided stimulus, but also its relative innovative potential.

While the stimulus conditions did not outperform the control condition on all dimensions of design metric outcomes, one explanation for this underperformance is the interruption endured during problem solving. Prior research is inconclusive regarding whether brief interruptions hurt or help problem solvers. For example, work by Gero et al. demonstrated that interruptions during design ideation are hurtful due to a cognitive shift from primary to secondary tasks [165]. Other works support this claim for problem solving in general, and have considered ways to mitigate the effects of disruptions [166–168]. Conversely, work by Sio et al. found that interruptions are beneficial [169]. Despite this, much of the previous research on the effects of design stimuli on design outcomes does not specifically study timing as a variable. Research from Tseng et al. found that stimuli are most helpful after the development of an open goal [145]. One thought is that the time when stimuli are typically provided during design studies may not occur during a period of deep problem solving (e.g., near the beginning of an experiment), and therefore do not cause this cognitive shift to occur. The timing of the stimulus intervention is not specifically studied in this chapter; however, it is an area in need of future investigation.

Another factor that may have impacted the inspirational stimuli conditions is team versus individual efforts. Providing example solutions halfway through problem solving may be analogous to two members on a team interacting or sharing ideas with each other at a set interval (i.e., independent work on the same design challenge with one opportunity to exchange current ideas/solutions). Because of deficiencies in group problem-solving, nominal teams (teams composed of individuals who do not collaborate during problem solving) have been shown to outperform teams in a variety of problem domains including brainstorming, conceptual design, configuration design, and verbalization tasks [41,44,80,170]. Under this theory, designers that did not receive an example solution should perform better because they did not collaborate with the computer team member (nominal teams). Those that did receive an example solution are hindered because of the interaction with the computer. To fully corroborate this theory, future work is needed to study this type of human-computer interaction of problem solving in "hybrid team" (human-computer) environments. The computational framework in this chapter provides promise for an effective design tool of the future. Forthcoming research can address how and when to intervene during design problem solving. More specifically, these open questions involve which modalities of interventions are best for designers, and when during the problem-solving process is most effective for them to be applied.

4.5 Summary

This chapter utilized Latent Semantic Analysis (LSA) to adaptively select relevant inspirational stimuli to aid designers during a cognitive study. Sixty designers were split into three conditions:

two conditions that modulated the distance of the provided inspirational stimulus and a control condition in which no stimulus was provided. The stimuli were selected based on the LSA comparison between the current status of the designers' output and a database of design solutions. One key contribution of this chapter is the adaptive determination of which stimulus to provide to a designer based on their current output of design activity. Results indicate that LSA is a viable technique to make interventions with inspirational design stimuli. Using a newly defined measure of design innovation, this chapter also investigates the impact of the inspirational stimuli intervention on design output. The overall innovativeness of the provided stimuli significantly correlated with the overall innovativeness of the designers' final design solutions. In fact, the overall innovativeness of a stimulus had a greater impact on a designer's output than the relative distance of the stimulus. This highlights the need to provide stimuli to designers not only at specific distances relative to the solution space, but also while assessing the innovative potential of the inspirational stimulus. While more work is needed to automate the process of providing designers with positively impactful inspirational stimuli, the real-time computational approach presented in this chapter is a critical step towards realizing this goal.

Chapter 5 : A Data-Driven Approach to Process Management

5.1 Introduction

Fundamental in nearly all facets of engineering practice, engineers work in teams [2,171]. Teams benefit from amalgamating diverse sets of technical skills, experiences, personalities, and perspectives for problem solving [172,173]. Teams, along with their corresponding attributes and characteristics, have been well studied in the context of engineering tasks, which often require the success of members within a team to work efficiently with each other [9,44,174]. Engineering problems can be challenging and complex, requiring multiple disciplines working together to exchange information regarding constraints, goals, and converge their expertise. Even process-oriented features, such as mere effort, have large impacts on team performance. With a theoretical basis in social psychology, social loafing, or when one team member contributes less due to being masked by the rest of the team [175,176], for example, can significantly hurt collective team performance. Thus, key factors such as communication, processes for engagement and information flow, and proper management can ultimately define the success or failure of a team [177,178].

Effective communication and coordination between interrelated roles are essential for solving collaborative engineering problems [179–181]. Although team members usually work on specific design tasks individually, team communication facilitates and stimulates design processes and exchange information across disciplines. Thus, from the design team's perspective, specialist design knowledge is usually embedded throughout the team and needs to be communicated to become valuable information for the design artifact to be produced [180]. Ideally, engineering teams will continue to effectively and efficiently communicate despite the problem's complexity.

However, even with the increased availability of information technology, engineering teams still struggle to communicate [180–183], leading them astray and thereby restricting their ability to collectively manage complexity in order to achieve a design solution.

Teams and communication are critical in designing complicated engineered systems, which often require managing coupled design parameters and multiple differing but interrelated factors, making the design process complex [184,185]. Artificial intelligence (AI) assistance methods have proven to be efficient in this area, supporting engineering teams in completing such challenging tasks rapidly and effectively. Engineers have used AI assistance tools to design products and explore the solution space more rapidly [186] and at different stages of the design process, including concept generation [187], concept evaluation [188], prototyping [189], and manufacturing [190], and concurrent-engineering design [191]. However, human-AI collaboration can also restrict team performance. Zhang et al. [192] reported that AI assistance hindered high-performing teams' success. Different authors have studied the impacts of AI assistance in other aspects of engineering design, including decision-making, optimization, and computational tasks [193,194], and its effects on mental workload, frustration, and effort [195,196], and its influence on the behavior of designers during the design process. While previous works cover the use of AI as assistive tools, there exists a lack of focus in the literature on the use of AI in a managerial role for the direction and guidance of teams.

Previous research highlights the power of process management on design teams, and a framework for understanding the role of real-time interventions. Gyory et. al. study the impacts of human process management on engineering design teams [91]. During a conceptual engineering design task, human managers intervene with a prescribed set of potential stimuli to affect their teams' process. This study shows that teams under this process management significantly outperformed unmanaged teams in the quality of their final design outcomes. The impacts extend to behavioral and process aspects as well, where the managed teams exhibit more engagement (contribution from all team members) and greater cohesion within their collective discourse. The current chapter seeks to explore whether such process management can be automated through an AI agent that could intervene in a similar way in real time.

To begin to enable the automation of the interventions, once such a need is identified with work by Goucher-Lambert et. al., by computationally adapting stimuli to provide aid in real time during problem solving [90]. Midway through a conceptual engineering design task, participants transcribe their best solution at that point in time, and then provided a tailored stimulus through automatic semantic comparisons via Latent Semantic Analysis (LSA) [65]. Modulating the semantic distance of the stimuli to their current design solution produces varying impacts on designers' ideation outcomes (e.g., design novelty, feasibility, usefulness, and overall innovative potential). More recent work by Gyory et. al. leverages the transcript data from the previously mentioned process managed teams and applies topic modeling techniques (including Latent Semantic Analysis and Latent Dirichlet Allocation) to computationally study design cognition and the impact of the interventions on communication [197]. The results show that with analyzing design discourse, the impact of the manager interventions can be detected, showing promise for the real-time implementation and detection via discourse information. While these previous works take steps towards real-time design process guidance, an automated system that identifies when interventions should take place (aka, triggers) and what an effective intervention at some point should be has not been automated or even considered. Advancing the power of AI opens the potential for dynamically tracking several team process measures and integrating them to determine applicable interventions to stimulate design teams' performance.

This chapter introduces an AI process manager agent to effectively guide the design process of engineering teams in real time, exploiting the performance results of team and behavioral outcomes from the human process managers demonstrated by Gyory, et al. [91]. In other words, this chapter demonstrates an effective AI agent that works in synergy with humans through interventions, resulting in a true AI-human hybrid team. The AI agent takes a data-driven approach to management, using real-time inputs to detect deficiencies in the team process and intervene at prescribed intervals. Trained on prior problem-solving team data on a similar drone design problem (that used the same experiential platform as this chapter), the AI induces ideal team process conditions over the course of the design problem. The inputs and measures tracked and integrated by the agent include team action and team communication data. To compare directly against human strategies, another experimental condition places teams under the guidance of a human process manager. Accordingly, while the AI agent takes a data-driven approach to management, this is compared to the more observational-based approach by the human process managers, though both have access to the same types of team inputs. Comparisons between the two types of management include the impact on team performance, intervention strategies, and team perceptions of effectiveness. Further insights can also be gained by the human process managers to identify the motivations (i.e., triggers) for intervening. Such insights can yield additional development opportunities of the AI process manager for real-time management.

5.2 Methodology

To study a data-driven approach to process management, this chapter develops an AI agent to process manage teams during a complex engineering design task. In a large-scale human study, two experimental conditions place teams under the guidance of either an AI agent or a human process manager. Results are analyzed via effects on overall performance and analysis of team process and manager intervention strategies. Moreover, post-study questionnaires collected from all individual team members and the human process managers assess the behavioral and perceived impacts on intervention effectiveness. Some of the more critical outcomes from the surveys relate to the relevance and helpfulness of the interventions, perceived effects on team performance, and an understanding of the rationales the human process managers identified to trigger an intervention with their teams. Prior to discussing the central artifact of this chapter, the AI process manager, the next two sections first explain the experimental framework to provide better context for the process managers and interventions.

5.2.1 Participants and Research Platform

Approved by the Institutional Review Board at Carnegie Mellon University, participants complete the experiment fully online, only able to interact with each other and the experimenter via the experimental platform. In total, 199 sophomore-to-senior engineering students at Pennsylvania State University in the United States participate in the study, recruited from two different mechanical and industrial engineering courses. Participants receive $20 in compensation for their time and effort. All participants read, agree to, and sign a consent form prior to engaging in any aspect of the experiment. Participants are randomly distributed among and across two team conditions – AI or human process management. Teams consist of five participants with one additional participant in the human managed team condition as the human manager. Data from six teams are removed due to technical issues with the platform and participants arriving late or leaving early without finishing the experiment. Altogether, data for 31 teams are obtained successfully during the experiment: 16 teams in the human process manager condition and 15 teams in the AI process manager condition.

The experimental research platform for the study, HyForm[4], simulates a drone delivery fleet design and path-planning problem [198]. Using an online collaborative design environment, the platform partners AI design agents and humans. The platform contains an embedded chat interface, allowing participants to communicate information and share their problem-solving outcomes through specific channels during the study. While this chapter describes only a high-level overview of HyForm, more in-depth details about the collaborative research platform and integrated design agents can be found in related work [185,199]. Note that the design agents in those references support the development of drones and path plans and are different from the process manager agent developed for this chapter. HyForm records all the communication, design actions, drone configurations, delivery routes, and performance metrics of each role within the teams, enabling complete reconstruction of a team's problem-solving process.

5.2.2 Experiment Overview – Experimental Timeline and HyForm Roles

Participants complete the 65-minute experiment outlined in Figure 5.1. First, they read and sign the consent form, then provided 12 minutes to complete the pre-study questionnaire, tutorials, and problem brief. The pre-study questionnaire consists of questions related to their drone design, operations, business planning, and computational design experiences, in order to control for and confirm similar levels of expertise in these domains. Participants read through two tutorials, one related to their specific role and respective HyForm interface (each role uses a different interface), and the other related to the communication tool and team structure. The problem brief lays out the

4 https://github.com/hyform/drone-testbed-server/releases/tag/2021-March-v2

mission of the company and goals, describes the team structure and roles, and provides a more in-depth explanation of their specific roles' objectives. After completion of the pre-session materials, the first 20-minute problem-solving portion begins. Throughout these sessions, an external [to the team] process manager can intervene to affect the problem-solving behaviors and processes of the team. More details related to these interventions and the process manager are discussed in the following sections. After the first session, a short, three-minute break presents participants with an opportunity to review the experimental materials (tutorials and problem briefs). Then, a second 20-minute problem-solving session commences. While similar to the first session in terms of overall objectives, the second session involves a "shock" to the customer market. After this second session, participants fill out a post-study questionnaire. For the members on the team (drone designers, operations specialists, and problem manager), this includes questions related to the perceived relevance and helpfulness of the interventions and their assessment of team performance. The human process managers fill out a different post-study questionnaire, which queries them on their strategy for intervening, the effectiveness of their interventions, and what additional types of interventions they would have liked to use.

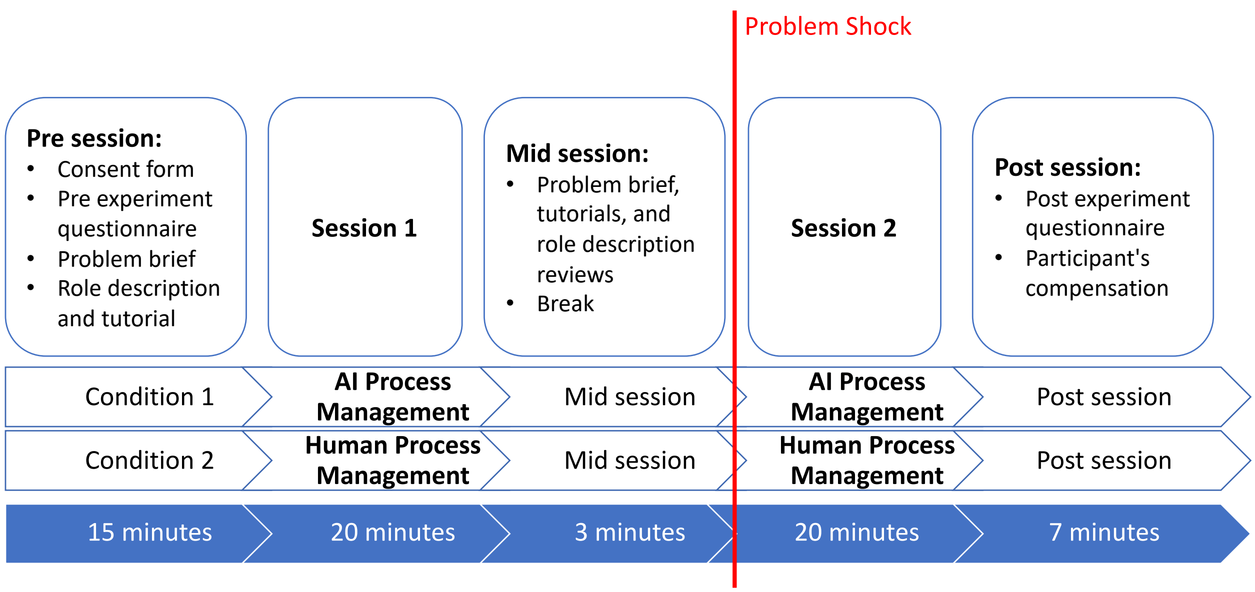


FIGURE 5.1: OVERVIEW OF THE EXPERIMENTAL TIMELINE

Using the collaborative research platform HyForm, teams design drone fleets and created delivery path plans to reach as many customers in the market as possible. A highly interconnected problem, each discipline (operations, design, and business) works together to achieve the objectives and optimize the overall profit for their team, given an initial budget of $15,000. The aforementioned problem shock refers to a change in the original market conditions to a COVID-19 scenario where more customers with low weight medical deliveries are added to the market, along with a 30% reduction in drone costs. Each team consists of five members and Figure 5.2 depicts the team structure. Each team contains two drone designers (design discipline), two operations specialists (operations discipline), and one problem manager (business discipline). The process manager is external to the team and is either a human or the AI agent. The team structure dictates the communication structure, shown by the arrows and dashed lines in the figure. The four distinct communication channels include: the design channel, the operations channel, the designer management channel, and the operations management channel. In the design channel, the two design specialists communicate with each other. Similarly, the operations channel permits the two operations specialists to interact with each other. The designer management channel allows the problem manager to communicate with both design specialists simultaneously, and the operations management channel permitted the problem manager to communicate with the two operations specialists simultaneously.

On the team, the drone designers carry out the design of the drones with different requirements for payload capacity, range, and cost. Provided with a base drone design to start, the design specialists build and modify drones by adding or removing different components (batteries, airfoils, nodes, rods, and propellers), varying their sizes and locations. Once created, the drone designs are sent to the operations specialists, who create delivery paths to reach customers in the market using the available drones' capabilities and operation costs. The problem manager is responsible for handling the company budget, choosing the customers in the market, and serves as the communication-bridging node between the design specialists and operations specialists (as depicted in Figure 5.2). Ultimately, the problem manager decides whether to approve or reject the final team plans for submission.

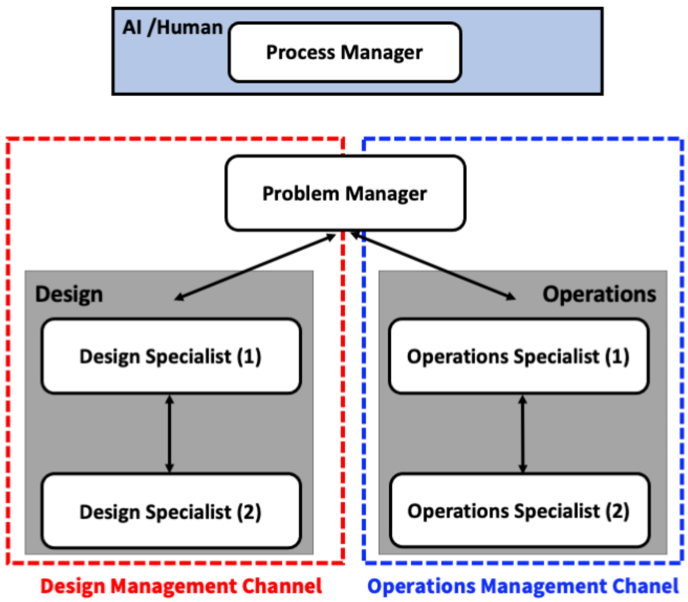


FIGURE 5.2: TEAM STRUCTURE. ARROWS INDICATE COMMUNICATION CHANNELS FOR DIRECT INTERACTIONS

5.2.3 Process Manager & Interventions

Throughout the two problem-solving sessions, an external process manager intervenes to affect the problem-solving behaviors of the team. The process manager observes features of the teams' process in real time and provides suggestions at specific times during the experiment. Being external to the team, the process manager cannot directly communicate to specific team members or help in directly solving the problem. Instead, the process manager guides with a set of prescribed, process-related interventions from a pre-defined list. Using a pre-defined list made it possible to control the types of interventions for consistency across managers, as well as reducing the additional variability induced by allowing for either a larger set or impromptu interventions. The process managers can intervene up to 12 distinct times across both twenty-minute problem-solving sessions. Figure 5.3 presents an overview of the specific timing for the interventions. Actions and communications are tracked and collected in five-minute intervals (integrated by the AI agent and shown to the human managers) and this information is considered by both experimental conditions to determine applicable interventions. The intervention opportunities occur at 2.5-minute intervals from each other, with the first available intervention starting at 5 minutes into each session (i.e., 5 minutes, 7.5 minutes, 10 minutes, 12.5 minutes, 15 minutes, and 17.5 minutes). This 2.5-minute time period for interventions balances a tradeoff between ensuring enough real-time information to collect for the process managers while also ensuring enough opportunities to intervene within each problem-solving session, in this case six times during each session. For this research, the two experimental conditions differ on whether the process manager is a human or the AI agent. Both human and AI agent process managers choose from among the same set of prescribed, process-related interventions. Therefore, for the entirety of the experiment, the teams never have any indication which type of process manager guides them.

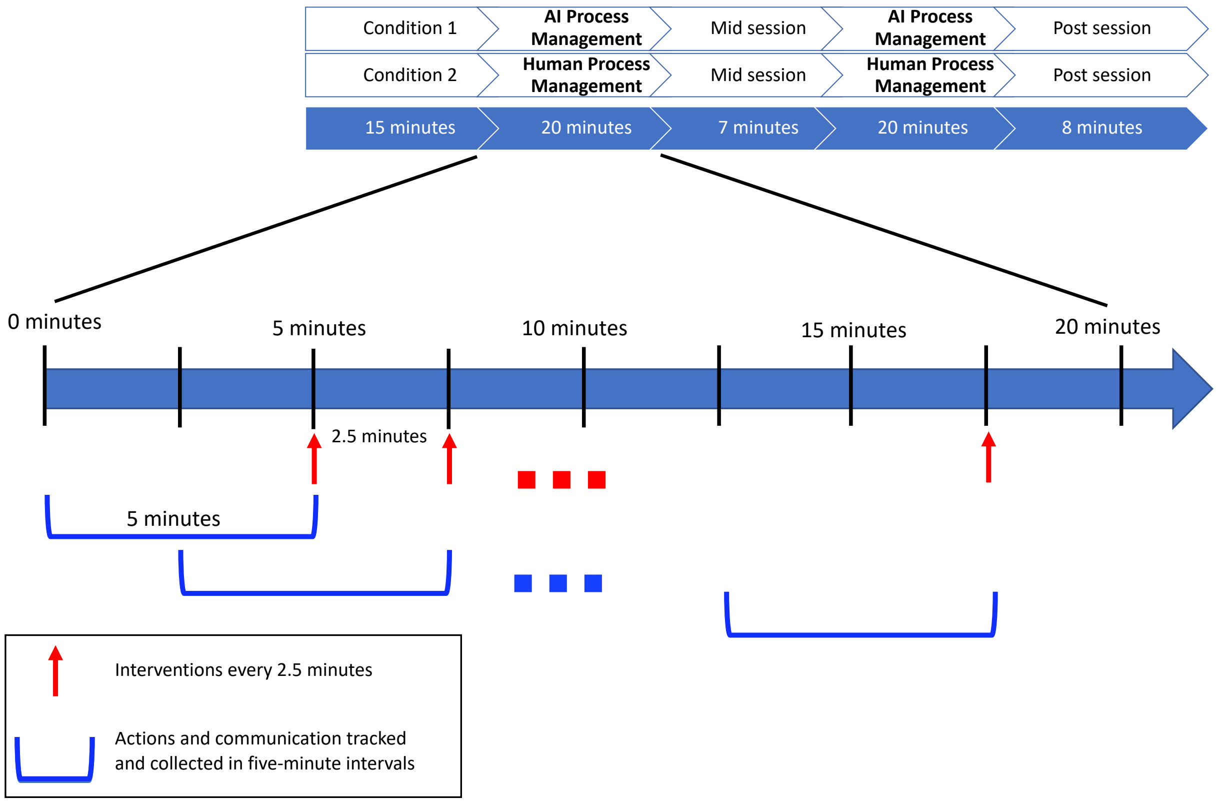


FIGURE 5.3: INTERVENTION TIMING

The human process managers have access to team information via a new mediation interface in HyForm (shown in Figure 5.4). Through the interface, the human managers observe in real time the team discourse occurring through all four communication channels, as well as the types of actions being performed by the drone designers and operations specialists. Pilot studies run prior to the actual data collection helped shape the design and improve the user experience of this interface. The right-hand side of the mediation interface lays out the set of prescribed interventions. A timer counts down until the next intervention (each 2.5-minute interval) and provides a buffer of 15 seconds to allow the human process manager to select a given intervention.

The process managers do not need to intervene at every interval. The intervention set includes a "*No Intervention*" option that does not send any message to the team. The human process managers decide to intervene based on their own assessment of the status of the team's problem solving based on the available real-time data, the same data that the AI manager has access to.

Table 5.1 and Table 5.2 show the interventions defined for this study, with action-based interventions in Table 5.1 and communication-based interventions in Table 5.2. Once the process managers choose an intervention, the intervention is delivered to the teams through a specific communication channel, shown in the right-hand column of the tables. "Design" indicates that the intervention goes to both design specialists, "Operations" indicates the intervention goes to both operations specialists and "Problem Manager" indicates that the intervention is only received by the problem manager. Three of the communication interventions go through all communication channels, and thus received by the entire team. While the AI agent implicitly identifies which team member(s) require an intervention, it does directly learn or track at the individual level, but rather at the team and discipline levels.

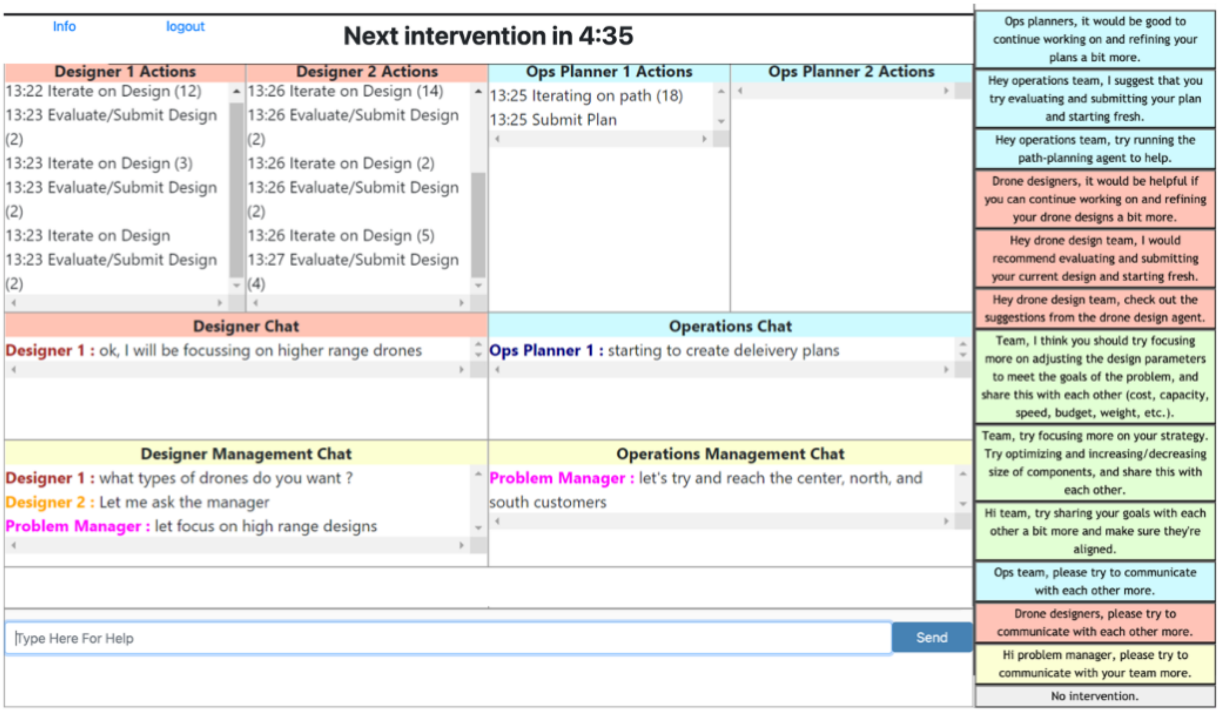


FIGURE 5.4: THE MEDIATION INTERFACE IN HYFORM FOR THE HUMAN PROCESS MANAGERS TO TRACK TEAM PROCESS (SCROLL BOXES ON THE LEFT) AND CHOOSE AN INTERVENTION (RIGHT-HAND COLUMN) AT DEFINED INTERVALS

TABLE 5.1: ACTION-BASED INTERVENTIONS, ALONG WITH THE CHANNEL WITH WHICH THEY ARE INJECTED TO THE TEAMS DURING PROBLEM SOLVING

Design Action Interventions	Communication Channels
Ops. planners, it would be good to continue working on and refining your plans a bit more.	Operations
Hey operations team, I suggest that you try evaluating and submitting your plan and starting fresh.	Operations
Hey operations team, try running the path-planning agent to help.	Operations
Drone designers, it would be helpful if you can continue working on and refining your drone designs a bit more.	Design
Hey drone design team, I would recommend evaluating and submitting your current design and starting fresh.	Design
Hey drone design team, check out the suggestions from the drone design agent.	Design

TABLE 5.2: COMMUNICATION-BASED INTERVENTIONS, ALONG WITH THE CHANNELS THROUGH WHICH THEY ARE INJECTED TO THE TEAMS DURING PROBLEM SOLVING

Communication Interventions	Communication Channels
Team, I think you should try focusing more on adjusting the design parameters to meet the goals of the problem, and share this with each other (cost, capacity, speed, budget, weight, etc.)	Design, Operations, and Problem Manager
Team, try focusing more on your strategy. Try optimizing and increasing/decreasing size of components and share this with each other.	Design, Operations, and Problem Manager
Hi team, try sharing your goals with each other a bit more and make sure they're aligned.	Design, Operations, and Problem Manager
Ops team, please try to communicate with each other more.	Operations
Drone designers, please try to communicate with each other more.	Design
Hi problem manager, please try to communicate with your team more.	Problem Manager

5.2.4 Artificial Intelligent (AI) Process Manager Computational Framework

During the problem-solving sessions, the AI process manager dynamically tracks several measures to determine the state of the team at a given point in time. These measures include communication frequency, communication semantics comprising similarity and content, and action frequency and diversity. Trained on prior team problem-solving data, the AI process manager induces the patterns of these measures temporally over the course of the problem-solving sessions from better performing teams, the goal being to reflect the behavioral dynamics of the teams, intervening when one of these measures significantly veers off course [199]. This section and Figure 5.5 present a more detailed description of the decision logic and conceptual framework for the underlying computation of the AI agent. As shown in the left-hand column of Figure 5.5, communication and action data represent the two main data input streams to the framework.

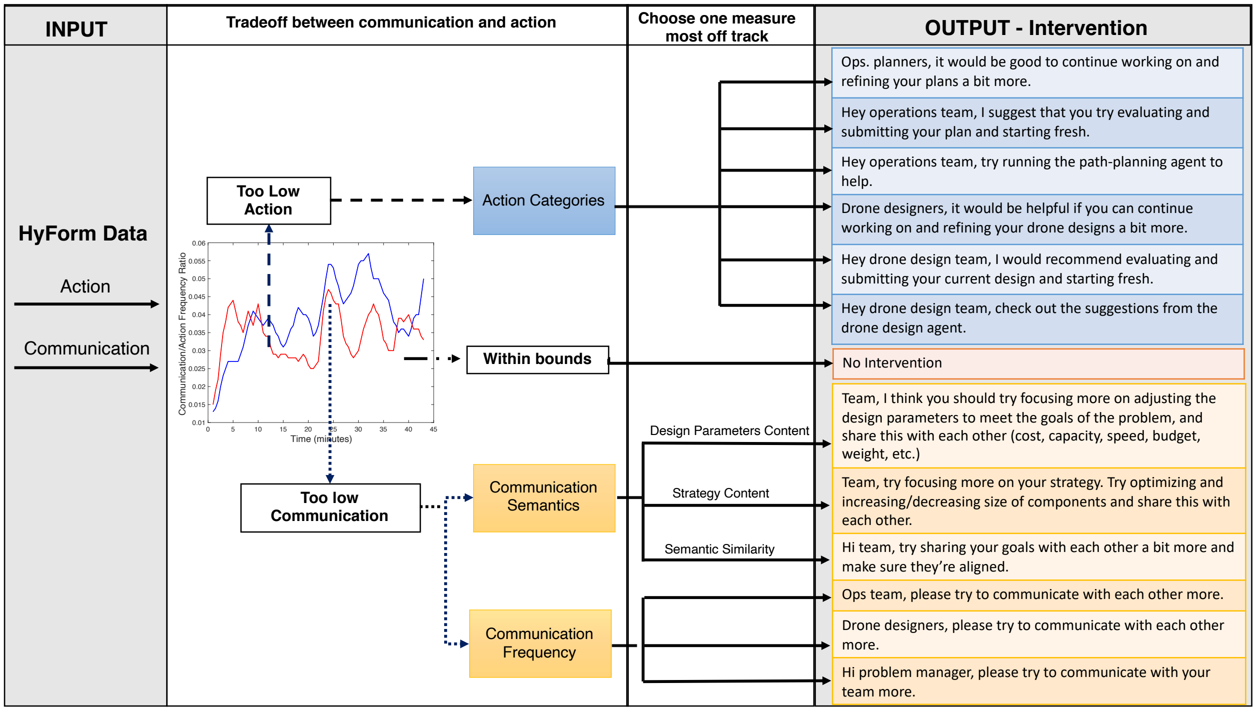


FIGURE 5.5: FLOW DIAGRAM SHOWING THE CONCEPTUAL FRAMEWORK FOR THE ARTIFICIAL INTELLIGENT PROCESS MANAGER

The AI agent utilizes the preceding five minutes of team data to determine appropriate interventions. Since the interventions occur every 2.5-minutes, the input represents a sliding window with 2.5-minutes of overlap of prior team data (as shown in Figure 5.3). As mentioned previously, these specific timings balance the tradeoff of maintaining an adequate amount of information for the AI agent to utilize as well as ensuring enough interventions throughout the experiment.

The first decision point in the computational framework (Figure 5.5) compares the overall team action frequency with the overall team communication frequency. Tradeoffs between effort spent on action versus spent on communication resulted in one of the more significant findings between the high and low performing teams from the previous HyForm experiment. Thus, this decision point leverages that finding, comparing the real-time, cumulative team communication and cumulative team action with that of the high performing teams. If the team's action frequency measure is too low then the AI agent enters the action branch (top branch in Figure 5.5), and if the team communication frequency measure is too low then AI agent enters the communication branch (bottom branch in Figure 5.5). If both action and communication frequencies are within bounds (± 1 standard error), then it chooses not to intervene.

To mathematically compute which of the branches in Figure 5.5 to enter, the AI process manager computes a weighted z-score for each communication and action (Equations 16 and 17, respectively):

$$com_i = |(z_{score_{c_i}} \times d_{c_i})|, \qquad (16)$$

$$action_i = |(z_{score_{a_i}} \times d_{a_i})|. \qquad (17)$$

In descriptive statistics, the z-score provides a quantitative and normalized approach to determine how many standard deviations a raw score lies from the population mean. In Equations 16 and 17, the subscripts c and a represent communication and action, respectively, the subscript i denotes the intervention number (i varies from $i \in [1, 12]$ for the twelve interventions across both problem-solving sessions), d represents an effect size, with the z-score calculated as shown in Equation 18,

$$z_{score} = \frac{x - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}, \qquad (18)$$

where x is the observed value taken from the real-time experiment, μ is the is population mean taken from the high-performing population data, σ is the standard deviation taken from the high-performing population data, and n is the total number of teams from the high-performing, prior data (in this case, $n = 11$). Once the z-score determines how large the difference between the sampled team communication/action data is from that of the high-performing team data, the z-scores are further weighted by an effect size ($z_{score_{c_i}}$ and $z_{score_{a_i}}$ in Equations 16 and 17). From the prior HyForm team data, the differences in action and communication frequency between the high and low performing teams considerably fluctuated over time (as shown in Figure 5.6). Figure 5.6 presents a moving average of the communication to action frequency ratio throughout that previous experiment.

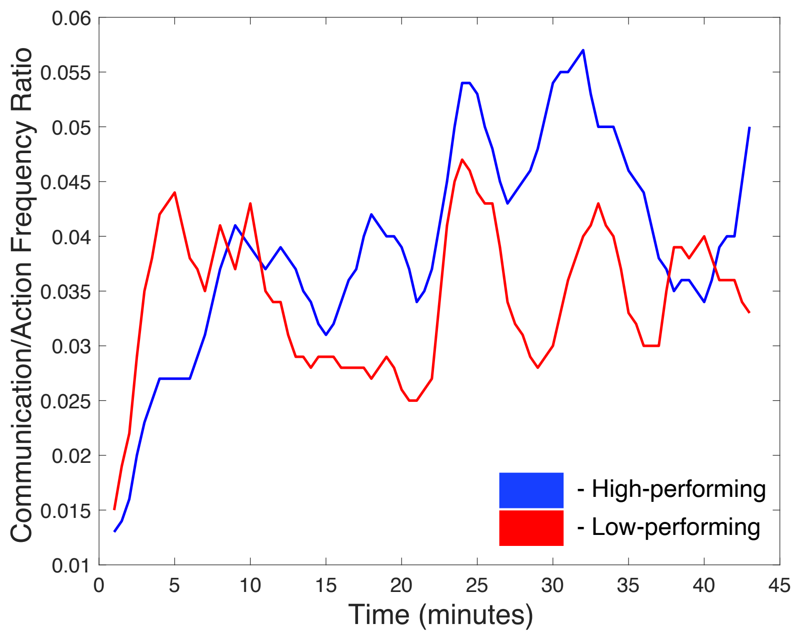


FIGURE 5.6: TRADEOFF IN ACTION AND COMMUNICATION BETWEEN THE HIGH-PERFORMING (BLUE) AND LOW-PERFORMING (RED) TEAMS FROM THE PRIOR HYFORM EXPERIMENT [189]

As shown, the differences between the high and low performing teams change over time. For example, in the latter half of the experiment, the high performing teams spend much more effort on communication. To take this additional dimension into account, the effect size (d) determines the extent of this discrepancy between the high and low performing team data over the 12 different 5-minute intervention intervals.

An additional benefit of utilizing z-scores is that they help determine whether the values are within ± 1 standard error from the high-performing data. Chosen as the threshold range for whether the sampled measure is within bounds, if both the action and communication frequencies from the sampled data are within one standard error from the high-performing teams' data, then the AI agent chooses not to intervene. Otherwise, the AI agent enters the branch with the lowest weighted z-score. The precise decision logic in Figure 5.7 determines which intervention branch to enter. Recall that the AI agent uses the preceding 5 minutes of data for computation (Figure 5.3). For example, when $i = 2$, at 7.5 minutes into the experiment (the first intervention occurs at 5 minutes), the AI agent uses communication and action data from 2.5 minutes up to 7.5 minutes. This 5-minute window holds for every intervention decision.

Once the AI process manager chooses a particular branch, similar, though unweighted, z-score calculations are computed on more specific team measures at the next decision point. These specific measures dictate the chosen intervention. For the action branch (top branch of Figure 5.5), z-scores are computed for each of the six action categories for the drone designers and the operations specialists. The sub-branches in the figure represent these six action categories. These categories include evaluating/submitting drone designs, iterating on drone designs, and running the drone design assistive agent for the drone designers; and submitting a path plan, iterating on a path plan, or running the assistive path planner agent for the operations specialists. Whichever of these action categories is most off from the high-performing team data determines the specific intervention to inject into the team. These are all discipline-level interventions, so either both drone designers or both operations specialists receive these interventions.

$for\ i \in [\,0, 12]$

$$z_{score_{a_i}} = \frac{x_{a_i} - \mu_{a_i}}{\left(\frac{\sigma_{a_i}}{\sqrt{n}}\right)}$$

$$z_{score_{c_i}} = \frac{x_{c_i} - \mu_{c_i}}{\left(\frac{\sigma_{c_i}}{\sqrt{n}}\right)}$$

$if\ \ (-1 < z_{score_{a_i}} < 1)\ \ \&\&\ (-1 < z_{score_{c_i}} < 1)$

[within bounds] – <u>No intervention.</u>

$elseif\ \ (z_{score_{a_i}} > 1)\ \ \&\&\ (z_{score_{c_i}} > 1)$

[no deficiencies in either communication/action] - <u>No intervention.</u>

$elseif\ (z_{score_{a_i}} < -1)\ \ \&\&\ \ (z_{score_{c_i}} > 1)$

[action deficiency] – <u>Enter action branch.</u>

$elseif\ (z_{score_{c_i}} < -1)\ \&\&\ \ (z_{score_{a_i}} > 1)$

[communication deficiency] – <u>Enter communication branch.</u>

$else\ (z_{score_{c_i}} < -1)\ \&\&\ \ (z_{score_{a_i}} < -1)$

$com_i = |(z_{score_{c_i}} \times d_{c_i})|$

$action_i = |(z_{score_{a_i}} \times d_{a_i})|$

$if\ com_i > action_i$

<u>Enter communication branch.</u>

$elseif\ action_i > com_i$

<u>Enter action branch.</u>

End

End

End

FIGURE 5.7: DECISION LOGIC FOR THE AI PROCESS MANAGER TO DETERMINE WHETER TO ENTER THE ACTION OR COMMUNICATION SETS OF INTERVENTIONS

If the AI agent chooses the communication branch (bottom branch of Figure 5.5), two types of communication measures are calculated: communication semantics and communication frequency. Communication semantics includes team semantic similarity and design parameter and design strategy content. Latent Semantic Analysis (LSA) computes the discourse similarity amongst the design team (treating each role as a distinct document in the model) using a singular value decomposition approach to reduce dimensionality within the discourse. Then, the average pairwise cosine similarity between role documents determines overall team semantic similarity. For the communication content, the AI agent counts sets of keywords for the problem-solving strategy and design parameter content. The keywords for design parameters include those related to the problem constraints and goals, including but not limited to velocity, payload, miles, houses, payload, profit; while the strategy keywords relate to how teams solve or adjust these design parameters and goals, such as increase, decrease, minimize, optimize, balance. The communication semantic interventions are team-level and thus received by the entire team. Communication frequency counts the total number of turns for each discipline (the three sub-branches off frequency in Figure 5.5 represent to the three disciplines of design, operations, and business). These three interventions are at the discipline-level and received by those respective roles. Once again, unweighted z-scores determine which metric is most off course. So, a z-score is calculated for all six dimensions and the largest z-scores determines which measure is most off course and induces the specific intervention.

5.3 Results

With the experimental methodology outlined in Section 5.2, Section 5.3 analyses the resulting data and compares the differences between the constructed AI agent and human process managers. These comparisons span several dimensions, including performance, intervention strategy, and perceived effectiveness. The maximum profit teams achieve across both sessions provides a measure of the overall performance. The intervention strategies between the AI and human process managers are examined by identifying the types and distributions of interventions used. Data from the post-study questionaries with the human process managers and team members provide further insight into manager strategies and ascertain how team members perceive their interventions, such as their relevancy and helpfulness.

5.3.1 Team Performance

In this chapter, team profit serves as the overall measure to identify how teams perform under the guidance of either the AI agent or a human process manager. Team profit combines the achievements of both the drone design discipline and the operations discipline, by totaling the weight of packages and food delivered within the customer market. Accordingly, due to the highly coupled and interdisciplinary nature of the problem, profit relies on the success of both disciplines: the types of drones designed and the path plans created. The problem manager can submit multiple plans throughout the experiment, though the best plan (i.e., the plan with the highest profit) serves as the primary performance measure for the team.

Figure 5.8 shows the average maximum profit the team conditions achieve in each of the experiment sessions. Recall that the second session presents a problem shock to the team; the customer market shifts to one for COVID-19 where more customers with low weight medical deliveries are added to the market, along with a 30% reduction in drone costs. Across both sessions, the two experimental conditions perform similarly ($p = 0.6$, $d = 0.14$), though a larger difference occurs in the second problem-solving session ($p = 0.1$, $d = 0.59$). While neither session presents a difference at the 5% significance level, the trend between sessions presents an interesting finding. The profit of teams guided by the AI agent marginally improves after the problem shock, while the profit of teams under human process management decreases. This trend indicates towards better adaptation in process strategies to the problem shock with the data-driven approach to management of the AI manager, which already produces similar performance levels as human management.

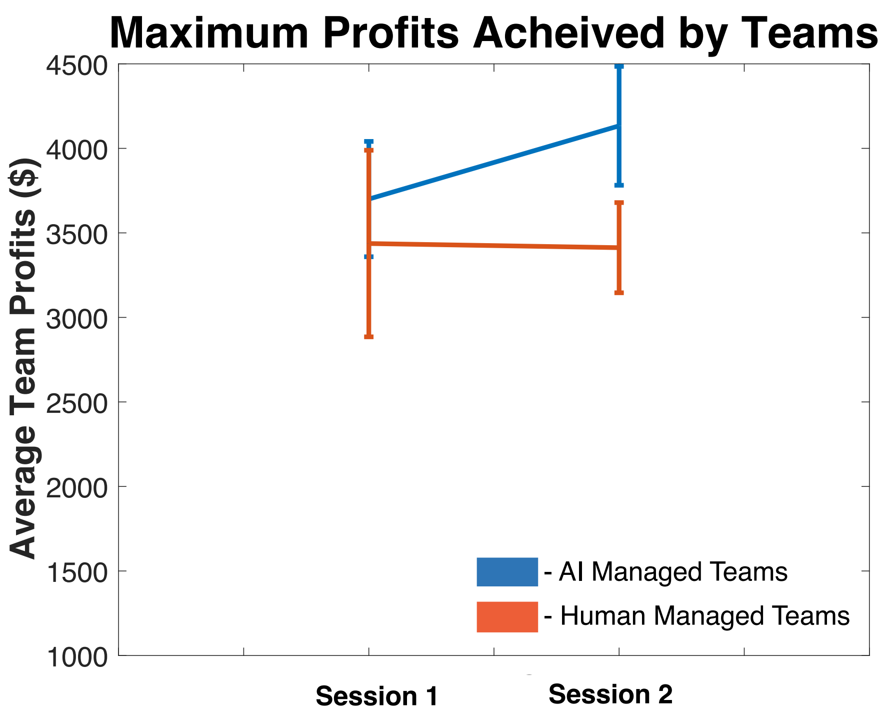


FIGURE 5.8: THE AVERAGE MAXIMUM PROFIT ACHIEVED BY TEAMS, ACROSS CONDITIONS, AND IN BOTH PROBLEM-SOLVING SESSIONS. (ERROR BARS SHOW ±1 S.E.)

Additionally, the post-study questionnaire queries participants on their perception of their teams' performance. On an interval scale from 0 to 100, with 100 being perfect, team members rate both the perceived quality of the performance of their team as well as their team's cohesion (cohesion describes "*how your team worked together,*" as defined to participants during the survey). Team member averages include the two drone designers, the two operations specialists, and the problem manager. Figure 5.9 shows that teams under guidance of the AI manager condition perceive the quality of their teams' performance significantly higher (p = 0.016, d = 0.41) and their teams' cohesion as higher in a marginally significant way (p = 0.053, d = 0.32). Furthermore, the human process managers perceive these quite differently than their respective team members. Figure 5.10 shows that in terms of both the quality of performance (p = 0.025, d = 0.63) and cohesion (p = 0.004, d = 0.81), the human process managers perceive these significantly better than the team itself. Thus, the analogous perception in performance levels hold from the viewpoint of the team itself, however, the process managers tend to relatively inflate these measures.

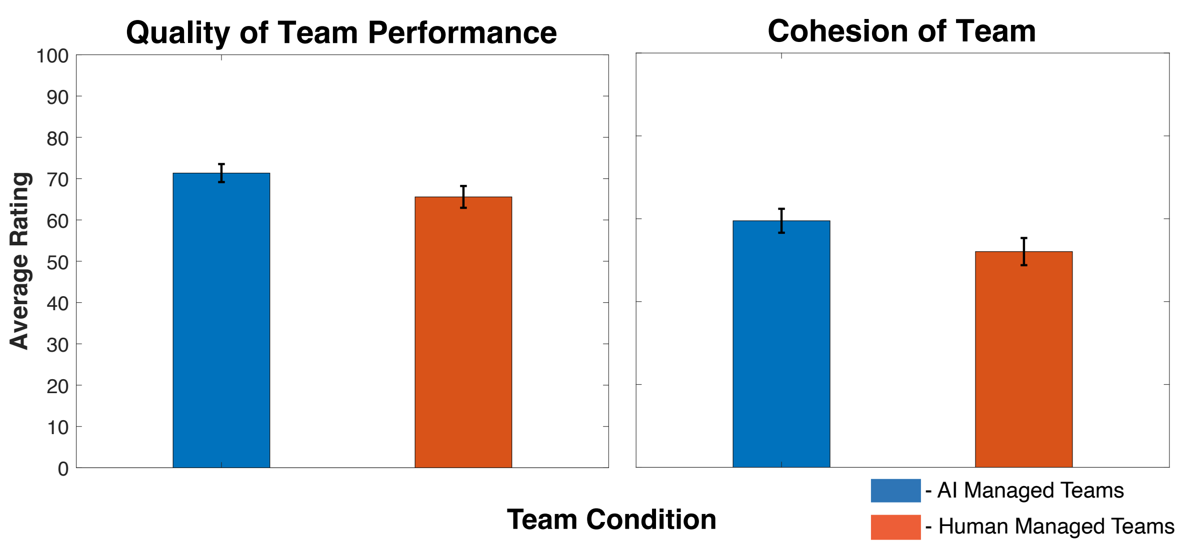


FIGURE 5.9: THE AVERAGE PARTICIPANT RATING OF QUALITY (LEFT) OF TEAM PERFORMANCE AND TEAM COHESION (RIGHT). (ERROR BARS SHOW ±1 S.E.)

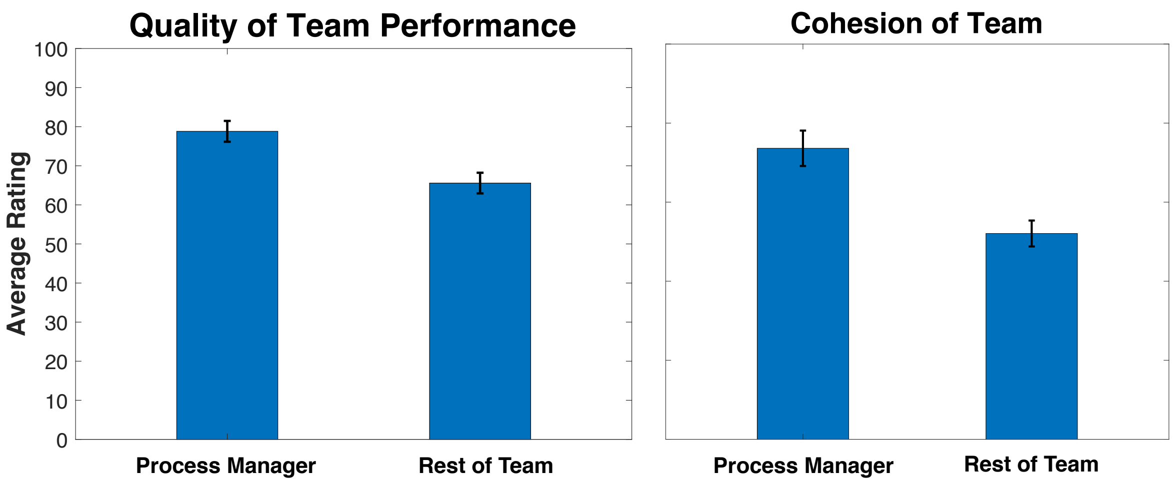


FIGURE 5.10: THE AVERAGE HUMAN PROCESS MANAGER RATING OF QUALITY OF TEAM PERFORMANCE (LEFT) AND TEAM COHESION (RIGHT). (ERROR BARS SHOW ±1 S.E)

5.3.2 Process Manager Strategies and Insights

Having demonstrated equivalent levels of performance between the human and AI process managed teams, the intervention strategies are next compared. This analysis identifies the types of injected interventions across sessions and disciplines, as well as a deeper understanding of the need for intervening from the human managers via the post-study questionnaires. From the questionnaires, the human process managers answer questions related to the effectiveness of their own interventions and the dynamics within the teams they oversaw that triggered them to offer guidance when they did.

In total, the human process managers intervene 127 times using all the interventions from the prescribed set, while the AI agent intervenes 167 times but only using 8 out of the prescribed interventions. While the AI process manager chooses to intervene more frequently, the human process managers tend to use more team-based interventions (37% of the humans' interventions compared to 29% of the AI agent's interventions). Figure 5.11 presents the distribution of interventions across team sessions and disciplines and the proportion of intervention categories.

The numbers next to each of the bars in the figure represent the raw counts. These are categorized by the two main intervention types: action-based interventions and communication-based interventions. Table 5.1 and Table 5.2 from the previous section present which interventions fall into the respective categories.

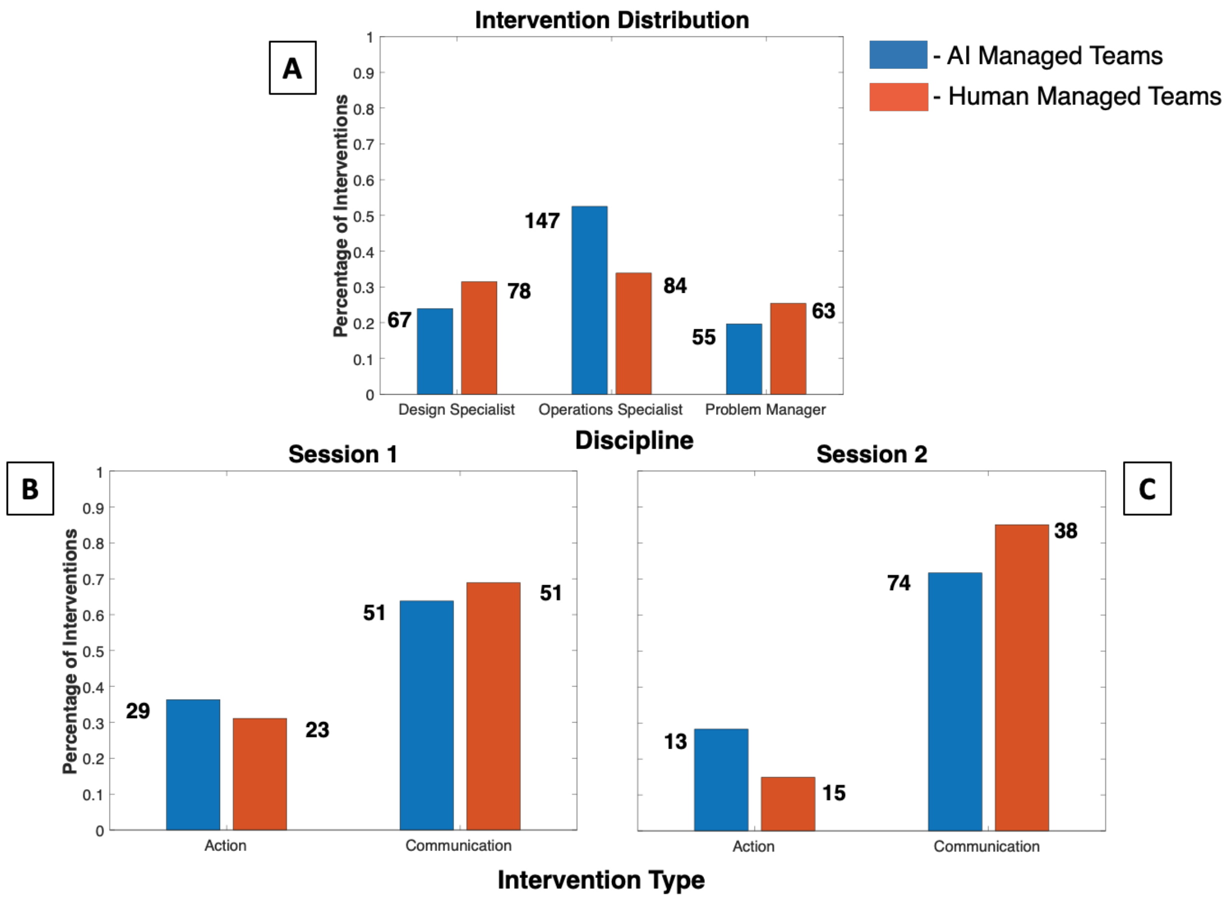


FIGURE 5.11: (A)THE DISTRIBUTION OF PROCESS MANAGER INTERVENTIONS TO DISCIPLINE AND PROPORTION OF INTERVENTIONS DURING THE (B) FIRST AND (C) SECOND PROBLEM-SOLVING SESSIONS. NUMBERS INDICATE RAW COUNTS.

Figure 5.11A shows the distribution of interventions to each team discipline (design specialists, operations specialists, and the problem managers). This includes team-level interventions; since each team-level intervention goes to all three disciplines, they are counted thrice in the figure. As shown, the AI process manager agent focuses more on the operations specialists than the human process managers, with the design specialists and problem managers seeing more equal levels of interventions. Figure 5.11B and Figure 5.11C also shows the proportions of interventions by category for the two sessions. Both experimental sessions show similar levels of action and communication interventions for the two conditions, though the second session shows more considerable differences in behavior. A distinct commonality across these results is the high degree of similarity between the two process manager types regarding the emphasis on communication-based interventions. Across both problem-solving sessions, the proportion of communication interventions is much greater than the action interventions, and when compared overall (combining results across both problem-solving blocks), the proportion is nearly identical, with 70% of the interventions being communication-based and 30% of the interventions being action-based.

The post-study questionnaire with the human process managers corroborates the emphasis on communication. A short answer prompt asks the process managers, "*What were some of the reasons you did/did not intervene with your team? – try to be as specific as possible*." Nearly all the process managers incorporate communication in their responses, both reasons for and against intervening. Several process managers note that they intervene when either there is a lack of communication across the entire team or specifically within disciplines (i.e., if the drone designers were not communicating with each other). When describing instances when they do not intervene, some of the process managers note that they are reluctant to intervene during critical communication (i.e., sharing of critical information such as goals), as they do not want to interrupt the flow of information, or when the problem manager is perceived as effective in their role. Recall that in the particular team structure (Figure 5.2), the process manager is responsible for bridging the communication between disciplines. Thus, the holistic intervention strategies across both the

AI agent and humans highlight the need for effective communication during problem-solving and its criticality as a measure for a process manager to track.

5.3.3 Intervention Effectiveness

Having demonstrated the similarities between the AI agent and human process managers' intervention strategy, the next questions examine the effectiveness of the interventions. However, one first needs to confirm if team members follow the interventions provided by the process manager. Figure 5.12 shows the percentage of team members rating their degree of compliance with the interventions received during the experiment. Asked during the post-study questionnaire, choices range on a categorical scale from "Always" to "Never." As Figure 5.12 shows, members in both conditions quote similar levels of obeyance with the interventions. This similarity holds across the entire range of options. Overall, team members are more likely than not to respond to the interventions, as 65% said they "always" or "most of the time" followed the provided interventions.

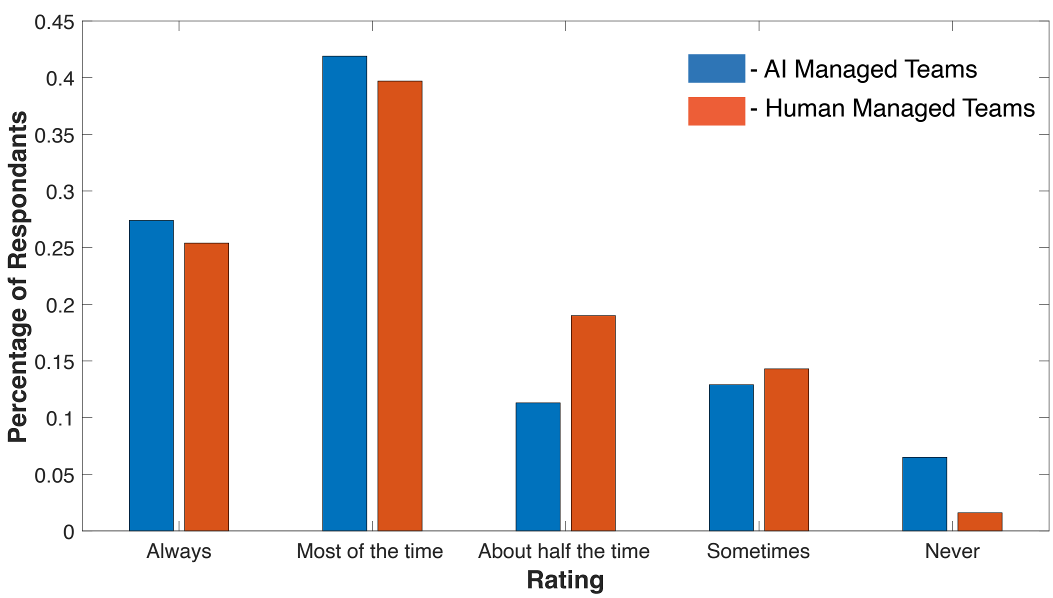


FIGURE 5.12: PERCENTAGE OF TEAM MEMBERS RATING THEIR ADHERENCE TO THE INTERVENTIONS PROVIDED BY THE PROCESS MANAGERS

While equally, and nearly all the time, willing to follow the interventions, team members next rate characteristics of the interventions they received. Again, these ratings range on an interval scale from 0 to 100, with 100 being perfect. Figure 5.13 tells a similar story in terms of the AI agent matching human performance. Regardless of whether the interventions come from a human or the AI process manager, teams rate both the relevance (p = 0.35, d = 0.17) and helpfulness (p = 0.11, d = 0.29) of the interventions similarly (a higher score indicates higher helpfulness and relevancy). Furthermore, teams rate the AI agent just as sensitive (p = 0.71, d = 0.05) to the needs of the team as the human process managers (higher score indicates higher sensitivity). Taken together, the results show that the AI agent process manager matches the human process managers, or human capabilities, in terms of providing interventions that are equally effective and relevant to the problem-solving process and satisfying the needs of the team.

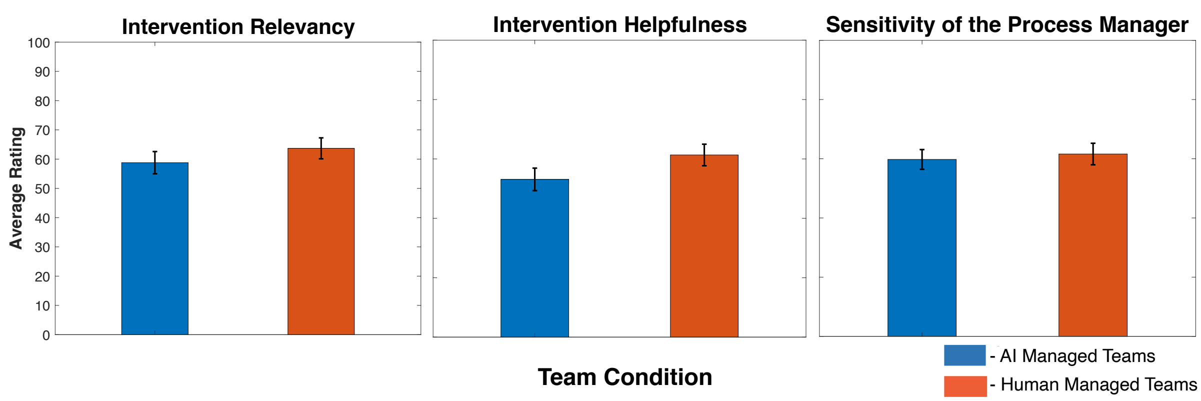


FIGURE 5.13: THE AVERAGE PARTICIPANT RATING OF RELEVANCE AND HELPFULNESS OF THE INTERVENTIONS AND THE SENSITIVITY OF THE PROCESS MANAGERS TO NEEDS OF THE TEAM. (ERROR BARS SHOW ±1 S.E)

5.4 Discussion

This research introduces an AI agent for real-time process management, comparing the behaviors and impacts to that of a human in the same role. Trained on prior team problem-solving data, the

AI process manager takes a data-driven approach. The training data involves the problem-solving characteristics of high-performing teams solving the same design task, also using the collaborative research platform, HyForm. Even though trained to produce prior team problem-solving behaviors, this chapter shows that the AI process manager matches the behaviors and capabilities of human management during new design sessions. The results throughout this chapter highlight this common theme between the two types of process managers, including team performance, intervention strategy, and the teams' perception of the effectiveness and helpfulness of the interventions. Interestingly, team perceive the interventions provided by the AI process manager as just as relevant and sensitive to the needs of the teams as those provided by human observing the process.

In terms of team output, teams perform similarly under the two types of process management. In the context of the problem used, the highest plan's profit a team submits measures the performance of the team. While the teams perform similarly in terms of this measure, the trends are somewhat different. Recall that between problem-solving sessions, a problem shock forces teams to adjust their strategy (the problem shock involves a market switch in customers and changes to the cost of drones). Following the problem shock between sessions, the teams guided by the AI agent increase their performance while the human teams remain stable. While not reaching significance at 5% with the current population size, this trend indicates possible higher robustness in the process of the teams guided by the AI agent. Future work can increase the population size of the study to see if these trends persist and reach significance. However, the AI agent performs at least as effectively as the human process managers in guiding team problem solving in real time. Members on the AI-managed teams also perceive the quality of their teams' performance significantly higher than those on the human managed teams. Furthermore, the human process managers perceive both the quality and cohesion of their teams as significantly better than the team members they managed. This could be due to the human process managers not being able to evaluate how the team is doing in terms of their performance. While the process managers have access to the process and interactions of the team, they do not receive feedback on performance progress. An intriguing direction of future work can implement this feedback on progress within the AI agent's framework and examine how this influences the managerial strategy and team performance.

While the AI agent is trained on prior team problem-solving behaviors, the holistic intervention strategy between the AI and human process managers turns out to be remarkably analogous. Both rely heavily on communication-based interventions rather than action-based interventions, even though the prescribed intervention set includes an equal number of each. Nearly 70% of the interventions by both process manager types are communication-based. This trend also holds across problem-solving sessions after the problem shock. This highlights the criticality of communication within teams as an effective measure for process managers to track during problem solving. Differences between the two process managers start to show in the number of total interventions and in the number of team-based interventions. Overall, the human process managers are more balanced in their intervention strategy: they intervene fewer times, choosing more times not to intervene, choose more team-level interventions, and equally distribute their interventions across disciplines. The AI process manager focuses a bit more, showing more interventions to the operations specialists. Even with these differences, the data-driven process management is broadly similar to that of the human strategy.

In refining the AI agent's intervention framework for future research, further insights are gained by the post-study questionnaires with the human process managers. Asked whether they felt constrained with the prescribed intervention set, nearly all (except one) felt constrained using the current set. This constraining of the intervention set reduced additional variability in the experimental design. Through a short answer question, they identified additional ways they would have liked to intervene with their teams. The desire for individualized interventions emerges as one of the main themes. The current set of interventions either go to the entire team or entire disciplines (i.e., both drone designers). Instead, future iterations of the AI process manager could identify deficiencies at the role-level and intervene with individual team members. Additionally, several process managers note that they would have liked custom interventions, coming up with specific interventions in real time, and directly interacting or chatting with the team. While the authors did consider this, this creates many additional layers of variability within the experimental design and would make it difficult to compare approaches. The process managers also comment on more interventions specific to the problem manager (in the current design, the only problem manager specific intervention is to increase their communication frequency), goal-specific interventions, and positive reinforcement. Regarding the latter, several participants indicate that instead of the "No intervention" option, it would be better to increase team morale and have an intervention to provide positive reinforcement such as, *"Keep up the good work."* These aspects can be implemented in future iterations of the intervention framework.

While still only in its first version, the AI process manager presents boundless opportunities as an experimental testbed for future research. In its current form, the AI agent tracks design actions and certain aspects of communication at the team and discipline levels. As noted by a few of the human process managers, further iterations can track these measures at the individual level within roles and inject the interventions to specific team members. Studied across different context, the timing of the intervention, or interruptions generally, could possibly lead to different impacts, either impeding or helping problem-solving [165,169,200,201]. Since timing is not a direct goal here, the experimental design followed a uniform timing approach to control for this, though different timing schemes, such as anachronistic scheduling, can be tested [202]. Additionally, in the current chapter, the manager tracks and facilitates the overall problem-solving *process* of the team. The manager might instead serve as a *problem* manager and mediate with interventions more relevant to the design task, goals, and constraints of the problem. In fact, the manager could be equipped with features of both and mediate with varying levels of problem- and process-related interventions. Since this research focuses on process, these specificities on task, goals, and constraints are not directly applicable to the process manager in this chapter. As well, the hope is that the method for process management is general, domain independent, and applies across problems, as will be explored in future work.

A large body of literature in human computer interaction and artificial intelligence studies human trust in AI agents. As AI becomes more capable, intelligent, and integrated, humans may become more skeptical and less willing to listen or utilize its power. These questions are also important for the application of AI within this research. As a process manager, or in any type of managerial role for that matter, those within the team must have trust in order to actually respond to the provided interventions. Fortunately, the results of this chapter show that the team members did comply. The last question on the post-questionnaire even askes participants whether they thought the process manager was an AI agent or a human. Regardless of condition, there was an equal 75% / 25% split of those that believe the process manager was an AI agent versus human, respectively. So even though most participants believe the process manager to be an AI agent, they still listen. Future work can examine this finding further, in addition to this question of trust, to see how these perceptions may affect their willingness to obey the process manager and if these perceptions change how they feel about the interventions they receive.

5.5 Summary

Process management brings profound benefits to aiding engineering teams and the engineering design process. Along this vein, this chapter creates an AI agent that manages the design process of teams in real time. This AI process manager dynamically tracks several action and communication-based features of team process, integrates them, and chooses an appropriate intervention. The problem context focuses on a highly interconnected drone design and path-planning task, one that requires effective interdisciplinary collaboration for success. While the AI process manager takes a data-driven approach to intervening (trained on previous team problem-solving data), this is compared to the impact and strategies of human process managers in the same role.

The results of this chapter show that the developed AI process manager matches the capabilities of the human process managers. These similarities hold across several dimensions, including overall team performance, intervention strategy, as well as the perceived impact on team performance, process, and intervention efficacy. Overall, communication deficiencies and inefficiencies stood out as guiding measures to elicit interventions by both the human and AI process managers. This highlights the criticality of effective communication management, particularly during a highly interconnected and interdisciplinary design problem such as the one presented in this chapter. Moreover, the underlying computational framework for the AI agent shows promise as an experimental testbed for future research in real-time management. Additional measures and intervention and decision-making strategies can be implemented and tested to better understand and enhance the impact of real-time process management during the design of complex engineering systems.

Chapter 6 : Impact of Real-Time Process Manager Interventions

6.1 Introduction

The goal of this chapter is to gain deeper insights on the impact of the process manager interventions from Chapter 5. In the preceding chapter, results compared the differences between the constructed AI process manager agent and human process managers regarding overall team performance, manager intervention strategy, and perceived effectiveness. Via a post-study questionnaire, team members perceived themselves as following the interventions provided by the managers, as shown in Figure 5.12. Here in Section 6.2, a set of before and after analyses corroborate whether the managers actually induce the intended changes in the process of the teams. Throughout this analysis, interventions are grouped by type and are studied across both manager conditions. For example, the impact of the three interventions related to communication frequency is computed in the same way, and thus these interventions are presented together and for teams in both manager conditions. Then Section 6.3 introduces trained regression models that identify the predictive nature of the interventions on team performance (i.e., whether some interventions possess a more beneficial impact than other interventions on the highest team profit achieved).

As a quick review of the overall usage of the interventions, in total, the human process managers intervene 127 times using all 12 interventions from the prescribed set, while the AI agent intervenes 167 distinct times, but only using 8 out of the 12 prescribed interventions. The interventions sets are shown in Table 5.1 and Table 5.2, delineated by action-based and communicated-based interventions, respectively. Figure 6.1A illustrates the distribution of interventions across these two categories. Both manager conditions (human and AI agent) focus

40% more on communication-based types than action-based interventions (Figure 6.1A). Figure 6.1B depicts the proportion of interventions across all managers, breaking communication-based interventions further down into those related to communication frequency and those related to communication content (and also includes "*No Intervention*", as this was a separate option). The next section uses this breakdown for analyzing the impact of the interventions on team process, as the impact is measured slightly differently depending on the intervention type.

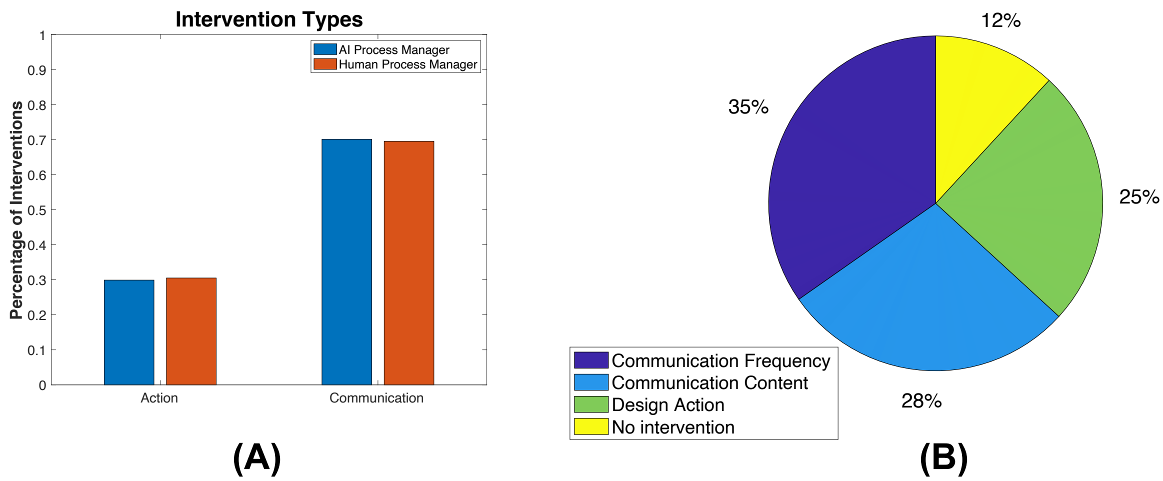


FIGURE 6.1: THE TYPES OF MANAGER INTERVENTIONS USED BY THE PROCESS MANAGERS. (A) PERCENTAGE OF ACTION AND COMMUNICATION INTERVENTIONS USED BY MANAGER TYPE, (B) OVERALL PROPROTION OF MANAGER INTERVENTION TYPES

6.2 Intervention Impacts on Team Process

6.2.1 Communication Frequency Intervention

The first set of interventions relate to communication frequency, or the last three interventions in Table 5.2. The process managers offer these interventions 117 times, constituting 35% of the total interventions. As these interventions are intended to increase the amount of communication in a specific discipline (drone designers, operations planners, or the problem manager), their impact is determined by measuring the communication frequency immediately before and after the intervention is provided. Figure 6.2 shows the total communication count across all 117 instances for 2.5 minutes, the entire time between consecutive interventions. As the communication data in general is sparser than the action data, 2.5 minutes is chosen to capture enough data (and used for the next analyses on communication content). The figure only reflects the communication within the specific disciplines that receive one of these interventions. For example, for a specific team, if the provided interventions states: "*Ops team, please try to communicate with each other more,*" only the communication in the operations channels is considered at that specific time, rather than the entire teams' discourse.

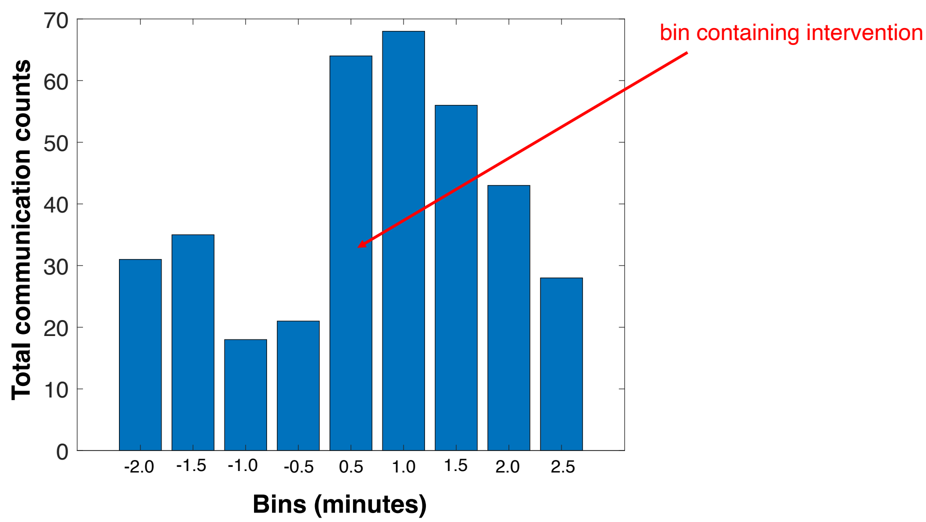


FIGURE 6.2: THE COMMUNICATION FREQUENCY COUNT OF TEAM DISCIPLINES, BOTH IMMEDIATELY PRIOR TO AND IMMEDIATELY FOLLOWING A COMMUNICATION FREQUENCY INTERVENTION

In Figure 6.2, the *'0.5'*- minute bin counts up to 30 seconds *after* the intervention, the *'1.0'* - minute bin counts the time from 30 seconds to 1-minute *after* the intervention, the *'-0.5'*- minute interval counts up to 30 seconds *before* the intervention, the '-1.0' minute bin counts between 1 minute and 30 seconds *before* the intervention, and etc. Overall, the figure shows a sharp, over threefold, increase in the communication frequency following these interventions. This increase in communication lasts across the entire 2.5-minutes and still does not fully fall back to the levels immediately prior to the intervention. This trend indicates a significant influence of this intervention type on the communication behaviors of the teams.

6.2.2 Communication Content Interventions

The next set of interventions studied includes those related to communication content, or the first three interventions in Table 5.2. The managers offer these interventions 96 distinct times, constituting 28% of the total interventions used. These concentrate on the content of the discourse amongst the team rather than just on the frequency, identifying discussions on the design parameters, constraints, goals, the teams' strategies, and the cohesiveness of their discourse. Again, in order to adequately measure the communication content, the 2.5-minutes prior to and after an intervention are used to measure their impact.

Measuring the impact for the first two interventions in Table 5.2 involves identification of specific keywords related to the design parameters and design strategy, respectively. Non-exhaustively, these include scanning for keywords such as: "*velocity*," "*payload,*" "*miles,*" "*houses,*" "*payload,*" and "*profit*" for the design parameters, and "*increase,*" "*decrease,*" "*minimize,*" "*optimize,*" and "*balance,*" for design strategy. As discussed in the previous chapter, earlier problem-solving studies conducted with HyForm motivated the identification of these keyword sets. The final communication content intervention ("*Hi team, try sharing your goals with each other a bit more and make sure they're aligned*") focuses on the cohesion of the team. The AI agent triggers this intervention through the natural language processing technique, Latent Semantic Analysis (LSA), to measure discourse similarity. Thus, LSA also analyses its impact, computing the similarity of communication among all members on a team before and after these interventions, again, in 2.5-minute intervals. To quantitatively assess the impact, the net change between the 2.5-minutes preceding and following an intervention is computed, with the variable, Δ, indicating this difference ($\Delta = after - before$). Consequently, a positive net change ($+\Delta$) indicates an increase in a measure while a negative net change ($-\Delta$) indicates a decrease. As with the communication frequency, the assumption is that a positive influence from the process manager will cause an increase, or positive net change, in these measures.

After computing the net change across all 96 instances of these three interventions, results indicate that there is not much of an impact on cohesion (showing a net increase in similarity of, $\Delta_{cosine} = +0.05$, in which the similarity can range between 0 and 1) or design strategy keywords (showing a net increase of keyword count of, $\Delta_{strat} = +0.00$ keywords). However, there exists a large impact from the design parameter interventions, with a total net increase in design parameter usage of, $\Delta_{param} = +104$ keywords. While the difference in these results is intriguing, the nature of how the different types of information need to be communicated, especially within the specific team structure, supports it. For example, the design parameters are concepts that are more likely needed to be, and more easily able to be, shared across the entire team. On the other hand, design strategy can be more easily understood on an individual basis. Since both the drone designers and operations planners focus on specific designs, design strategy is not as critical to share across the entire team.

6.2.3 Design Action Interventions

The last set of interventions analyzed are the design action interventions (all shown in Table 5.1) and offered by the process managers 80 distinct times, constituting 25% of the total interventions. These interventions are anticipated to elicit specific actions from team members. For example, the intervention: "*Hey operations team, try running the path-planning agent to help*" intends to push the operations discipline to run the assistive, path-planning agent in HyForm. Accordingly, whether or not a specific action occurs within a 1-minute time period after an intervention determines the effectiveness of the intervention. The exception to this involves the two interventions related to drone/path iteration (the interventions in row one and four in Table 5.1). Instead of looking for one specific action, the impact of these two interventions is measured in a manner similar to the communication frequency interventions, with an aggregate count of all actions related to design iteration. Since design actions occur more frequently than team communication, 1-minute intervals are used as opposed to the 2.5-minute intervals used for the communicated-based interventions.

Results across the set of action-based interventions differ quite dramatically. First, Figure 6.3 shows the overall action count for the iteration-based interventions just discussed. The designation of time intervals on the x-axis in the figure follows the same nomenclature as that in Figure 6.2. As Figure 6.3 shows, there is no significant increase in the aggregate of design iteration actions following these specific interventions. In fact, the trend across time actually shows a decrease in the aggregate, indicating that these interventions are not well followed by teams. Again, the figure only reflects the actions within the specific disciplines that receive one of these interventions rather than the entire teams.

Table 6.1 presents the results for the remaining four design action interventions. Since these interventions are intended to induce one specific type of design action, the impact of these is considered in a binary sense (i.e., did or did not the intervention occur within the one-minute threshold period). Moving down the rows, the table presents the total number of intervention instances (first row), the total number of effective instances (whether the intended action occurs within the one-minute time period) (second row), and the total number of effective instances via percentage (third row). Out of these four, the only intervention showing high adherence prompts the drone designers to evaluate and submit their designs (column 1). Otherwise, the interventions show lower levels of process adherence from the teams, even though team members claimed to follow them.

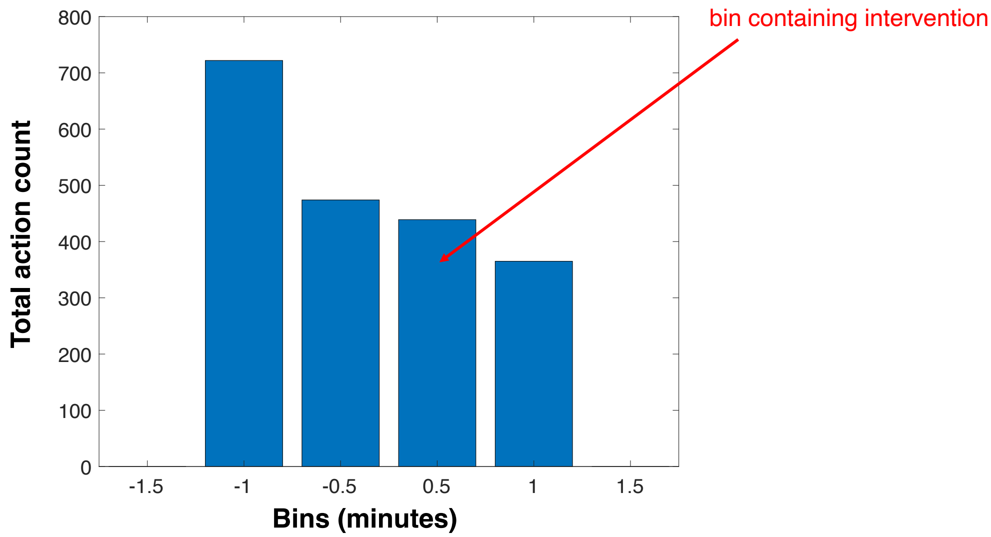


FIGURE 6.3: THE DESIGN INTERATION ACTION FREQUENCY COUNT, BOTH IMMEDIATELY PRIOR TO AND IMMEDIATELY FOLLOWING A DESIGN ITERATION INTERVENTION

TABLE 6.1: THE TOTAL FREQUENCY AND FREQUENCY OF EFFECTIVENESS FOR FOUR DESIGN ACTION INTERVENTIONS

	"Hey drone design team, I would recommend evaluating and submitting your current design and starting fresh."	*"Hey drone design team, check out the suggestions from the drone design agent."*	*"Hey operations team, I suggest that you try evaluating and submitting your plan and starting fresh."*	*"Hey operations team, try running the path-planning agent to help."*
Total #	10	5	7	9
Effective #	10	2	1	4
Percent Effective	100%	40%	14%	44%
Intervention Timing	35.72 seconds	178.10 seconds	221.95 seconds	93.96 seconds
Total Average (across all teams)	49.20 seconds	235.58 seconds	218.40 seconds	142.71 seconds

In addition to quantifying these interventions in a binary sense, the time-to-action is also considered. The time-to-action measures the time between when the process manager provides the intervention to the time that the intended action actually occurs following it. The idea assumes that in order to deduce whether these actions are induced by the process managers, as opposed to by random or natural occurrence, the time-to-action will be shorter during the intervention time periods where one of these four specific interventions are provided. Accordingly, rows four and five of Table 6.1 provide this additional insight. Row four shows the time-to-action for the specific time intervals for these specific four interventions, and row five provides the time-to-action across all intervention intervals. Overall, the time-to-action is shorter during the specific instances of these interventions (row four) at least for three out of the four interventions, and thus more likely to be attributed to the process managers. However, it should be noted that statistical analysis of this data presents challenging, as some of these include only one or two data points. Thus, while these results are descriptive, more data should be collected to allow for more definitive, inferential assessments.

6.3 Regressing Interventions and Team Performance

In order to determine the impact of the interventions on team performance, linear regression models are trained on the data. The overarching hypothesis is that certain interventions may be more effective (or predictive) of team performance than others. The predicator variables in the model are the counts of the number of times managers use each of the 12 interventions for a team, including the "*No intervention*" option, for a total of 13 independent variables. Since the maximum profit achieved by a team designates the team performance in this chapter, maximum profit represents the response variable. A separate linear regression model is trained for each problem-solving session and each team condition, resulting in four models.

Table 6.2 presents the model statistics results from the four linear regression models. For both process manager conditions, the models for the first problem-solving session are significantly more predictive of overall performance than the trained models for the second problem-solving session, indicated by the larger *R*-squared values in the table. Due to this significant difference, these are the two models analyzed to gain further insights into the effects of interventions on team performance and differences between he human and AI process managers.

TABLE 6.2: MODEL STATISTICS FOR EACH OF THE FOUR TRAINED LINEAR REGRESSION MODELS, ONE FOR EACH TEAM CONDITION (ROWS) AND PROBLEM-SOLVING SESSION (COLUMNS)

	Session 1	Session 2
AI	$R^2 = 0.5260$ F = 3.891, $P = 0.038$	$R^2 = 0.0669$ F = 1.0004, $P = 0.52$
Human	$R^2 = 0.9877$ F = 93.28, $P = 0.0016$	$R^2 = -0.2894$ F = 0.66, $P = 0.74$

Table 6.3 offers more details for the two regression models from the first problem-solving session. The first column shows each predictor variable (the 12 possible interventions, including "*No intervention*"), the second and third columns present the resulting estimates (i.e., coefficients) on each of the dimensions for the two models, columns four and five show the resulting p-values on each of these estimates for the two models, and the final column shows which specific intervention each dimension represents. Comparing the two models in this way highlights some interesting insights between the among the interventions as well as differences between the AI agent and human process managers.

To analyze the parameter estimates, a positive value signals a positive correlation between estimate and team profit (i.e., the more times an intervention is used the greater the team profit, or vice versa), whereas a negative correlation indicates a negative impact (the more times an intervention is used the lower the team profit). For both models, dimension x_{13}, ("*No Intervention*") has a small impact on team performance, indicated by the small magnitude of their estimates relative to the other estimates within their respective regression model. While interesting, this provides a validation of the "*No Intervention*" option. Dimensions x_{11} and x_{12} highlight an additional insight. Both of these represent the dimensions focused on communication content, design parameters and design strategy, respectively. These dimensions also have relatively large and positive magnitudes on their parameter estimates, indicating that they have a greater positive impact on team performance. In fact, dimension x_{11} has one of the greatest values for each model and, as seen from the previous before and after analysis, had a large impact on team process behaviors. The five interventions not used by the AI process manager throughout the first problem-solving session result in estimators of 0 (x_1, x_4, x_5, x_6, and x_7). Comparing across manager conditions, three of these interventions (x_4, x_6, and x_7,) significantly hurt teams in the human process manager condition, resulting in large, negative magnitudes of the parameter estimates. Integrating this with results from the before and after analyses from the previous section, four out of the five interventions not used by the AI process managers (the design actions) also did not induce behavioral impacts on the problem-solving process of the teams. Thus, these interventions might not have been impactful because they did not induce the associated behavioral impacts intended by the process managers.

Comparing across the human and AI process manager models reveals interventions that yield different effects on team performance. Significant differences here are identified by computing the 95% confidence intervals on the parameter estimates and identifying those that do not overlap. The interventions resulting in these differences include x_3, x_8, x_{10}, and x_{11}. While the latter three of these interventions all correlate positively with performance, x_3 does not. In the human-managed teams it negatively correlates with performance while in the AI-managed teams it positively correlates with performance. In fact, overall, the AI process manager model has only one intervention that negatively correlates with performance, x_2, whereas the human process manager model has six. Thus, there are fewer interventions that hurt the team in the AI process manager condition. This is a strong indication of a possible overall better strategy, or selection of interventions, by the AI process manager agent. While this did not necessarily produce better overall team performance for the AI team (Figure 5.8), other factors could have been at play that mitigated this effect, such as the ordering, timing, or adherence to the interventions. More analyses should be conducted to fully understand the nuances of these implications.

TABLE 6.3: COMPARISON OF THE TWO REGRESSION MODELS FOR FIRST PROBLEM-SOLVING SESSION

Dimension	Estimate, β (AI)	Estimate, β (Human)	P-value (AI)	P-value (Human)	Intervention
x_1	0	307.67	-	0.2104	*Drone designers, it would be helpful if you can continue working on and refining your drone designs a bit more.*
x_2	-84.69	-914.17	0.8859	0.0123	*Drone designers, please try to communicate with each other more.*
x_3	1541.1	-3742.7	0.0956	0.0058	*Hey drone design team, I would recommend evaluating and submitting your current design and starting fresh.*
x_4	0	-503.45	-	0.0959	*Hey drone design team, check out the suggestions from the drone design agent.*
x_5	0	1207.9	-	0.107	*Hey operations team, I suggest that you try evaluating and submitting your plan and starting fresh.*
x_6	0	-4229.1	-	0.00006	*Hey operations team, try running the path-planning agent to help.*
x_7	0	-2966.5	-	0.0142	*Hi problem manager, please try to communicate with your team more.*
x_8	759.84	3179.3	0.2303	0.0017	*Hi team, try sharing your goals with each other a bit more and make sure they're aligned.*
x_9	394.94	999.97	0.1805	0.0838	*Ops planners, it would be good to continue working on and refining your plans a bit more.*
x_{10}	610.55	4390.9	0.0444	0.0025	*Ops team, please try to communicate with each other more.*
x_{11}	1501.3	3701.2	0.0964	0.0013	*Team, I think you should try focusing more on adjusting the design parameters to meet the goals of the problem, and share this with each other (cost, capacity, speed, budget, weight, etc.).*
x_{12}	1329.5	1472	0.05	0.0061	*Team, try focusing more on your strategy. Try optimizing and increasing/decreasing size of components and share this with each other.*
x_{13}	203.04	-97.71	0.4717	0.4726	*No Intervention*

6.4 Summary

The results from this chapter provide deeper insights into the effects of interventions from the developed AI process manager agent. While team members report they act in accordance with the interventions provided, this chapter presents before and after analyses to corroborate this. In fact, results show mixed adherence to the interventions, depending on intervention type. While the interventions focused on communication frequency and design parameter content show high influence on team process, others such as those focused on design actions show less impact, at least immediate impact. Measuring the time-to-action shows a tendency for teams to perform a certain action more quickly under a relevant intervention, though more data should be collected in future work to further validate this trend. In addition to analyzing the direct impact of the interventions on team process, trained regression models show which interventions yield higher prediction to team performance. Comparing the coefficients on the intervention variables within and across process manager conditions yields further findings and implications on the most impactful, both positive and negative, ones. Overall, the results from this chapter can be used to guide further insights to support the effective, process management of teams.

Chapter 7 : Conclusions, Contributions, and Areas for Future Work

7.1 Overview

Teams are fundamental across many fields, especially in engineering, to solve large, complex problems that require a breadth of domain expertise. However, it has been identified in social and cognitive psychology, as well as recently in the field of engineering design, that teams may not always perform as effectively as possible, are prone to pitfalls, and can even underperform compared to individual performance. Stimulus methods are one technique in the area of engineering design methods to facilitate the problem-solving process of designers. Within this field, research shows that stimuli can induce differing effects depending on the characteristics of said stimulus/intervention and even when the stimulus is provided during problem solving. Therefore, in order to be most impactful, there needs to be mechanism(s) to dynamically track designers and design teams to determine their state/progress to provide interventions in a manner that reflect that state. Accordingly, to address the aforementioned issue, this dissertation presents the following thesis statement:

> *Real-time process management, via the monitoring of design cognition and discourse, can adapt to the state and dynamics of the designers and design progress, thereby facilitating the overall problem-solving design process.*

This thesis is supported through a set of frameworks that motivate and culminate in the development of an artificial intelligent (AI) process manager that intervenes in real-time during problem solving to affect the behavioral processes of teams. The advantage of an AI agent in this research application is twofold: its ability to be able to track multiple metrics simultaneously and over time, and its ability to ascertain underlying patterns within data, such as team communication, that may not be perceivable via direct human inspection. To reach this goal, several exploration studies are conducted to analyze the efficacy of different interventions, strategies, and computational methodologies. These motivate the development of a data-driven approach to process management, which is tested in real-time during an interdisciplinary, complex engineering design problem through an online platform. The AI process manager framework can serve as a valuable testbed for future studies in the process management of engineering design teams.

In order to pursue the aforementioned thesis, first, Chapter 2 presents a large-scale, behavioral study with human process managers overseeing and intervening with design teams. This initial study explores the impact of real-time process management and different strategies for intervening. Furthermore, post-hoc interviews reveal the process managers' motivations for intervening when and with what they did. Though stand alone in the insights revealed, this motivating study is also necessary in order to fully develop an AI manager that can emulate the strategies and nuances of human. Most critically, results show significant impact of the process managers on team performance, measured via final design output, with managed teams performing better than unmanaged teams, though still not quite as well as individual performance (enabled for comparison via the creation of nominal teams). In regard to team behaviors, the process-managed teams show higher team contribution and more cohesion in their communication. Analyzing the interventions provided over time, the overarching managerial strategy shows an exploratory-to-convergent style, with design strategies making up the largest types of interventions early on in the design process before moving to more physical design components near the end of problem-solving. Post-study interviews with the human process managers revealed that a desire to invoke topic shifts in the teams' discourse served as a critical motivation for their interventions. These motivations serve as one of the primary inspirations for utilizing topic modeling as a computational approach to enable the tracking of team discourse, as presented in the next chapter, Chapter 3.

Team communication can provide valuable insight into the cognition of the designers and the design progress. Consequently, Chapter 3 studies whether team members' discourse can be leveraged to computationally detect the impact and effects of the managerial interventions and produce the intended topic shifts in engineering design teams. Accordingly, topic models are trained on the transcript data from the behavioral team study from Chapter 2. Results show that the two team conditions significantly differ in a number of the extracted topics from the topic model, and in particular, those topics that most pertain to the manager interventions. Overall, managed teams' discourse focus on topics much more related to design functions than the unmanaged teams, and a temporal look at numerous intervals during problem solving reveals the largest impacts occur during the latter half of problem solving. Finally, a before and after analysis of interventions reveal that the process manager interventions significantly shift the topics of the team members' discourse toward that of the interventions immediately after they are provided.

Following these more exploratory chapters, Chapter 4 starts to move towards a computational framework that provides interventions in real-time during the problem-solving process. During the middle of a conceptual engineering design task, designers are provided with an inspirational stimulus. This stimulus, however, reflects their current design progress as the stimulus that is provided is either semantically "near" or "far" from their current design at that point in time. In regard to the impact on design performance, results highlight that near and far design stimuli have differing impact on the characteristics of final ideation outcomes. Furthermore, the overall innovativeness of the provided stimuli (i.e., the overall quality of the stimuli) significantly correlates with the overall innovativeness of the designers' final design solutions. In fact, the overall innovativeness of a stimulus has a greater impact on a designer's output than the relative distance of the stimulus. This emphasizes the need to provide stimuli to designers not only at specific distances relative to the solution space, but also while assessing the innovative potential of the inspirational stimulus.

Chapter 5 then presents the development of an AI process manager to examine and compare the effects between a data-driven approach and human mediation. The computational framework for the AI process manager is derived by training on behaviors of teams from previous experiments. Several measures are tracked in near real-time to determine the state of the team and choose an appropriate intervention, including communication frequency, similarity, and content, and action frequency and type. Spanning across several dimension, results show that the AI process manager matches the capabilities of human process managers. These include team performance (i.e., team profit), intervention strategy, and perceived effectiveness. Furthermore, regardless of manager type (human or AI), team members note that they follow the interventions provided by the managers. The succeeding chapter of this dissertations presents additional analyses to determine if this is reflected within team process and behavior changes.

Chapter 6 presents further analyses from the AI process manager experiment from Chapter 5. These analyses more precisely uncover the impact of specific interventions on team performance and process. Before and after analyses corroborate whether the managers induce the intended changes in the process of the teams. Additionally, regression models identify the predictive nature of the interventions on team performance. Results show mixed adherence to the interventions, particularly depending on the intervention type, in terms of the intended effect on team process. The interventions showing the highest impact relate to communication frequency and communication content of design goals/constraints. The design action interventions show less of an immediate impact on team process. Measuring the time-to-action indicates a tendency for teams to perform a certain action more quickly given a relevant intervention. In addition to analyzing the direct impact of the interventions on team process, trained regression models show which interventions yield higher prediction to team performance, with each intervention as an independent variable in the model. Comparing the coefficients on the intervention variables, within and across process manager conditions, yields further findings and implications on the interventions that are most predictive of team performance. For instance, the interventions related to communication content, particularly those on design goals, are not only well followed but also show one of the highest correlations with team performance.

7.2 Contributions

The work incorporated in this dissertation adds several contributions to the research in engineering design theory, methodology, and automation. These can be summarized as follows:

- *Gain insight on the strategies of real-time, adaptive process management on design teams by analyzing the patterns and motivations exhibited by human managers*

- *Introduce a topic modeling framework to study the impact of manager intervention in team cognition in real-time during the problem-solving process*

- *Utilize real-time team discourse to computationally detect topic shifts and better understand the impact of process management during the problem-solving process*

- *Develop and implement a framework to modulate the distance of design stimuli, through semantic similarities, to the current state of a designer in real-time*

- *Identification of metrics to automatically track design team state dynamically during problem solving*

- *Develop an artificial intelligent process manager agent, that is trained on previous problem-solving behaviors, to intervene in real-time to affect the problem-solving process of teams*

- *Identify the perceived and team process-induced impacts of a data-driven approach to process management*

7.3 Areas for Future Work

In an effort to mitigate some of the shortcomings within teams, this dissertation demonstrates the strength of the impact of process management on engineering design teams. However, common with human research based on a limited number of behavioral experiments, future work is needed to fully understand the nuances of the results presented in this dissertation. The tools and implications throughout the preceding chapters lay out many directions and ideas for this future work. The non-exhaustive possibilities presented in this section can be roughly broken down into two distinct categories: those focused on the generality of the findings' contexts and the further development and refinement of the AI process manager and underlying computational frameworks.

7.3.1 Extension of Problem Contexts

Throughout this dissertation, the experiments focus on a limiting number of contexts. These contexts include those related to the problem structure (e.g., problem type and domain), team structure, and time frame of the experiments and problem-solving sessions. As a quick recap, in Chapter 2 (and for the topic modeling framework in Chapter 3), the design problem focuses on a 30-minute, conceptual engineering design problem task with 2 distinct problem shocks. In Chapter 4, the problem again, consists of a conceptual engineering design task, but was 20-minutes in length. The experiment in Chapter 5 (and the corresponding analyses in Chapter 6), presents a more complex, 40-minute engineering design task with one problem shock midway. While each of these experiments are already a bit different in their own right, the aforementioned contexts can be further extended to provide a more robust validation and generality of the findings presented.

Two of the experiments consist of conceptual engineering design tasks, which form the early stages of the design process, requiring the tasks of concept generation and concept selection. The latter experiment on drone design and operations planning can be categorized as a configurational engineering design task. While together these two problem structures show generality, additional problem types and complexities can further validate the findings. The goal would be to move the experimental problems closer to those reflected by engineers in practice, which often poses challenging within a human-subjects, experimental setting. Additionally, some of the interventions in each of the experiments correspond to the specific problem structures and domains. For example, several of the interventions in Chapter 2 & 3 relate to peanut de-shelling and the action interventions in Chapter 5 & 6 correspond to drone design and path planning. Thus, extending the problem types will also enable the extension of possible interventions. The challenge here will be shifting away from the problem-specific interventions and/or having an AI agent identify these prior to problem-solving. Furthermore, engineering problems can encompass multiple meetings, days, weeks, or months. While time periods can also be a limiting factor within human subject studies, as the ones presented in this dissertation, it would be interesting to emulate the process of teams across longer time frames and identify how this impacts what types of interventions would be needed. These longer time periods may also affect team processes. For example, the culture and dynamics of teams can change over time, again, impacting the types of interventions that may be needed and even the overall strategies of process managers. With longer time periods additional challenges emerge, such as the need of accessing real-time communication amongst teams.

Finally, the teams used within each of the respective studies are homogeneous in terms of their experiences and skillsets. Future studies can look at heterogeneous teams, made up of individuals with different expertise, experiences, and skillsets, and how that may affect the interventions and strategies required. The teams in the studies in Chapters 1 – 4 lack any significant structure, while the teams in the studies in Chapters 5 and 6 present a strict hierarchical structure. Understanding the intervention strategies required amongst other team structures would add to the generality of this work. Of course, the overarching goal is that the methodologies and frameworks for automating process management remain general, domain independent, and applies across problem contexts.

Since the culminating AI process manager in this dissertation is trained on previous problem-solving data, in order to extend across contexts in the aforementioned ways, the goal will be to start forming and testing theories. In this way, such an AI process manager agent will not need to be "trained" prior to each new implementation. Instead, the AI agent can sense its current environment (novel problem contexts, team structures, time frames, etc.) and apply these theories to choose an appropriate intervention. The ultimate goal is that the automation and computational frameworks for process management is general, domain independent, and applies across problems.

7.3.2 Further Development of AI Process Manager

The AI process manager and corresponding tools for automation developed for this dissertation can be utilized as a testbed for further analyses. Currently, the AI process manager tracks team communication (including frequency, aspects of its content, and similarity), as well as the frequency of specific types of actions. These all can be extended and built upon to administer different manager strategies and refine the team process measures that are tracked dynamically over time. For example, more individualized or customized interventions that are specific to individuals rather than to multiple players or the entire team can be administered. The latter would involve identifying deficiencies at the role level rather than the at the discipline level. Another interesting aspect is the notion of positive reinforcement. Instead of having the manager not intervene when they feel that a team is on the right path, the manager intervenes with a sentiment to boost team morale. This type of positive reinforcement could have tremendous impact on team attitude, efforts, and ultimately performance. Furthermore, the AI presently follows a set timing for the interventions, though this was a specific decision chosen in the experimental design to mitigate confounding variables. However, the manager can instead intervene more freely (or not intervene more freely) whenever they feel necessary. As noted by several of the human process managers in the post-study questionnaire, many chose not to intervene during important and critical conversations amongst the teams. Thus, these ad-hoc interventions can be further studied to provide further insights into the identification of appropriate timings.

The majority of the interventions and overall intervention strategy currently implemented in the AI process manager can be interpreted as convergent or goal directed. While two of the interventions ask designers to evaluate and submit their designers and start fresh, the remaining interventions do not explicitly push for exploration of the design space or other types of divergent thinking. As this divergent-convergent search of the design space is not only a well-studied strategy within the engineering design literature, but also one that was witnessed among the process managers in Chapter 2, this would be an interesting strategy to emulate next within the AI framework and compare the impact to other strategies. In order to achieve this, this requires extending the intervention set to include ones that promote such an exploratory push of the design space, perhaps even modeling those from the initial behavioral experiment.

In addition to extending the types and timings of interventions, the team measures that are tracked by the process managers can also be adjusted. It its current form, the manager tracks the *process* of the team, so the manager is not intended to help in directly solving the problem, the reason why the AI only tracks communication and action behaviors. Instead, or in addition to these process measures, the manager can serve as a *problem* manager. In this way, different aspects of the team's design state can be tracked, and the manager can mediate with interventions more relevant to the design task, goals, and constraints of the problem. This notion relates to aspects of the experiments presented in Chapters 1 and Chapter 4, where some of the provided interventions more directly aid in solving the design problem itself. Fundamentally, the AI can track where designers are within the design space and help push them in a certain direction, either exploring other regions or converging on specific design ideas. Also, this new conceptualization of the AI framework does not necessarily have to be mutually exclusive. The AI manager framework can integrate both of these notions of problem- and process-based management and serve these roles together by providing interventions related to both.

Another interesting avenue of future research is implementing an additional feedback mechanism into the AI process manager agent. For example, the analyses presented in Chapter 6 (specifically, the before and after analyses) are conducted to identify whether teams actually followed the interventions. This information can be reinserted back into the process manager as additional information to use in deciding the next intervention. Often, in the current framework with the restricted set of possible interventions, the AI process manager would use the same interventions multiple times in a row. Perhaps the AI can use this additional feedback to mitigate this and implement a different wording or phrasing to get the same information across (this reoccurrence of the same intervention multiple times in a row was noted by many of the participants in post-study questionaries). Further, in consideration of human interaction with the AI manager, timing and appropriateness of information should be considered to prevent the AI content or frequency from becoming annoying to the human user. Finally, different underlying computational frameworks can be implemented in the design of the AI agent. For example, latent semantic analysis is used as the natural language processing (NLP) technique to measure the similarity amongst the team communication. Other, perhaps more sophisticated and state-of-the-art, NLP approaches such as short text topic modeling (STTP) can be used to measure this. Moreover, a probabilistic approach can be taken within the decision-making framework of the AI (i.e., the mechanism on how the AI actually chooses which intervention to implement).

In the end, the methodologies and tools this dissertation presents create promising opportunities for future research in automating process management. While teams are useful to problem-solving, there are underlying pitfalls that may cause them to underperform. This dissertation tasks useful steps towards this mitigation of pitfalls by proposing different computational approaches for real-time interventions that can track and adapt dynamically during the problem-solving process. The overarching research goal is to extend this management across contexts and capturing the right characteristics of teams to provide a rich set of potential interventions. This will ultimately facilitate the problem-solving process and cause teams to perform as efficiently and effectively as possible.

Chapter 8 : References

[1] Laughlin, P. R., 2011, *Group Problem Solving*, Princeton University Press.

[2] Hoffman, L. R., 1966, "Group Problem Solving," *Advances in Experimental Social Psychology*, pp. 99–132.

[3] Mills, J. E., and Treagust, D. F., 2003, "Engineering Education - Is Problem-Based or Project-Based Learning the Answer?," Australas. J. Eng. Eng., **3**(Janurary), pp. 2–16.

[4] Dutson, A. J., Todd, R. H., Magleby, S. P., and Sorensen, C. D., 1997, "A Review of Literature on Teaching Engineering Design Through Project-Oriented Capstone Courses," J. Eng. Educ., **86**(1), pp. 17–28.

[5] Seat, E., and Lord, S. M., 1999, "Enabling Effective Engineering Teams: A Program for Teaching Interaction Skills," J. Eng. Educ., **88**(4), p. 385.

[6] Turkalj, Z., Fosic, I., and Dujak, D., 2008, "Conflict Management in Organization," Interdiscip. Manag., (917340826).

[7] Tesluk, P. E., Quigley, N. R., and Tekleab, A. G., 2009, "A Longitudinal Study of Team Conflict, Conflict Management, Cohesion, and Team Effectivenss," Gr. Organ. Manag., **34**(2), pp. 170–205.

[8] Kim, M. S., 2007, "Analysis of Team Interaction and Team Creativity of Student Design Teams Based on Personal Creativity Modes," Proc. ASME 2007 Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf., pp. 1–13.

[9] Dong, A., Hill, A. W., and Agogino, A. M., 2004, "A Document Analysis Method for Characterizing Design Team Performance," J. Mech. Des., **126**(3), pp. 378–385.

[10] Maier, N. R., 1967, "Assets and Liabilities in Group Problem Solving: The Need for an Integrative Function," Psychol. Rev., (1), p. 45.

[11] Karau, S. J., and Williams, K. D., 1993, “Social Loafing: A Meta-Analytic Review and Theoretical Integration ,” J. Personal. Soc. Psychol. , **65**(4), pp. 681–706.

[12] Alnuaimi, O. A., Robert, L. P., and Maruping, L. M., 2010, “Team Size, Dispersion, and Social Loafing in Technology-Supported Teams: A Perspective on the Theory of Moral Disengagement ,” J. Manag. Inf. Syst. , **27**(1), pp. 203–230.

[13] Ransom, H. H., 1974, “Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes. By Irving L. Janis. (Boston: Houghton Mifflin Co., 1972. Pp. Viii, 277. $4.50 Paper.) ,” J. Polit. , **36**(1), pp. 218–220.

[14] Schmidt, A., 2013, “Groupthink,” Encycl. Br. Online.

[15] Myers, D. G., 1976, “The Group Polarization Phenomenon,” Psychol. Bull., **83**(4), pp. 602–627.

[16] Wilson, J. O., Rosen, D., Nelson, B. A., and Yen, J., 2010, “The Effects of Biological Examples in Idea Generation,” Des. Stud., **31**(2), pp. 169–186.

[17] Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K., and Kotovsky, K., 2011, “On the Benefits and Pitfalls of Analogies for Innovative Design: Ideation Performance Based on Analogical Distance, Commonness, and Modality of Examples,” J. Mech. Des., **133**(8), p. 081004.

[18] Goucher-Lambert, K., Moss, J., and Cagan, J., 2019, “A Neuroimaging Investigation of Design Ideation with and without Inspirational Stimuli: Understanding the Meaning of near and Far Stimuli,” Des. Stud., **60**(January), pp. 1–38.

[19] Hu, Y. and Taylor, M.E., A Computer-Aided Design Intelligent Tutoring System Teaching Strategic Flexibility. In: 2016 ASEE Annual Conference & Exposition Proceedings. ASEE Conferences. Epub ahead of print 2016. DOI: 10.18260/p.27208.

[20] Rahman, M.H., Schimpf, C., Xie, C. and Sha, Z., 2019. A Computer-Aided Design Based Research Platform for Design Thinking Studies. *Journal of Mechanical Design*, *141*(12), p.121102.

[21] Roll, I., Wiese, E.S., Long, Y., Aleven, V. and Koedinger, K.R., 2014. "Tutoring self-and co-regulation with intelligent tutoring systems to help students acquire better learning skills." *Design recommendations for intelligent tutoring systems*, 2, pp.169-182.

[22] Higgins, P.G., 1999. "Job shop scheduling: Hybrid intelligent human-computer paradigm." University of Melbourne, Department of Mechanical and Manufacturing Engineering.

[23] Liew, C., 2018. "The future of radiology augmented with artificial intelligence: a strategy for success." *European journal of radiology*, 102, pp.152-156.

[24] Dellermann, D., Ebel, P., Söllner, M. and Leimeister, J.M., 2019. "Hybrid intelligence." *Bus Inf Syst Eng*.

[25] Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S. and Yang, G.Z., 2019. XAI—Explainable artificial intelligence. Science Robotics, 4(37).

[26] Ezer, N., Bruni, S., Cai, Y., Hepenstal, S.J., Miller, C.A. and Schmorrow, D.D., 2019, November. Trust Engineering for Human-AI Teams. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 63, No. 1, pp. 322-326). Sage CA: Los Angeles, CA: SAGE Publications.

[27] van den Bosch, K. and Bronkhorst, A., 2018. Human-AI cooperation to benefit military decision making. NATO.

[28] Yang Q, Steinfeld A, Rosé C, et al., 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. DOI: 10.1145/3313831.3376301.

[29] Toft, Y., Howard, P. and Jorgensen, D., 2003. "Human-centered engineers—a model for holistic interdisciplinary communication and professional practice. "International journal of industrial ergonomics,31(3), pp.195-202.

[30] Sonnenwald, D.H., 1996. "Communication roles that support collaboration during the design process." Design studies,17(3), pp.277-301.

[31] Hendry, D.G., 2004, August. "Communication functions and the adaptation of design representations in interdisciplinary teams." In Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques(pp. 123-132).

[32] Kratzer, J., Leenders, R.T.A. and Van Engelen, J.M., 2010. "The social network among engineering design teams and their creativity: A case study among teams in two product development programs." International Journal of Project Management, 28(5), pp.428-436.

[33] Wu, Q. and Cormican, K., 2016. "Shared leadership and team creativity: A social network analysis in engineering design teams." Journal of technology management & innovation,11(2), pp.2-12.

[34] Safoutin, M.J. and Thurston, D.J., 1993. A communications-based technique for interdisciplinary design team management. *IEEE Transactions on Engineering Management*, *40*(4), pp.360-372.

[35] Bellamy, L., Evans, D. L., Linder, D.E., McNeill, B.W., Raupp, G., 1994. Teams in engineering education. Report to the National Science Foundation on Grant Number USE9156176, Arizona State Uni-versity, Tempe, AZ

[36] Faure C (2004) Beyond Brainstorming: Effects of Different Group Procedures on

Selection of Ideas and Satisfaction with the Process. J Creat Behav 38:13

[37] Rietzschel EF, Nijstad BA, Stroebe W (2006) Productivity is not enough: A comparison of interactive and nominal brainstorming groups on idea generation and selection. J Exp Soc Psychol 42:244–251. doi: 10.1016/j.jesp.2005.04.005

[38] Putman VL, Paulus PB (2009) Brainstorming, Brainstorming Rules and Decision Making. J Creat Behav 43:23–39

[39] Taylor DW, Berry PC, Block CH (1958) Does Group Participation When Using Brainstorming Facilitate or Inhibit Creative Thinking? Adm Sci Q 3:23. doi: 10.2307/2390603

[40] Diehl M, Stroebe W (1987) Productivity loss in brainstorming groups: Toward the solution of a riddle. J Pers Soc Psychol 53:497–509. doi: 10.1037/0022-3514.53.3.497

[41] Sio UN, Kotovsky K, Cagan J (2018) Silence is Golden: The Effect of Verbalization on Group Peformance. J Exp Psychol Gen

[42] Lee CS, Huggins AC, Therriault DJ (2014) A measure of creativity or intelligence?: Examining internal and external structure validity evidence of the remote associates test. Psychol Aesthetics, Creat Arts 8:446–460. doi: 10.1037/a0036773

[43] Mittal S, Frayman F (1989) Towards a Generic Model of Configuration Tasks. In: IJCAI Proceedings of the 11th International Joint Conference on Artificial Intelligence - Volume 2. pp 1395–1401

[44] McComb C, Cagan J, Kotovsky K (2017) Optimizing Design Teams Based on Problem Properties: Computational Team Simulations and an Applied Empirical Test. J Mech Des 139:041101. doi: 10.1115/1.4035793

[45] Wielinga B, Schreiber G (1997) Configuration-Design Problem Solving. AI Des 49–56

[46] Liu S, Boyle IM (2009) Engineering design: Perspectives, challenges, and recent advances. J Eng Des 20:7–19. doi: 10.1080/09544820802670914

[47] McComb C, Cagan J, Kotovsky K (2015a) Rolling with the punches: An examination of team performance in a design task subject to drastic changes. Des Stud 36:99–121. doi: 10.1016/j.destud.2014.10.001

[48] Moss J, Kotovsky K, Cagan J (2011) The Effect of Incidental Hints When Problems Are Suspended Before, During, or After an Impasse. J Exp Psychol Learn Mem Cogn 37:140–148. doi: 10.1037/a0021206

[49] Fu K, Cagan J, Kotovsky K (2010) Design Team Convergence: The Influence of Example Solution Quality. J Mech Des 132:111005. doi: 10.1115/1.4002202

[50] Linsey JS, Markman AB, Wood KL (2012) Design by Analogy: A Study of the WordTree Method for Problem Re-Representation. J Mech Des 134:041009–041009. doi: 10.1115/1.4006145

[51] Jansson DG, Smith SM (1991) Design fixation. Des Stud 12:3–11. doi: 10.1016/0142-694X(91)90003-F

[52] Sio UN, Kotovsky K, Cagan J (2015) Fixation or inspiration? A meta-analytic review of the role of examples on design processes. Des Stud 39:70–99. doi: 10.1016/j.destud.2015.04.004

[53] Agogué, M., Kazakçi, A., Hatchuel, A., Le Masson, P., Weil, B., Poirel, N., and Cassotti, M., 2014, “The Impact of Type of Examples on Originality: Explaining Fixation and Stimulation Effects,” *J. Creat. Behav.*, **48**(1), pp. 1–12.

[54] Brun J, Ezzat H, Weil B Managing the impacts of non-verbal devices on idea generation: a new challenge for creative leaders. In: European Academy of Management - EURAM.

Varsovie, Poland

[55] Ezzat H, Agogué M, Weil B, Cassotti M (2018) Specificity and Abstraction of Examples: Opposite Effects on Fixation for Creative Ideation. J Creat Behav 0:1–8. doi: 10.1002/jocb.349

[56] Rickards T, Moger S (2002) Creative Leadership Processes in Project Team Development. Br J Manag 11:273–283. doi: 10.1111/1467-8551.00173

[57] Hey, J., Linsey, J., Agogino, a. M., and Wood, K. L., 2008, "Analogies and Metaphors in Creative Design," *Int. J. Eng. Educ*., **24**(2), pp. 283–294.

[58] Goldschmidt G, Casakin H (1999) Expertise and the use of visual analogy: implication for design educatio. Des Stud 20:153–175

[59] Yilmaz, S., and Seifert, C., 2010, "Cognitive Heuristics in Design Ideation," *11th Int. Des. Conf. Des*, pp. 1–11.

[60] Isaksen, S. G., 2013, "Facilitating Creative Problem-Solving Groups" [Online]. Available: http://www.cpsb.com/research/articles/creative-problem-solving/Facilitating-CPS-Groups.html.

[61] Linsey JS, Tseng I, Fu K, et al (2010) A Study of Design Fixation, Its Mitigation and Perception in Engineering Design Faculty. J Mech Des 132:041003. doi: 10.1115/1.4001110

[62] Gyory JT, Cagan J, Kotovksy K (2018a) The Effect of Managerial Intervention on Engineering Design Team Performance. In: Gero J (ed) Design Computing and Cognition '18. Como, Italy

[63] Shah, J. J., Smith, S. M., and Vargas-Hernandez, N., 2003, "Metrics for Measuring Ideation Effectiveness," Des. Stud., **24**(2), pp. 111–134.

[64] Dong, A., 2005, “The Latent Semantic Approach to Studying Design Team Communication,” Des. Stud., **26**(5), pp. 445–461.

[65] Landauer, T. K., Foltz, P. W., and Laham, D., 1998, “An Introduction to Latent Semantic Analysis,” Discourse Process., **25**(2–3), pp. 259–284.

[66] Harman D (1986) An experimental study of factors important in document ranking. SIGIR ’86 Proc 9th Annu Int ACM SIGIR Conf Res Dev Inf Retr 186–193. doi: 10.1145/253168.253206

[67] Cicchetti, D. V., 1994, “Guidlines, Criteria, and Rules of Thumb for Evalauting Normed and Standardized Assessment Instruments in Psychology,” Psychol. Assess., **6**(4), pp. 284–290.

[68] Brooks LR (1968) Spatial and verbal components of the act of recall . Can. J. Psychol. Can. Psychol. 22:349–368

[69] Hsi S, Linn M, Bell J (1997) The Role of Spatial Reasoning in Engineering and The Design of Spatial Instruction. J Eng Educ 151–158

[70] Thompson G, Lordan M (1999) A review of creativity principles applied to engineering design. Proc Inst Mech Eng Part E J Process Mech 213:17–31. doi: 10.1243/0954408991529960

[71] Liu YC, Bligh T, Chakrabarti A (2003) Towards an “ideal” approach for concept generation. Des Stud 24:341–355. doi: 10.1016/S0142-694X(03)00003-6

[72] Stempfle J, Badke-Schaub P (2002) Thinking in design teams - An analysis of team communication. Des Stud 23:473–496. doi: 10.1016/S0142-694X(02)00004-2

[73] Comer DR (1995) A Model of Social Loafing in Real Work Groups. Hum Relations 48:647–667. doi: 10.1177/001872679504800603

[74] Jones PE, Roelofsma PHMP (2000) The potential for social contextual and group biases in team decision-making: Biases, conditions and psychological mechanisms. Ergonomics 43:1129–1152. doi: 10.1080/00140130050084914

[75] McComb C, Cagan J, Kotovsky K (2015b) Lifting the Veil: Drawing insights about design teams from a cognitively-inspired computational model. Des Stud 40:119–142. doi: 10.1016/j.destud.2015.06.005

[76] Siggelkow N, Rivkin JW (2005) Speed and Search: Designing Organizations for Turbulence and Complexity. Organ Sci 16:101–122. doi: 10.1287/orsc.1050.0116

[77] Kavadias S, Sommer SC (2009) The Effects of Problem Structure and Team Diversity on Brainstorming Effectiveness. Manage Sci 55:1899–1913. doi: 10.1287/mnsc.1090.1079

[78] Girotra K, Terwiesch C, Ulrich KT (2010) Idea Generation and the Quality of the Best Idea. Manage Sci 56:591–605. doi: 10.1287/mnsc.1090.1144

[79] Ezzat H, Camarda A, Cassotti M, et al (2017) How minimal executive feedback influences creative idea generation. PLoS One 12:1–11. doi: 10.1371/journal.pone.0180458

[80] Gyory JT, Cagan J, Kotovsky K (2018b) Should Teams Collaborate During Conceptual Engineering Design?: An Experimental Study. In: ASME International Design Engineering Technical Conference. Quebec City, Canada

[81] Ball, L. J., and Christensen, B. T., 2019, "Advancing an Understanding of Design Cognition and Design Metacognition: Progress and Prospects," *Des. Stud.*, **65**(2018), pp. 35–59.

[82] Cross, N., 2001, "Design Cognition: Results from Protocol and Other Empirical Studies of Design Activity," *Design Knowing and Learning: Cognition in Design Education*, C.

Eastman, W. Newstatter, and M. McCracken, eds., Elsevier, Oxford, UK, pp. 79–103.

[83] Den Otter, ad, and Emmitt, S., 2007, “Exploring Effectiveness of Team Communication: Balancing Synchronous and Asynchronous Communication in Design Teams,” *Eng. Constr. Archit. Manag.*, **14**(5), pp. 408–419.

[84] Lloyd, P., Lawson, B., and Scott, P., 1995, “Can Concurrent Verbalization Reveal Design Cognition?,” *Des. Stud.*, **16**(2), pp. 237–259.

[85] Martin, M. J., and Foltz, P. W., 2004, “Automated Team Discourse Annotation and Performance Prediction Using LSA,” *Proceedings of Human Language Technology and North American Association for Computational Linguistics Conference 2004: Short Papers. Association for Computational Linguistics*, pp. 97–100.

[86] Dong, A., Kleinsmann, M. S., and Deken, F., 2013, “Investigating Design Cognition in the Construction and Enactment of Team Mental Models,” *Des. Stud.*, **34**(1), pp. 1–33.

[87] Song, B., Srinivasan, V., and Luo, J., 2017, “Patent Stimuli Search and Its Influence on Ideation Outcomes,” *Des. Sci.*, **3**, p. e25.

[88] Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C., and Wood, K., 2013, “The Meaning of ‘Near’ and ‘Far’: The Impact of Structuring Design Databases and the Effect of Distance of Analogy on Design Output,” *J. Mech. Des.*, **135**(2), p. 021007.

[89] Fu, K., Cagan, J., Kotovsky, K., and Wood, K., 2013, “Discovering Structure in Design Databases Through Functional and Surface Based Mapping,” *J. Mech. Des.*, **135**(3), p. 031006.

[90] Goucher-Lambert, K., Gyory, J. T., Kotovsky, K., and Cagan, J., 2020, “Adaptive Inspirational Design Stimuli: Using Design Output to Computationally Search for Stimuli That Impact Concept Generation.,” *ASME J. Mech. Des.*, pp. 1–8.

[91] Gyory, J. T., Cagan, J., and Kotovsky, K., 2019, "Are You Better off Alone? Mitigating the Underperformance of Engineering Teams during Conceptual Design through Adaptive Process Management," *Res. Eng. Des.*, **30**(1), pp. 85–102.

[92] Mohr, J. W., and Bogdanov, P., 2013, "Introduction-Topic Models: What They Are and Why They Matter," *Poetics*, **41**(6), pp. 545–569.

[93] Blei, D. M., 2012, "Probablistic Topic Models," Commun. Acm, **55**(4), pp. 77–84.

[94] Rosen-Zvi, M., Griffiths, T., Steyvers, M., and Smyth, P., 2004, "The Author-Topic Model for Authors and Documents," *Conference on Uncertainty in Artificial Intelligence*, pp. 487–494.

[95] Blei, D. M., and Lafferty, J. D., 2007, "A Correlated Topic Model of Science," *Ann. Appl. Stat.,* **1**(1), pp. 17–35.

[96] Ball, Z., and Lewis, K., 2020. "Predicting Design Performance Utilizing Automated Topic Discovery." ASME. *J. Mech. Des*. 142(12): 121703.

[97] Ahmed, Faez, Mark Fuge, and Lev D. Gorbunov., 2016, "Discovering diverse, high quality design ideas from a large corpus." *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. Vol. 50190. American Society of Mechanical Engineers.

[98] Ahmed, F., and Fuge, M., 2018, "Creative Exploration Using Topic-Based Bisociative Networks," *Des. Sci.*, **4**(12), pp. 1–30.

[99] Bhowmik, T., Niu, N., Savolainen, J., and Mahmoud, A., 2015, "Leveraging Topic Modeling and Part-of-Speech Tagging to Support Combinational Creativity in Requirements Engineering," *Requir. Eng.*, **20**(3), pp. 253–280.

[100] Joung, Junegak, and Harrison M. Kim., 2020, "An LDA-based Approach for Product

Attribute Identification from Online Customer Reviews."

[101] Suryadi, Dedy, and Harrison M. Kim., 2019, "A Data-Driven Approach to Product Usage Context Identification From Online Customer Reviews." *Journal of Mechanical Design* 141.12.

[102] Nijstad, B. A., and Stroebe, W., 2006, “How the Group Affects the Mind: A Cognitive Model of Idea Generation in Groups,” *Personal. Soc. Psychol. Rev.*, **10**(3), pp. 186–213.

[103] Gyory, Joshua T., Jonathan Cagan, and Kenneth Kotovsky., 2018, "An Exploration of the Effects of Managerial Intervention on Engineering Design Team Performance." *International Conference on-Design Computing and Cognition*. Springer, Cham.

[104] Blei, D. M., Ng, A. Y., and Jordan, M. I., 2003, “Latent Dirichlet Allocation,” *J. Mach. Learn. Res.*, **3**, pp. 993–1022.

[105] Scott Deerwester, Richard Harshman, Susan T, George W, Thomas K, Deerwester, S., Dumais, S. T., Furnas, G. W., and Landauer, T. K., 1990, “Indexing by Latent Semantic Analysis,” *J. Am. Soc. Inf. Sci.*, **41**(6), pp. 391–407.

[106] Hofmann, T., 2017, “Probabilistic Latent Semantic Indexing,” *ACM SIGIR Forum*, **51**(2), pp. 211–218.

[107] Ramos, J., 2003, “Using TF-IDF to Determine Word Relevance in Document Queries Juan,” *Proceedings of the First Instructional Conference on Machine Learning*, Piscataway, NJ, pp. 133–142.

[108] Kim, T. Y., Min, M., Yoon, T., & Lee, J. H., 2010, "Semantic Analysis of Twitter contents using PLSA, and LDA." *SCIS & ISIS SCIS & ISIS 2010*. Japan Society for Fuzzy Theory and Intelligent Informatics.

[109] Rosa, K. D., Shah, R., Lin, B., Gershman, A., & Frederking, R., 2011, "Topical clustering of tweets." *Proceedings of the ACM SIGIR: SWSM* 63.

[110] Mehrotra, R., Sanner, S., Buntine, W., & Xie, L., 2013, "Improving lda topic models for microblogs via tweet pooling and automatic labeling." *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*.

[111] Nguyen, V. -A., Boyd-Graber, J., and Resnik, P., 2012, SITS: A hierarchical nonparametric model using speaker identity for topic segmentation in multiparty conversations. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics:* Long Papers –Vol. 1. Stroudsburg, PA, USA: Association for Computational Linguistics, 78–87.

[112] Purver, M., Kording, K., Griffiths, T. L., and Tenenbaum, J. B., 2006, Unsupervised topic mod- ¨ elling for multi-party spoken discourse. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics.* Stroudsburg, PA, USA: Association for Computational Linguistics, 17–24.

[113] Crain S.P., Zhou K., Yang SH., Z. H., 2012, "Dimensionality Reduction and Topic Modeling: From Latent Semantic Indexing to Latent Dirichlet Allocation and Beyond.," *Mining Text Data*, Springer, Boston, Massachusetts, USA.

[114] Handler, A., Denny, M., Wallach, H., and O'Connor, B., 2016, "Bag of What? Simple Noun Phrase Extraction for Text Analysis," *Proceedings of 2016 EMNLP Workshop on Natural Language Processing and Computational Social Science*, Austin, TX, pp. 114–124.

[115] Loper, E., and Bird, S., 2004, “NLTK: The Natural Language Toolkit,” *Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions, Association for Computational Linguistics*.

[116] Huang, L., Ma, J., and Chen, C., 2017, “Topic Detection from Microblogs Using T-LDA and Perplexity,” *Proceedings - 2017 24th Asia-Pacific Software Engineering Conference Workshops (APSECW)*, pp. 71–77.

[117] AlSumait, L., Barbará, D., and Domeniconi, C., 2008, “On-Line LDA: Adaptive Topic Models for Mining Text Streams with Applications to Topic Detection and Tracking,” *Proceedings - IEEE International Conference on Data Mining, ICDM*, IEEE, pp. 3–12.

[118] Heinrich, G., 2005, “Parameter Estimation for Text Analysis,” *Tech. Rep*.

[119] Merriam-Webster, M. W. S. C., 1977, *Dictionary*, G & C Merriam Company, 830., Springfield, Massachusetts, USA.

[120] Kullback, S. and Leibler, R. A., 2010, “On Information and Sufficiency,” *Ann. Math. Stat.*, **22**(1), pp. 79–86.

[121] Pérez-Cruz, F., 2008, “Kullback-Leibler Divergence Estimation of Continuous Distributions,” *IEEE International Symposium on Information Theory - Proceedings*, pp. 1666–1670.

[122] Shlens, Jonathon., 2014, "Notes on Kullback-Leibler divergence and likelihood." arXiv preprint arXiv:1404.2000.

[123] Kapoor, R., Gupta, R., Jha, S., & Kumar, R. , 2018, "Boosting performance of power quality event identification with KL Divergence measure and standard deviation." *Measurement 126*: 134-142.

[124] Foulds, James, Levi Boyles, Christopher DuBois, Padhraic Smyth, and Max Welling., 2013, "Stochastic collapsed variational Bayesian inference for latent Dirichlet allocation." In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 446–454. ACM.

[125] Hoffman, Matthew D., David M. Blei, Chong Wang, and John Paisley., 2013, "Stochastic variational inference." *The Journal of Machine Learning Research* 14, no. 1: 1303–1347.

[126] Griffiths, Thomas L., and Mark Steyvers., 2004, "Finding scientific topics." *Proceedings of the National academy of Sciences* 101, no. suppl 1: 5228–5235.

[127] Asuncion, Arthur, Max Welling, Padhraic Smyth, and Yee Whye Teh., 2009, "On smoothing and inference for topic models." In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pp. 27–34. AUAI Press.

[128] Teh, Yee W., David Newman, and Max Welling., 2007, "A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation." In *Advances in neural information processing systems*, pp. 1353–1360.

[129] Airoldi, E. M., Blei, D., Erosheva, E. A., & Fienberg, S. E., 2014, eds. Handbook of mixed membership models and their applications. CRC press.

[130] Newman, D., Baldwin, T., Cavedon, L., Huang, E., Karimi, S., Martinez, D., Scholer, F., and Zobel, J. (2010a). Visualizing search results and document collections using topic maps. *Web Semantics 8*: 169–175.

[131] Newman, D., Lau, J. H., Grieser, K., and Baldwin, T. (2010b). Automatic evaluation of topic coherence. In *Proceedings of Human Language Technologies: The 2010 Annual*

Conference of the North American Chapter of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 100–108.

[132] Blei, David M., and John D. Lafferty., 2006, "Dynamic topic models." *Proceedings of the 23rd international conference on Machine learning*.

[133] Pujara, Jay, and Peter Skomoroch., 2012, "Large-scale hierarchical topic models." *NIPS Workshop on Big Learning*. Vol. 128.

[134] Liu, Y., Liu, Z., Chua, T. S., & Sun, M., 2015, "Topical word embeddings." *Twenty-ninth AAAI conference on artificial intelligence*.

[135] Goucher-Lambert, K., and Cagan, J., 2019, "Crowdsourcing Inspiration: Using Crowd Generated Inspirational Stimuli to Support Designer Ideation," Des. Stud., **61**, pp. 1–29.

[136] Linsey, J. S., Wood, K. L., and Markman, A. B., 2008, "Modality and Representation in Analogy," *AI EDAM*, pp. 85–100.

[137] Moreno, D. P., Hernandez, A. a., Yang, M. C., Otto, K. N., Holtta-Otto, K., Linsey, J. S., Wood, K. L., and Linden, A., 2014, "Fundamental Studies in Design-by-Analogy: A Focus on Domain-Knowledge Experts and Applications to Transactional Design Problems," Des. Stud., **35**(3), pp. 232–272.

[138] Cagan, J., Kotovsky, K., and McComb, C., 2018, "Are Teams Teams? And Other Musings on Humpty Dumpty's Use of Language In The Design World," *ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. 1–3.

[139] Gentner, D., 1983, "Structure-Mapping: A Theoretical Framework for Analogy," Cogn. Sci., **7**(2), pp. 155–170.

[140] Forbus, K. D., Gentner, D., and Law, K., 1995, "MAC/FAC: A Model of Similarity-

Based Retrieval," Cogn. Sci., **19**(2), pp. 141–205.

[141] Krawczyk, D. C., McClelland, M. M., Donovan, C. M., Tillman, G. D., and Maguire, M. J., 2010, "An FMRI Investigation of Cognitive Stages in Reasoning by Analogy.," Brain Res., **1342**, pp. 63–73.

[142] Chan, J., Dow, S. P., and Schunn, C. D., 2015, "Do the Best Design Ideas (Really) Come from Conceptually Distant Sources of Inspiration?," Des. Stud., **36**(C), pp. 31–58.

[143] Moss, J., Kotovsky, K., and Cagan, J., 2007, "The Influence of Open Goals on the Acquisition of Problem-Relevant Information.," J. Exp. Psychol. Learn. Mem. Cogn., **33**(5), pp. 876–891.

[144] Moss, J., Kotovsky, K., and Cagan, J., 2011, "The Effect of Incidental Hints When Problems Are Suspended before, during, or after an Impasse.," J. Exp. Psychol. Learn. Mem. Cogn., **37**(1), pp. 140–148.

[145] Tseng, I., Moss, J., Cagan, J., and Kotovsky, K., 2008, "The Role of Timing and Analogical Similarity in the Stimulation of Idea Generation in Design," Des. Stud., **29**(3), pp. 203–221.

[146] Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., and Yaniv, I., 1995, "Demystification of Cognitive Insight: Opportunistic Assimilation and the Prepared-Mind Perspective," *The Nature of Insight*.

[147] Shi, F., Chen, L., Han, J., and Childs, P., 2017, "A Data-Driven Text Mining and Semantic Network Analysis for Design Information Retrieval," J. Mech. Des. Trans. ASME, **139**(11), pp. 1–14.

[148] Han, J., Shi, F., Chen, L., and Childs, P. R. N., 2018, "A Computational Tool for Creative Idea Generation Based on Analogical Reasoning and Ontology," Artif. Intell. Eng. Des.

Anal. Manuf. AIEDAM, **32**(4), pp. 462–477.

[149] Han, J., Shi, F., Chen, L., and Childs, P. R. N., 2018, "The Combinator-a Computer-Based Tool for Creative Idea Generation Based on a Simulation Approach," Des. Sci., **4**, pp. 1–34.

[150] Pennington, J., Socher, R., and Manning, C., 2014, "Glove: Global Vectors for Word Representation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543.

[151] Gilon, K., Ng, F. Y., Chan, J., Assaf, H. L., Kittur, A., and Shahaf, D., 2018, "Analogy Mining for Specific Design Needs," *Conference on Human Factors in Computing Systems - Proceedings*.

[152] Chan, J., Chee Chang, J., Hope, T., Shahaf, D., and Kittur, A., 2018, "SOLVENT: A Mixed Initiative System for Finding Analogies between Research Papers," *Proc. ACM Hum.-Comput. Interact.*, pp. 1–21.

[153] Gyory, J. T., Goucher-Lambert, K., Kotovsky, K., and Cagan, J., 2019, "Exploring the Application of Network Analytics in Characterizing a Conceptual Design Space," Proc. Des. Soc. Int. Conf. Eng. Des., **1**(1), pp. 1953–1962.

[154] Ahmed, F., Ramachandran, S. K., Fuge, M., Hunter, S., and Miller, S., 2019, "Interpreting Idea Maps: Pairwise Comparisons Reveal What Makes Ideas Novel," J. Mech. Des. Trans. ASME, **141**(2).

[155] Hu, W.-L., and Reid, T., 2018, "The Effects of Designers' Contextual Experience on the Ideation Process and Design Outcomes," J. Mech. Des., **140**(10), p. 101101.

[156] Kennedy, C., 2013, "Two Sources of Subjectivity : Qualitative Assessment and Dimensional Uncertainty Two Sources of Subjectivity : Qualitative Assessment and

Dimensional Uncertainty," Inquiry, **56**(2–3), pp. 258–277.

[157] Dieter, G. E., and Schmidt, L. C., 2013, *Engineering Design*, McGraw-Hill, New York, NY.

[158] Verhaegen, P.-A., Vandevenne, D., and Duflou, J. R., 2012, "Originality and Novelty: A Different Universe," *International Design Conference - Design 2012*, Dubrovnik - Croatia, pp. 1961–1966.

[159] Nieto, M. J., and Santamaría, L., 2007, "The Importance of Diverse Collaborative Networks for the Novelty of Product Innovation," Technovation, **27**(6–7), pp. 367–377.

[160] Stephen, G. G., Gavin, M., and Lynda, A.-S., 1995, "Assessing a Multidimensional Measure of Radical Technological Innovation," **42**(3), p. 232.

[161] Goucher-Lambert, K., and Cagan, J., 2017, "Using Crowdsourcing to Provide Analogies for Designer Ideation in a Cognitive Study," *International Conference on Engineering Design*, Vancouver, B.C, pp. 1–11.

[162] Miller, S. R., Bailey, B. P., and Kirlik, A., 2014, "Exploring the Utility of Bayesian Truth Serum for Assessing Design Knowledge.," Human-Computer Interact., **29**(5–6), pp. 487–515.

[163] Viswanathan, V. K., and Linsey, J. S., 2013, "Design Fixation and Its Mitigation: A Study on the Role of Expertise," *Journal of Mechanical Design*, p. 051008.

[164] Wilson, J. O., Rosen, D., Nelson, B. a., and Yen, J., 2010, "The Effects of Biological Examples in Idea Generation," Des. Stud., **31**(2), pp. 169–186.

[165] Gero, J. S., Jiang, H., Dobolyi, K., Bellows, B., and Smythwood, M., 2015, "How Do Interruptions During Designing Affect Design Cognition," *Design Computing and Cognition '14*, pp. 119–133.

[166] Cades, D. M., Boehm-Davis, D. A., Trafton, J. G., and Monk, C. A., 2011, “Mitigating Disruptive Effects of Interruptions Through Training: What Needs to Be Practiced?,” J. Exp. Psychol. Appl., **17**(2), pp. 97–109.

[167] Hess, S. M., and Detweiler, M. C., 1994, “Training to Reduce the Disruptive Effects of Interruptions,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pp. 1173–1177.

[168] Ratwani, R. M., and Trafton, J. G., 2011, “A Real-Time Eye Tracking System for Predicting and Preventing Postcompletion Errors,” Human-Computer Interact., **26**(3), pp. 205–245.

[169] Sio, U. N., Kotovsky, K., and Cagan, J., 2017, “Interrupted: The Roles of Distributed Effort and Incubation in Preventing Fixation and Generating Problem Solutions,” Mem. Cogn., **45**(4), pp. 553–565.

[170] Mullen, B., Johnson, C., and Salas, E., 1991, “Productivity Loss in Brainstorming Groups: A Meta-Analytic Integration Productivity Loss in Brainstorming Groups,” Basic Appl. Soc. Psych., **12**(1), pp. 3–23.

[171] Paulus, P. B., Dzindolet, M., and Kohn, N. W., 2012, “Collaborative Creativity-Group Creativity and Team Innovation,” *Handbook of Organizational Creativity*, Elsevier Inc., pp. 327–357.

[172] Horwitz, S. K., and Horwitz, I. B., 2007, “The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography,” J. Manage., **33**(6), pp. 987–1015.

[173] Dahlin, K. B., Weingart, L. R., and Hinds, P. J., 2005, “Team Diversity and Information Use,” Acad. Manag. J., **48**(6), pp. 1107–1123.

[174] Tribelsky, E., and Sacks, R., 2011, “An Empirical Study of Information Flows in Multidisciplinary Civil Engineering Design Teams Using Lean Measures,” Archit. Eng. Des. Manag., **7**(2), pp. 85–101.

[175] Karau, S. J., and Wilhau, A. J., 2019, “Social Loafing and Motivation Gains in Groups: An Integrative Review,” *Individual Motivation within Groups: Social Loafing and Motivation Gains in Work, Academic, and Sports Teams*, Elsevier Inc., pp. 3–51.

[176] Gardner, H. K., 2012, “Performance Pressure as a Double-Edged Sword,” Adm. Sci. Q., **57**(1), pp. 1–46.

[177] Cohen, G.P., 1993. The virtual design team: an information-processing model of design team management. (Doctoral dissertation, Stanford University).

[178] Kunz, J.C., Cohen, G.P. and Levitt, R.E., 1993. Modeling effects of organizational structure and communication tools on design team productivity. *Prietula, Michael, editor*.

[179] De Montjoye, Y. A., Stopczynski, A., Shmueli, E., Pentland, A., and Lehmann, S., 2014, “The Strength of the Strongest Ties in Collaborative Problem Solving,” Sci. Rep., **4**(1), pp. 1–6.

[180] den Otter, A., and Emmitt, S., 2008, “Design Team Communication and Design Task Complexity: The Preference for Dialogues,” Archit. Eng. Des. Manag., **4**(2), pp. 121–129.

[181] Senescu, R. R., Aranda-Mena, G., and Haymaker, J. R., 2013, “Relationships between Project Complexity and Communication,” J. Manag. Eng., **29**(2), pp. 183–197.

[182] Heisig, P., Clarkson, P. J., and Vajna, S., 2010, *Modelling and Management of Engineering Processes*, Springer.

[183] Senescu, R. R., and Haymaker, J. R., 2009, “Specifications for a Social and Technical

Environment for Improving Design Process Communication," *Proc. of the 26th Int. Conf. on IT in Constr*, pp. 227–237.

[184] Zurita, N. F. S., and Tumer, I. Y., 2017, "A Survey: Towards Understanding Emergent Behavior in Complex Engineered Systems," *Proceedings of the ASME Design Engineering Technical Conference*, American Society of Mechanical Engineers (ASME).

[185] Song, B., Soria Zurita, N. F., Zhang, G., Stump, G., Balon, C., Miller, S. W., Yukish, M., Cagan, J., and McComb, C., 2020, "Toward Hybrid Teams: A Platform to Understand Human-Computer Collaboration During The Design Of Complex Engineered Systems," Proc. Des. Soc. Des. Conf., **1**, pp. 1551–1560.

[186] Koch, J., and Paris-Saclay, I., 2017, "Design Implications for Designing with a Collaborative AI," *AAAI Spring Symposium Series* .

[187] Camburn, B., Arlitt, R., Anderson, D., Sanaei, R., Raviselam, S., Jensen, D., and Wood, K. L., 2020, "Computer-Aided Mind Map Generation via Crowdsourcing and Machine Learning," Res. Eng. Des., **31**, pp. 383–409.

[188] Camburn, B., He, Y., Raviselvam, S., Luo, J., and Wood, K., 2020, "Machine Learning-Based Design Concept Evaluation," J. Mech. Des. Trans. ASME, **142**(3).

[189] Dering, M. L., Tucker, C. S., and Kumara, S., 2018, "An Unsupervised Machine Learning Approach to Assessing Designer Performance during Physical Prototyping," J. Comput. Inf. Sci. Eng., **18**(1).

[190] Williams, G., Meisel, N. A., Simpson, T. W., and McComb, C., 2019, "Design Repository Effectiveness for 3D Convolutional Neural Networks: Application to Additive Manufacturing," J. Mech. Des. Trans. ASME, **141**(11).

[191] Jin, Y., Levitt, R.E., Christiansen, T.R. and Kunz, J.C., 1995. The virtual design team: modeling organizational behavior of concurrent design teams. *AI EDAM*, *9*(2), pp.145-158.

[192] Zhang, G., Raina, A., Cagan, J., and McComb, C., 2021, "A Cautionary Tale about the Impact of AI on Human Design Teams," Des. Stud., **72**, p. 100990.

[193] Rao, S. S., Nahm, A., Shi, Z., Deng, X., and Syamil, A., 1999, "Artificial Intelligence and Expert Systems Applications in New Product Development - a Survey," J. Intell. Manuf., **10**(3), pp. 231–244.

[194] Raina, A., Cagan, J., and McComb, C., 2019, "Transferring Design Strategies from Human to Computer and across Design Problems," J. Mech. Des. Trans. ASME, **141**(11).

[195] Maier, T., Zurita, N. F. S., Starkey, E., Spillane, D., Menold, J., and McComb, C., 2020, "Analyzing the Characteristics of Cognitive-Assistant-Facilitated Ideation Groups," *Proceedings of the ASME Design Engineering Technical Conference*.

[196] Maier, T., Abdullah, S., McComb, C., and Menold, J., 2021, "A Query Conundrum: The Mental Challenges of Using a Cognitive Assistant," SN Comput. Sci., **2**(3), p. 194.

[197] Gyory, J., Kotovsky, K., and Cagan, J., 2021, "The Influence of Process Management: Uncovering the Impact of Real-Time Managerial Interventions via a Topic Modeling Approach," J. Mech. Des., pp. 1–46.

[198] "HyForm™ · GitHub" [Online]. Available: https://github.com/hyform/drone-testbed-server/releases/tag/2021-March-v2. [Accessed: 23-Apr-2021].

[199] Zhang, G., Soria Zurita, N. F., Stump, G., Song, B., Cagan, J., and McComb, C., 2021, "Data on the Design and Operation of Drones by Both Individuals and Teams," Data Br., **36**, p. 107008.

[200] Cades, D. M., Boehm-Davis, D. A., Trafton, J. G., and Monk, C. A., 2011,"Mitigating Disruptive Effects of Interruptions Through Training: What Needsto Be Practiced?," J. Exp. Psychol. Appl.,17(2), pp. 97–109.

[201] Hess, S. M., and Detweiler, M. C., 1994,"Training to Reduce the Disruptive Effects of Interruptions," Proc Human Factors Ergonom. Soc. Annual Meet.,38(18), pp. 1173–1177.

[202] Lock S. Anachronistic intervention: Performative collaborative design in the wild. CoDesign. 2013 Mar 1;9(1):17-36.