LESSONS LEARNED IN THE SPACE SECTOR: AN INTERACTIVE TOOL TO DISSEMINATE LESSONS LEARNED TO SYSTEMS ENGINEERS

by

Kimberly A. Rink

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

School of Aeronautics and Astronautics

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Karen Marais, Chair
School of Aeronautics and Astronautics
Dr. Colin Gray
Department of Computer Graphics Technology
Dr. David Spencer
School of Aeronautics and Astronautics

Approved by:

Dr. Weinong Chen
Head of the Graduate Program

ACKNOWLEDGMENTS

I would like to thank my family and friends for supporting me throughout my educational journey. Thank you for encouraging me to pursue my dream and for giving me the strength to carry on through difficult times. Thank you mom for firmly planting my feet upon the ground, and thank you dad for keeping my head in the stars.

I would like to thank everyone that helped make this work possible. I especially want to thank Dr. Colin Gray, who introduced me to powerful qualitative analysis techniques; Theodore Sorg, whose undergraduate research assistance improved both my project and my mental health; and Zachary Yater, whose programming skills and UX design expertise helped realize the network.

Thank you to everyone who made every day in the office a little bit brighter. To my graduate school mentor, Diane Aloisio: thank you for being my example of an empathetic and fearless engineer. To the members of the VRSS research group: thank you for being a welcoming, challenging, and endlessly supportive group of friends.

Finally, to my adviser, Dr. Karen Marais: thank you for helping me become a better researcher, author, and communicator. But most importantly, thank you for supporting the passions and aspirations of all your graduate students.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Rink, Kimberly, A. MSAAE
Institution: Purdue University
Degree Received: May 2019
Title: Lessons Learned in the Space Sector: An Interactive Tool to Disseminate Lessons Learned to Systems Engineers.
Committee Chair: Karen Marais

Organizations, like individuals, are expected to learn from their mistakes. Companies that successfully rely on past knowledge to inform programmatic decisions use knowledge management tools to capture and disseminate this information, often in the form of lessons learned databases. However, past mistakes continue to happen in the aerospace industry, including NASA. Although NASA has taken measures to stress the importance of lessons learned in organizational culture, relatively little work has been done to develop the user interface of their lessons learned database. Encouraging engineers to review lessons only goes so far when the interface itself is outdated and difficult to use. We propose that an interactive network tool is an effective way to disseminate lessons learned to novice systems engineers.

In this thesis, I begin by developing a model to represent spacecraft anomaly narratives and applying this model to the Jet Propulsion Laboratory's publicly available lessons learned database. I then create an interactive network tool and populate it with the set of modeled lessons. Then, I design an experiment to determine how novice engineers use two different knowledge management tools—the interactive network and the NASA database. I use transcripts of users' thought processes, verbalized to me during the experiment, to create a mental model of how users with access to knowledge management tools respond to engineering scenarios. From the mental model, I identify the functional strengths and weakness of both the interactive network and the NASA database. Finally, I discuss the results of the experiment and recommend future improvements to the interactive network tool.

We found that the interactive network was a better resource for users to make connections between topics, and that the NASA database was a better resource for users to search for specific information. Using the interactive network over the NASA database correlated with an increase in

performance for the majority of the experiment, but data we collected do not provide enough evidence for us to conclude that the interactive network is a better dissemination tool than the NASA database in all scenarios. We found that receiving lessons learned from either of the tools takes time because each tool's functionality elicits new tasks from the user. Finally, we found that the top performers in the experiment used each of the tool's strongest features.

CHAPTER 1. INTRODUCTION

1.1 Motivation

Mistakes are inevitable. Some mistakes arise from a complex and unpredictable series of events, some mistakes are due to human error, and some mistakes seem painfully preventable. In the event of a significant mistake, engineers use a patchwork of accident models to understand what went wrong. But sometimes, the only thing we as individuals can do is to learn from past mistakes and make sure the same situation is never repeated.

Like individual people, companies are also expected to have a memory of institutional knowledge and learn from their own experiences. Successful organizations do just that—use past mistakes and knowledge to inform programmatic decisions. Consider for example NASA's Juno mission. In 2016, the Juno spacecraft inserted into orbit around Jupiter to investigate the planet's formation and evolution. The Juno operations team at the Jet Propulsion Laboratory (JPL) learned heavily from two planetary missions that came before it—avionics knowledge was inherited from the Mars Reconnaissance Orbiter (MRO), and records from the Galileo mission helped Juno overcome Jupiter's harsh radiation environment. According to the Juno project manager, the team relied more on information handed down from MRO and Galileo than from Juno's own mission data to resolve major anomalies (NASA Jet Propulsion Laboratory, 2017). Juno was successful in part because information was retained from previous projects, made available to the new operations team, and accepted by the new team; in other words, a framework was in place to share information and people were willing to use it.

How do companies capture critical information for later use? Institutional experience is often recorded (in part) in knowledge management systems, such as lessons learned databases. According to a 2002 United States General Accounting Office (GAO) audit of NASA's knowledge management practices, in the context of space exploration a *lesson learned* is defined as "knowledge or understanding gained by experience […] A lesson must be significant in that it has a real or assumed impact on operations; valid in that it is factually correct; and applicable in that it identifies a specific design, process, or decision that reduces or eliminates the potential for failures and mishaps, or reinforces a positive result." NASA created the publicly-available web-based Lessons Learned Information System (LLIS) as the main interface for agency-wide collection and sharing of its extensive lessons learned database (GAO, 2002).

Are these tools effective in helping organizations learn from experience? This question is difficult to answer because few knowledge management systems are publicly accessible and most are rarely reported on, as private companies tend to keep mistakes close to the vest. However, two observations give insight into the usefulness of these tools. First, we know that the same mistakes keep happening. Previous work by Aloisio & Marais identified factors that contributed to project failures and accidents across multiple industries. The factor "did not learn from failure" appeared in 50% of the accidents they studied, including prominent examples in the space industry such as the Hubble Telescope (Aloisio, 2015). Second, internal and external audits continue to find flaws in knowledge management systems. Over the past decade, several reports were published that criticized NASA's knowledge management process in particular. In addition to the aforementioned GAO report, a NASA Inspector General audit in 2012 concluded that NASA's lessons learned database was a key component in its overall knowledge management strategy, but the database itself was too under-used and under-funded to be useful. The fact that companies often make the same mistakes together with the negative reviews of NASA's knowledge management systems provide evidence that the process of exploiting lessons learned, especially in the space sector, has significant room for improvement.

What part of the knowledge management process in particular needs improvement? We attempt to answer this question through the lens of NASA's knowledge management practices. Li et al. (2016) identify three organizational levels at which knowledge management practices tend to fail: from an interaction-oriented perspective, a people-oriented perspective, and a system-oriented perspective. For each level, we provide a definition, cite examples of its existence in NASA's knowledge management practices, and discuss whether and how NASA has responded to it. The first level is the *interaction-oriented perspective*, which explains that resistance to knowledge management systems comes from organizational culture and social context. Several reports provide evidence of cultural resistance to knowledge management at NASA. According to the GAO report (2002), "respondents indicated that LLIS, NASA's primary method for disseminating lessons learned agency-wide, is not the primary source for lessons learning”. Ten years after this report, an audit by the NASA Inspector General (2012) found that “the Chief Engineer’s overall strategy for knowledge management, lessons learned, and LLIS is not well defined. Consequently, LLIS has been marginalized in favor of other NASA knowledge sharing system components and is of diminishing and questionable value”. Additionally, Dennehy et al. (2010) observe that “program and Project managers, together with Chief Engineers, need to continue to find ways to embed these lessons in NASA [sic] design & development processes.” In an attempt to shift the culture, NASA has responded to these critiques by creating new rules that ensure more eyes pass over lessons learned and institutional information. For example, JPL has incorporated lessons learned into Design Principles and Flight Project Practices, which are sets of mandatory rules for projects. Many rules contained in these documents can be directly traced back to a JPL lesson learned. Additionally, NASA has created requirements for managers to review and contribute to the database. However, introducing new rules also has its downsides. Synthesizing a nuanced and complex lesson into a sentence-long flight rule allows the issue to lose resolution, and the more institutional rules created, the more projects tend to waive rules that seem inapplicable. Nevertheless, NASA consistently addresses challenges from an interaction-oriented perspective and is working to make learning from past mistakes a key pillar in its safety culture (Columbia Accident Investigation Board, 2003)

The second level of resistance according to Li et al. is the *people-oriented perspective*, which explains that resistance to knowledge management systems comes from factors internal to the working group. At NASA, the necessary rules are in place, but managers’ lack of engagement and reinforcement may reduce the effectiveness of these practices. The NASA Inspector General (2012) found that “NASA’s project managers do not routinely use LLIS to search for lessons identified by other projects or routinely contribute new information to LLIS”. This is also emphasized by Dennehy et al. (2010), who state that “although the GN&C engineering practitioners across the Agency are often reminded of the importance of (and in some organizations, the requirement of) applying relevant lessons learned to their individual day-to-day tasks, there is little in the way of specialized education, training, and materials made available to help those engineers do a better job of managing critical knowledge and capturing lessons learned”. NASA has responded to critiques from the people-oriented perspective by creating and incentivizing programs that familiarize local working groups with lessons learned. NASA has implemented *pause and learn* sessions, which are described as structured meetings where team members reflect upon and share lessons learned from specific project events. NASA has found pause and learn sessions to be "an effective low-impact way to: a) identify and spread local best practices, b) implement on-the-spot individual and team learning, c) build a team approach to problem solving, d) build team morale, and e) increase likelihood of project success" (Dennehy et al., 2010). Similar to the interaction-oriented perspective, challenges from a people-oriented perspective have been acknowledged by NASA and remain a work in progress.

The third and final level of resistance is the *system-oriented perspective*, which explains that resistance to knowledge management systems comes from design of the system itself. Reports often cite the poor usability of the LLIS. According to the GAO report (2002), "one reason LLIS is not widely used, according to the center official, is because its lessons cover so many topics that it is difficult to search for an applicable lesson. Another respondent indicated that it is difficult to weed through all the irrelevant lessons to get to the few 'jewels' that you need to find." Additionally, "although the system is viewed as providing a useful repository for storing lessons, officials agreed with managers' concerns about the difficulties involved in searching the system and finding relevant lessons, the inconsistent quality of information contained in the system, and the lack of lessons about positive project experiences". However, NASA's response to critiques of the actual system design has been minimal—the LLIS has roughly stayed the same since the website was introduced in 2005. Of the three perspectives that explain resistance to knowledge management, the system-oriented perspective is the one that NASA has addressed least. To us, this suggests that the mechanism of disseminating lessons learned to employees—the LLIS—is in need of improvement. **Is there a better mechanism to present lessons learned directly to engineers that can benefit from them?**

1.2 Approach

To answer this question, we created our own interactive tool to disseminate lessons learned by altering a previous knowledge management tool, the development of which we describe in this section. Previous research in our group by Aloisio & Marais (2018) identified 23 high-level causes that contribute to accidents and project failures in the aerospace, infrastructure, and energy industries. Additionally, they identified 16 high-level recommendations from accident investigation reports and meta-analyses of the same accidents and project failures. Then, they analyzed case studies of 33 project failures and 30 accidents to identify how often these causes and recommendations manifested, and they created a nodal network to visualize the resulting data, shown in Figure 1.1. The network consists of red and blue nodes that represent causes and recommendations, respectively. Nodes that are connected appear in at least one case study together, and the line weight of the connection indicates the intersectional probability between the two nodes. The location of the nodes is unconstrained; connected nodes attract each other, and disconnected nodes repel each other. This results in nodes with many connections gravitating towards the center of the network and nodes with fewer connections gravitating towards the perimeter.

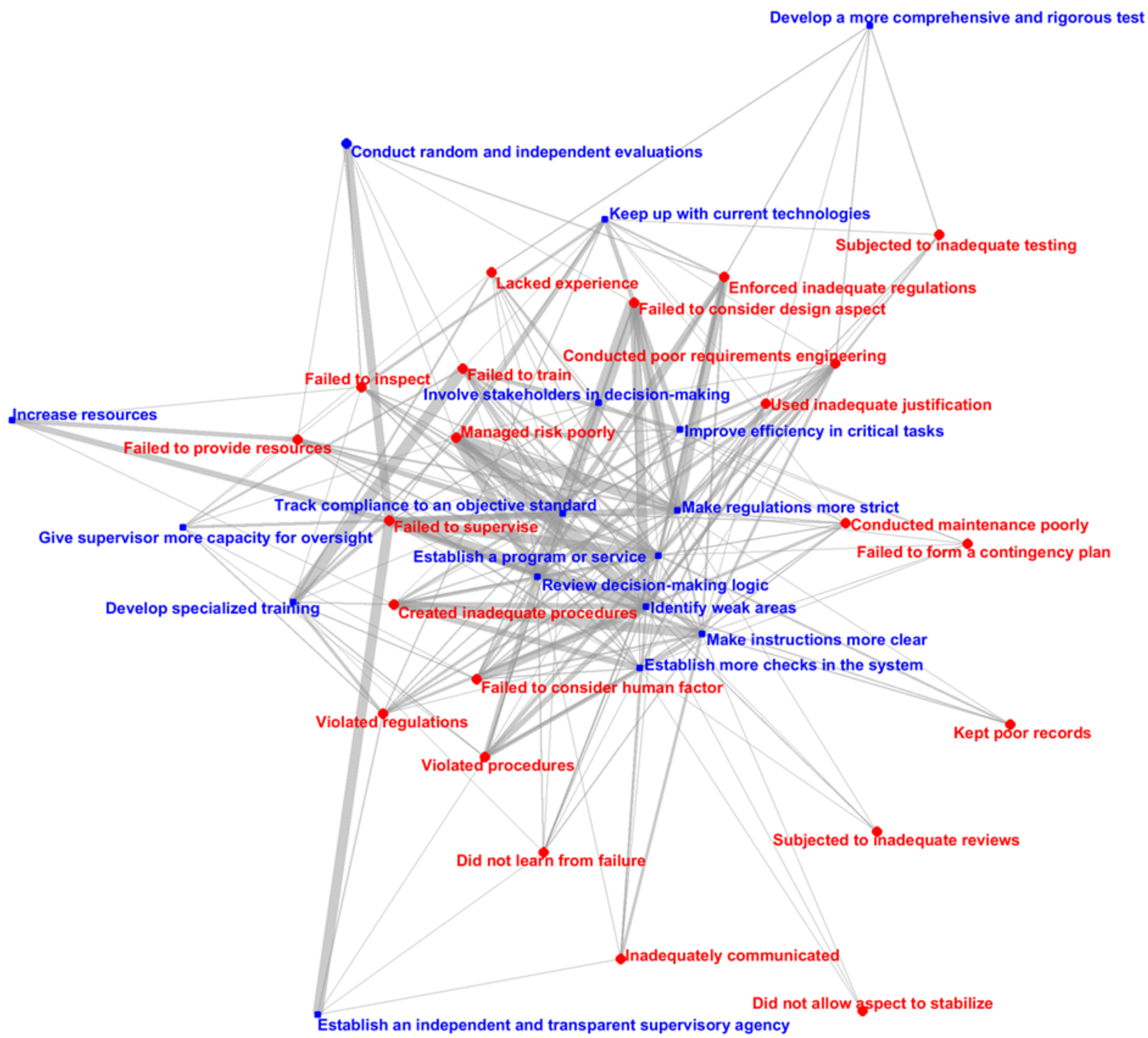


Figure 1.1: Static Cause/Recommendation Network (Aloisio & Marais, 2018)

With the findings represented visually, could learning about these common causes and recommendations from the network help improve engineers' performance in systems engineering scenarios? To answer this question, Aloisio & Marais (2018) turned the cause/recommendation network into a web-based tool with an interactive user interface to disseminate the findings to systems engineers, as shown in Figure 1.2. The structure of the tool is as follows: accident and project failure summaries below the network give examples of how the 23 causes and 16 recommendations were present in each case study. Upon clicking a certain node, a definition of the associated cause or recommendation appears on the left, connections to other nodes are highlighted, and relevant case studies appear below the network. Drop down menus at the top of the page let the user filter data by actors and by industries. Hereafter, this tool is referred to as the "cause/recommendation network".

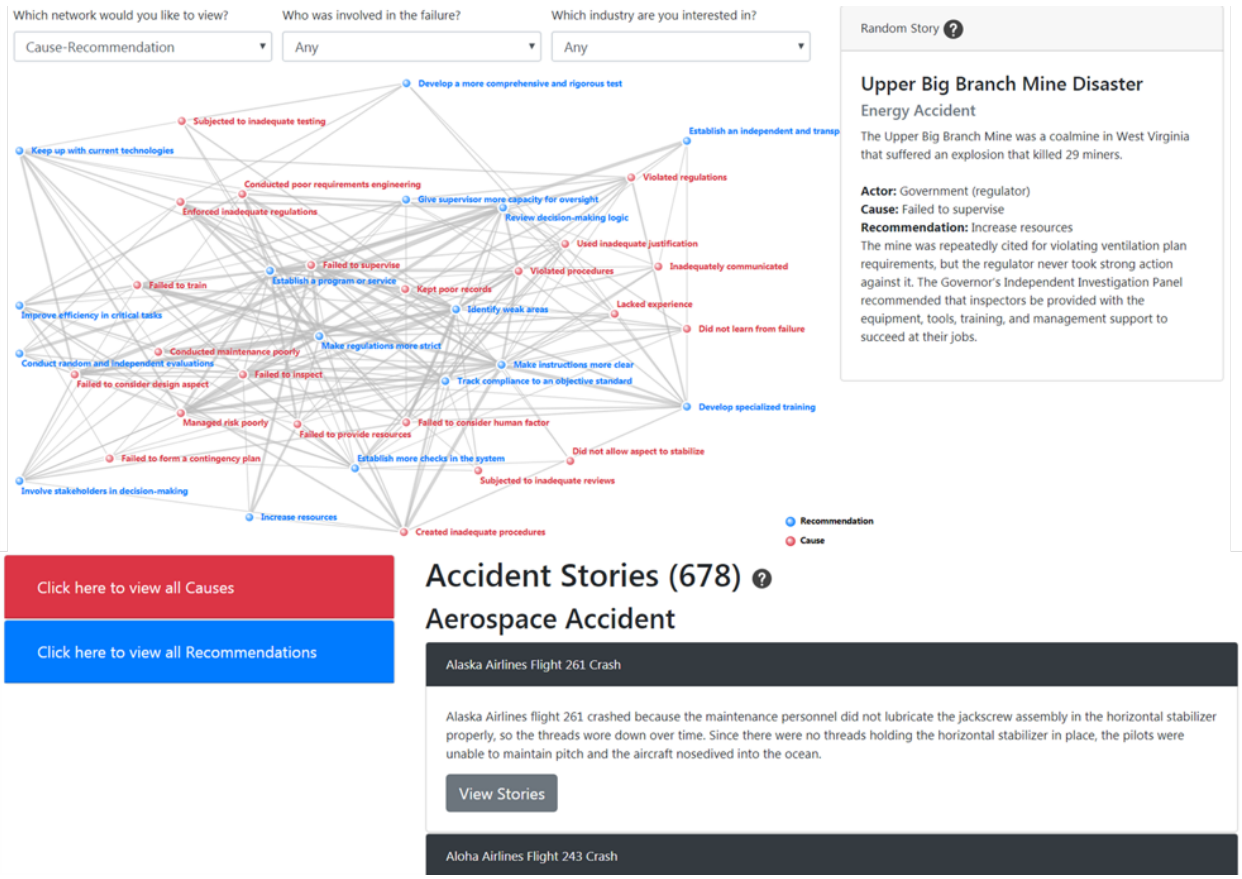


Figure 1.2: Interactive Cause/Recommendation Network Tool[1]

To determine whether the cause/recommendation network is effective at disseminating information about systems engineering failures, Aloisio (2018) designed and performed an experiment that sought to determine whether the network was useful for systems engineers forming remediation measures for aerospace projects. The experiment required participants to answer questions about two failure narratives in the aerospace industry—one aircraft related and one spacecraft related. Participants were given one of four tests that permuted their exposure to the tool (access to the tool for both questions, access to the tool for question 1 only, access to the tool

[1] The cause/recommendation network can be accessed here: https://engineering.purdue.edu/VRSS/research/force-graph/index_html.

for question 2 only, or no access to the tool). The interviewee population consisted of four “expert” engineers who work for a large aerospace company and 17 “novice” engineers who were Purdue Aeronautics and Astronautics graduate students. The results of the experiment proved that each population answered the questions and responded to the tool differently. The experts were not interested in using the tool to answer the questions and relied almost exclusively on prior experience. Unlike the experts, the novices could distinguish between subtleties of the questions and used the tool to provide multiple relevant recommendations for each question. Additionally, the novices used the tool heavily for the reportedly unfamiliar scenario of the two (the spacecraft scenario). According to Aloisio (2018), an explanation for the results is that “novices were used to encountering problems they had no experience with and used all of the tools at their disposal.” The conclusion of the experiment was that the tool was useful for novice engineers when recommending solutions to engineering failures.

At the end of the experiment, the participants gave feedback on the cause/recommendation network tool. The four experts primarily left feedback that critiqued *how* the tool should be used, speculating that the tool may not fit seamlessly into their established work practices and suggesting alternate applications for the tool, whereas novices focused more on critiquing the tool’s specific features. The most common feedback from experts was that they believed the tool could be more helpful for younger engineers with less systems experience. Additionally, several experts wondered if the tool could be used to disseminate a company’s own lessons and institutional knowledge (Aloisio, 2018).

As discussed in the motivation, an opportunity exists for a different user interface other than NASA’s LLIS to disseminate internal lessons learned to engineers. According to Aloisio & Marais’ experiment, the cause/recommendation network tool is most useful for novice engineers in unfamiliar scenarios, and according to the feedback, there is clear interest in populating a nodal network-type tool with a company’s internal data. It is possible that a redesigned nodal network tool populated with NASA lessons learned information can better disseminate these lessons to novice systems engineers than the NASA LLIS. However, the only way to know whether these tools are effective is to first understand *how* they are used and what role they play in a novice engineer’s work practice.

This leads us to our research question: **What role do knowledge management tools with lessons learned information play in the systems engineering decisions of novice engineers?** This thesis attempts to answer the question through the creation of an alternate nodal network-based knowledge management tool and the execution of an experiment to determine how novice engineers use both the new tool and the original NASA LLIS.

Improvements in knowledge management processes are desired because just as engineers of all experience levels are expected to learn and grow from their mistakes, organizations must too. But a common belief is that *systems engineering cannot be taught—it must be experienced*. It is reasonable to surmise though that it would be beneficial to educate novice engineers on mistakes that have happened in the past, especially mistakes at their own organization. Perhaps depicting how these knowledge management tools are used may uncover the strengths and weaknesses of each tool, which could address the deficiencies of the NASA LLIS mentioned in Section 1.1. A more interactive tool may increase novice engineers' participation and interest in learning from their organization's past successes and failures. By attempting to improve the user interface of knowledge management systems, we are focusing on the individual's role (especially the novice engineer) in that pursuit of gaining systems engineering experience, even if they have not necessarily "lived it" yet.

1.3 Thesis Outline

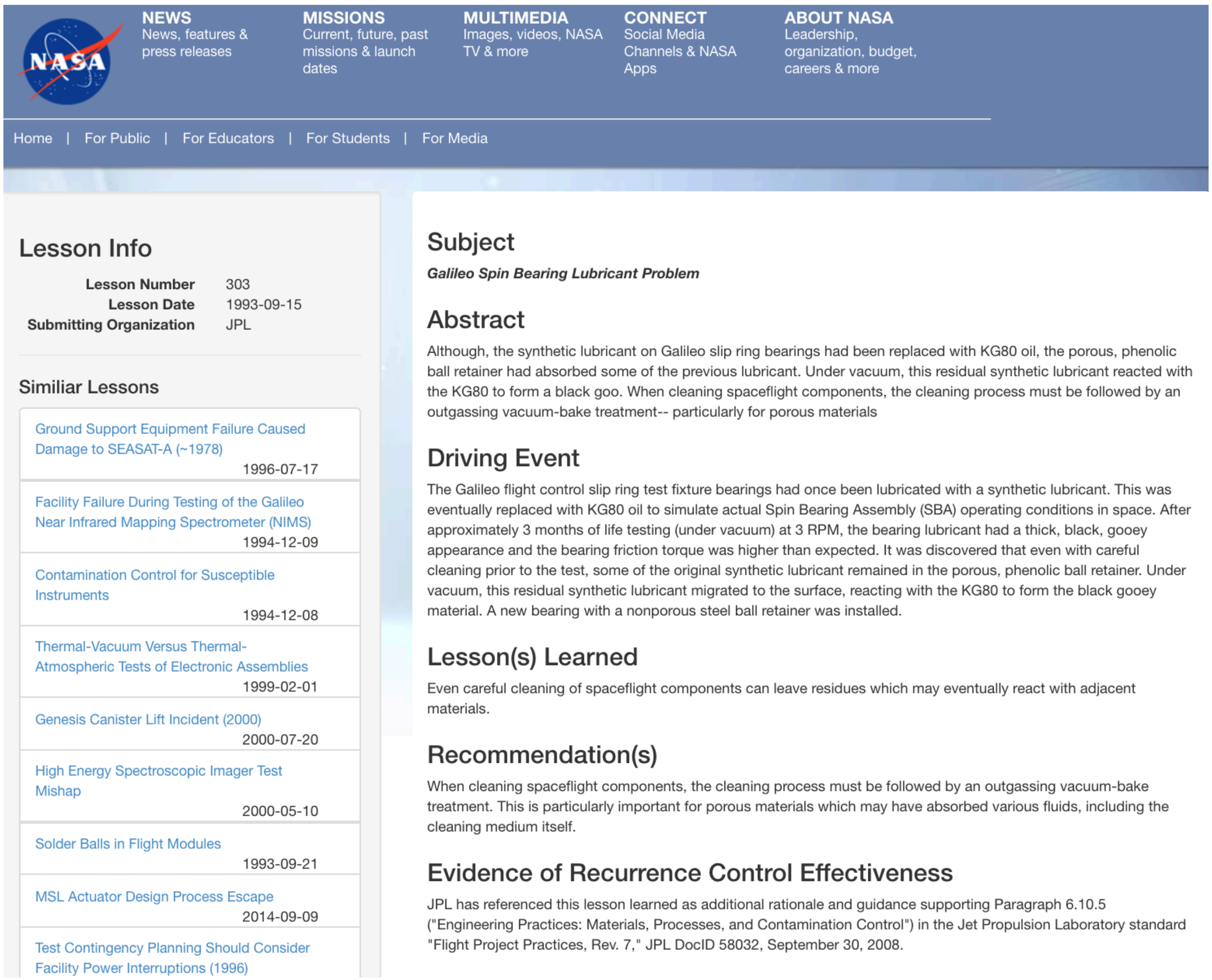
This thesis first describes how we developed an alternative knowledge management tool to disseminate NASA lessons learned, then it describes how we determined how potential users interacted with two different tools—the new tool and the NASA LLIS. In Chapter 2, we develop a model to represent the narrative of each individual lesson and a coding scheme to identify factors that occur within the model. In Chapter 3, we update the design of the existing cause/recommendation network tool to facilitate our model, populate the new tool with coded NASA lessons learned data, and describe and justify other changes and usability features introduced to the new tool. In Chapter 4, we design an experiment to determine how novice engineers use both the NASA LLIS and the new tool to respond to engineering scenarios. In Chapter 5, we analyze the results from our experiment to identify users' demands on the tools and the strengths and weaknesses of each tool. Chapter 6 concludes the thesis.

CHAPTER 2. DEVELOPING A MODEL TO REPRESENT LESSONS LEARNED

In this chapter we describe the model we created to represent both the relevant technical and organizational information within a lesson learned. First, we provide a description of the NASA lessons learned database, the types of lessons that it captures, and our rationale for only considering a subset of these lessons. Next, we detail how we developed a model to represent the necessary components that describe a spacecraft anomaly, along with how we developed a coding scheme within the model to identify common factors between lessons. We provide an explanation for each type of code, and we conclude this chapter with an example of how to code a lesson from the database.

2.1 Background on NASA Lessons Learned Database

We used the NASA lessons learned database (the information contained within the LLIS) as the data that populates our tool, and we used the LLIS (the interface for accessing the information) as the baseline knowledge management system to which we compare our tool. Why are we using the NASA's knowledge management systems as opposed to another organization's? The first reason is that the LLIS is publicly available. Although there is wide agreement that capturing lessons learned is beneficial, very few opportunities exist for an outsider to access an organization's knowledge management tools. The LLIS is one of the only publicly available systems—private companies rarely make mistakes public, and other government agencies require credentials to access their knowledge management tools (Birkinshaw, 2001). Likely in part due to this accessibility, the LLIS is widely critiqued and frequently audited, as previously mentioned in the motivation. Second, the LLIS is easily accessible. It is web-based and open, therefore, anyone with an internet connection can access the information without log-in credentials. The information within the LLIS is relatively accessible as well because the user interface provides rudimentary ability to navigate topics and search keywords. Next, the LLIS follows a loose but consistent format. Each lesson consists of four categories: *Abstract*, *Driving Event*, *Lesson(s) Learned*, and *Recommendation(s)*, which made it easier to parse each lesson for relevant findings. Many lessons have a wealth of information about a specific event or practice. Finally, the LLIS is within our field of study. Our experiment participants (described in Section 4.2) are undergraduate and graduate aerospace engineering students that already have or will gain early-career industry experience in the aeronautics/astronautics field; therefore, most concepts in the NASA lessons learned database are familiar to them. Figure 2.1 gives an example of an individual lesson within the LLIS.


NEWS
News, features & press releases
MISSIONS
Current, future, past missions & launch dates
MULTIMEDIA
Images, videos, NASA TV & more
CONNECT
Social Media Channels & NASA Apps
ABOUT NASA
Leadership, organization, budget, careers & more

Home | For Public | For Educators | For Students | For Media

Lesson Info

Lesson Number	303
Lesson Date	1993-09-15
Submitting Organization	JPL

Similiar Lessons

Ground Support Equipment Failure Caused Damage to SEASAT-A (~1978) 1996-07-17

Facility Failure During Testing of the Galileo Near Infrared Mapping Spectrometer (NIMS) 1994-12-09

Contamination Control for Susceptible Instruments 1994-12-08

Thermal-Vacuum Versus Thermal-Atmospheric Tests of Electronic Assemblies 1999-02-01

Genesis Canister Lift Incident (2000) 2000-07-20

High Energy Spectroscopic Imager Test Mishap 2000-05-10

Solder Balls in Flight Modules 1993-09-21

MSL Actuator Design Process Escape 2014-09-09

Test Contingency Planning Should Consider Facility Power Interruptions (1996)

Subject

Galileo Spin Bearing Lubricant Problem

Abstract

Although, the synthetic lubricant on Galileo slip ring bearings had been replaced with KG80 oil, the porous, phenolic ball retainer had absorbed some of the previous lubricant. Under vacuum, this residual synthetic lubricant reacted with the KG80 to form a black goo. When cleaning spaceflight components, the cleaning process must be followed by an outgassing vacuum-bake treatment-- particularly for porous materials

Driving Event

The Galileo flight control slip ring test fixture bearings had once been lubricated with a synthetic lubricant. This was eventually replaced with KG80 oil to simulate actual Spin Bearing Assembly (SBA) operating conditions in space. After approximately 3 months of life testing (under vacuum) at 3 RPM, the bearing lubricant had a thick, black, gooey appearance and the bearing friction torque was higher than expected. It was discovered that even with careful cleaning prior to the test, some of the original synthetic lubricant remained in the porous, phenolic ball retainer. Under vacuum, this residual synthetic lubricant migrated to the surface, reacting with the KG80 to form the black gooey material. A new bearing with a nonporous steel ball retainer was installed.

Lesson(s) Learned

Even careful cleaning of spaceflight components can leave residues which may eventually react with adjacent materials.

Recommendation(s)

When cleaning spaceflight components, the cleaning process must be followed by an outgassing vacuum-bake treatment. This is particularly important for porous materials which may have absorbed various fluids, including the cleaning medium itself.

Evidence of Recurrence Control Effectiveness

JPL has referenced this lesson learned as additional rationale and guidance supporting Paragraph 6.10.5 ("Engineering Practices: Materials, Processes, and Contamination Control") in the Jet Propulsion Laboratory standard "Flight Project Practices, Rev. 7," JPL DocID 58032, September 30, 2008.

Figure 2.1: NASA Lessons Learned Database (LLIS)

Within the NASA lessons learned database, we only considered lessons submitted by the Jet Propulsion Laboratory (JPL). Considering only JPL lessons—418 out of the 2070 in total—provides us with a set of manageable size. JPL's contributions make up 20.2% of the database, which is the second-largest contribution from a single NASA center. Also, JPL is the only center that consistently contributed 10 or more lessons per year over a six-year period (NASA Office of Inspector General, 2012). Out of this set of 418 JPL lessons, we analyzed 413. We did not analyze five of the lessons because the content of these lessons was redacted from the publicly available database due to International Traffic in Arms Regulations (ITAR) restrictions.

2.2 Lessons Learned Model and Coding Scheme

When creating a model to represent lessons learned, we must keep in mind the purpose of the tool: a resource for engineers to learn more about a particular topic or workplace scenario. Our previous cause/recommendation network consists of high-level causes and recommendations that address the organizational side of systems engineering pitfalls rather than the technical side. However, we need a way to capture technical information as well as organizational information. The majority of JPL lessons describe a spacecraft anomaly or mishap (the minority of lessons refer to a preferred practice or a general observation not associated with a particular event). To create our model, we parsed through JPL lessons and identified five categories necessary to describe the narrative of a spacecraft anomaly.

Event: ***The proximate cause of the spacecraft anomaly.*** This facet of a spacecraft anomaly describes what happened to the spacecraft, e.g., component failure, loss of data, or accident.

Component: ***The part of the spacecraft that suffered the proximate cause.*** This facet of a spacecraft anomaly describes the part, assembly, or subsystem that experienced the failure mode, e.g., solar array, flight software, or pyrotechnic hardware.

Technical factor: ***Root causes of the anomaly that have no specific actor.*** This facet of a spacecraft anomaly captures all of the physical phenomena that contributed to the anomaly, e.g., part fatigue, voltage settings, or interference from other subsystems.

Organizational factor: ***Root causes of the anomaly inherent to a specific actor.*** These codes are the 23 high-level causes of systems engineering failures identified in previous work by Aloisio & Marais (2018), e.g., inadequately communicated, managed risk poorly, or failed to consider human factors.

Recommendation: ***Ways to mitigate or prevent the anomaly.*** Sixteen of these codes are the high-level recommendations for systems engineering failures (Aloisio & Marais, 2018). We supplemented this set with technical recommendations. Codes were developed from the recommendations explicitly given in the *Lesson(s) Learned* and *Recommendation(s)* sections of the lesson—they were not inferred, e.g., develop specialized training, add functional redundancy, or requalify heritage systems.

More potential categories of a spacecraft anomaly exist (e.g., actor or mission phase), but the five aforementioned categories were present in every spacecraft anomaly narrative that we parsed. We chose to represent five categories because any fewer would lose key narrative information and any more would jeopardize network clarity and legibility. Similar to how the cause/recommendation network had two types of nodes with corresponding colors—causes in red and recommendations in blue—the new lessons learned network will have five.

Once we identified the five categories, the next steps were to create a set list of factors that fall under each category, then code each lesson with the identified factors. We performed both of these tasks in parallel by constantly redefining the set list of factors while we coded all 413 lessons. We followed a similar process to our group's previous factor identification work: we individually parsed each lesson for relevant "findings" that fall under one of the five categories. We then associated each finding with a code. If no code existed to describe the finding, we created a new code. About halfway through the set of 413 lessons, we reached the desired level of code saturation because we were no longer identifying any new codes. Another aspect of the coding process involved determining the level of granularity, i.e., how to define factors that are general enough to be common across lessons without losing the resolution of each individual lesson. To address this, we performed a trade between specificity and generality for each code. At the beginning of the process, our codes tended to be very specific (e.g., "Rigorously test lower TRL technology"), but as we identified more codes, we iteratively refined and merged them to achieve the desired level of generality (e.g., "Vet new technology" was the final version of this code. "Test technology" is too general). Because each code would be represented as an individual node in the tool, regarding network clarity and legibility, fewer codes is better. A list of codes for each category is located in Appendix A.

Table 2.1 shows how we coded the lesson "Provide Adequate Maintenance and Hazard Response for UPS Units". The lesson describes an incident where an uninterrupted power supply (UPS) battery experienced a thermal runaway. Several technical and organizational factors contributed to the event, including improper voltage settings and a lack of routine maintenance (NASA Jet Propulsion Laboratory, 2014). We repeated this coding process for the remaining 412 JPL lessons with accessible content.

Table 2.1: Findings and Associated Codes for Lesson #12101

Finding	Type	Code
"On the morning of Monday July 14, 2014, **overheated** batteries in UPS-1 were found to be **emitting a very strong odor and toxic fumes**."	Event	Fire/overheating
"On the morning of Monday July 14, 2014, overheated **batteries** in UPS-1 were found to be emitting a very strong odor and toxic fumes."	Component	Battery
"It was found that the charger was **set to several volts higher** than the specified setting. This would result in the remaining 39 batteries receiving a charge voltage of 0.132 VDC per battery higher than normal… The battery manufacturer indicates that this **~0.1 V higher than the maximum recommended voltage value could have contributed to the thermal runaway**."	Technical factor	Voltage settings
"The UPS-1 system was **not equipped with adequate fault sensor safeguards**, such as cell temperature and voltage measurements, to warn of cells in thermal runaway. Also, the system **lacked malfunction or over-voltage alarms** for timely recognition of malfunctions, and it **lacked automatic shutoffs** for battery overcharging."	Technical factor	Inadequate hazard protection
"Where it is not clear whether purchase of facility equipment was funded by the project or by the institution (or where it was funded by a defunct project), then responsibility for maintenance may also be unclear. In the case of the UPS-1 system, **it had not been maintained** over the two years since it was installed."	Organizational factor	Conducted maintenance poorly
"Battery defects are not easy to predict. There were **no procedures in place** for facility personnel and facility users to respond to the symptoms of battery overheating."	Organizational factor	Created inadequate procedures

"**Equip all UPS units with visible and audio warning systems** for timely recognition and remote monitoring of malfunctions…. **Equip all UPS battery chargers to automatically stop battery charging** and **automatically isolate batteries upon over-temperature detection**."	Recommendation	Add hazard protection to spacecraft design
"**Train laboratory personnel** and other building occupants to recognize warning signals and hazardous conditions."	Recommendation	Develop specialized training

CHAPTER 3. CREATING A TOOL TO DISSEMINATE LESSONS LEARNED

In this chapter we describe how we created a new lessons learned dissemination tool. First, we describe the contextual design process we followed and how development of and experimentation with our group's previous knowledge management tool contributed to our iteration of contextual design. Next, we detail the steps we took to create a new iteration of the tool, focusing on how we used feedback from the previous tool to both change the scope of and add functionality to the new tool. Finally, we describe the final product and highlight its main features.

3.1 Contextual Design Process

Our goal was to design a new knowledge management tool that provides more paths than the LLIS to access the same technical information contained within the NASA lessons learned database. To develop the new tool, we followed Beyer & Holtzblatt's (1997) steps of contextual design, which are depicted in Figure 3.1. Contextual design is an iterative user experience design process grounded in customer feedback. According to Beyer & Holtzblatt, contextual design "makes data gathered from customers the base criteria for deciding what the system should do and how it should be structured". We treated the development of a new nodal network as an iteration of the contextual design process with the cause/recommendation network as a starting point. Although they did not necessarily execute specific contextual design steps, Aloisio & Marais (2018) developed the cause/recommendation network to the penultimate step of contextual design, which is *mock-up and test* with customer*s*. The feedback they collected regarding expert and novice engineers' opinions on the cause/recommendation network enabled us to implement the customer-centered contextual design process to create our new network. It was not necessary for us to start a new iteration from the first step of contextual design—because we wanted to change the application of the tool, redefine the customer, and change the data that populates the tool while retaining the core nodal network design, we started from the *work redesign* step. We used the feedback from the cause/recommendation network extensively to inform the redesign of the tool.

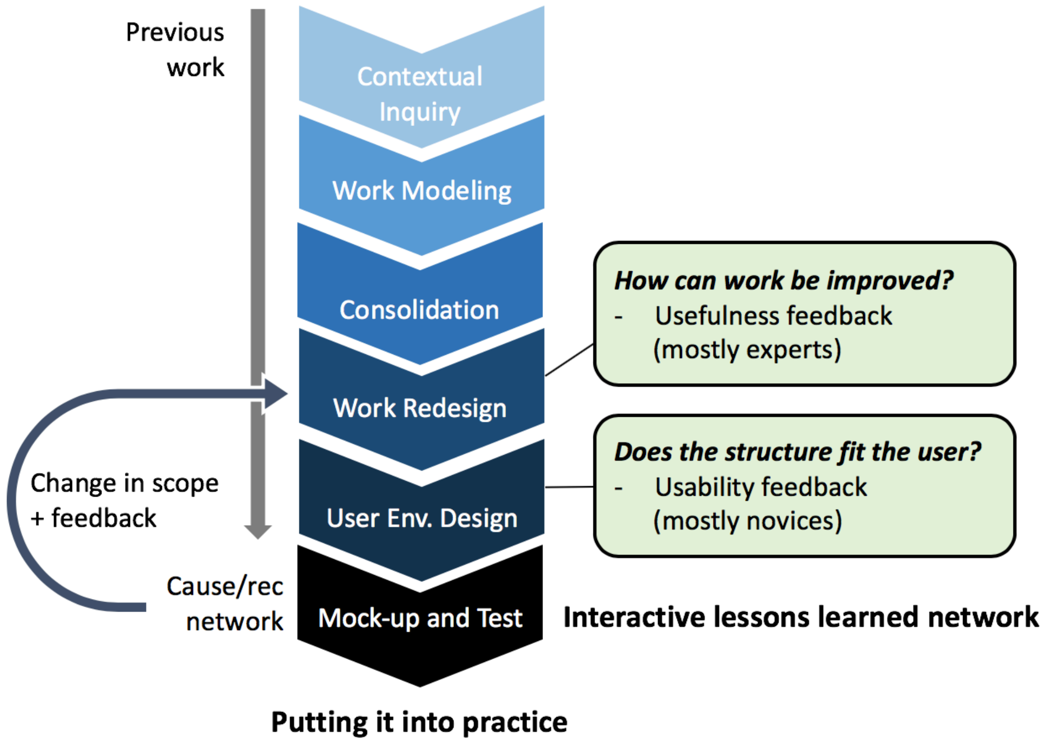


Figure 3.1: Steps of the Contextual Design Process (Beyer & Holtzblatt, 1997)

3.2 Cause/Recommendation Network Tool Redesign

Work redesign was our starting point in the contextual design process. According to Beyer & Holtzblatt, (1997) work practice refers to the general objective that must be met, and work redesign is where "the team uses the consolidated data to drive conversations about how work could be improved and what technology could be put in place to support the new work practice." In this step, the vision for improving the work practice drives the system definition. In our case, the work practice that can be improved is the dissemination of information to novice engineers, and an improved way to structure the work is to present the information in a more interactive way that focuses on commonalities between topics. The vision for a solution, which we discussed in Section 1.2, is what drove the changes to the previous tool to define the new system. These changes include giving it a new application and audience while retaining the basis of our tool, the nodal network layout.

Although the changes in scope of the new tool were inherent to our motivation for creating a new tool, they were also supported by feedback on the previous cause/recommendation network. The feedback Aloisio (2018) collected was split into two categories: usability and usefulness. Usefulness feedback consists of higher-level suggestions that relate to work practices, whereas usefulness feedback consists of lower-level recommendations regarding the details and features of the tool. The usability feedback, shown in Table 3.1, supports the changes to the cause/recommendation tool that we previously identified. For example, one participant recommended that the tool should be used to "train new systems engineers" which supports the change in target customer from all engineers to specifically novice engineers. The recommendations "populate tool with company-specific data" and "disseminate information on [company's] own internal failures" support the change in application of the tool from providing accident data across many industries to populating it with lessons learned data of a single organization. As mentioned in Section 1.2, expert engineers provided much more usefulness feedback than usability feedback because they have a better understanding of how a knowledge management tool could possibly fit into their established work practices.

Table 3.1: Cause/Recommendation Network Usefulness Feedback (Aloisio, 2018)

User	**Feedback**
Expert	"Use this tool as a teaching tool in a classroom environment. Not solely at academic institutions, but also for training at companies for new systems engineers."
Expert	"He would use the tool if it was populated with their own company-specific data [like a lessons-learned database]. They have volumes of their own data and go through it rigorously."
Expert	"The tool did not help with developing the design [e.g., early in the design cycle], but it helped identify and address potential issues. It's in a unique category of tools: identify potential pitfalls and potholes. There may not be many tools out there that are doing this. It may be valuable." "Provide the tool platform as a blank template that a company could use to disseminate information on their own internal failures." "Show the gap between what the accident investigators recommended and what the company actually implemented based on the recommendations, and the recommendations' impact on safety or performance."
Novice	"Give the user an opportunity to input information from their own project. Different insight could be useful (i.e. FMEA paperwork criticisms). He was concerned that the body investigating the accidents would not criticize their own processes."

Novice	“He should be able to add his own recommendations into the tool easily. Then he could click on all the things the recommendation could be applicable to and the network consumes it.”

Once we identified the application, customer, and scope of the new tool during the work redesign step, we considered what specific features will make the tool excel at disseminating lessons learned. These features were identified during the next step of the contextual design process, which is *user environment design*. User environment design is where “the parts and how they are related to each other from the user's point of view [are shown]” (Beyer & Holtzblatt, 1997). To make sure the tool adequately supports the work practice of as many users as possible, it should have multiple avenues to access the same information. In our case, a user interacting with the tool must have access to all the same technical information found in the NASA lessons learned database but have more ways to access and absorb it than provided by the LLIS. The usability feedback, shown in Table 3.2, helped us identify specific features that achieve that goal. The main features we decided to implement, described in Section 3.3, were chosen by how often they appeared in the feedback and by their ease of implementation without changing the basic structure of the network. Novice engineers focused more on critiquing specific features of the tool rather than commenting on *how* the tool could fit into their engineering projects. The most common desired functionalities include search capability, a tool tutorial or manual, and the ability to observe interactions between more than two nodes.

Table 3.2: Cause/Recommendation Network Usability Feedback (Aloisio, 2018)

User	**Feedback**
Expert	“Multiple cause selections would be useful. Some findings had multiple types of causes that would be useful to see together in the tool.”
Novice	“If you play around with the tool you can get through a lot of the information but a tutorial video or instruction manual may be useful. It also may scare some people away if they don’t want to read a 100-page instruction manual. Develop some way of introducing the user to the tool without scaring them.”
Novice	“The links between causes aren’t currently filtered by “who was involved in the failure”. He filtered the causes down by “personnel (operations)”, then clicked on “inadequately communicated” (which had stories under Buncefield). This cause had a link to “managed risk poorly”, which did not have a corresponding story under Buncefield because all of the “managed risk poorly” stories for this accident were for operations management.”

Novice	"It would be good if there was a search function. If the user could input 'any recommendation about engines' for example would help the user whittle the tool down to their own context."
Novice	"Have the list of recommendations reduce down to what is connected to the cause you clicked on, maybe in a different tab up at the top."
Novice	"Link the random story to its place in the list. Also, potentially provide a link to more information on the failure (the user could just google, but that's another step), potentially to the Wikipedia page of the failure. It was also unclear how severe the failure was, as in how many people died." "The significance of line weight is not very obvious." "There is a disconnect because the cause/recommendation in the story doesn't have a link to the network. It wasn't clear that the alphabetical list of causes/recommendations was interactive. The student thought they would have to search through the network to find the right node to get more information."
Novice	"Consider introducing a new feature: selecting multiple causes at once to see the recommendations they have in common. Also consider showing the specific corresponding map for each accident when you click on 'view stories'." "The lines connecting nodes make it difficult to read the node names, consider making them a bit lighter. It really matters to see the line weight when you click on a node to see its connections. Consider having two versions of the tool: one with no lines and the other with all the connections."
Novice	"There could be a search function where a user could input a keyword. 'Physics based failure'. Once the database becomes more populated a search query would be very helpful." "Consider using some form of picture representation, like a tree. If you make a certain set of decisions, what could be the outcome? If you pick two causes to improve upon, for example, what could happen?"
Novice	"Consider changing node size or font size to make it more obvious which nodes were connected more frequently."
Novice	"Allow the user to select more than one node to see for example what recommendations two causes have in common."
Novice	"Many of the systems people were not interested in reading the stories because they were inspired by the network itself and the connections. Maybe consider including some sort of help that shows them they can refer to the stories below to get more ideas/details."
Novice	"Include some sort of search bar." "He would search for 'design failure' or something more specific in the accident stories. 'wing failure' or 'propeller failure'. Adding more cases would be helpful for this because then there would be more specific things to search for. Consider adding key words for each accident. (Alaska: maintenance, lubrication, horizontal stabilizer; Aloha: corrosion)"

Novice	“Consider displaying some weight on the nodes to indicate how prevalent they are in the data.”
Novice	“A back button would be useful to see your previous selections.”

3.3 Final Product

We worked with a software engineer to build the interactive lessons learned network in HTML and JavaScript. The final tool, referred to as the “interactive lessons learned network”, is shown in Figure 3.2. The network has a consistent five-color scheme corresponding with the five categories of factors of our lessons learned model, described in Section 2.2. All factors of a lesson are represented as nodes in the network with the color corresponding to its category, and the legend below the network doubles as buttons that filter nodes by category. The general anatomy of the network is the same as the cause/recommendation network, where nodes that appear in the same lesson are connected, and connected nodes attract each other while disconnected nodes repel each other. Nodes with more connections appear towards the middle of the network, and nodes with less connections appear towards the outside of the network. The shade of the connecting line indicates the strength of the connection between the two nodes. A new feature introduced in this tool is that node size indicates the presence of that factor in the dataset—larger nodes indicate that the corresponding factor appears in many stories. Although 413 of the total 418 lessons are actually coded (see Section 2.1), all 418 lessons show up the tool. Because the lesson title itself conveys information, empty lessons remain in the tool to keep consistency with the database. The network consists of more data and more nodes than the previous cause/recommendation network. This may potentially introduce issues of legibility and speed when it is tested, depending on the demands of the user.

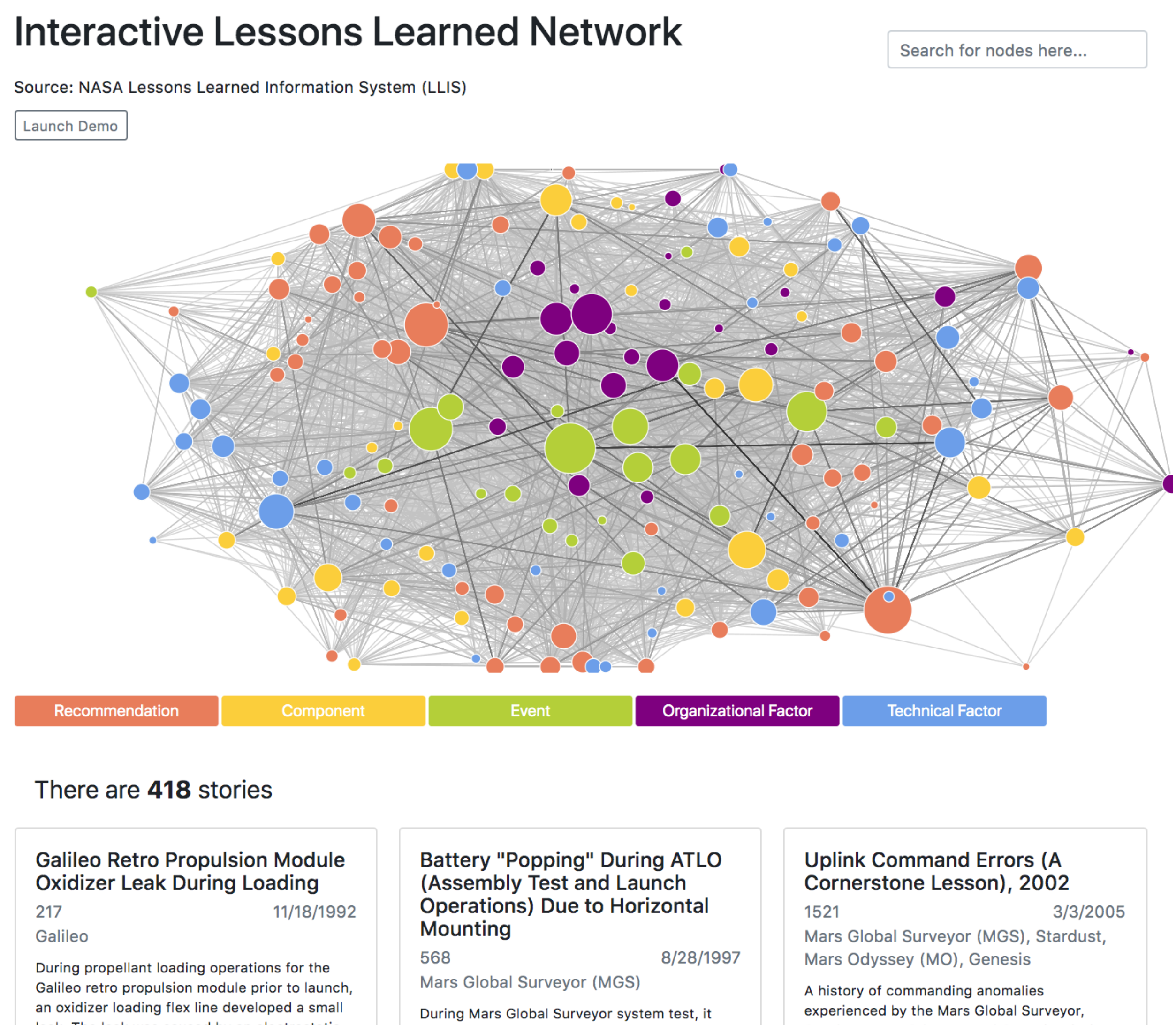


Figure 3.2: Interactive Lessons Learned Network[2]

The following are key features identified during the user environment design step that were added to the tool:

Click logic: Holding down 'Ctrl' while clicking multiple nodes allows the user to isolate and view the information the selected nodes have in common. Connections between the selected nodes are highlighted in the network, and the lessons that have the selected nodes in common appear below. Tiles for each selected node appear below the legend and can be individually deselected by clicking

[2] The interactive lessons learned network can be viewed here: https://engineering.purdue.edu/VRSS/research/lessons-learned-network/index_html.

the “x” on the tile. When no nodes are selected, all 418 stories appear. This feature is discoverable by executing the tool tip and by hovering over a selectable node.

Tool tip: Clicking the “Launch Demo” button in the top-left corner initiates an interactive tutorial of the network that explains the anatomy of the network and walks the user through the main features of the tool. The tutorial can be exited at any time.

Search bar: Typing a word or phrase into the search bar highlights the node that matches the search term. Suggested search completions also appear when the cursor is in the search bar.

Button to read original lesson: Clicking the “Read the whole story” button at the bottom of every lesson navigates the user to the original lessons learned page within the LLIS. This allows the user to access the exact same information as the LLIS, which is necessary if they want to read more detailed technical information about the lesson.

The final design of the interactive lessons learned network meets our original design goal of retaining the basis of the nodal network while providing more paths to access the same technical information found in the NASA lessons learned database. How will users respond to and interact with the tool, and does it have an advantage over the NASA LLIS? These questions are answered during the penultimate step of the contextual design process, *mock-up and test* with customers. Chapter **Error! Reference source not found.** describes the experiment we developed to determine how users interact with knowledge management tools and to collect feedback to potentially inform the next iteration of the contextual design process of a nodal network tool.

CHAPTER 4. DESIGNING AN EXPERIMENT TO DETERMINE HOW TOOLS ARE USED

In this chapter we describe the experiment that we developed to determine how novice engineers use both the interactive lessons learned network and the NASA LLIS. First, we describe the key questions the experiment aims to answer and what data is necessary to answer them. Next, we describe the three-part experiment that consists of a think-aloud protocol, case-based reasoning questions, and feedback collection. We include an explanation of how each part of the experiment addresses a key question and how each part was developed. Lastly, we describe the experiment procedures and recruitment efforts, and we discuss measures we took to reduce experimental bias.

4.1 Experiment Development

Once we developed an alternate knowledge management tool, we considered what we needed to answer our research question: **What role do knowledge management tools with lessons learned information play in the systems engineering decisions of novice engineers?** To answer this question, we sought to determine how users approach and interact with each knowledge management tool. Do the tools compel the user to engage?[3] To obtain further insight into the comparative strengths and weaknesses of each tool, we added a performance element to the experiment—similar to Aloisio & Marais's experiment with the cause/recommendation network, we wanted to determine if these tools help novice engineers develop meaningful solutions to scenarios. By developing a mixed-method experiment that bundles qualitative and quantitative analyses, our goal was to determine how the tools are used and whether they are helpful for novice engineers.

To explore these points, we developed an experiment with specific, attainable objectives that follows the same procedure for both tools. We split our initial research question into three key questions with objectives that when answered will provide us with meaningful qualitative and quantitative information about each tool. The three key questions are given below, followed by what information is required to answer the question.

[3] A note on terminology: Hereafter, "network" refers to the interactive lessons learned network, "database" refers to the NASA LLIS, and "tool" generally refers to either the network or the database, whichever one the participant is given.

1. **How do novice engineers use each tool, and do the tools have the proper functionality to support the demands of the user?** To answer this question, we need a way to prompt the user to interact with a tool and a way to record what a user is thinking throughout the process.
2. **Do these tools help users craft answers to engineering questions?** To answer this question, we need a way to objectively measure a user's engineering performance when using a certain tool.
3. **What are users' opinions on the tools?** To answer this question, we need a mechanism for gathering a user's honest feedback.

We designed a three-part experiment where each part addresses one of the three key questions by providing the corresponding need. For this experiment we randomly assigned a participant to either the database or the network and asked them the same series of prompts and questions.

4.1.1 Part 1: Think-Aloud Protocol

The first part of the experiment addresses the question: **How do novice engineers use each tool, and do the tools have the proper functionality to support the demands of the user?** To capture what the user is thinking, we first deployed the interviewing method of think-aloud protocol. Think-aloud protocol was developed to obtain insights into participants' thoughts, assuming that these insights will help the researcher better understand the topic. It is often used to understand cognitive processes when solving puzzles or performing repetitive tasks (Eccles & Arsal, 2017). In Section 5.3, we use the participants' verbalized thoughts to identify the different tasks they performed when answering prompts and to identify the features of the tool they used to perform these tasks.

Regarding procedure, the participant is first given access to one of the two knowledge management tools. The participant is then prompted to verbalize every thought they have when using the tool to answer open-ended engineering prompts. This protocol is based on the work of Ericsson & Simon (1993) which is widely considered standard think-aloud practice. Our experiment also implements their recommended best practices which include instructing the participants not to plan out what they say, offering participants a warm-up prompt, sitting to the side of the participant when observing, not letting them talk for more than a few minutes, and instructing them to speak (without leading) when they fall silent.

4.1.2 Part 2: Case-Based Reasoning

The second part of the experiment addresses the question: **Do these tools help users craft answers to engineering questions?** To determine whether the tools are helpful or not, we must have an objective way to establish each participant's relative engineering performance. The second part of our experiment uses case-based reasoning to objectively measure a participant's performance on objective engineering questions.

Case-based reasoning is the concept of relying on prior knowledge to solve a familiar problem (Kolodner, 1992). We created three open-ended engineering questions that deploy modified case-based reasoning, where we provide participants with a situation similar to one they may have already experienced. Even if participants have not experienced these particular scenarios, they will likely encounter a similar scenario at a future workplace setting—ideally a setting where one would have access to an institutional knowledge management tool. We asked the participant to provide a free-response answer to each of the three questions. Each question is a scenario with a correct answer based on information from the database; Figure 4.1 shows one of the questions.

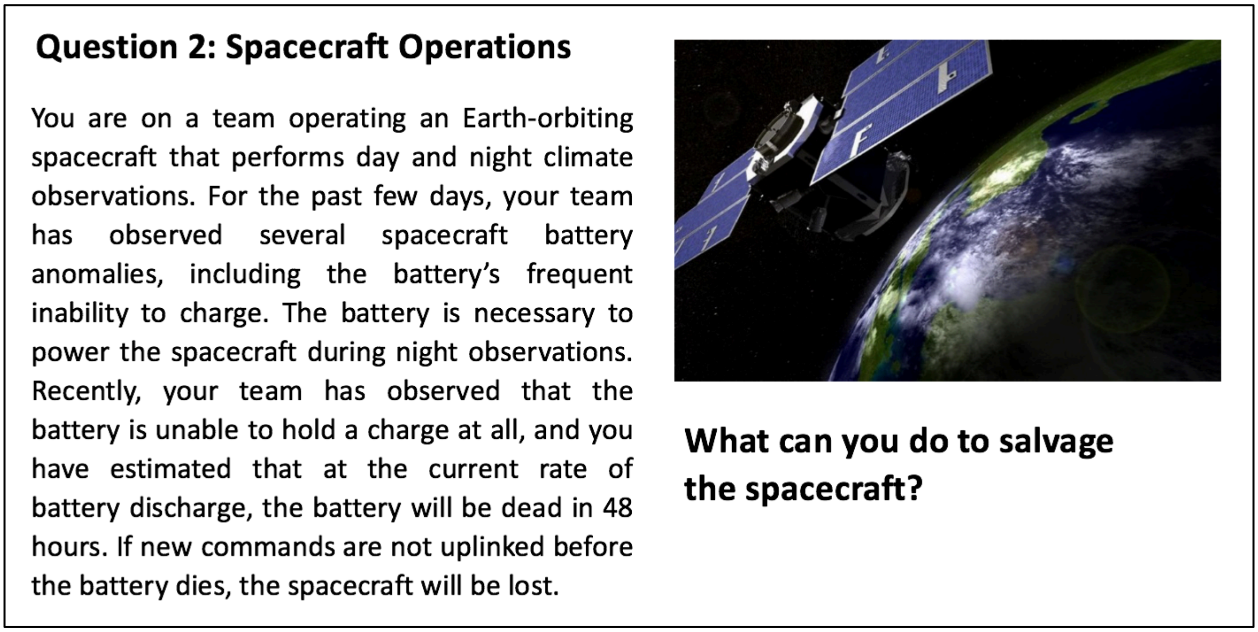


Figure 4.1: Question 2 as it Appears to the Participant

Each of the three questions is not necessarily representative of a certain type of case-based scenario. However, each question had its own sub-motivation that gave us more insight into if and how the participants trust knowledge management tools. One of the motivations behind the questions is the concept of transfer, which is defined by Detterman (1993) as the degree to which a behavior will be repeated in a new situation. We are concerned with two different types of transfer: near and far. *Near transfer* is the concept of applying knowledge to situations that are identical to the original with only a few key differences, whereas *far transfer* is applying knowledge to situations that are different that the original by varying degrees. The rationale and sub-motivation of each question, including the type of transfer that tool reliance would elicit, is discussed below.

1. **Battery Question**: This question asks the user to discuss several risks to which an experimental battery could be subjected. The user should be more familiar with this topic because batteries are studied in the curriculum and because battery issues are mentioned several times during the think-aloud protocol in Part 1. Using the tool to answer this question would lie on the spectrum of far transfer, as the question asks the participants to make new connections between factors, albeit for a topic with which they should be

familiar. The sub-motivation of this question is to answer: *will they turn to the tool for information on a familiar topic?*

2. **Spacecraft Operations Question**: This question asks the user to respond to a spacecraft in-flight anomaly (see Figure 4.1). The rationale for this question is that the exact situation is described in a specific lesson—the CloudSat battery anomaly. A participant who uses the tool to successfully answer this question would deploy near transfer because the situation described in the question is nearly identical to a lesson that can be accessed by either tool. The sub-motivation of this question is to answer: *are they able to navigate to a specific lesson?*
3. **Contamination Question**: The final question asks the user to discuss several ways they would address possible spacecraft contamination. This question addresses a specific flight hardware topic to which most students have never experienced first-hand. Using a tool to answer this question would also deploy far transfer because the question asks the participants to make new connections between factors for a more unfamiliar topic. The sub-motivation of this question is to answer: *will they turn to the tool for information on an unfamiliar topic*?

We also created an accompanying rubric for each question to objectively score the participants' written answers. The rubric scores each response out of six points: three points are given for the correct answers identified in the database, one point is given for mentioning another relevant factor, one point is given for correctly answering the question, and one point is given for clarity and legibility. To assess rubric objectivity, two independent graders scored each response. The results of the inter-rater agreement are discussed in Section 4.3. Table 4.1 shows an example of a graded rubric from Question 2.

Table 4.1: Rubric and Participant's Response to Question 2: Spacecraft Operations

Desired Quality	**Explanation**	**Example Response**	**Grader 1 Score**	**Grader 2 Score**
Provides correct answer (1)	Explicitly mentions 1 of the following: CloudSat mission, daylight-only operations (DO-Op), dropping/descoping night observations	"*Night-time observations should be halted immediately. The spacecraft will have to give up some or all night-time operations but will able to maintain enough charge at night to survive and resume daytime operations each orbit.*"	**1**	**1**
Provides correct answer (2)	Directly or indirectly mentions another power source onboard the spacecraft. e.g. "Divert power from the spacecraft's RTG" or "Point the spacecraft towards the sun"	*"Solar panel power generation should also be maximized."*	**1**	**0**
Provides correct answer (3)	Recommends 1 of the following: develop operational procedures/ constraints, consult subject matter experts, add hazard protection/ redundancy/margin	✗	**0**	**0**
Mentions another relevant factor in discussion	Mentions a relevant topic not mentioned above, provides a recommendation not mentioned above, or mentions a stakeholder.	*"Batteries of the same sort as the spacecraft uses should be gathered up for testing on earth. A potential solution may be found by modifying voltage or timing of battery charging or draining them to a lesser degree than before."*	**1**	**1**
Answers the question asked	Response is an obvious attempt at a solution.	✓	**1**	**1**
Clearly communicates response	Provides a response with minimal confusion or conjecture. This does not necessarily mean that the response is completely spelling- or grammar-error free, but it must be legible.	✓	**1**	**0**
		Total:	**5/6**	**4/6**

For each response, the end product for analysis from Part 2 is the response score from each grader, the time taken for the participant to answer, and the categorical level of tool usage based on observation. In Section 5.2 we discuss whether a particular tool helped participants score higher on the case-based reasoning questions. The quality of a participant's answer may depend on factors including how much each participant used the tool, how long it took them to answer the questions, and how much of their answer was based on prior knowledge versus information from the tool. The effect of these factors is addressed with statistical analyses described in Section 5.2.2.

4.1.3 Part 3: Demographics and Feedback

The third and final part of the experiment addresses the question: **What are users' opinions on the tools?** To answer this question, we developed a web survey for participants to take at the end of the experiment. The survey captures demographic information for each participant, such as the participant's academic year and their amount of self-reported professional experience in the space, military, aviation, and defense industries. Next, the survey asks participants to rate the tool they had access to from 1–10 in four categories—look and feel, navigation, level of detail, and usefulness—with 1 being the lowest score and 10 being the highest. Finally, they are asked for feedback on the tool, including whether they would refer to a similar resource for help in a workplace situation.

4.2 Experiment Deployment

We recruited for our experiment by sending an email to eligible participants, which included undergraduate and graduate students in aeronautical and astronautical engineering. To maintain confidentiality and avoid potential distractions, the experiment setting consisted of a closed-door office where participants had access to a computer. The participant was not allowed to navigate outside the given web-based tool or the survey materials. Once the participant read the consent form and expressed verbal consent, the experiment commenced. We conducted the think-aloud protocol first, the reason being that it helped familiarize the participant with the tool in a non-threatening setting. We also wanted to record their first impressions and initial thought process. The case-based reasoning questions were conducted second—once familiar with the tool, we were interested in whether the participant would turn to it to solve problems reminiscent to what they would experience in the workplace. The feedback survey was performed last because a participant’s opinions on the tool may change throughout the experiment. Once the experiment was finished, the audio recording of the think-aloud portion was transcribed and destroyed to maintain participant confidentiality. The case-based reasoning questions were later graded in batches by two independent graders.

4.3 Discussion on Possible Sources of Experimental Bias

We took several measures to eliminate a participant’s bias in their responses. A possible source of bias is that the participant was afraid of their answers being linked to their identity, which could have possibly dissuaded them from providing truthful answers. To address this, the interviewer read a consent form to every participant before the experiment, which assured the participant’s confidentiality and that a breach in confidentiality would not impact their safety or reputation. Another possible source of bias was that participants with access to the network were aware of the interviewer’s involvement in its development, which could have affected their feedback. When presenting the network to the participant, the interviewer avoided possessive language (e.g. “our tool”). To signal privacy during the feedback portion, the interviewer turned her back while the participant filled out the survey.

A possible source of bias in the results could come from the rubric. We designed the rubric to be as objective as possible for each of the three questions. All three rubrics have the same general response qualities, and the explanations of each response quality gives the correct answer to the question as given in the NASA lessons learned database. Another measure we took to eliminate bias was having two independent graders score each of the responses. For every question, we used the average of the two graders’ scores in the analysis. As an additional assessment of rubric objectivity, once all responses were scored, the graders discussed the differences in their scores and attempted to come to a consensus over each discrepancy. Out of 90 responses (30 participants answering three questions each), 39 responses had a discrepancy in scores between graders. Out of these 39 discrepancies, there were three instances of a two-point difference and only one instance of a three-point difference between the graders. The two graders were able to come to consensus for all 39 discrepancies, which provided adequate proof of rubric objectivity.

CHAPTER 5. ANALYZING THE RESULTS

In this chapter we describe how we analyzed the results of our experiment. Section 5.1 discusses the participant demographics and observations, including who participated in our experiment, how much they used the tools, and their feedback and preferences. Section 5.2 analyzes the quantitative results of the case-based reasoning questions including statistical analyses to determine which factors influenced a participant's score. In Section 5.3, we qualitatively analyzed the think-aloud transcripts to identify the tasks that participants execute using the tools. Lastly, we discuss our overall findings including the intersection of quantitative and qualitative results in Section 5.4.

5.1 Participant Demographics

Before we analyze the results, we first present who participated in the experiment. The demographics information in this section was collected during the feedback survey (Part 3 of the experiment), and the tool usage data in this section was observed during the case-based reasoning questions (Part 2 of the experiment).

5.1.1 Who Participated

Thirty people responded to our recruitment email and fully completed all three parts of the experiment. The majority of participants were undergraduates, and the majority of participants had some prior industry experience in the industries they are likely to enter upon graduation, shown in Figure 5.1. The collective sample does not have overwhelming experience, however—the majority of participants have less than one year of industry experience, and no PhD students participated in the experiment. Considering the participants' level of industry experience is valuable because similar to how experts provided the majority of usefulness feedback for the cause/recommendation network, participants with more industry experience may be able to better visualize how the tool fits into industry work practices. Additionally, they may already be familiar with other organizations' knowledge management tools and provide informed feedback on desired functionality. Only one participant reported that they had previously interacted with the database.

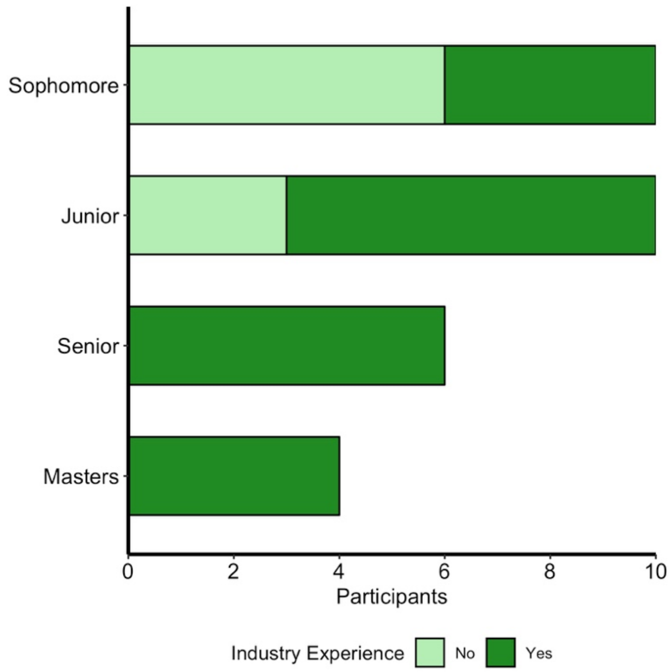


Figure 5.1: Participants' Academic and Industry Experience

5.1.2 Usage

Because a knowledge management tool is only as effective as how much it is used, we observed how much participants actually used the tool they were given. First, we defined a categorical variable called *usage* that describes how much the participant relied on the tool to answer the case-based reasoning questions. The value was assigned by the interviewer when observing the participant as they answered each question. The possible categorical values that the variable *usage* can assume are defined as follows:

High: Participant used the tool to inform the majority of their answer. The most common "high" usage scenario was the participant initially using the tool and consistently switching back and forth between using the tool and writing down their answer to the question.

Medium: Participant used the tool to inform less than half of their answer. The most common "medium" usage scenario was the participant initially answering the question without the tool, using the tool to look something up, and then finishing their answer.

Low: Participant used the tool to inform little to none of their answer. The most common "low" usage scenario was the participant not using the tool at all.

Participants were randomly assigned to one of the two tools in equal numbers, and Figure 5.2 shows the number of responses that fall under each usage category, broken up by question and by tool (90 responses in total: 30 responses for each question and 45 responses for each tool). Participants tended to use the network more than the database, but the difference is negligible. The difference of usage categories across all three questions was negligible as well, however, individual participants did not necessarily use the tool consistently across the three questions (e.g., one participant used the tool "high" for Question 1, but they used it "low" for Questions 2 and 3).

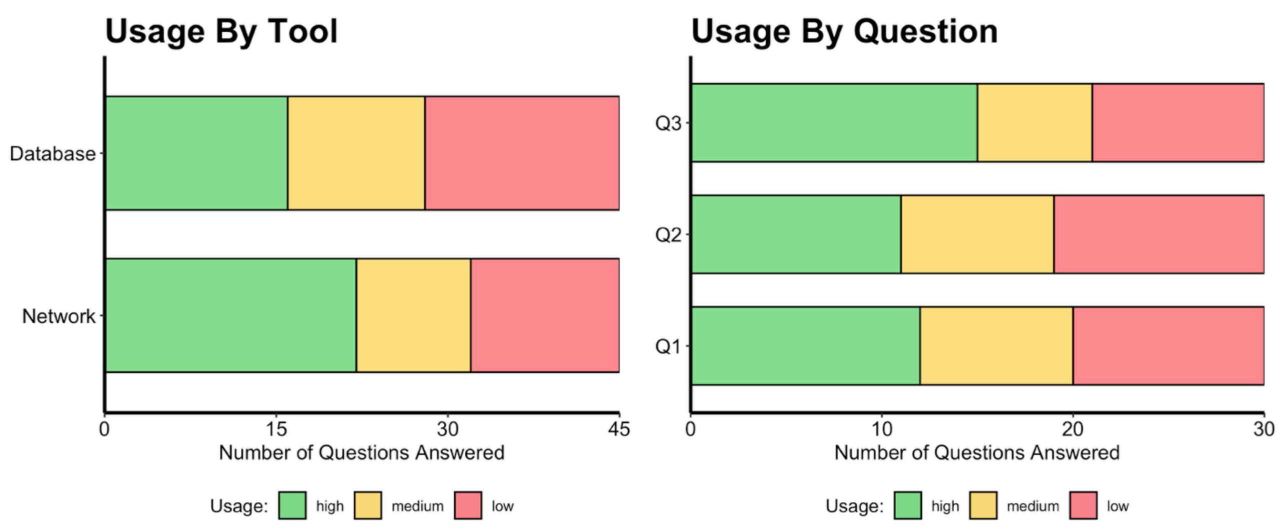


Figure 5.2: Categorical Observation of Usage by Tool and by Question

5.1.3 Preference

Now that we know that the tools were used at similar levels, what were participants' opinions on each of the tools? After the think-aloud protocol and case-based reasoning questions were answered, we asked participants to rate the tool they used in four categories, from 1 being worst to 10 being best. A box plot of the ratings of the four categories is shown in Figure 5.3. From these results, we observed that participants with the network favored its "look and feel" more than participants with the database, and they found the network more useful than did participants with the database. Participants with the database were more satisfied with the level of detail provided by the database. However, these data are skewed towards favorability for both tools and do not exhibit much variance.

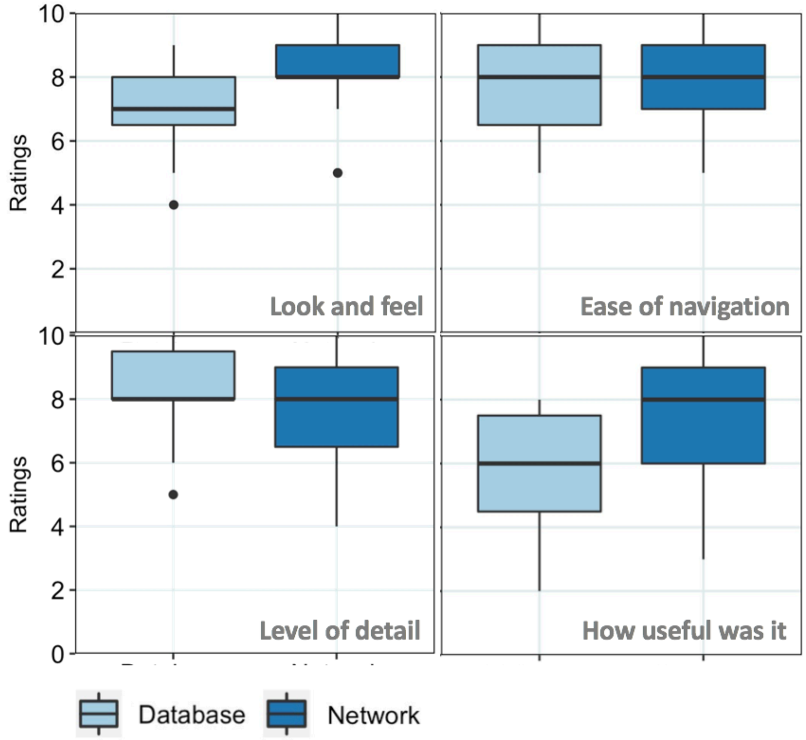


Figure 5.3: Participants' Ratings for Four Usability Categories

The final structured question on the feedback survey states, “Would you consult this resource for help in an unfamiliar situation at work?” The spread of answers to this question is shown in Figure 5.4. The participants’ answer to this question was overwhelmingly positive, with all but three participants stating that they would at least “probably” use a knowledge management tool similar to the database or the network at their workplace.

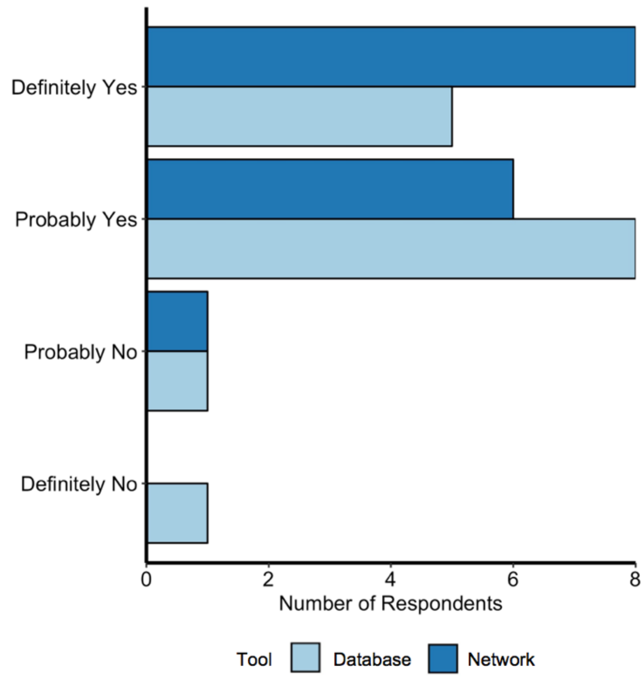


Figure 5.4: Responses to “Would you consult this resource at work?”

Feedback on a particular tool and usage of the tool during our experiment are not necessarily indicative of how much a participant would actually implement the tool into their work practices. For example, several participants rated their tool a score of 2 (with 1 being lowest and 10 being highest) for usefulness but also responded “Definitely Yes” when asked if they would use a similar tool in a work setting. Additionally, a few participants did not use the tool at all but also responded “Definitely Yes” to the same question. It is possible that some participants would use a knowledge management tool to support their actual work practices even if they did not use it or find it useful during our experiment.

Why might preferences be overwhelmingly favorable for both the network and the database? It is possible that participants did not give honest, critical feedback because the experiment has no implications on themselves or their reputation, which is stressed at the beginning of the experiment during the consent agreement. It is also possible that participants *did* feel pressure to rate the tools favorably due to the interviewer’s presence in the testing area, but this is not likely due to measures we took to eliminate a participant’s potential bias in their responses, described in Section 4.3.

5.1.4 Summary of Participant Demographics

To summarize this section, the thirty participants of the experiment were primarily undergraduate students with some previous industry experience in aerospace or related industries. How much a participant used their tool was independent of which tool they were given. Participants with the network found it more useful during the experiment than did participants with the database. Finally, both experimental groups responded positively to their given tool, and all but three participants reported that they would at least "probably" use a similar tool in a workplace setting.

5.2 Quantitative Analysis of Case-Based Reasoning Answers

Now that we know who participated in the experiment, how much they engaged with the tools, and their opinions on the tools, how did they perform for each case-based reasoning question? During this section, we quantitatively analyze the performance of participants with the network versus participants with the database to determine which tool has an advantage over the other and to determine whether or not knowledge management tools are helpful when responding to engineering scenarios. The data in this section was collected during Part 2 of the experiment.

5.2.1 Overview of Results

First, we looked at the raw scores and times of the case-based reasoning questions, broken down by tool group. Table 5.1 gives the average scores and times for each group. According to the table, participants with the network scored higher overall and on average took more time to answer the questions—the average score was 1.02 points higher for participants who had access to the network instead of the database, but their average time was three minutes and 27 seconds longer. Figure 5.5 also depicts each participant's responses on a graph of score versus time. From this figure, time seems to correlate positively with score. We further analyze this relationship in Section 5.2.2.1.

Table 5.1: Average Score and Time for Each Tool

Tool	**Average Score** (out of 6)	**Average Time** (min)
Network	4.30	8.92
LLIS	3.28	5.47

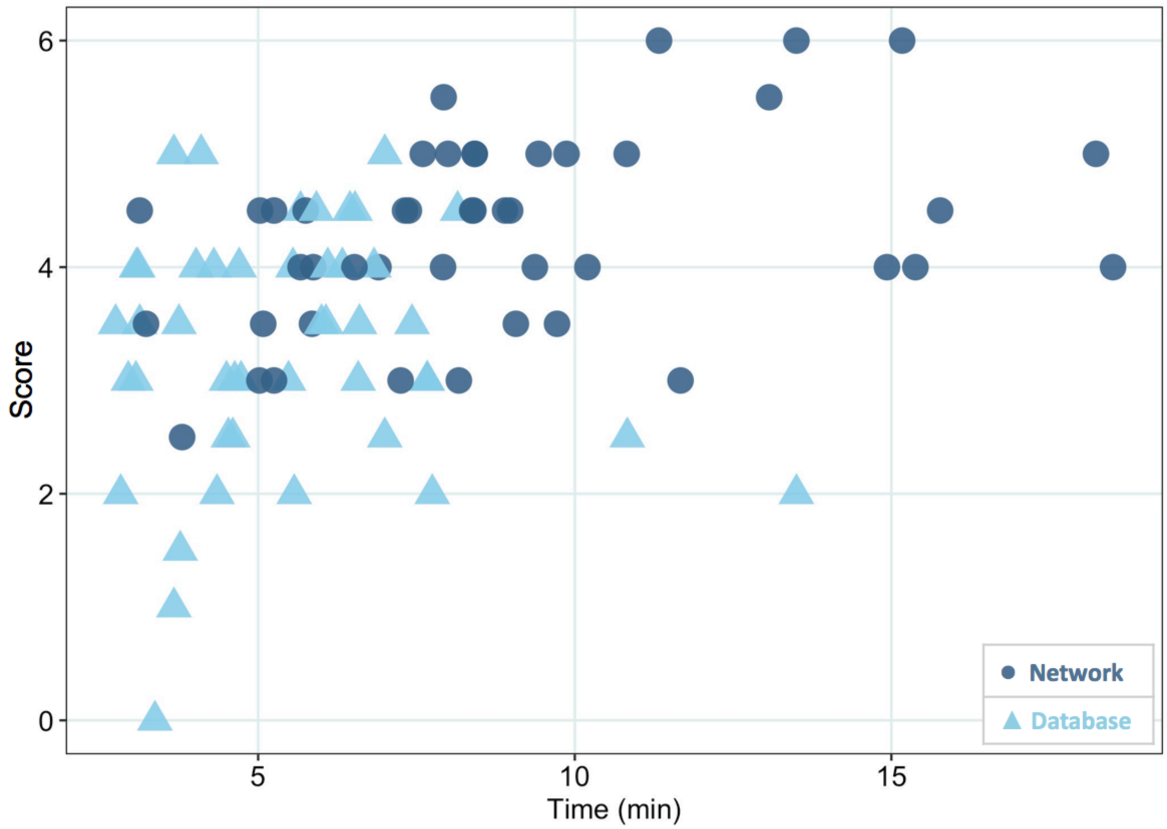


Figure 5.5: Score vs. Time for All Responses, Filtered by Tool

Next, we break participants' case-based reasoning scores down by each question. Table 5.2 shows both the average scores and times of participants, broken down by individual question and by which tool they used. For each question, participants with the network produced a higher average score than participants with the database. Additionally, Question 2 took less time than Questions 1 and 3 but had the highest average score. However, it is difficult to make hard conclusions from the average scores across questions because they are very similar. Each individual data point is also shown graphically in Figure 5.6.

Table 5.2: Average Scores and Times Broken Down By Question

		Q1	Q2	Q3	Tool Total
Network	Average Score (out of 6)	4.17	4.20	4.53	**4.30**
	Average Time	10:40	6:50	9:16	**8:55**
LLIS	Average Score (out of 6)	2.83	3.80	3.20	**3.28**
	Average Time	5:41	4:56	5:47	**5:28**
Question Total	Average Score (out of 6)	**3.50**	**4.00**	**3.87**	
	Average Time	**8:10**	**5:53**	**7:32**	

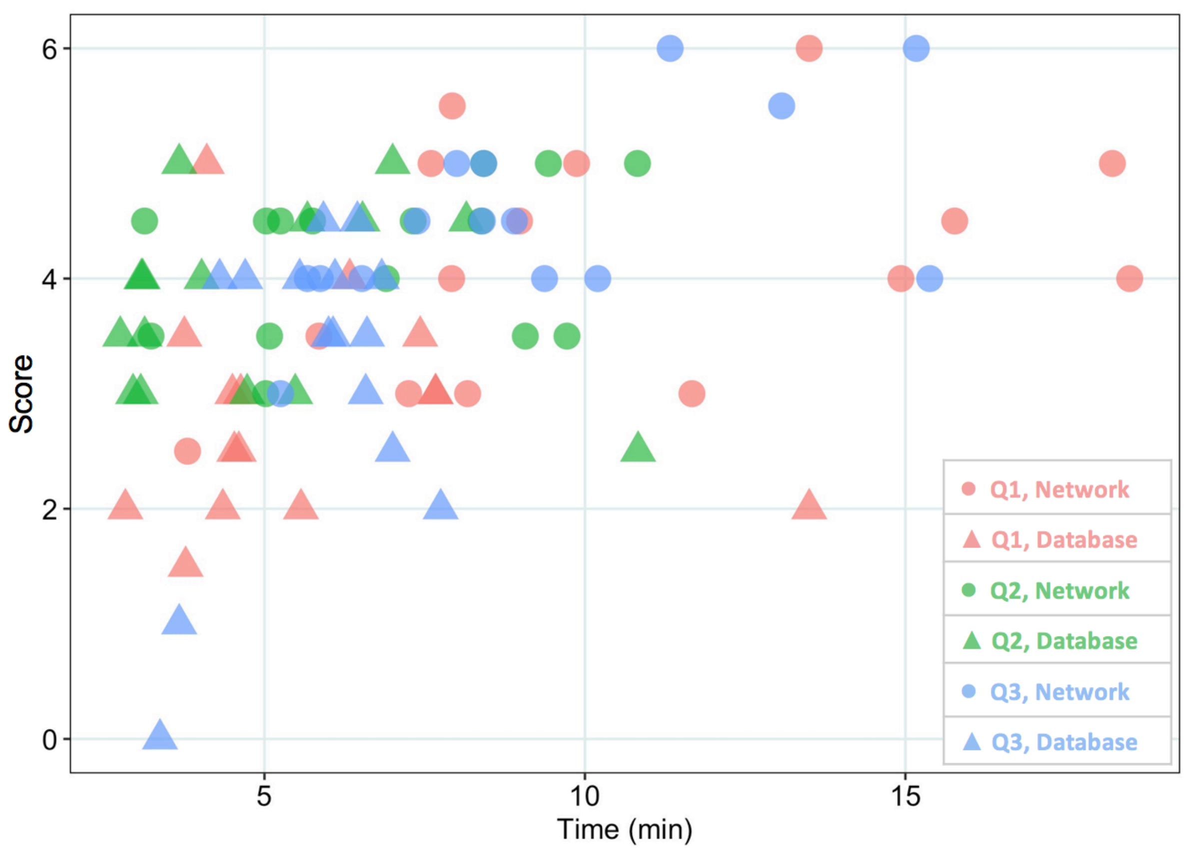


Figure 5.6: Score vs. Time for All Responses, Filtered by Tool and Question

The scores and times of the case-based reasoning questions are also broken up by usage category, shown in

Table 5.3. Similar to the breakdown by question, participants with the network produced a higher average score than participants with the database for each usage category. For both tools, high usage took a longer amount of time than medium and low usage. In other words, users who did not use either tool tended to arrive at their answers more quickly. High usage of the network resulted in a higher average score than using the network at a medium or low level, but for the database, the average score for each usage category was about the same. *High network* usage also resulted in the highest average score overall. These results are also shown graphically in Figure 5.7.

Table 5.3: Average Scores and Times Broken Down by Usage

		High	Medium	Low	Tool Total
Network	Average Score (out of 6)	4.48	4.15	4.11	**4.30**
	Average Time	10:29	8:37	6:31	**8:55**
LLIS	Average Score (out of 6)	3.32	3.38	3.18	**3.28**
	Average Time	7:07	5:29	3:56	**5:28**
Usage Total	Average Score (out of 6)	**3.99**	**3.73**	**3.58**	
	Average Time	**9:04**	**6:55**	**5:03**	

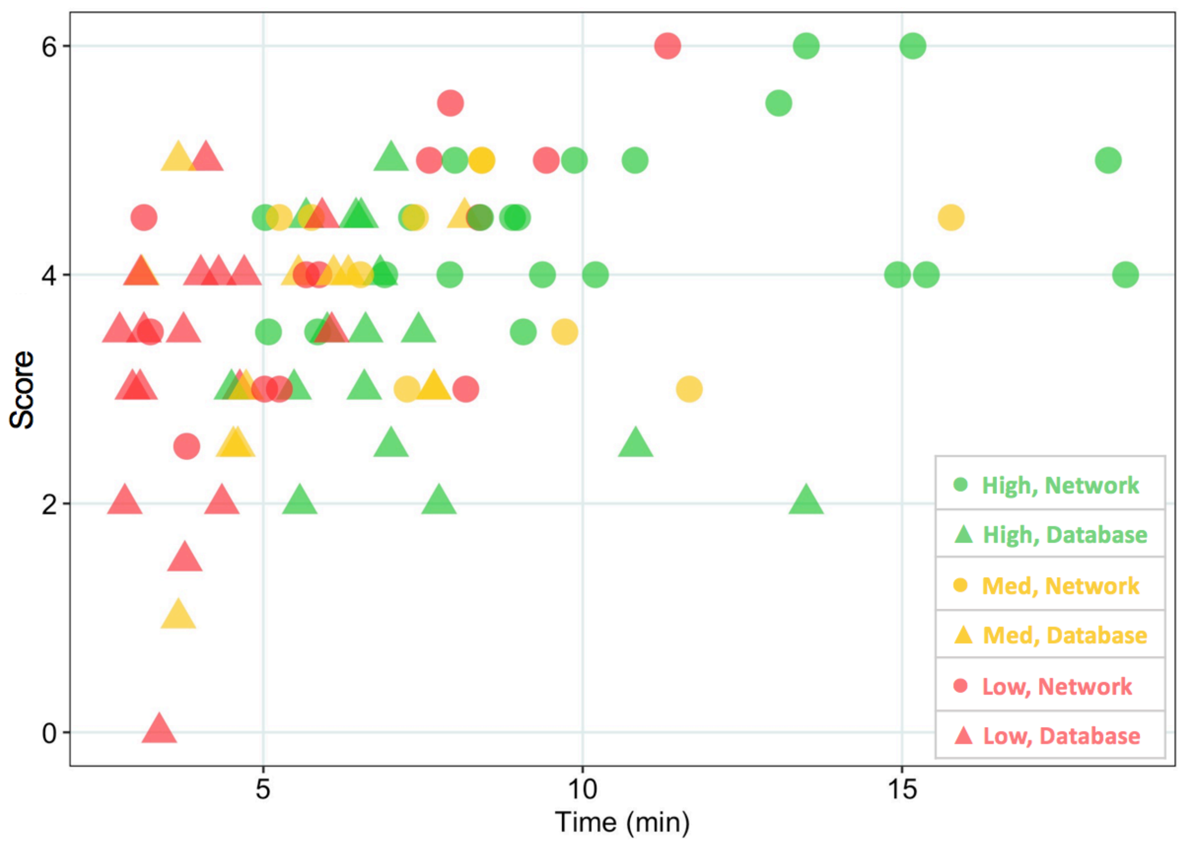


Figure 5.7: Score vs. Time for All Responses, Filtered by Tool and Usage

5.2.2 Statistical Analysis

Using the network resulted in a higher average score on the case-based reasoning questions, but is one tool stronger than the other for all scenarios? Is a participant's score dependent on any other factors, and how do they all relate? To answer these questions, we performed statistical analyses using correlation tests, linear regression, and a two-sample t-test.

5.2.2.1 Correlation Tests

Figure 5.5 indicates that the score on the case-based reasoning questions and time taken to answer each question may be positively correlated. To provide evidence of this relationship, we performed a correlation test between time and score. The correlation test requires the assumption that each observation is independent. Because we cannot assume independence between a participant's three answers, we could not analyze all 90 datapoints together (30 participants with three answers each); therefore, we split the correlation test up by question. We used the "cor" function in R to perform a correlation test on the data from each of the three questions. The results are shown in Table 5.4.

Table 5.4: Correlation Test Results for Questions 1-3

Question	**Correlation Coefficient (R)**	**Conclusion**
(1) Battery Question	0.478	Weak positive correlation
(2) Operations Question	0.238	No correlation
(3) Contamination Question	0.598	Moderate positive correlation

A correlation coefficient of $R = 1$ describes a perfect positive correlation, $R = -1$ describes a perfect negative correlation, and $R = 0$ describes no correlation between the two variables. In our case, time and score have a weak positive correlation for Question 1 ($0.3 < R < 0.5$), a moderate positive correlation for Question 3 ($0.5 < R < 0.7$), and no discernable correlation for Question 2 ($R < 0.3$). Although we cannot prove a definitive positive correlation, we have enough evidence to conclude that the values of time and score *cannot* be considered to be independent. The variables time and score on the case-based reasoning questions both provide us with valuable insight into how participants interacted with knowledge management tools, but as we move forward with our statistical analysis, which one is more important to us? Because we care more about whether knowledge management tools are helpful versus how efficiently participants use them, score is the metric of success for our statistical analyses.

Now that we have disproved independence between time and score, does how much a participant used their tool influence how long it took them to answer the questions? To determine whether there is a correlation between time and usage, we performed a one-way analysis of variance (ANOVA) to first determine whether or not there is a significant difference in a mean value between different categories. In this case, we compared the mean times for each usage category ("low", "medium", and "high"). Similar to the correlation test, ANOVA requires the assumption of independence between observations; therefore, we analyzed the results from each question individually. We used the "aov" function to perform the ANOVA analysis in R, and the results are shown in Table 5.5. The p-value describes the statistical confidence that there is a significant difference between the means of any two groups. For example, if the p-value is 0.05, then based on our data there is evidence to suggest that we are 95% confident that a statistically significant difference in mean time exists between "high-low", "high-medium", or "medium-low" groups. In general, we conclude that p-values less than 0.05 are significant. For all three questions, the ANOVA test proved that there was a significant difference between the mean times for at least one group combination but did not give information as to which groups were different or the values of the differences.

Table 5.5: One-Way ANOVA Results for Questions 1-3

Question	**P-Value**
(1) Battery Question	0.005 **[4]
(2) Operations Question	0.022 *
(3) Contamination Question	0.014 *

To quantify the mean time difference between each usage group, we performed a Tukey honest significant differences (HSD) test, which makes pairwise comparisons between the means of different groups with 99% confidence. The output of the "aov" function was passed to the "TukeyHSD" function in R to calculate statistically significant differences between the means for each group combination, along with the p-value adjusted for multiple comparisons between groups. The results of the analysis are shown in Table 5.6.

Table 5.6: Tukey HSD Test Results for Questions 1-3

Question	**High-Medium**		**High-Low**		**Medium-Low**	
	Mean Time Difference	Adjusted P-Value	Mean Time Difference	Adjusted P-Value	Mean Time Difference	Adjusted P-Value
(1) Battery Question	2:30	0.329	5:38	0.005 **	3:05	0.224
(2) Operations Question	1:09	0.530	2:52	0.017 *	–1:43	0.253
(3) Contamination Question	2:46	0.095 .	3:13	0.020 *	–0:26	0.946

[4] "**" indicates very significant p-value ($0.001 < p < 0.01$), "*" indicates significant p-value ($0.01 < p < 0.05$), and "." indicates marginally significant p-value ($0.05 < p < 0.1$)

An example of interpreting the results is as follows: participants that used the tool “high” will take on average 5 minutes and 38 seconds longer to answer Question 1 than participants that used the tool “low”. Because the mean time difference between “high” and “low” usage categories has statistically significant differences for all three questions, the results of the Tukey HSD test prove that we *cannot* assume independence between time and usage. Similar to the relationship between time and score, because we cannot assume independence between time and usage, we must choose one of the two factors to analyze further with our regression analysis. In our case, we care more whether usage affects score rather than time. Additionally, we already know that there is at least a weak linear relationship between time and score. Again, the question of *how* participants use tools versus how *efficiently* participants use tools guides us to care more about usage and score rather than time.

5.2.2.2 Linear Regression

Participants who used the network tool received, on average, higher scores in the case-based reasoning questions. The average increase in scores is not necessarily a direct result or effect of using the network—the amount that the tool was used may also affect the score. To verify that our tool positively correlates with an increase in score for the case-based reasoning questions, we performed a linear regression analysis. Linear regression is a statistical model that describes a linear relationship between a dependent variable, $\hat{Y}$, and independent variables, $\boldsymbol{x}$. In our case, the dependent variable is the participant’s score on a case-based reasoning question and the independent variables are which tool they used and how much they used it. We describe this linear model in Equation 1:

$$\widehat{Y_\iota} = \alpha + \boldsymbol{\beta x} + \varepsilon \tag{1}$$

Where $\widehat{Y_\iota}$ is the score for question i, α is an intercept term, $\boldsymbol{\beta}$ is a vector with the coefficients for each independent variable x, and $\varepsilon \sim N(0, {\sigma_\varepsilon}^2)$ is the observation-specific random error. To complete the linear regression, we use the pairs of observed variables (x,Y) to find the coefficients, $\boldsymbol{\beta}$, of the independent variables.

Linear regression requires several assumptions:

- The explanatory variables $\boldsymbol{x}$ are independent of each other. In our case, that means that we assume that how much a participant used a particular tool (“low”, “medium”, “high”) is independent of which tool they were using. We showed in Section 5.1.2 that this assumption is valid.
- Participants’ responses are independent with each other. In our case, the subjects did not interact with each other during the experiment, nor, as best we know, did they discuss their answers with each other before or after the experiment.
- The residual term ε is normally distributed.

Table 5.7 describes the variables in our regression model. We did not consider any interaction variables in our analysis due to the small sample size—introducing more variables will decrease the ability of the model to calculate the coefficient values.

Table 5.7: Regression Analysis Variables

Variable	**Database**	**Values**		
$\hat{Y}$	Predicted score	Continuous: [0, 6]		
x_{tool}	Whether the participant used the database over the tool	Network	Database	
		0	1	
x_{low}	If the participant’s observed tool usage was low	Low	Medium	High
		1	0	0
x_{med}	If the participant’s observed tool usage was medium	0	1	0

Due to the necessary regression assumption of independence between observations, we cannot analyze all 90 data points together. Because of this, we performed a regression analysis on each participant’s average score across the three questions, followed by analyses of the scores for each individual question (similar to the correlation tests in the previous section). First, to analyze the factors that correlate with a participant’s average score, we must determine their average usage across all three answers as well. To do this, we make a fourth assumption that the usage categories are separated by regular intervals, i.e., the difference between “low” and “medium” usage is the same as the difference between “medium” and “high” usage. With this assumption, we assigned a corresponding numeric value to each category—a value of 0 to “low”, 1 to “medium”, and 2 to "high"—and calculated the average for each participant. For example, a participant that used the tool "medium" for Question 1, "low" for Question 2, and "high" for Question 3 was assigned an average usage category of "medium". A value ending in 0.5 was treated as a fuzzy intersection and rounded down.

$$[medium, low, high] \Rightarrow avg(1,0,2) = 1 \Rightarrow medium \tag{2}$$

$$[low, high, high] \Rightarrow avg(0,2,2) = 1.5 \Rightarrow medium \tag{3}$$

Once we calculated values for average score and average usage for each participant, we used the "lm" function in R to perform a linear regression analysis on the average score data. The resulting regression equation to predict a participant's average score is:

$$\hat{Y}_{avg} = 4.10 - 1.01x_{tool} - 0.19x_{low} - 0.19x_{med} \tag{4}$$

To interpret the model, we look at the values of each coefficient. We interpret this equation as follows: Using the database over the network ($x_{tool} = 1$) will decrease the average score by 1.01 points, while keeping the other independent variables fixed. Using the tool at a low level ($x_{low} = 1$, $x_{med} = 0$) will decrease the average score by 0.19 points, while keeping the tool fixed. Similarly, using the tool at medium ($x_{low} = 0$, $x_{med} = 1$) will also decrease the average score by 0.19 points, while keeping the tool fixed.

We now know how tool and usage correlate with the average score based on the coefficient values, but we must consider the statistical significance to make any conclusions. We show the linear regression output along with associated p-values for each coefficient in Table 5.8. The p-value describes the statistical confidence that a particular coefficient is zero. For example, if the p-value is 0.05, then based on our data there is evidence to suggest that we are 95% confident that the associated coefficient is not zero. For our analysis, we set the p-value significance level at 0.05, meaning we focus on the coefficients that have p-values lower than the set level.

Table 5.8: Linear Regression Results for Average Score

Variable	**Estimate, $\boldsymbol{\beta}$** (Std. Error)	**P-Value**
Intercept, α	4.10 (0.24)	3.19e–16 ***[5]
x_{tool}	–1.01 (0.28)	0.001 **
x_{low}	–0.19 (0.34)	0.58
x_{med}	–0.19 (0.33)	0.56

From Table 5.8, x_{tool} is the only variable with significance. We conclude that a participant's average score across all three case-based reasoning questions will increase by 1.01 points when the network is used over the database (if usage is kept fixed).

The same analysis was performed for Questions 1-3, and the results are given in Table 5.9.

Table 5.9: Linear Regression Results for Questions 1-3

Question	**Variable**	**Estimate, $\boldsymbol{\beta}$** (Std. Error)	**P-Value**
(1) Battery Question	Intercept, α	4.11 (0.27)	2.03e–14 ***
	x_{tool}	–1.26 (0.38)	0.003 **
	x_{low}	–0.17 (0.31)	0.59
	x_{med}	0.16 (0.34)	0.64
(2) Operations Question	Intercept (α)	4.22 (0.25)	1.73e–15 ***
	x_{tool}	–0.38 (0.27)	0.162
	x_{low}	–0.24 (0.31)	0.45
	x_{med}	0.22 (0.34)	0.52
(3)	Intercept (α)	4.65 (0.35)	4.25e–13 ***

[5] "***" indicates extremely significant p-value ($p < 0.001$)

Contamination Question	x_{tool}	–1.32 (0.41)	0.004 **
	x_{low}	–0.25 (0.48)	0.61
	x_{med}	–0.24 (0.55)	0.66

From Table 5.9, x_{tool} is statistically significant in Questions 1 and 3. We conclude that a participant's answer to Question 1 will increase by 1.26 points when the network is used over the database (if we keep usage fixed), and that a participant's answer to Question 3 will increase by 1.32 points when the network is used over the database (if we keep usage fixed). For all three models, the usage variables (x_{low} and x_{med}) show no statistical significance. There are two possible explanations for this result: either our data set is small and does not include enough evidence that usage matters, or usage is truly not correlated with the average score.

5.2.2.3 Two-Sample T-Test

We cannot say that one tool is always more effective than the other from the linear regression results, but does *exposure* to one tool produce a higher average score than the other, especially for Question 2? We use the word *exposure* because this analysis no longer factors in how much the participant used the tool. We answered this question by performing a two-sample t-test. Our null hypothesis, H_0, is that the mean of the case-based reasoning scores for the network were less than or equal to the mean scores for the database. Our alternate hypothesis, H_1, is that the mean scores were higher for the network. Rejecting the null hypothesis will statistically prove that the network produces higher average scores, not considering any other factors of the experiment. The t-test requires the assumption that the data is normally distributed.

$$H_0: \mu_{network} \leq \mu_{database} \tag{5}$$

$$H_1: \mu_{network} > \mu_{database} \tag{6}$$

We used the "t.test" function in R to perform a Welch two-sample t-test on the scores for each of the three case-based reasoning questions. The results of the tests, including the t-statistic, p-value, and conclusion of the test are shown in Table 5.10. We were able to reject H_0 if the p-value was less than 0.05.

Table 5.10: T-test Results

Question	T-Statistic	P-Value	Conclusion
(1) Battery Question	3.74	4.24e-4	Reject H_0
(2) Operations Question	1.51	0.072	Fail to reject H_0
(3) Contamination Question	3.33	0.001	Reject H_0

Because we were able to reject H_0 for Questions 1 and 3, we concluded that exposure to the network results in a higher average score for these two questions, which corroborated but did not add to the linear regression results.

5.2.3 Summary of Quantitative Results

In summary, participants with access to the network produced a higher average score and required more average time for *all* combinations of Questions 1, 2, 3 and usage categories “high”, “medium”, and “low”. The average case-based reasoning score of participants with the network was higher than the average score of participants with the database by more than one point (out of a possible six). For both breakdowns of scores by question and by usage, participants with access to the network *consistently* scored higher and took more time than participants with access to the database. In general, high use of the network may increase the score of the participant’s answer, but neither tool helps participants come up with a quality answer faster than without the use of a knowledge management tool.

Through correlation tests, we concluded that there is a weak positive correlation between time taken to answer a question and the score received; therefore, we cannot assume that time and score are independent. Although both time and score are useful indicators of a participant’s performance, score is our preferred metric of success because we are focusing on a particular role of knowledge management systems—we care more about how participants use these tools versus how efficiently they are used. Allowing a user to efficiently reach an answer is also a desired trait of a knowledge management tool but is beyond the scope of this project. Using a one-way ANOVA analysis paired with a Tukey HSD test, we were able to determine that a positive correlation exists between time to answer case-based reasoning questions and tool usage; therefore, we also cannot assume independence between the variables time and usage.

Using linear regression, we were able to conclude that using the network over the database increased a participant's predicted average score and their score on Questions 1 and 3. In particular, using the network (versus using the database) increases a participant's average score by 1.01 out of six possible points, by 1.26 out of six possible points for Question 1, and by 1.32 out of six possible points for Question 3. The results of the two-sample t-test corroborated these results, concluding that exposure to the network tool, not considering other factors, increased the mean scores for Questions 1 and 3. However, linear regression nor the two-sample t-test produced significant results regarding Question 2.

We were unable to find significance in the results for Question 2, due to either a small sample size or because there are genuinely no relationships between variables for Question 2. Therefore, we cannot statistically conclude that the network is more effective in all scenarios.

These results make sense because Questions 1 and 3 were more similar in nature than Question 2—Question 2 deployed near transfer by asking participants to consider a single event in the database, as opposed to Questions 1 and 3 which deployed far transfer by requiring participants to make connections to other topics.

5.3 Qualitative Analysis of Think-Aloud Protocol

From the quantitative analysis, we know that the average scores for Questions 1 and 3 differ based on which tool the participant was given. Does the presence or absence of significant features between the two tools help shed light on this result, and do any specific features help or hinder a participant's performance? Does the network score higher and take longer than the database for a reason? And most importantly, how do participants interact with each of the tools? To answer this question visually, we analyzed the think-aloud transcripts from Part 1 of the experiment to create a mental model of how participants responded to engineering scenarios, including interaction (and lack of interaction) with the given tool.

5.3.1 Mental Model

We created a mental model from the think-aloud transcripts of how novice engineers respond to the open-ended engineering prompts. A mental model is a visual representation of a user's motivations and thought processes broken down by specific tasks they execute to solve a problem. Mental models allow designers to identify strengths and gaps in their tool functionality based on the number of tasks that can or cannot be fulfilled by the tool (Young, 2008). In our case, we used a mental model to determine the functional strengths of each tool and whether they had an impact on the results of our experiment.

The first step in creating the mental model was choosing which transcripts to analyze. We strategically chose our sample by assembling a set of transcripts that was useful, diverse, and consistent. The majority of transcripts we chose were useful, meaning that the participant followed the protocol of the experiment by describing their thought process to the interviewer. We focused on transcripts where the participant describes *how* they answer the question, as opposed to just providing different answers *to* the question (we did include one of these transcripts for sample diversity). Next, we chose transcripts that increased the diversity of the sample with regards to academic and industry experience. We also chose transcripts that varied in Part 2 performance statistics, including usage, average times, and scores. Lastly, we maintained consistency between tools. We used six transcripts in total, three from each tool, and we included the participant with the highest average score across the three case-based reasoning questions for each tool. A breakdown of participants included in the sample is given in Table 5.11.

Table 5.11: Sample of Transcripts Used to Create Mental Model

Participant	Tool	Year	Experience	Avg. Score	Avg. Time	Avg. Usage
A	Database	Senior	1-2 years	4.67	4:26	medium
B	Network	Senior	1-2 years	5.50	12:28	high
C	Database	Senior	1-2 years	2.33	4:07	medium
D	Database	Junior	1-2 years	3.33	3:43	low
E	Network	Sophomore	None	5.00	9:14	medium
F	Network	Junior	2+ years	3.00	9:32	low

The second step in creating the mental model was combing through each transcript and identifying tasks that the participants executed to answer the seven prompts in Part 1. Tasks relay how the participant attempts to answer the question instead of what their answer actually is, therefore, we sought useful transcripts because they are more task-dense. According to Young (2008), tasks can be explicit actions, implied actions, third-party actions, philosophies, and feelings. Similar to the process of identifying factor codes within the NASA lessons learned database (described in Section 2.2), it was important to initially retain all of the detail of a task before combining similar tasks. Once all transcripts were parsed for tasks, we combined tasks in standard mental model fashion. First, we grouped tasks that can be accomplished in similar ways into *towers*. Then we grouped towers into *mental spaces*, groups of towers that describe a distinct thought process. We then ordered the mental spaces based on the sequential thought processes that most participants followed, as observed by the interviewer.

The third step to creating the mental model was identifying the functionality of each tool. We accomplished this in part during the *user environment design* step of the contextual design process in Section 3.2. During this step, we created a list of knowledge management tool functionalities and indicated whether each tool had this functionality or not. The list also includes functionalities that have not been addressed by either tool, for example, providing a link to external sources outside of the NASA lessons learned database that discuss the event. We then matched these functionalities to towers of tasks that they can accomplish.

The general anatomy of a mental model is as follows: mental spaces consisting of towers (and towers consisting of tasks) are arranged atop the centerline. Functionalities are placed beneath the towers that consist of tasks they can accomplish. It is important to note that not all tasks need be accomplished by a functionality, and functionalities can be matched to more than one tower (Young, 2008).

The final step to creating our mental model was determining what other dimensions of information were useful and organizing how the information was displayed. Every mental model is different and can represent dimensions of information based on the author's needs. The specific rules for our mental model are as follows: the most frequent tasks appear on the top of the tower, and tasks decrease in frequency moving down the tower (by frequency, we mean how many participants performed a certain task, not necessarily the total number of times it was executed). The color scale is used to indicate frequency as well. We prioritized the most frequent tasks because the more frequent they are, the more important they are to address. If the tool does not have the functionality to address a very frequent task, almost every user will encounter the same problem. It is possible that some tasks arise from using a particular tool, but the *other* tool possesses the functionality to address it—to visualize this, we placed colored "jewels" in the corner of tasks to indicate whether that task was only associated with a particular tool. Finally, the functionality tiles are color coded by which tools have said functionality—both tools, one, the other, or neither. Our mental model is shown in Figure 5.8.

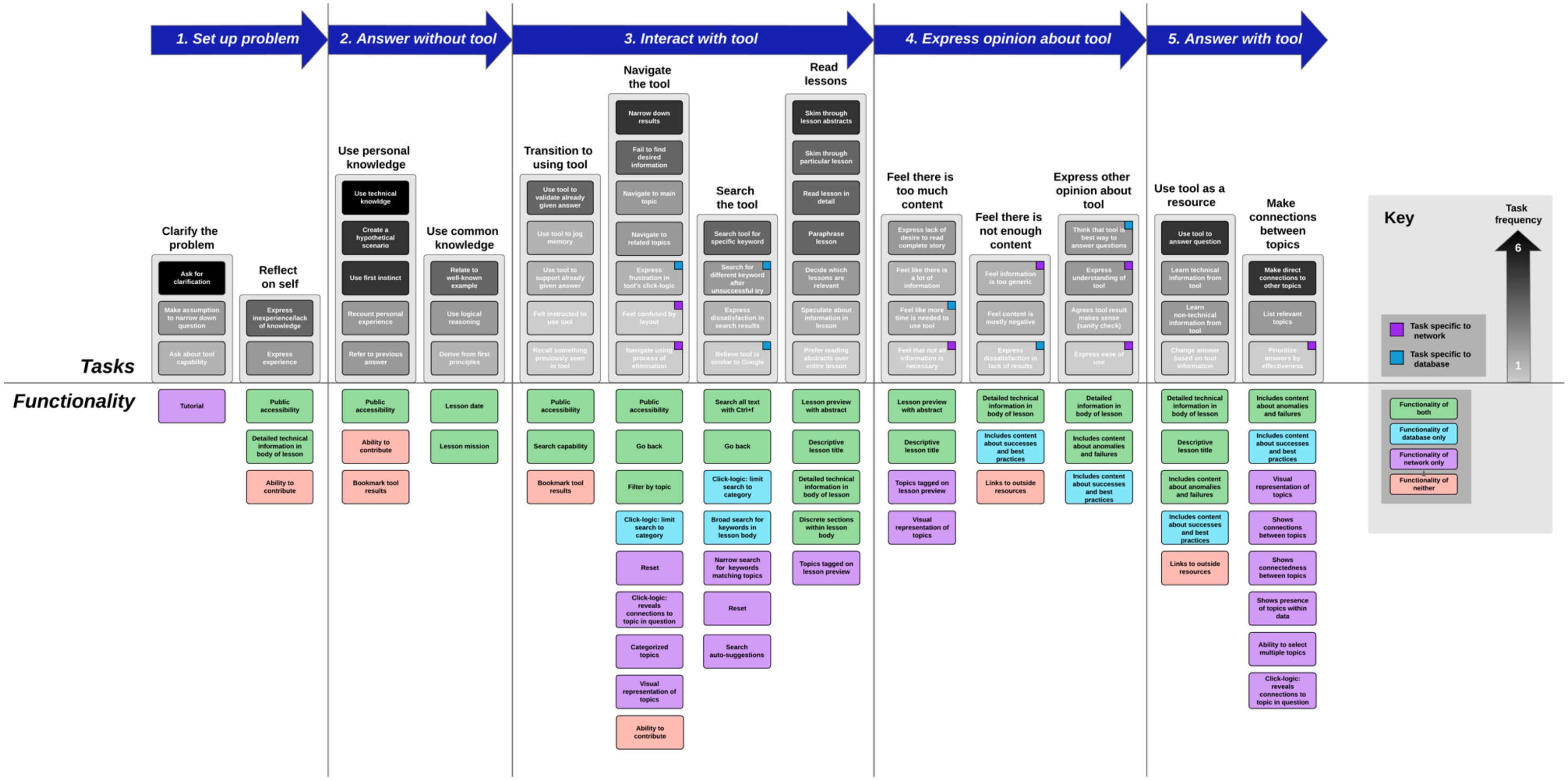


Figure 5.8: Mental Model: How Novice Engineers Respond to Engineering Scenarios

The order of mental spaces is significant in our model. Most participants (including ones not included in our mental model set) followed the same sequential thought process to answering the prompts: they state what they know first, state a reason to turn to the tool, interact with it, express an opinion on it, and refine their answer based on information they received. Participants in general followed this same order, but the ones that did not use the tool ended their response after the second mental space, *Answer without using tool.* According to the mental model, the network contributes more functionality tiles than the database. This is not surprising, as the network was contextually designed to provide more usability features.

Several towers were not completely met by the functionality of either tool. For example, the tower *Use personal knowledge* was only marginally satisfied by both tools. Functionalities not provided by either tool include *Ability to contribute to tool*, *Bookmark tool results*, and *Provide links to outside resources*. From the model, we can also identify the functional strengths of each tool. We observed that the network provided most functionality for the towers *Navigating the tool* and *Make connections between topics*. Both tools had around the same number of distinct functionalities for the tower *Search the tool*. However, the network had a narrower search capability which allowed users to only search for topics that were already identified as a node. The database had a search function similar to that of a search engine, where it searched for keywords within the lesson itself. From the mental model, we learned that the functional strength of the network is the ability to make connections between different topics, and the functional strength of the database is its broad search capability. Although the model helped us identify several gaps in tool functionality, in general, both knowledge management tools supported nearly all of the tasks that we identified for answering an engineering question, and both tools lacked very few obvious functionalities.

5.3.2 Use Cases

Once we created the mental model, we focused on tasks accomplished by two specific participants to illustrate how an individual's responses fit into the mental model. Participant A was the highest average scorer across all three case-based reasoning questions with access to the database, and Participant B was the highest average scorer with access to the network. Table 5.12 provides the demographics and statistics of both participants. Both participants were seniors in aeronautical and astronautical engineering with previous industry experience, and both their transcripts, which are found in Appendix E, are included in the sample we used to build the mental model.

Table 5.12: Case-Based Reasoning Performance for Participants A and B

	Participant A	**Participant B**
Tool	Database	Network
Year	Senior	Senior
Experience	1-2 years	1-2 years
Average Score	4.67	5.50
Average Time	4:26	12:28
Usage	[low, medium, medium	[high, high, high]

The specific tasks that Participant A performed are projected onto the mental model in Figure 5.9. From the figure, Participant A executed several tasks in the towers *Search the tool* and *Read lessons*. We interpret the figure as follows: Participant A followed the general process of initially answering a prompt based on their own knowledge, turning to the database to validate their answer, searching for specific keywords, and reading technical information in the body of lessons to add to their answer. Besides stating that the database was helpful in responding to the engineering prompts, Participant A did not express many strong opinions about the tool during the think-aloud protocol.

The specific tasks that Participant B performed are projected onto the mental model in Figure 5.10. Participant B executed the most tasks in the towers *Navigate the tool*, *Read lessons*, and *Make connections between topics*. In this case, Participant A followed the general process of immediately using the network to answer each prompt, navigating the nodal network to filter lessons, reading the technical information in the body of the lesson, and answering the prompt based on the technical information and on the connections between nodes in the network. With the occasional frustration, Participant A also expressed positive opinions about the network's usefulness.

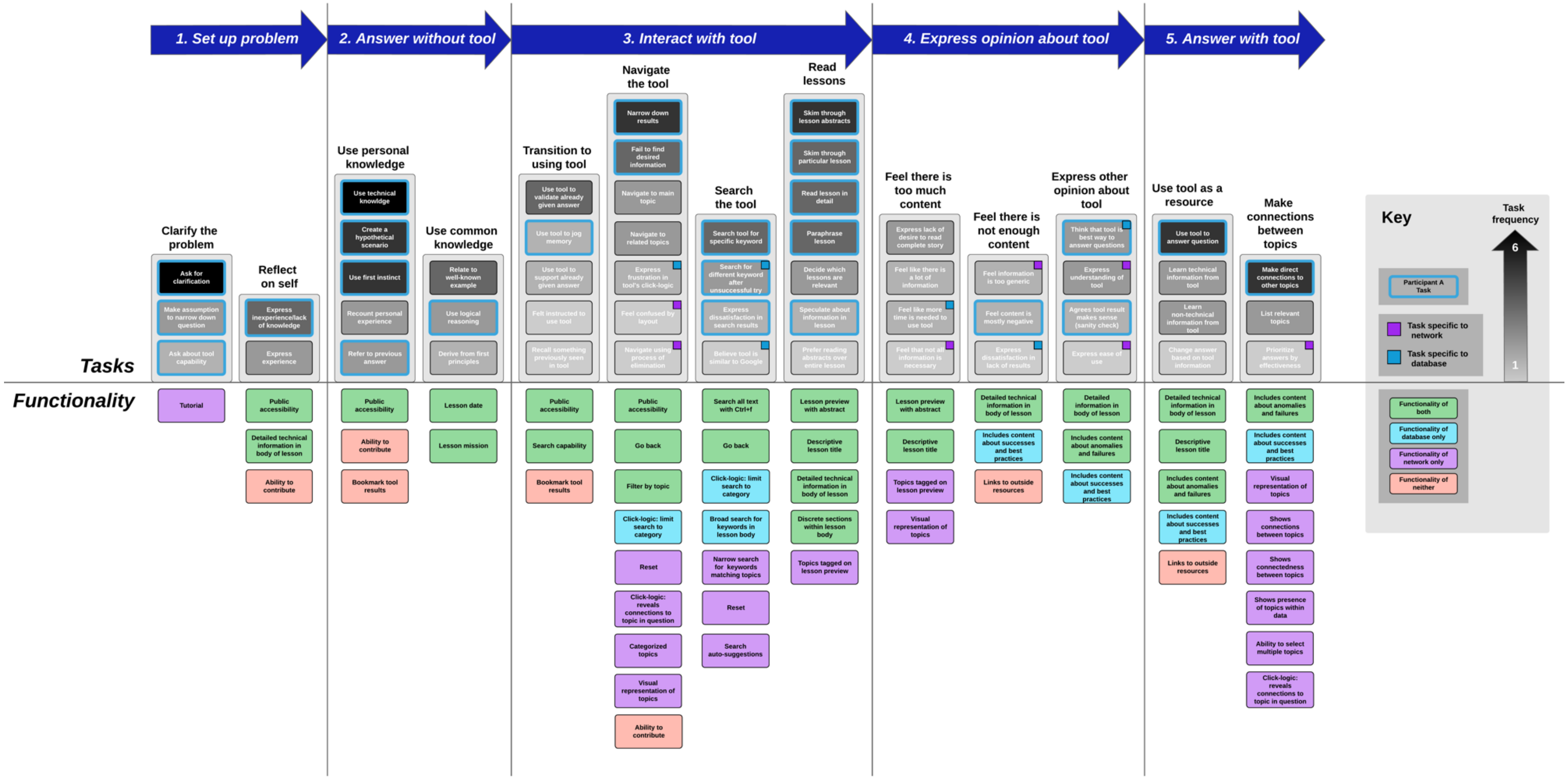


Figure 5.9: Mental Model with Highlighted Tasks Accomplished by Participant A

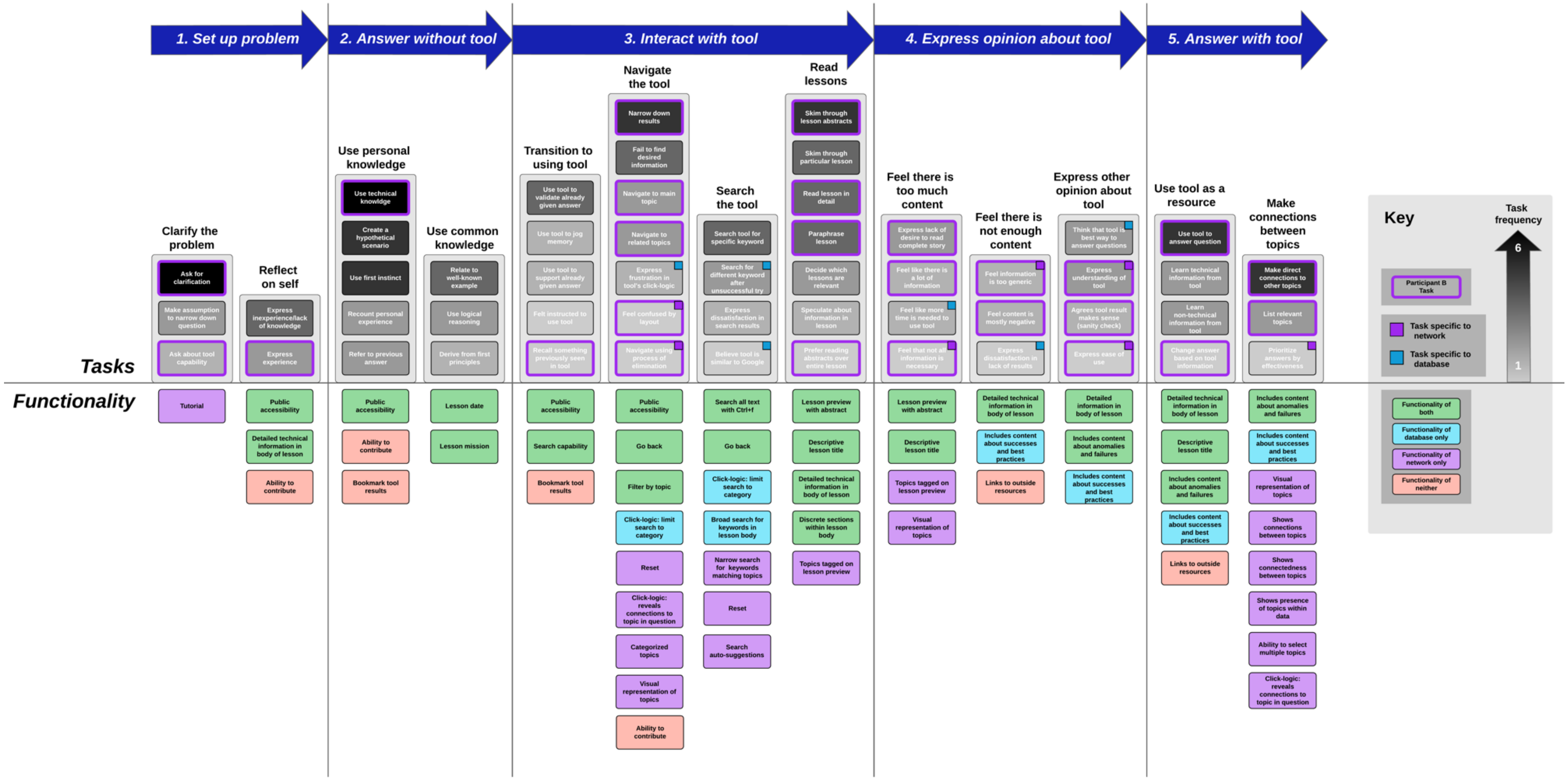


Figure 5.10: Mental Model with Highlighted Tasks Accomplished by Participant B

The purpose of performing the think-aloud protocol before the case-based reasoning questions was to give participants time to familiarize themselves with their tool before they were graded on their responses. If a participant used the tool during the think-aloud protocol, we assume that they used the tool in a similar manner during the case-based reasoning questions. From these use cases, we observe that the two top scorers on the case-based reasoning questions were participants who gravitated to the functional strength of their given tool during the think-aloud protocol. Participant A relied heavily on the search function of the database, which we identified from the model as its functional strength. They responded to the engineering prompts by searching for keywords in the prompt, quickly scanning lesson abstracts, and reading a lesson that seemed relevant—we can assume that Participant A followed the same process to provide high-scoring answers to the case-based reasoning questions. Participant B relied heavily on the nodal network, the functional strength of the network tool, which allowed them to make connections to many related topics. Participant B answered the prompts by clicking on nodes related to the given prompts, observing the connections made to other nodes, and thoroughly reading lessons associated with the selected node. We can also assume that Participant B followed the same process to provide high-scoring answers on the case-based reasoning questions. Both use cases provide qualitative evidence that relying on the strength of the given knowledge management tool will likely help a participant effectively respond to engineering scenarios.

5.3.3 Summary of Qualitative Results

In summary, both tools have the functionality to support most of the tasks we identified. The network has more functionality, which is not surprising because it was designed to have more routes to access the same information contained within the database. However, there is room for improvement for both tools, including a way to easily contribute to the tools and a method of providing supplemental technical information from an outside source. Adding more functionalities that support the most frequent tasks could further increase participation and interaction with the network tool. We determined from our mental model analysis that the functional strength of the database is its broad search capability and the functional strength of network is the ability to make connections between relevant topics. Successful participants relied heavily on the functional strength of the tool they received. Because different features are attractive to different users, the network possesses a functional advantage over the database because its multiple ways of accessing the same data can better cater to users with a variety of learning styles.

5.4 Discussion

In light of our quantitative and qualitative analyses, can we make any more conclusions from the results? In general, participants with the network performed better than participants with the database due to the network's functionality. Our linear regression analysis concluded that a participant's average score across all three case-based reasoning questions increased by 1.01 points if the network was used over the database, keeping all things constant. Additionally, the group of participants that had access to the network and used it "highly" were on average the highest-scoring group of participants. This result can be explained in the context of our mental model—the network has more functionality than the database to facilitate the participants' many tasks that went into answering engineering prompts. Additionally, the participant use cases provided evidence that successful participants used the same functionalities for both the think-aloud protocol and for the case based reasoning questions. Therefore, it is reasonable to surmise that participants that explored the network used the functionalities they encountered during the think-aloud protocol to provide high-scoring answers to the case-based reasoning questions. This is not a surprising result because the network was created using the steps of contextual design to implement features commonly requested by previous users, i.e., the network was intentionally designed with the deficiencies of the database and of the previous cause/recommendation network in mind.

Next, the case-based reasoning questions each played to the functional strength of a certain tool. From our linear regression analysis, we concluded that a participant's predicted score on Questions 1 and 3 would improve by more than one point if they used the network over the database, keeping other factors constant. However, we were unable to prove that a particular tool was better for Question 2. With our newfound knowledge of tool functionality from the mental model, we can see that Questions 1 and 3 played to the strength of the network, and Question 2 played to the strength of the database. Question 2 was based on near transfer, requiring a participant to find a specific lesson within the tool that directly applied to the question. According to the mental model, the database's functional advantage of a broader search capability made this an easier task for participants with the database because searching for a keyword in the database will provide more results than searching within the network. Additionally, Questions 1 and 3 were based on far transfer, requiring a participant to use the tool to find recommendations and attempt to apply them to the questions. According to our mental model, the network's functional advantage of making connections between relevant topics made far transfer an easier task for participants with the network. Participants with access to the network also may have scored higher because according to the rubric, participants were given one point if they "mention another relevant factor", a task better suited by a tool that enables far transfer.

Finally, using either of the tools as opposed to not using a knowledge management tool at all increases the time taken to answer a question because the abundance of functionality elicits new tasks from participants. As discussed previously, time taken to answer questions is also an indicator of success because efficiency is desired in almost all processes, including knowledge management. However, we narrowed the scope of our study of knowledge management tools by choosing score as our metric of success. This decision was consciously made during the contextual design process, where we decided to design a tool that provides many ways to access technical information to disseminate as much of the lesson as possible to the user. We did not focus on providing functionality that allows the user to access information faster. This decision is consistent with our understanding of how NASA's knowledge management tools fit into its organizational culture. They are not meant to be an "emergency" resources or decision support systems, where the goal is to improve efficiency when making decisions. Rather, they are used as a supplemental and information rich resource that should be reviewed at any time throughout the spacecraft lifecycle, not just when an anomaly occurs. This decision is also consistent with our results—participants used both knowledge management tools to score higher in a longer amount of time.

CHAPTER 6. CONCLUSIONS

6.1 Summary

In Chapter 2, we introduced the NASA Lessons Learned Information System (LLIS) as the baseline comparison to our new lessons learned dissemination tool. Next, we created a model that captures the important information of an individual lesson within the NASA lessons learned database. This model identifies five categories of a lesson—Component, Event, Technical Factors, Organizational Factors, and Recommendations—to describe the narrative of a spacecraft anomaly. Next, we identified codes for each facet of our model while concurrently applying it to the NASA lessons learned database. Finally, we coded 413 out of the 418 lessons learned contributed by the Jet Propulsion Laboratory (JPL).

In Chapter 3, we introduced Beyer & Holtzblatt's (1997) steps of contextual design as the process we followed to create a new lessons learned dissemination tool. We provided the feedback Aloisio (2018) collected on the previous cause/recommendation network tool, and we used this feedback to inform our steps of contextual design. To create a new tool from the cause/recommendation network, we performed the step *work redesign* by using customer usefulness feedback to inform the change in scope and customer. Next, we performed the step *user interface design* by using customer usability feedback to inform new features of the tool. Lastly, we introduced the final version of the interactive lessons learned network tool, populated it with the coded lessons learned data, and described its main functionalities.

In Chapter 4, we designed a three-part experiment to determine how novice engineers use each of the knowledge management tools (hereafter referred to as "network" and "database"). First, we split our research question up into three smaller key questions and identified what data must be collected in order to answer each. We described the first part of our experiment that deploys the interviewing technique of think-aloud protocol to answer the first key question. Then we described the second part of our experiment that deploys case-based reasoning to answer the second key question. We described the third part of our experiment that uses a survey to answer the third key question and to collect feedback for a possible third iteration of the contextual design process to improve the network tool in the future. Finally, we described the test procedures and discussed possible sources of bias within the experiment.

In Chapter 5, we discussed the results of the experiment. First, we observed that the majority of our participants were undergraduates, and the majority had at least some amount of industry experience. Next, we discussed the participants' preference for the tools. We found that in general, more participants found the network helpful than the database during the experiment, but both groups indicated that they would at least "probably" use a similar knowledge management tool in a workplace setting. We then discussed the amount of usage each tool received from the participants, observing that both tools were used at equal levels. We discussed the participants' scores on the case-based reasoning questions, where we found that for every question and for every use category, participants with access to the network had a higher average score than participants with access to the database.

To test our observation that the network was possibly more effective at disseminating lessons learned than the database, we performed linear regression and a two-sample t-test on the results. The regression analysis predicted that using the network highly (as opposed to using the database highly) would increase a participant's score by more than one point out of a possible six points for their total average score and for their scores on Questions 1 and 3. Neither the regression analysis nor the t-test found significance in the Question 2 data.

We finally described our qualitative analysis, where we analyzed a sample of think-aloud transcripts to create a mental model of how participants respond to engineering scenarios. From the mental model, we identified strengths and weaknesses in both of the knowledge management tools. Next, we focused specifically on two participants' transcripts and showed how their individual responses fit within the model. From these two use cases, we found that relying on the given tool's functional strength during the think-aloud protocol may lead to the participant scoring highly on the case-based reasoning questions. Lastly, we discussed our quantitative results in light of our qualitative results and found that case-based reasoning Questions 1 and 3 played to the functional strength of the network, which may explain why participants with the network scored higher on average for these two questions.

6.2 Key Findings

We began this research with four questions. First, **how do novice engineers use each tool, and do the tools have the proper functionality to support the demands of the user?** Novice engineers respond to engineering scenarios through a variety of tasks shown in Figure 5.8, and both knowledge management tools had the functionality to support nearly all of these tasks. The network possesses more functionality overall, which was expected because it was developed using a customer feedback-centered contextual design process. Each tool has its own functional strengths—the network helps engineers make connections between topics, and the database enables engineers to search for specific information. Both tools also have room for improvement, such as a way to easily add the user's own lessons to the data and a way to access outside technical information on a subject. Finally, we observed that using either knowledge management tool increases the time it takes for a participant to answer a question because interacting with a tool elicits more tasks from the user.

Do these tools help users craft answers to engineering questions? Due to our small sample size, we were not able to prove statistically which tool helps participants craft better answers for every scenario. We did observe that participants with access to the network performed better on average, and we were able to statistically prove for Questions 1 and 3 that using the network over the database results in a higher predicted score. Additionally, we observed that the two highest scorers on the case-based reasoning section of the experiment gravitated to their tool's functional strength during the think-aloud protocol.

What are users' opinions on the tools? On average, participants thought the network was more useful and more aesthetically pleasing than the database during the experiment. Both tools in general were received positively by participants, and an overwhelming majority indicated that they would use a similar knowledge management took in a workplace setting. However, we have evidence to believe that self-reported preference about a tool is not necessarily an indication of if or how much they will use the tool. It is possible that some participants would use a knowledge management tool to support their actual work practices even if they did not use it nor find it useful during our experiment.

And finally, **what role do knowledge management tools with lessons learned information play in the systems engineering decisions of novice engineers?** Our mental model identified the tasks that novice engineers execute when using knowledge management systems. The tasks that users perform vary widely based on the user's learning style and based on which knowledge management system to which the user has access. However, almost all participants followed the same sequential process to respond to systems engineering prompts. We have enough quantitative and qualitative evidence to recommend continuing development of a nodal network tool to disseminate lessons learned information to novice systems engineers. With a bit more user-centered development, ideally using the feedback we collected to inform another iteration of the contextual design process, the network tool could serve as an effective resource to disseminate lessons learned to novice systems engineers in the aerospace industry.

6.3 Limitations and Potential Improvements

What are some possible sources of bias in our experiment? One possible source of bias is that one of the two graders was able to determine which tool was used for each response. However, the inter-rater agreement sufficiently addressed this source of bias. Another possible source of bias is the questions were crafted to be easier for a specific tool. This is not the case, but the contextual design of the network *did* craft it to be effective for workplace scenarios similar to the ones described in the case-based reasoning questions, and it was designed for it to be easier to access general information about certain spacecraft concepts. Therefore, it is not unexpected that the network would perform better. The language used in the case-based reasoning questions is not specific to a certain tool, and both tools are populated with the exact same data.

What are some ways to improve the experiment for these particular tools? More conclusive results may be attainable with more experiment participants. Increasing the sample size may also increase the significance of our results. Because each experiment took around an hour to complete (not including the time it took to transcribe the recordings and score the answers), this is a challenge to accomplish. An online survey, as opposed to an in-person experiment, may help gather feedback on the tools, but it would not help identify a user's tasks nor obtain objective answers to the case-based reasoning questions. Because the participants' feedback ratings were skewed towards favorability, perhaps allowing them to rate the tool on a scale from 1-5 versus 1-10 would collect more accurate feedback. To gain more insight into participants' interaction with the tools, other strategies that we can deploy with this particular experiment include monitoring participants' search history, collecting clickstream data, video recording the think-aloud protocol, and asking questions after the experiment is completed, or retro-reporting.

Have we identified any ways to improve the tool during the next iteration of contextual design? First, adding a broader search capability to the interactive lessons learned network is a minimal-effort task that would greatly improve the functionality of the network. Next, it is possible that five categories are too much for a visual tool. According to Shneidermann & Plaisant (1990), four consistent themes with associated colors is the maximum number that users can easily handle. Also, another iteration of contextual design should rely heavily on feedback we collected from the Part 3 survey and opinions expressed verbally in the think-aloud protocol transcripts. One issue addressed in the feedback is legibility of the network. Legibility can be improved by addressing the small font size and the overlapping node labels on the network. Another issue was that the navigation and click logic of the network was not intuitive for some users. For both tools, there were several times where a user navigated to a state that they did not intend and could not figure out how to navigate back to the previous state. Another point of confusion for both tools was the lack of buttons—several users suggest that the tool would be more intuitive if every element acted as a button to filter information. One participant reported that the network was slow at times, which may be due to the large amount of data in the network. A possible solution would be to narrow down the number of lessons in the data. A final suggestion is to fix a few bugs in the network so it can keep up with demanding users.

Can our interactive network tool be easily implemented at other NASA centers and aerospace companies, and would it be as effective? Because the participants of our experiment were not collectively affiliated with any particular aerospace organization (besides the Purdue University School of Aeronautics & Astronautics), the network tool would likely work for any application of disseminating lessons learned to novice engineers. However, because we developed the coded factors to describe spacecraft anomalies from JPL lessons learned data, the specific set of codes would be more effective at JPL than at other industries. The codes would likely work well for organizations that provide the same services as JPL, such as NASA Goddard and the Applied

Physics Laboratory (APL). However, any company implementing a similar tool may find it helpful to alter the codes to match company-specific jargon. The codes would also potentially work for other aerospace companies, but they would need to supplement the codes with more specific ones related to their particular projects (e.g. JSC would need more codes relating to human space flight anomalies).

Finally, besides making improvements to the network tool, what are some other future projects that relate to our experiment? A useful complimentary project would be to create a GUI for a user to quickly and easily code and upload their own lessons into the network. Although having a user code their own lessons would not be very time-consuming, codifying an entire database consisting of hundreds of lessons learned is not a trivial task to accomplish by hand. This process would be greatly aided by using artificial intelligence to identify and create codes specific to an organization then apply these codes to the organization's pre-existing lessons. A final related project would be to apply lessons learned to model-based systems engineering (MBSE). MBSE is used in the aerospace industry to holistically represent every aspect of a project lifecycle. Lessons learned or best practices could be represented directly inside a project's model, and designers could be notified when their project violates one of these best practices. Instead of going through a time-consuming waiver process, it would allow the designers, management, and waiver review board to understand a requested design deviation within the context of the project.

APPENDIX A. FACTOR CODES

Components

Table 6.1: Component Codes and Presence Count

Components	**Presence Count**
Circuit components	36
Flight hardware	35
Communication hardware	30
Flight software	25
Instrument	18
Attitude control hardware	15
Lab facilities and equipment	15
Cables and lines	14
Communications link	14
Ground support equipment (GSE)	14
Spacecraft design	14
Antenna	13
Battery	12
Gimbal/actuation assembly	12
Camera/spectrometer	11
Pyrotechnic hardware	11
Flight computer	10
Propellant tanks	10
Spacecraft structure	10

Components	**Presence Count**
Command interface	9
Power system hardware	9
Power system hardware	9
Propulsion system	9
Pressurization system	6
Relays, valves, and sensors	6
Spacecraft exterior	6
Solar array	5
Magnetometer	4
Launch vehicle	2

Events

Table 6.2: Event Codes and Presence Count

Events	**Presence Count**
Component failure	47
Degraded performance	45
Accident	32
Malfunction	32
Risk increase	30
Component damage	28
Contamination	21
Test failure	21
Electrical short	18
Cost/schedule overrun	16
Operation outside allowable conditions	11
Fire/overheating	10
Loss of data	10
Dynamic perturbation	9
Corrupted data	7
Hangup/trip	7
Operations difficulty	7
Overcurrent event	7
Deviation from requirements	6
Loss of telemetry	6
Rupture	3

Technical Factors

Table 6.3: Technical Factors Presence Count

Technical Factors	**Presence Count**
Material property effects	35
Failed weld/poor workmanship	29
Interference from other subsystems	18
Defective/counterfeit part	17
Ground/storage environment	17
Component layout	16
Inadequate hazard protection	16
Inaccurate models	15
Incorrect parameter	15
Temperature effects	15
Space environment effects	13
Voltage settings	13
Electrostatic discharge (ESD)	12
Heritage	12
Logic error	12
Test configuration	12
Electromagnetic interference (EMI)	11
Part fatigue	11
Transient effects	11
Sneak path	10
Vibration effects	10

Technical Factors	**Presence Count**
Late design change	9
Electric arc	7
Torque effects	7
RF breakdown	6
Single-event upset (SEU)	6
Demated connector	5
Single point failure	5
Combustion product leakage	4
Inadequate margins	4
Part shelf life	4
Natural disaster	3
Software incompatibility	3
Design creep	2
Excited structural resonance	2
Propellant migration	2

Organizational Factors

Table 6.4: Organizational Factor Codes and Presence Count

Organizational Factors	**Presence Count**
Failed to consider design aspect	30
Subjected to inadequate testing	26
Created inadequate procedures	23
Failed to provide resources	17
Subjected to inadequate reviews	17
Kept poor records	15
Violated procedures	14
Inadequately communicated	14
Conducted maintenance poorly	13
Used inadequate justification	10
Conducted poor requirements engineering	10
Failed to inspect	9
Failed to supervise	9
Did not learn from failure	8
Lacked experience	7
Failed to form a contingency plan	6
Failed to consider human factor	6
Did not allow aspect to stabilize	4
Lost crucial knowledge	4
Managed risk poorly	3
Failed to train	3

Organizational Factors	**Presence Count**
Enforced inadequate regulations	2
Violated regulations	1

Recommendations

Table 6.5: Recommendation Codes and Presence Count

Recommendations	**Presence Count**
Establish more checks in the system	48
Make instructions more clear	48
Develop a more comprehensive and rigorous test	36
Conduct additional analysis early on	29
Establish a program or service	28
Add hazard protection to spacecraft design	24
Develop specialized training	24
Record and report key information	24
Add hazard protection to test setup	22
Monitor component's environment or behavior	22
Develop operational procedures and constraints	21
Increase resources	21
Consult and update institutional resources	20
Identify weak areas	20
Design for robustness or resilience	19
Give supervisor more capacity for oversight	19
Improve efficiency in critical tasks	19
Involve stakeholders in decision-making	19
Conduct subsystem test before integration	18
Test as you fly	18
Develop a contingency plan	17

Recommendations	**Presence Count**
Conduct piece-part analysis	16
Consider operating environment	16
Requalify heritage systems	16
Review decision-making logic	15
Track compliance to an objective standard	15
Use trusted distributors	15
Add features to the spacecraft to reduce risk	14
Consult subject matter experts	14
Define and update requirements	14
Verify components before test	14
Consider all operating modes in the design	12
Design for simplicity and testability	12
Increase margins	12
Test with prototype hardware	12
Add functional redundancy	11
Conduct end-to-end testing	11
Communicate key parameters	10
Add logic to spacecraft design	8
Conduct post-test analysis	8
Update models	8
Consider contributions of all subsystems	7
Use new or high reliability hardware	7
Vet new technology	7

Recommendations	Presence Count
Conduct random and independent evaluations	5
Keep up with current technologies	4
Make regulations more strict	2
Sterilize spacecraft rigorously	2
Establish an independent and transparent supervisory agency	1

APPENDIX B. PART 1: THINK-ALOUD PROTOCOL PROMPTS

Practice Prompt: “How many windows are there in Armstrong Hall?”

1. “What parts of the spacecraft could be impacted by an electrical short?”
2. “What factors would you consider when working with an experimental battery?”
3. “What problems could be posed by inexperienced personnel?”
4. “How would you mitigate a high-gain antenna failure?”
5. “What hazards could be posed by a part with an expired shelf life?”
6. “How could you improve the performance of a communications link?”
7. “How would you prevent a spacecraft’s electronics from experiencing electromagnetic interference?”

APPENDIX C. PART 2: CASE-BASED REASONING QUESTIONS

Question 1: Battery

Your boss has put you in charge of testing and storing an experimental rechargeable battery. While the battery is charging between tests, you notice that the battery's shelf life expired three months ago. Upon inspection of the battery's specifications, you realize that there are no guidelines or procedures that specify the proper voltage setting during charge. According to the NASA Systems Engineering Handbook, risk is characterized by the following three elements:

1. The scenario(s) leading to degraded performance in one or more performance measures
2. The likelihood(s) of those scenarios
3. The consequence(s), impact, or severity of the impact on performance that would result if those scenarios were to occur

Based on this definition of risk, discuss two risks to which the battery may be subjected. How might you mitigate these risks?

Table 0.1: Question 1 Grading Rubric

Desired Quality	**Explanation**	**Point Value**
Provides correct answer (1)	Identifies 2 of the following as risks: voltage settings, shelf life/expiration, layout, ground/storage environment, hazard protection, material properties	**1**
Provides correct answer (2)	Identifies 2 of the following as consequences: fire/overheating, battery failure, degraded performance, overcurrent, malfunction	**1**
Provides correct answer (3)	Identifies 2 of the following as recommendations: test with prototype hardware, monitor/track components, use trusted distributors, design for robustness/resilience, develop operational procedures/constraints, develop better tests/instructions/checks/training, add hazard protection/redundancy/margin	**1**
Mentions another relevant factor in discussion	Mentions a relevant topic not mentioned above, provides a recommendation not mentioned above, or mentions a stakeholder.	**1**
Answers the question asked	Discusses 2 risks with scenario, likelihood, and consequence for each.	**1**
Clearly communicates response	Provides a response with minimal confusion or conjecture. This does not necessarily mean that the response is completely spelling- or grammar-error free, but it must be legible.	**1**

Question 2: Spacecraft Operations

You are on a team operating an Earth-orbiting spacecraft that performs day and night climate observations. For the past few days, your team has observed several spacecraft battery anomalies, including the battery's frequent inability to charge. The battery is necessary to power the spacecraft during night observations. Recently, your team has observed that the battery is unable to hold a charge at all, and you have estimated that at the current rate of battery discharge, the battery will be dead in 48 hours. If new commands are not uplinked before the battery dies, the spacecraft will be lost. **What can you do to salvage the spacecraft?**

Table 0.2: Question 2 Grading Rubric

Desired Quality	**Explanation**	**Point Value**
Provides correct answer (1)	Explicitly mentions 1 of the following: CloudSat mission, daylight-only operations (DO-Op), dropping/descoping night observations	**1**
Provides correct answer (2)	Directly or indirectly mentions another power source onboard the spacecraft. e.g. "Divert power from the spacecraft's RTG." e.g. "Point the spacecraft towards the sun."	**1**
Provides correct answer (3)	Recommends 1 of the following: develop operational procedures/ constraints, consult subject matter experts, add hazard protection/redundancy/margin	**1**
Mentions another relevant factor in discussion	Mentions a relevant topic not mentioned above, provides a recommendation not mentioned above, or mentions a stakeholder.	**1**
Answers the question asked	Response is an obvious attempt at a solution.	**1**
Clearly communicates response	Provides a response with minimal confusion or conjecture. This does not necessarily mean that the response is completely spelling- or grammar-error free, but it must be legible.	**1**

Question 3: Contamination

You are in charge of integration and test of NASA's newest Mars probe. Launch is approaching fast, and today your team finished mounting the science instrument payload onto the spacecraft bus in the clean room. Immediately after this activity, you notice spots of brown residue on the side of the spacecraft. Outside of the clean room, you mention this to a more experienced co-worker who theorizes that the residue is a sign of contamination. Historically, spacecraft contamination has led to a host of mission-ending problems, including rising spacecraft temperatures, clouding of instrument optics, and even unexpected electronics behavior. Now, you must figure out the source of the contamination before it potentially spreads to the expensive payload. **What are some potential causes of contamination on your spacecraft? List them in the order you would investigate them, and explain why.**

Table 0.3: Question 3 Grading Rubric

Desired Quality	**Explanation**	**Point Value**
Provides correct answer (1)	Explicitly identifies 2 of the following as causes: material properties, temperature, space environment, incorrect parameter, hazard protection, poor workmanship, ground/storage environment	**1**
Provides correct answer (2)	Provides rationale for each item's order in the list.	**1**
Provides correct answer (3)	Recommends 1 of the following: make instructions more clear, establish checks in the system, add hazard protection, develop specialized training, develop contingency plan, monitor component, develop operational procedures and constraints, sterilize spacecraft	**1**
Mentions another relevant factor in discussion	Mentions a relevant topic not mentioned above, provides a recommendation not mentioned above, or mentions a stakeholder.	**1**
Answers the question asked	Provides an ordered list of AT LEAST 3 possible factors.	**1**
Clearly communicates response	Provides a response with minimal confusion or conjecture. This does not necessarily mean that the response is completely spelling- or grammar-error free, but it must be legible.	**1**

APPENDIX D. PART 3: FEEDBACK SURVEY QUESTIONS

1. **What is your academic year?**

   (Radio buttons: Sophomore, Junior, Senior, Master's, PhD)

2. **What is your professional experience? What industries have you worked in and for how long?**

   Space: (Radio buttons: No experience, Less than 1 year, 1-2 years, 2+ years)

   Aviation: (Radio buttons: No experience, Less than 1 year, 1-2 years, 2+ years)

   Defense: (Radio buttons: No experience, Less than 1 year, 1-2 years, 2+ years)

   Military: (Radio buttons: No experience, Less than 1 year, 1-2 years, 2+ years)

   Other: (Radio buttons: No experience, Less than 1 year, 1-2 years, 2+ years)

3. **Have you used this resource before?**

   (Radio buttons: Yes, No)

4. **On a scale from 1 (worst) to 10 (best), rate the look and feel of the tool you used.**

   (Choose rating)

   1 2 3 4 5 6 7 8 9 10

   Comments on look and feel:

   (Free response)

5. **On a scale from 1 (worst) to 10 (best), rate the ease of navigation of the tool you used.**

   (Choose rating)

   1 2 3 4 5 6 7 8 9 10

   Comments on navigation:

   (Free response)

6. **On a scale from 1 (worst) to 10 (best), rate the amount of detail provided by the tool you used.**

   (Choose rating)

   1 2 3 4 5 6 7 8 9 10

   Comments on amount of detail:

   (Free response)

7. **On a scale from 1 (worst) to 10 (best), how useful was the tool in answering the questions?**

   (Choose rating)

   1 2 3 4 5 6 7 8 9 10

8. **What did you like about the tool?**

   (Free response)

9. **What parts of the tool could be improved?**

   (Free response)

10. **Would you consult this resource for help in an unfamiliar situation at work?**

    (Radio buttons: Definitely Yes, Probably Yes, Probably No, Definitely No)

11. **Any other comments?**

    (Free response)

APPENDIX E. THINK-ALOUD TRANSCRIPTS

Participant A

K: Okay. Do you have any questions? No, okay, then we're going to get started. So again remember to think aloud whenever I ask you a question. So our first question is what—what parts of the spacecraft could be impacted if there was an electrical short on the spacecraft?

A: All right, if we're thinking just generic spacecraft, I mean, an electrical short on the spacecraft so that could be anywhere, so that could be it could be just a specific... I mean, if we're just talking about what could be impacted, I guess we're thinking worst-case scenario like the main—I forget what it's called—but like the main mission overseeing unit of the spacecraft that like plans and controls every part of it is short circuited. So that would affect pretty much all… okay. So now I'm just thinking obviously every piece of electrical equipment on the spacecraft. So that would include communications equipment such as antennas and not just the actual communication. But, like, configuring the antennas to be pointing in the right direction and to be open at the right frequency and all that, any sort of control software that controls reaction thrusters or propulsion systems could be impacted, anything that controls cooling or heating of the spacecraft could be impacted, anything that powers it so solar panel power generation technology. If it uses like a nuclear reactor to generate propulsion or energy that could also be impacted, anything—any computer that is logging data to be related later could also be impacted. Trying to think... I'm sure there's other random pieces of hardware that I'm not even thinking of right now. I'm just going through here. communications is the big thing. Any sort of sensors, I guess. So anything that senses gravitational fields, pressure, magnetic fields, anything like that. Those are big things that are standing out to me. Yeah actuators which sort of applies to everything propulsion systems any sort of calculations that the spacecraft might be making to adjust itself could be impacted. That's everything I got. Okay.

K: Yeah, you can end your answer any time.

A: Okay.

K: Okay. What factors would you consider when working with an experimental battery?

A: An experimental battery? Am I allowed to Ctrl+f?

K: Yes, you may.

A: Right. So this is the only one with the word battery in it. So I'm going to see if skimming through this has anything. I see there's an experimental battery. Perfect. All right, so it's [inaudible] their annual maintenance is very… should be maintenance should have alarmed systems for timely warning of malfunctions. Trained personnel, so could you repeat the question again?

K: So yeah, what factors would you consider when you're working with an experimental battery?

A: Okay, so... Just reading this give some ideas. So having some sort of notification system to know when something goes wrong with the battery. And when something does go wrong making sure that as it says here the mishap scene is preserved. So there's enough left of it that you can figure out what went wrong and why it happened, making sure that the people who are—or at least somebody somewhere is trained to deal with mishaps to investigate them to see where they went wrong. Essentially building the battery so that if it fails, it's relatively easy to take it apart and figure out why it failed. Don't make—not making it too complicated or convoluted for people to see what connects to what. I think a big part of it would be just the system design of it. Make sure that humans can understand it enough to understand why it failed. That's all focusing on just failure. I guess I mean also doing lots of study into the reliability of a doing rigorous, rigorous tests exposing it to the worst possible conditions to make sure that you have that you see all the ways it could fail and have built-in ways to mitigate that and if it does happen ways to mitigate the effects of that. When working with an experimental, yeah. I mean mainly I feel like the only thing I'm talking about his failure, but that's the only thing that's coming to mind is just being aware of failure, being able to investigate failure, and being able to mitigate the effects of any failures that do occur. And also making sure that it's actually necessary and worth looking into if it is experimental and new making sure that the benefits it's giving are worth the risk.

K: Um, okay what problems could be posed if you had experience—no if you had inexperienced personnel working on a project?

A: Inexperienced personnel... Biggest problems I think would be assumptions that they might make that would turn out not to be true. That could be anything from assuming that other people know things that they don't or assuming that other people are better at something than they actually are assuming that that you know, assuming if I'm of these personnel assuming that I understand something better than I do which is a very vague sort of notion. Inexperienced personnel... Trying to see if that most things in here are technical things. I'm not seeing many things about personnel, but the... Getting people who are too attached to a certain idea, I would say, like someone gets an idea or the here and idea and they say decide that's their favorite idea for how to implement a system and they sort of put the blinders on and stop listening to any other ideas and even if tests and studies show that their idea isn't a very good one, they still sort of stick to that one because they haven't—they're not used to hearing other ideas or not used to having to throw away ideas that aren't good. Yeah, that's all that comes to mind for me. Okay?

K: Okay. How would you mitigate a high-gain antenna failure?

A: All right. Perfect. All right. Galileo, let's see… really high gain antenna is damaged during... "Be cautious of informal environmental testing to validate analytical models. It is typically conducted without the strict test controls and unambiguous assignment of responsibility that's required for testing of flight subsystems and spacecraft." Could you repeat the question again?

K: How would you mitigate a high-gain antenna failure?

A: Okay, how do you mitigate the high-gain antenna failure? So I'm seeing two components to that one is predicting failures and the other is implementing systems that would prevent—that would actually make it. So realizing what failures you need to mitigate and then actually implementing systems to mitigate those failures. So and it seems like one lesson that's coming from this particular study... Skim through the recommendations a bit more. So you need to make sure that the test environment that you are putting your system through is both rigorous and well controlled and also actually represents the risks that it would experience out in the real world, which is kind of a tightrope to walk because the real world is not tightly controlled. Let's see... Understand the results, especially those include detailed physical or functional inspection nicely. Okay. So basically, make sure that if something goes wrong during a test, you can tell exactly what went wrong and exactly what caused it to go wrong before you go jumping to conclusions about what happened. So make sure that every single part of the testing and assessing failures and handling the hardware is assigned to a specific person. Everyone knows what their role is. And then kinematic structures should test prototype hardware... Okay, so from what I can tell, it's saying if you can't do a very detailed analysis on actual flight hardware, you should test it with prototype hardware. Not totally sure what kinematically indeterminate means but I guess that means volatile or unstable and predictable parts of the hardware. Tested with prototype rather than flight hardware... It seems like a very obvious recommendation, but I'm guessing what that's saying is you might break it so don't use the actual hardware you're planning to fly, which seems like a silly thing to have to put in a paper, but I guess the mistakes do happen. Yeah, that's my answer.

K: All right. So what hazards could be posed if you used a part that have like an expired shelf life or has passed a shelf life?

A: Okay, depends a lot on the part. Life-limiting life expectancy... It seems like the only useful thing here. So this is mainly about electrical parts, it looks like. So over these will come life-limiting items. Okay. So this isn't so much about doing this doesn't really apply to the question. This is just about doing things wrong makes things have a shorter life. So just relying on my engineering intuition. So the question was what could go wrong if you have an expired part?

K: If you use an expired part.

A: Use an expired part, yeah. A whole lot of things depending on the part. It also depends what it means by expired. It could be like for some reason that part is no longer legal to use, maybe it uses a material that you're no longer allowed to use in spacecraft because of some new discovery. If it's an electrical component, it could be as benign as it just works more slowly than you want it to if it's like a computer. Also, if it's some sort of chemical, like it depends what you mean by part because if part can be like... The propellant that you're using, I mean that could be that could, you know, it could explode your entire spacecraft. If you're using an if you're using just like an expired, you know sort of clunky motor maybe your system still works, but it doesn't work... It doesn't work as long as you want it to, and it starts seizing up and breaking down more quickly. So it decreases the lifespan of your overall system. It could be that perhaps the part... Maybe something about the environment that you're sending the system into has changed since the part was created. And so it's not prepared for all of the hazards that will be exposed to whether that's radiation or winds or impact or anything like that. Yeah, I mean that's I feel like there's sort of infinite questions depending which part is—infinite answers depending which part is expired. But those are the only ones that come to mind.

K: Okay. Great. Okay, how could you improve the performance of a communications link?

A: Improve the performance of a communications link. I'm guessing there's something in here that would actually be useful for this. So it seems to be just in the title... Ground data... Actuator design... Skimming these abstracts because we have to go off of... I'm sort of communications link between Cassini probe radio frequencies would not adjust for Doppler effect... All right, so remember the Doppler effect, end-to-end testing, High Fidelity, referencing Doppler effects, power switching [inaudible] matrix, two relays could inadvertently actuate ones commanded... A simple system control protocol should be established or by independently requesting that relay actuations are executed in predictable manner... Contractor proposals for attention for cost sheets to favor upgrading the unit... Resulting in some performance reliability. Okay. Mutual interference from the Viking orbiters can be adequately analyzed... Okay, so, can you repeat the question?

K: Yeah, how could you improve the performance—the performance of a communications link?

A: Okay. So for my vague understanding a communications Link and my vague skimming of those abstracts, the main things that come to mind are when designing it make sure to predict and test for thing—basically everything that was just there. So the Doppler effect, making sure to test the independence of controlling different pieces of the communications link, making sure trying to control one doesn't affect the other. I mean a lot of it just comes down to the same basic lessons of remember what has failed in the past. We try to be very creative when brainstorming just like... Spend a While brainstorming ways that things can go wrong and test and design for all of that very explicitly. That's the main thing. It was just what to consider when designing a communications link?

K: How would you improve the performance.

A: Improve the performance, right. All right, so part of it—I mean part of it would just be getting new or better technology but also making sure that the benefits of that new technology as far as increased performance are good enough to accept any po—any lower reliability because of the newer the technology the less reliable and will be in general making sure that that is explicitly considered and tested for and that nobody forgets to consider that newer technologies, even if they have better performance metrics, could also be less reliable, have a shorter lifespan. Also considering just the objectives of the mission. So is it even—if your mission can work with older less reliable technology and if you don't need really fancy like laser to communications to achieve your objectives, then don't do it because they're more expensive. And they're less reliable and it's better to have slowly transmitted data than no data at all if that's an option. That's all I can really think of. I don't have a lot of intuition as far as Publications like this.

K: That's fine. Okay, last one. How would you prefer—excuse me—how would you prevent a spacecrafts electronics from experiencing electromagnetic interference?

A: Interference... Sort of read that one from electromagnetic interference. Alright, so basically these two. Doesn't really seem to say anything other than that bad stuff happens, but you can test for bad stuff. So, I mean, that's pretty—pretty obvious test for electromagnetic interference, test for all the ways that it could happen again. Just try to think of all the ridiculous situations that could occur and make sure that those are explicitly accounted for and that could include just saying this is so unlikely that we're not going to design for it. But at least acknowledge that that's what you're doing. This is more about what EMI can do and not so much about how you EMI happens. This or this was like what?

K: How could you prevent a spacecraft's electronics from experiencing EMI?

A: How can I prevent them from experiencing EMI? Yeah, okay, so not how to get them to be okay when they do experience it, how to prevent them from experiencing it. I mean the... Since I don't have a lot of my own knowledge to go off of, if I'm going off what's here. The best info I can find is... Finding just that think of all the ways that EMI could occur and then test for those and if you're going to not design a system to mitigate those that make sure you're explicit the acknowledging that you are doing so. Beyond that vague high-level notion, I don't think I can confidently say anything else.

K: All right. That is it.

Participant B

K: So what parts of your spacecraft could be impacted by an electrical short?

B: This is where I'm trying to like find information related to that.

K: Yes.

B: Okay. So if I see if I look at electrical short, so these are all things that are very... Whatever they call it. But there they mention it in there. It's like if I look at 18 stories involving electrical short... So I'm looking for the story that would involve that. We'll see if I click on another one... We have... Electrical short and something?

K: Oh, I was saying electrical short or short circuit.

B: Okay, always saying, yeah. So this is a short event that everything that's going to connect to it is something that... Is it something that mentions it, there is way more than 18 connected here... Recommendation. Okay, that makes sense what that means. Component... Budget... Okay, so it's got like the outcome. Mostly negative. I don't want to have to go read through each of these, just skim them. Can you read the question one more time?

K: Yeah, what parts of your spacecraft could be impacted by an electrical short?

B: We're looking at parts, ok. Electrical... There's a lot of things actually. Especially anything okay. So like if we're talking about Parts, I'm looking at all the elements and components, something like cables and lines, hardware computer systems, communication hardware, valves and sensors, any of those parts, but I would say because those are the only ones that seem to connect to an electrical short and specific components of what you mentioned. Yeah, okay even list each of them down here. So I would say any of those could or would be impacted and that kind of makes sense the more you think about it.

K: Okay. Yeah, when you're done you can just say "I'm done". Okay. So next question. What factors would you consider if you were working with an experimental battery?

B: If I'm working with an experimental battery, okay. Let's see. So we're talking about... One more time?

K: What factors would you consider when working with an experimental battery?

B: Factors would I consider. Okay, so let's see if I go to component, if they have anything related to specific battery. Ok so battery. So factors that I would consider when working with it might be like a technical Factor. So looking at stories, the factor working with the battery I would want to consider any of these like I would say voltage settings, I would say electric Arc, maybe parameters, interference electromagnetic interference maybe... And some of these are not necessary for battery, but these will include it so... Yeah, per shelf life actually almost looks like it's easier just to read with a list in here. Voltage settings I said that. Logic of course always. Yeah, alright.

K: Okay next one. So what problems could be posed if you are working with inexperienced people on your project?

B: Inexperienced people.

K: Yeah.

B: I might look at it from the last ending like okay, organizational factor if you're working with experienced people, Maybe things were like in adequately communicated. So looking back so problems—problems that I would face working with inexperience. Okay, so we're looking up problems see... We're going to say organizational Factor. This is just one example, I guess the same there was inadequate communication. Organizational factor... Lacked experience, how about that? Lacked experience so problems that we might could just be any of these three events you could have degraded performance, malfunction, or an accident. I guess it's kind of the Baseline to think of it. Obviously there might be some more little things that lead up to those. These are the main points...Yeah, so they talked about this one employee was unfamiliar with this device or whatever and it caused an accident.

K: Okay. Next one. What could you do to mitigate a high gain antenna failure?

B: Mitigate it. Okay. So, let's see. Let's see if there's any antenna failure. General test failure... Anything under recommendations for that? Here's a high gain antenna story right here and their recommendations are designed for robustness or resilience, consider operating environment, and design for simplicity and testability. I think that's specifically high gain antenna. But you can look at some other antenna examples like include like early analysis and consulting experts, okay.

K: Okay. Okay, what hazards could be posed if you're using a part that had expired over like a past its shelf life.

B: So an expired part. And then what was the first part of that?

K: What hazards could be posed by that?

B: Yeah. Okay. So say we had a part failure. Let's just do to use component failure there's a lot because obviously that can happen a lot so hazards. Which section would have that? Recommendations, no. Component, not necessarily. Read the question one more time.

K: What hazards could be posed by a part with an expired shelf life?

B: Yeah. So maybe not component failure then. Looking under organizational factor just to see if there's anything specific about that. Technical failure, technical factor. I see part fatigue, I'll probably click on that if I don't see anything else. Yeah, okay, so see part fatigue. Looking for some hazards related to that. Okay. Yeah, so maybe an event so here's events related to that. I actually found out that I kind of like reading examples better. So hazards. I like the little descriptions. So I see some of these they have events so you can say like one Hazard could just be an accident or an electric short, but also some of these have multiple technical factors. So I don't know if the event would be related to both or one of the other. Because looking at like the DD and power supply failure one. It's like technical factors. Okay like part fatigue, maybe that's what we're looking at. But also saying like vibration effects under the event electrical short, so if I read it and see it says... I'm not really sure. Read the question more time.

K: What hazards could be posed by a part with an expired shelf life?

B: So part fatigue could be something. There actually might be another factor in here. Expired shelf life—here's part shelf life. Look under that. And this could be one of those things where you know, it's not a super straight-forward or it can be related to that. Four stories involving certain events saying malfunction, component failure, test failure, and degraded performance. That kind of seems to align more with like expired parts. So I think I'm going to make that my answer instead of my last one where I was looking at—was I looking at—I was looking at part fatigue. I would say part shelf life and I would say those four. Okay. So make more sense to me.

K: Okay, we have two more. So how could you improve the performance of a communications link with your spacecraft?

B: Improve performance?

K: How could you improve the performance of a communications link with your spacecraft?

B: Okay, so we're looking for ways. So we consider the last thing to be under recommendations. So maybe I'll start there. There's kind of a lot, so this might not be the best place to start. I want to improve communications... Okay... Organizational factors... Read it one more time.

K: How could you improve the performance—excuse me—how could you improve the performance of a communications link with your spacecraft?

B: Improve communications link recommendations. Sometimes it helps to go back through and look at each one again... Communications link, let's click on that. So the question is, how could you prove it? Yeah, so that would be any recommendations. So 14 stories involving communications link. One talks about updating requirements, making sure they're well-defined. Design with testability in mind, program or service maybe. Test as you fly that's the classic fly as you test. Make instructions clear. And I like operational procedures as well, so they know what they're doing. Read the question one more time.

K: How could you improve the performance of a communications link with your spacecraft?

B: Okay, that works. Update requirements, design for simplicity, test as you fly, and then I really liked develop operational procedures.

K: Okay, last one. How could you prevent your spacecraft from experiencing electromagnetic interference?

B: Preventative... Electromagnetic. Okay, so I remember seeing this somewhere. Talked about E&M interference. So you said yes. So ways to prevent that?

K: Correct.

B: Okay. Usually a lot of stuff like that in the recommendations. Let's see. I kind of know a little bit about this already so I'll see which ones I like. A lot of testing before, like testing is going to be a recommendation for a lot of things not just this. I kind of like monitor components environment or behavior. There's no assurance that materials are at their lowest magnetic state. Magnetized components could result... Yeah, you don't want to lose your data. The stories do involve EMI. Okay. So this is one of the stories mentions like a like specific circuit analysis code. Specifically like optical wires cables and connectors which yeah like at all. Because they're affected by that. Still scrolling through all the old stories here. They're kind of the same. They're like... Test knowing the limits of all of your like electrical components and how they're going to react to this. But I guess more specifically. The recommendations are to monitor its environment or its behavior of each component, piece part analysis. And then a bunch of different types of testing. Like subsystem test, early analysis... Seems like it was a pretty simple way to do that, so I guess that's my answer.

K: Okay, great.

APPENDIX F. USER FEEDBACK

NASA Lessons Learned Information System (LLIS)

Table 0.4: User Feedback on the NASA Lessons Learned Information System (LLIS)

Feedback Question	**Response**
Any specific comments on look and feel?	"The site itself looks outdated, and when searching for a specific problem, it felt like you would have to be well aware of the NASA naming convention to be able to effectively find what you're searching for, else it just seems like you're swimming in a sea of random stuff."
	"When I go back to the previous page, the website will go back to the top of the page. While, I hope the page could stay at the original location that I clicked into the link."
	"Not super smooth, but not ugly either."
	"It seemed like a classic NASA color scheme and was not distracting from use of the tool."
	"A very functional database with all information well categorized and searchable."
	"Standard database, wasn't extremely better or worse than normal databases in terms of aesthetics and feel."
	"GUI is partially old and can sometimes be hard to navigate."
	"When I go back to the previous page, the website will go back to the top of the page. While, I hope the page could stay at the original location that I clicked into the link."
Any specific comments on ease of navigation?	"It was easy to get through, just finding the exact article you want was a pain."
	"I want the option of filtering out keywords and be shown documents with exactly the word I choose and I want it now."

Feedback Question	Response
	"I ignore the sort button because it is too small."
	"Aside from the reset button, the tool was very easy to use."
	"Reasonably intuitive."
	"Needing to reset the entire search every time you filter for results and want to go back is not helpful. The filters provided for the JPL documents were not all in alphabetical order and not always very specific."
	"Seemed to be easy enough, and lead to a lot of valuable information."
	"Try to condense categories on the left-hand-side."
	"Very easy to navigate, many topics to choose from and search bar was prominent."
	"Sometimes you would wish they had an advance search tool. Some keywords can have numerous entries making it harder to find entries that are related to what you are searching. Having an advanced search option would allow the user to be more specific and reduce the number of entries they would have to skim."
Any specific comments on amount of detail?	"Each article went into great detail about whatever it was talking about."
	"A bit overwhelming with detail."
	"There was a vast number of articles with even more information than just what was provided in each article's abstract."
	"I wish I could know who was originally involved with the issue/who entered the lesson into the database for further inquiry."
	"None in particular."
	"Decent amount of detail to separate topics, couldn't really add more otherwise the page would have been clustered."

Feedback Question	**Response**
	"Some articles seemed very detailed, while others seemed to be more broad and talked about the problem and solution at a higher level than the detailed versions. The long versions are good for when reading further, but can be hard to identify if the abstract is not conclusive."
What did you like about the tool?	"The tool searched quickly through the database and was minimalistic since everything was provided in orderly fashion. It also had articles on topics that I would not find on google itself."
	"It provided subcategories which helped in identifying certain aspects of a device or an issue."
	"From what I could find answers to, it gave very in depth answers, as well as giving ideas for possible answers if not explicitly stating the answer."
	"Filters by relevance."
	"The format of every article is the same, which makes people easy to track the keywords they are looking for. Also, the order of the subtitles follows the hierarchy rule well."
	"It has a collection of previous problems which makes it easier to assess the probability of an issue."
	"Relatively intuitive, with a lot of details if desired."
	"I liked the abstract provided with each article as well as each article's title having information that made it easy to figure out what the issue experienced was."
	"I liked the association of lessons learned with applicable topics. Reading the topics helped me narrow down my search, especially when I knew little about what I was working with."
	"Lots of interesting information."
	"The records are very detailed and have a consistent setup."

Feedback Question	Response
	"I'm already a bit of a nerd, and have had experience with civilian space simulation software, so it wasn't too out of the ordinary for me."
	"Both the category and search bar options."
	"Had a large amount of articles and sources to draw information from."
	"The large amount of entries that cover technical mistakes and non-technical mistakes. This helps in terms of improve the planning process, or taking a look at engineering mistakes from a systems engineering perspective."
What parts of the tool could be improved?	"The searched words could be highlighted or made bold after the search since that would make it easier to read the right sentence of the article."
	"Search bar."
	"The topic index."
	"More advanced searching abilities and filters to sort documents by."
	"Maybe increase the font size of the words below the 'Subject' section a little bit."
	"It seems robust as is."
	"Good keyword search could be helpful; often difficult to find specific cases related to a certain topic (e.g. spacecraft contamination)."
	"Having all of the filters in alphabetical order. Making the filters easier to go back and forth between. Making more useful filters i.e. "battery" "contamination" and other problems commonly experienced across missions."
	"The filtering mechanisms for lessons learned is a little unwieldy simply due to the sheer amount of options available. Perhaps adding access to options conditionally could help."

Feedback Question	Response
	"Design."
	"Find a way to make the reader have an easier time understanding the records."
	"Just UI. The information provided already seemed to be well compiled."
	"How specific the tool interprets your search."
	"User-friendliness, have articles related to search word (not just articles containing search word)."
	"This use of abstracts could be enforced to make skimming and identifying articles easier. Advanced search options would also allow the use to be more specific in their search."

Interactive Lessons Learned Network

Table 0.5: User Feedback on the Interactive Lessons Learned Network

Feedback Question	**Response**
Any specific comments on look and feel?	"Good graphic organization but not text organization."
	"I liked the multiple selection feature. Color coding was nice, maybe the five categories could be different, I felt like I didn't use some of them."
	"It was very easy to use."
	"Some nodes overlap."
	"Some of the label is not shown properly, especially on the edge of the screen."
	"It looks very nice, but does not feel good to use."
	"Some items overlapped each other (esp. the text), making them difficult to distinguish."
Any specific comments on ease of navigation?	"Navigation is easy to use but oversimplified."
	"The search bar provided no issues with navigation. Sometimes I would have a node linked to another node and when I would select both of them no stories would pop up. Not sure why this is."
	"Once you had read the demo is was easy but the demo was poorly located and should have had a different color background so it would stand out as something that was suggested to do before using the tool."
	"Search bar seemed useful."
	"Very user-friendly."
	"The 'search' box should be more clear."

Feedback Question	Response
	"There should be a way to show connections of one note to all nodes that has the same color."
	"Overlapping words."
	"Worked well when it didn't bug out. I'd like to be able to search for multiple keywords. ex: 'battery' and 'telemetry'."
Any specific comments on amount of detail?	"Pretty thorough and detailed."
	"The summary was useful."
What did you like about the tool?	"It was easy to navigate and find the things I needed."
	"I liked the ability to connect multiple 'nodes' to narrow down the results you were looking for."
	"I liked how easy it was to get to very specific situations that have happened in the past."
	"Connected search items with other, related items."
	"How easy it was to navigate and find useful information."
	"I like the web idea and the combination of multiple parts of it."
	"Visually pleasing and color coding helped categorize the content."
	"Very relevant and informative."
	"I didn't really use it but it appears to be super organized."
	"It provides examples of similar situations, it relates several aspects of an issue, giving a broad perspective for problem solving."

Feedback Question	Response
	"It shows connection between the series of event."
	"Shows related topics to the subject of interest."
	"Looks nice. Good summaries of the articles."
	"The tool allows me to select one or a few related issues and see all connected reports and possibilities highlighted. Possible relationships between any components and events can be examined."
	"It very clearly showed how the different causes, mitigations, and effects were related to each other."
What parts of the tool could be improved?	"More topics."
	"Add lists of 'textbook' explanations for common events. The past examples were good for lessons learned, but not for answering basic questions for a certain situation."
	"The part I mentioned before about when selecting multiple nodes that are shown to be linked no stories popping up."
	"Organizing search results and avoiding repeat topics."
	"Words from each bubble would overlap making it difficult to read and in turn find specific information."
	"The articles / stories were very large and looked like pintrest in the structure and should have been listed for quicker usage."
	"Demo could be more intuitive."
	"I liked the way it was. I can't think of any better way to easily visualize the content the way it is."

Feedback Question	**Response**
	"There's so many stories on the tool that it can get overwhelming."
	"Distribution of nodes."
	"Data filter."
	"Make it in a list format to easily see all the main failures/issues."
	"The search function sometimes bugs. Searching for multiple keywords."
	"I saw no need of improvement."
	"Provide a definition/explanation for the scope of the items to be accessible within the tool."

REFERENCES

Aloisio, D. (2015). Accidental Knowledge: Using Accidents and Other Project Failures to Inform Research in System Engineering. West Lafayette, IN.

Aloisio, D. (2015). *Accidental Knowledge: Using Accidents and Other Project Failures to Inform Research in System Engineering.* West Lafayette, IN.

Aloisio, D. (2018). *Lessons from Systems Engineering Failures: determining why systems fail, the state of systems engineering education, and building an evidence-based network to help systems engineers identify and fix problems on complex projects.* West Lafayette, IN.

Aloisio, D., & Marais, K. (2018, July 7-12). Instant-Expertise in Failure Causation: Developing and Presenting a Network of Causes and Recommendations Extracted from Past Failures. *20th Annual INCOSE International Symposium.* Washington, D.C.

Beyer, H., & Holtzblatt, K. (1997). Chapter 1. In *Contextual Design: Defining Customer-Centered Systems.* San Francisco: Morgan Kaufmann.

Birkinshaw, J. (2001). Why is Knowledge Management So Difficult? *Business Strategy Review*, 11-18.

Columbia Accident Investigation Board. (2003). *Report Volume I.* Washington, D.C.: Government Printing Office.

Dennehy, C. J., Labbe, S., & Lebsock, K. L. (2010). The Value of Identifying and Recovering Lost GN&C Lessons Learned: Aeronautical, Spacecraft, and Launch Vehicle Examples. *American Institute of Aeronautics and Astronautics*.

Detterman, D. K. (1993). The Case for the Prosecution: Transfer as an Epiphenomenon. In D. K. Detterman, & R. J. Sternberg, *Transfer on Trial: Intelligence, Cognition, and Instruction* (pp. 1-24). Norwood, New Jersey: Ablex Publishing Corporation.

Eccles, D. W., & Arsal, G. (2017). The think aloud method: what is it and how do I use it? *Qualitative Research in Sport, Exercise and Health*, 514-531.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data, revised edition.* Boston: The MIT Press.

Kolodner, J. L. (1992). An Introduction to Case-Based Reasoning. *Artificial Intelligence Review*, 3-34.

Li, J., Liu, M., & Liu, X. (2016). Why do employees resist knowledge management systems? An empirical study from the status quo bias and inertia perspectives. *Computers in Human Behavior, 65*, 189-200.

NASA Jet Propulsion Laboratory. (2014). Provide Adequate Maintenance and Hazard Response for UPS Units (LLIS #12101). Pasadena, CA.

NASA Jet Propulsion Laboratory. (2017). Projects Benefit from Lessons Learned on Previous Spaceflight Projects (LLIS #21501). Pasadena, CA: NASA Lessons Learned Information System.

NASA Office of Inspector General. (2012). *Review of NASA's Lessons Learned Information System.* Washington, D.C.

Shneiderman, B., & Plaisant, C. (1990). Chapter 2. In *Designing the User Interface.* Boston: Pearson Addison-Wesley.

United States General Accounting Office. (2002). *Better Mechanisms Needed for Sharing Lessons Learned.* Washington, D.C.

Young, I. (2008). *Mental Models: Aligning Design Strategy with Human Behavior.* Brooklyn: Rosenfeld Media, LLC.