Title
Design of a Data-Driven Micro-Display for Situation Awareness in Bursty Environments (When Not Much Is Happening Most of the Time)

Permalink
https://escholarship.org/uc/item/7p12r63s

Author
Holloway, Alexandra

Publication Date
2015

License
CC BY-SA 4.0

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA
SANTA CRUZ

DESIGN OF A DATA-DRIVEN MICRO-DISPLAY FOR
SITUATION AWARENESS IN BURSTY ENVIRONMENTS
(WHEN NOT MUCH IS HAPPENING MOST OF THE TIME)

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

by

Alexandra Holloway

June 2015

The Dissertation of Alexandra Holloway
is approved:

___________________________________________
Dr. Darrell D. E. Long, Chair

___________________________________________
Dr. Alex Pang

___________________________________________
Dr. Lynne Cooper

___________________________________________
Tyrus Miller
Vice Provost and Dean of Graduate Studies
# Table of Contents

List of Figures vi
List of Tables viii
Abstract ix
Dedication x
Acknowledgments xi

1 Introduction 1

2 How the Stamp Caught Fire 6

3 Review of Literature 15
   3.1 User-Centered and Participatory Design, Design Research, and Design Synthesis 15
   3.2 How Automation Affects Awareness 18
      3.2.1 Norman’s Cycle of Perception and Action 18
      3.2.2 Mental Models 21
      3.2.3 Situation Awareness 22
      3.2.4 Automation and Human Out of the Loop 24
   3.3 Visual Display of Information 26
      3.3.1 Data Ink, Chartjunk, and Clutter 26
      3.3.2 Partitioning Space and Designing for Glanceability 29

4 Methods 31
   4.1 Domain Selection 32
   4.2 Design Research, Validation, and Synthesis 34
   4.3 Bias in Data Collection 35
   4.4 Data Collection 39
5 The Perfect Analogy: Why It Is (And Isn’t) an Issue of Situation Awareness

6 Field Work Resulting in Findings About Operators’ Work
6.1 Five Phases of Operation
6.2 Operations Work Is “Bursty”
6.3 The Sequence of Events (SOE) Forms Operator SA Needs
   6.3.1 Reading Through the SOE To Construct a Mental Model
   6.3.2 Marking the SOE to Verify, Project, and Follow Along
6.4 Displays Are Pre-Set Based on the SOE
6.5 Situation Awareness for Link Control Operators
   6.5.1 Level 1 SA: Determining “Is Everything OK?”
   6.5.2 Level 2 SA: Comparing Actual, Expected, and Predicted Values
   6.5.3 Level 3 SA: Predicting the Future
6.6 Data Requirements

7 Results

8 Discussion and Analysis
8.1 Opportunities for Automation
   8.1.1 Comparing Actuals, Expecteds, and Predicteds
   8.1.2 Deciding What to Do
   8.1.3 Why the Human Will Never Be Replaced
8.2 On Simulation
   8.2.1 Create Experiences, Not Simulations
   8.2.2 Simulate the Data for Data-Driven Simulation
   8.2.3 Don’t Try To Make It Fun (It’s Not)
8.3 Measuring Performance, Workload, and Situation Awareness
8.4 Advancements in Situation Awareness Theory
8.5 Malleable Attention and Underload

9 Conclusion and Implications for Research and Practice
9.1 Implications for Design for Situation Awareness With Bursty Workload
9.2 Design Guidelines

A Technical Appendix: Building the Postage Stamp
A.1 User-Centered Design of a Data-Driven Prototype
   A.1.1 Designing summary information display on paper
   A.1.2 Creating a command language for data-driven display
A.2 First user test (June): Trust in data sources
   A.2.1 Nominal and off-nominal scenarios: MOM and Voyager
   A.2.2 Distractor task
   A.2.3 Results: Trust as related to data validity and reliability (and source)
A.3 Incorporating real telemetry data into the postage stamp
# List of Figures

2.1 Road on the way to Goldstone Deep Space Communications Complex . . . 7  
2.2 Goldstone Deep Space Communications Complex room . . . . . . . . . . 8  
2.3 Goldstone Deep Space Communications Complex workspace . . . . . . . 8  
2.4 Eight simultaneous postage stamp displays . . . . . . . . . . . . . . . . 12  

3.1 Norman’s task action sequence . . . . . . . . . . . . . . . . . . . . . . . 20  
3.2 Norman’s task action sequence with Endsley’s 3-level model of situation  
   awareness and Endsley & Kaber’s categories for automation . . . . . . . . . 27  

4.1 Mixed media interactive postage stamp prototype . . . . . . . . . . . . . . 33  
4.2 Timeline of work, Oct 2013–Nov 2014 . . . . . . . . . . . . . . . . . . . 41  

6.1 Operator workstation at Goldstone . . . . . . . . . . . . . . . . . . . . . 48  
6.2 Panoramic photograph of the inside of the Darkroom, JPL, Pasadena, CA . 48  
6.3 Operator workstation with postage stamp information overlay . . . . . . 51  
6.4 Flow analysis for link control operator tasks . . . . . . . . . . . . . . . . 54  
6.5 Bursty workload: Perceived operator workload for one track . . . . . . . . 55  
6.6 Harold design activity setup . . . . . . . . . . . . . . . . . . . . . . . . . 59  
6.7 Harold design activity marked-up sequence of events (first trial) . . . . . 60  
6.8 Harold design activity result (first trial) . . . . . . . . . . . . . . . . . . . 61  
6.9 Harold design activity result (second trial) . . . . . . . . . . . . . . . . . 62  
6.10 Goal-directed task analysis for link control operations . . . . . . . . . . . 63  

7.1 Postage stamp: Design evolution of a micro-display with information for  
   operators to determine whether a support is “okay” (i.e., performing nom-  
   inally) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 70  
7.2 Tiled postage stamp micro-displays created and used by operator at Gold-  
   stone . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71  

8.1 Norman’s task action sequence, reprise . . . . . . . . . . . . . . . . . . . 73  
8.2 Operator as central between perception/comprehension and projection of  
   future states . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 75
8.3 Automation opportunity: Comparing the user’s projection of the system response to the actual and/or expected real-time telemetry stream
8.4 Automation opportunity: Goals, intentions, and actions
8.5 Automation opportunities at both acting on the system and evaluation of the system state
8.6 Automation opportunities at both the decision- and the data-gathering points, but with additional user feedback
8.7 Automation opportunities at both the decision- and data-gathering points, but with the operator in a monitoring-only role
8.8 Automation drives the entirety of the system

9.1 Design opportunities for bursty environments

A.1 What’s inside a postage stamp?
A.2 Postage stamp state diagram for track
A.3 Postage stamp state diagram for antenna
A.4 Postage stamp state diagram for downlink
A.5 Postage stamp state diagram for uplink
A.6 Postage stamp state diagram for command mode and ranging
A.7 Postage stamp state diagram for two-way
A.8 Distractor task: Asteroid Speedway
A.9 Postage stamp setup for Goldstone visit (June 3, 2014)
A.10 Setup for link control operator user test (September 2014)
A.11 Link control operator user test (September 2014)
A.12 Stamp roll
List of Tables

3.1 Endsley’s three-level model of situation awareness .................. 24
3.2 Endsley and Kaber’s main categories for automation opportunities ... 26
4.1 Characteristics of a complex dynamic system ....................... 34
4.2 General and targeted user-centered design research techniques used in this study .................................................. 36
4.3 Major milestones of data collection ................................... 40
5.1 Deep space link control operators and air traffic controllers have a lot in common ......................................................... 43
6.1 Data requirements for Deep Space Network operations ............... 68
A.1 Postage Stamp User Test: NASA-TLX metrics of mental demand, temporal demand, performance, effort, and frustration ........... 121
Abstract

Design of a Data-Driven Micro-Display for Situation Awareness in Bursty Environments (When Not Much Is Happening Most of the Time)

by

Alexandra Holloway

Operation of a complex system places high task demands on the human operator, who must maintain an intricate mental model highly aligned with the changing system state. Because this continuously-operating system consists of numerous interconnected components with feedback separated in time and space, the expert operator must maintain situation awareness to be able to provide reliable real-time monitor and control of the system. One instance of this complex environment is the Deep Space Network, a sophisticated worldwide network of telecommunication equipment used by NASA. Operators integrate distributed system information in real time, where operational failure can harm personnel, equipment, mission-critical data, and/or spacecraft, resulting in an operational challenge of situation awareness. In-depth research in operability led to the characterization of the operators’ “bursty” work environment in which long periods of low cognitive demand were punctuated by bursts of high activity. To address situation awareness in the bursty environment, I applied user-centered and participatory design techniques to elicit information requirements and build a data-driven micro-display called the Postage Stamp, paying specific attention to enabling effective transition between low and high levels of workload. In the field, operators grasped the Postage Stamp’s function and utility, and memorized it for future use. Based on the experiences, the contributions of this dissertation are a new type of display, driven by data and designed for the operator’s situation awareness needs; and design guidelines for building a display to increase situation awareness in bursty environments.
Dedication

I dedicate this thesis to my grandmother, who asked me daily for seven years:
“Are you done yet?” Finally the answer is Yes.
Acknowledgments

This work was supported in part by NASA Jet Propulsion Laboratory and California Institute of Technology in Pasadena, CA under the supervision of Caltech Institutional Review Board (IRB) and by the President’s Dissertation Year Fellowship (2014–15).

This work could not have been completed without the unwavering support of my thesis advisor, Dr. Darrell D. E. Long. Darrell has guided me through various stages of my career for the past decade as a mentor, colleague, and friend. When things seemed dark, Darrell would be the one to turn the light back on. I would like to thank Dr. Lynne Cooper, who started as my writing mentor and became my dissertation mentor at JPL, for moving this project from an idea to a completed work with persistence and appropriate criticality. Lynne helped me find the information in the data. I thank Dr. Alex Pang for serving on my dissertation reading committee and providing thoughtful, thorough feedback in the months leading up to the completion of this work.

I would like to thank my colleagues at the Jet Propulsion Laboratory in Pasadena, CA who provided access to the domain, designed with me, and taught me about Deep Space Network operations: my colleagues Dr. Scott Davidoff, E. Jay Wyatt, and Mike Levesque; and my collaborators at the Deep Space Network operations facilities including Johnathon Munoz, John Mason, and all of the operators who generously donated their time to explaining their work and sharing their ideas with me. This work could not have been completed without the technical contribution of Bryan Duran and the design sessions with Jesse Kriss, Erin Murphy, Garrett Johnson, Guy Pyrzak, and Jon Blossom.

I would like to thank the faculty who supported me early in my academic career. Dr. Sri Kurniawan at UC Santa Cruz supported me in the first years of my thesis research, when I was still learning about user-centered design and human-computer interaction. When I was an undergraduate, Dr. Richard Hughey inspired me to pursue
my doctorate. I am grateful for having worked in his Kestrel massively parallel processor lab as an undergraduate researcher. That was a taste of the future.
Chapter 1

Introduction

Operation of complex systems poses an important challenge of maintaining operator situation awareness. A complex system is a sociotechnical system that requires the integrated operation of hardware, software, people, resources, communications, and data, with some or all of these components distributed and/or automated, to perform precision research, science, engineering, and technology functions. Each component presents its own challenge of complexity, and the integration of them results in interconnected relationships. The system and its components is real-time, meaning that the state of the system changes over time, such as a result of changing data coming through the sensor sources. The system is dynamic, meaning it cannot be paused. A large number of different sensors and data sources commingle, creating causal relationships between system variables, the relationships of which may be distant in time and space. What’s more, the systems must remain operational at all times.

Operating such a system requires the operator to have a deep understanding of how the system behaves as the whole, how each variable affects the others, and the effects of automation on the system. Thus, only highly-trained, expert operators provide support for these complex systems. Part of the operator’s training involves forming and learning to maintain a mental model of the system. A mental model is an internal
representation of the operator’s understanding of how the system and its components work together. Certification programs and continued training ensure operator expertise, and procedures and safeguards help prevent failure.

The fidelity and correctness of the operator’s mental model influences how well he or she can predict what the system will do next. In order to maintain a correct mental model, operators must remain cognizant of real-time information. How well they can perform their operational duties correlates with the amount of situation awareness. Situation awareness is defined loosely as having three levels: perceiving and system’s real-time data and state, comprehending what the system is doing, and making predictions about the system’s behavior going forward [16].

Though the systems are complex and repercussions for user error are dire, we routinely rely on these systems. When designed well, they work perfectly most of the time. Situation awareness is a research area in which people study the effects of a system’s design on an operator’s perception, comprehension, and prediction of the system’s current and future state. The challenge of designing for situation awareness for operators working with complex, real-time, dynamic systems exists across many domains. For example, air traffic controllers must monitor and give direction to aircraft flying over a specific sector, and coordinate handovers as aircraft move between sectors [74]. Pilots of the aircraft must pay attention to dozens of variables related to the airplane’s proper function [10]. Nuclear power plant operators must maintain the optimal operating conditions for the reactor [55]. In these domains, the operator must perceive, interpret, draw conclusions from, make predictions about, and corroborate hundreds of data sources in real time with limited information.

Roughly speaking, the complexity of a system is the amount of time and effort needed to describe the system to someone else [25]. A complex system consists of a large aggregation of hardware, software, and human behavior [61], making it a sociotechnical system in which humans work together with the technology to produce results. This
system has a great variety of states it can assume [61]. The system has feedback loops between system variables, which in turn have causal relationships which may be distant in time and space [28]. The system typically has partial automation to remove the rote work from the operator [51, 52] and relies on data from a variety of sensor sources [8].

The tasks operators perform as part of the function of supporting operations of complex systems are highly technical [18], involving an intricate operator mental model of how the system works. Operations requires timely decision-making [18] in a high variety of task situations [61]. The tasks themselves are primarily physical or perceptual tasks [18], often with required simultaneous secondary tasks [50]. Often, operators are under high task load [50]. Operator error in this class of complex systems may result in the loss of life or valuable equipment. The work that these operators do is intellectually complex, and operators need to understand both the domain of the system they operate and the system itself. These systems operate under high stakes. Error at any point can have devastating consequences.

The Deep Space Network (DSN) is an example of such a complex system, having many different components, each of which is individually complex and interacts with others in complex ways. The system has been operational for 50 uninterrupted years while having incremental changes and upgrades. Each component of the system has been crafted, not mass-produced; even equipment which must be identical has idiosyncrasies that an operator must accommodate. This makes an operator’s fluid mental model of how the system behaves and evolves difficult to codify.

The DSN provides uninterrupted communication to the world’s deep space and radio science missions, both within our solar system and beyond. A network of antennas and telecommunications equipment placed around the globe send and receive data to spacecraft such as Voyager and the Mars rovers like Curiosity. Operators monitor the DSN’s equipment around the clock to ensure proper functioning of all of the components of the system. The operator’s on-shift workload is minimal, such that most operators
pursue secondary activities like learning a new programming language, reading, or socializing. But the minimally-intense work is punctuated by bursts of demand. I call this a bursty environment. An operator may experience bursts when multiple spacecraft rise in the sky at the same time and require support, or when a single spacecraft has a particularly complex support need.

The DSN is a high-stakes system: it is a unique asset with sites distributed around the globe, requiring collaboration between three nations. This work fulfills the objective of coordinating precious data between Earth and missions located in deep space, as well as performing radio science experiments and observations, and furthering low-signal telecommunications as a field [67]. The entirety of the sociotechnical system, including each operator, is under immense pressure to perform reliably. Errors can place multi-million dollar equipment in jeopardy, and can result in the loss of irrecoverable science data sent by billion-dollar missions. However, the program is under budgetary pressure to operate with lower cost. My entry into this domain began when the DSN requested applied research to see if there were ways to improve cost effectiveness without negatively impacting reliability.

This research focuses on the human-computer interfaces for the operators, the user interactions and experiences (UX) when using the system. Highly-trained expert operators staff the three operations facilities around the clock using specialized software to provide communications support. Operators monitor the equipment telemetry and control the system using over thousands of individual displays. To complicate matters further, future plans of daytime-only operations with remote control call for increasing each operator’s workload without affecting reliability and performance.

This dissertation describes the creation and introduction of a new type of micro-display called the postage stamp, addressing situation awareness and based on a design analogy to air traffic control. My team and I used data-driven prototyping techniques to determine requirements and iterate on the design. Operators at multiple DSN
operations facilities contributed to the design, created prototypes, and evaluated the micro-display, consistent with a participatory design approach [60]. Operator scenarios helped to establish a realistic experience of simulation of the system.

This dissertation contributes to the area of computer science. Specifically, it introduces the concept of a micro-display for situation awareness in bursty environments (when not much is happening most of the time) for expert operators performing monitoring tasks and assessing system state. The research integrates concepts from the fields of human factors, human-computer interaction, design of information displays, and computer-supported cooperative work. I describe the work leading up to the postage stamp’s design, including the problem statement, field work, and anthropologically-situated design methodologies.

This dissertation introduces new technologies designed using user-centered methods, grounding those technologies in theory, and incorporating participatory methods with domain operators performing specific tasks in a laboratory environment and in the field, and applying new theories to the lessons learned in the user-centered design work with operators.
Chapter 2

How the Stamp Caught Fire

Managers of the Deep Space Network (DSN) presented the problem like this: Link control operators, the men and women responsible for connecting to spacecraft in deep space and sending and receiving data, pay attention to a ton of things – and, with the next upgrade to operations, their attention will divide further. Over the next several years, operators will go from watching one or two antenna-spacecraft connections at a time, with the equipment local to their operations facility, to watching three or more connections, some of which may be remote. Is this a good idea? How does the DSN maintain reliability and performance?

I started by observing link control operators in their natural habitat. I used ethnographic and anthropological techniques: I became an ethnographer, embedding myself in the link control operators’ workflows, habits, and terminology. It took a long time to learn the language of link control operations. Now I talk shop with the operators in the Darkroom or at Goldstone or Canberra about receivers and exciters and the optimal measured carrier power for a specific antenna-spacecraft connection.

Driving out to Goldstone from the Jet Propulsion Laboratory with my colleagues Scott and Jesse took three hours. As we drove, slowly the foliage of Los Angeles disappeared and the landscape of the barren Mojave desert stretched out to the horizon.
Figure 2.1: “Goldstone” is visible on the side of a hill in the Mojave desert as one drives up to the Goldstone Deep Space Communications Complex. One of the site’s five working antennas is seen in the foreground.

We entered a military base. One yellow caution sign directed civilian vehicles to yield for tanks. Other roadside signs warned vehicles about turtles and burros crossing the road: both animals come to the base in search of water. We passed massive round antennas, each on a high pedestal, as we wound through the base. Figure 2.1 shows the landscape of the barren Mojave desert near Barstow, CA en route to the Goldstone Deep Space Network complex.

The Goldstone Deep Space Network complex lay nestled in a valley between rolling San Bernadino mountains. The operations facility, adjacent to the giant 70-meter antenna, housed all operators. It took another thirty minutes of driving to arrive after entering the base. The other four antennas stood scattered around the base. You could only walk to two of the antennas from the operations room.

Inside the operations room, five operators sat at computer workstations with six monitors stacked two high. Where once one station controlled one antenna, now, software allowed the spacecraft-antenna combinations to be controlled from any of the workstations in the room. Figure 2.2 shows the inside of the Goldstone Deep Space Communications Complex room, where link control operators provide support to the world’s deep space missions. And Figure 2.3 shows an operator at work behind one of the consoles.

The operators I shadowed had an uncanny sense of time. They knew exactly when the next event would happen, without even looking at a clock. Between events, they were relaxed and talkative. They told stories and showed off personal projects.
Figure 2.2: Goldstone Deep Space Communications Complex room, in which an operator sits behind a semi-circular desk, facing the front of the room.

Figure 2.3: An operator working at the Goldstone Deep Space Communications Complex.
They argued and debated. They explained what they do when things go wrong, how the organization shifts to accommodate an anomaly and to resolve it as quickly as possible. I asked many questions about what they do when things go wrong. A lot of it was similar to what astronauts do when something goes wrong aboard the International Space Station. They collaborate with each other locally and between their site and the Darkroom, where their supervisors sit, on a solution. Many pairs of eyes shift to troubleshooting and research. Communication between operators and the Darkroom supervisors becomes crucial. We talked a lot about how operators anticipate what will go wrong, how they address it, and what to do afterwards.

It occurred to me that we were spending a lot of effort thinking about what operators do when things go wrong. But most of the time, things are going swimmingly. Things are nominal most of the time. An operator’s work is bursty. An operator bursts into action to set up the support. This is important. This is where the operator selects the right program to run, checks all parameters, and prepares for the support by learning what to expect in the coming hours. Once the track is set up, there is a lot of passive monitoring. If everything is configured correctly and the project makes no changes to the schedule, the track supports itself. At mode changes, the operator pays attention and makes sure that the telemetry shows the support getting back into a lock state. Then a lot of nothing again. Then the support is over, and the operator tears everything down, running a program to stow the antenna and cool the transmitter. So, most of the operator’s day is watching things go well.

So then I asked a very simple question: “How do you know everything is okay?” I was looking over an operator’s shoulder as he stared at six monitors’ worth of tiny numbers and colors. It was like looking at a pointillist painting: greens and oranges and reds and purples in dots, on a grey background, with small black numbers, all stationary around the operator’s head.

The operator began explaining. Here, he said, you know the carrier is in lock,
so you look at the subcarrier, and finally, at the symbols. “You’ve got telemetry,” he said, and concluded that downlink was all right. Then he worked through the uplink in the same way, pointing out things that made each track, each support unique. If you have this spacecraft with this antenna, you expect this result. But if it’s another spacecraft with another antenna, you will see something different.

“Wait... What are you looking at?” I asked. The operator pointed all around the six screens. There and here and over here too, and sometimes here.

“It’s not just about the state indicator. It’s about the values on these signals,” he explained. An operator looks at the numbers for specific signals, and does some mental math, calculating whether or not the signal shows nominal values. It’s the same thing the software display does, but smarter. The software display mainly calculates whether a telemetry channel value is within a static range. But the operator is taking into account everything he or she has ever seen or heard about this specific support combination. The operator takes into account history, the environment, and a wider range of inputs.

The weather plays a huge role. “Your signal has to get through all this atmosphere,” one operator said, pointing at the wall where a window could have been. “The more atmosphere, the more moisture. The more moisture, the worse the signal. And you just have to keep that in mind when you look at this number here. It’ll get worse if there’s clouds or rain.”

Each operator I talked to had been doing this forever. They have been here, operating the Deep Space Network, for twelve years. Twenty years. After the Air Force. Hired just out of college, in the ’80s. To them, looking all around and knowing where to look is as simple as commuting to work every morning. There is no way a new operator could pick this up quickly.

“We train new guys for six months to a year,” an operator barked as I watched him work. He described the on-the-job apprenticeship-like training provided to five
LCOs – none of whom worked out. “One guy, we trained up, and two months later he left. Maintenance. It pays more, and it’s better than this.” He laughed wryly, looking at his monitors, at this and that and over there too.

What if all that stuff were grouped together? If you could look at a single object and know if everything is okay? I imagined the stamp to be supplementary to whatever the operators were using at the time. I imagined a future in which an operator monitored all supports going on at the same time, worldwide.

Then I began designing what would become the stamp. We called it the postage stamp because of its size and aspect ratio. It looked like something you could scale way down, print out, and affix to an envelope. In the design process, I identified the subsystems (uplink, downlink, and ranging, at first, to match the operators’ program options), chose colors (muted versions of the bright color scheme in existence), and began piecing these together into a design. I tried arrows. I tried a globe. I tried with iconography. I scrapped all of that. And then the stamp was born.

The initial stamp design hinged on three constraints:

1. Data drives the stamp. Moreover, the right data drives the stamp. I identified the data to use to drive the stamp, and how it should be displayed.

2. Anyone can see the stamp if they’re facing it, whether they are seated at the station or standing at the back of the room. Another way to codify this constraint is that you can fit a dozen or more stamps on your screen in a small area. To do this I used crisp fonts and design techniques to make each drop of ink on the screen mean something. I removed any non-essential line. I tested the design by standing across the room and asking people to tell me what they see.

3. All the elements must be familiar. I use terms, numbers, and colors that the operators use. That’s one of the reasons I ditched the iconography and fancier designs.
Figure 2.4: I instructed operators, “Narrate what you see here in these Postage Stamp displays.” And they did! Correctly!

The stamp went through a number of iterations when I showed it to my design colleagues – and finally, we presented it to operators. I used data that I cobbled together from various sources, massaged, cleaned, and spoon-fed to the stamp. I defined a command language that the prototype could understand to run the sequence of events.

After a 30-second introduction to the stamp (I pointed out the parts of the stamp, identified numbers used, and explained what the colors meant), I asked the operators to narrate what they see (Figure 2.4).

“I’m going to speed up time – to make things more interesting.” Keeping time constant meant waiting a very long time between events. “Watch the stamp and tell me what you think is happening with the support.”

With just that introduction, with just that prompt, and with no other displays, operator after operator correctly identified the track’s events. Even when a spacecraft dropped downlink lock to go two-way (an expected event that looks like a problem), the operator was right there with the stamp.

Two days later, an operator from Goldstone shot me an e-mail.
“Yesterday I got bored stiff so I made a quick mockup of your Stamp,” he wrote. “I saw that I had a MOM [spacecraft] and you guys love MOM displays so I decided to give it a try. It’s not as nice as yours. I did it from memory and quickly.” He wrote that he tracked two spacecraft, the Mars Orbital Mission (MOM) being one of them, without any other display, and added: “The tempo of the passes somehow was very relaxed ... I can see it succeeding with multiple links.”

Wow! An operator from Goldstone re-created my stamp – from memory – and used it to track! That’s crazy!

On the next operator visit, the operator said that for nominal supports, he would want no other displays open. This exchange occurred:

INTERVIEWER: So you’re saying you would want to track just based on the postage stamps.
SUBJECT: That’s correct.

And now, whenever I see him, an operator who works in the Darkroom here at JPL asks when he can start using the stamp. “Can you make me a mobile version?” he says.

After the last operator visit, I redesigned the stamp to accommodate two key things that were missing from the original design: multiple spacecraft tracked by one antenna, and an antenna tracking multiple objects. I made the changes, and sent them out in an A/B test to several operators to sanity check the design. Then I handed over the design to my colleague who writes all the code to bring the stamp to life.

“I’m really digging the new stamp design,” my colleague said.

Feedback to date has been extremely positive. For example:

1. The operators who were able to narrate the stamp after 30 seconds of training;
2. One operator considered the stamp valuable enough to re-create the design based on his memory of it, and used his stamp to track;
3. That operator now wants to use only the stamp to track;
4. Another operator in the darkroom can’t wait for a stamp he can use; and

5. The software engineer “digs” the design.

And that is how the stamp caught fire.
Chapter 3

Review of Literature

This research builds upon two related research streams: human-computer collaboration to understand the phenomenon observed in the field and visual design of information displays to support the design of an intervention. Prior work in the field shows how automation and visual design affects awareness and situation awareness in human-computer collaborative work. The design of a new interface is User-centered and participatory design methods drive the design research of the intervention.

3.1 User-Centered and Participatory Design, Design Research, and Design Synthesis

User-centered design methods [45] focus the design on the user’s experiences. The user-centered design approach is a cycle of design, prototyping, and evaluating with the user to maintain user engagement and involvement in the product, but user involvement is often limited to the evaluation stages of the approach.

Participatory design is a technique originally used to involve factory workers in the design and construction of their factory [60], and has since moved to the computing workplace to let the users of the tools have direct impact into the tools. In some cases,
users find it hard to see the big picture because they only know what they work on; participatory design fails when the design depends on a user’s restricted vision.

Combining the contextual approach with the participatory and user-centered design approach involves users through all of the stages of the design and provides an understanding of the user’s socio-technical system. Contextual design methods developed by Beyer and Holtzblatt (1999) allow the researcher to investigate the user’s work experiences through targeted interviewing, shadowing, observation, and other techniques [3]. When the users are high-skilled experts in their field, the work of investigating the users becomes more complex. First, the researchers must develop a deep understanding of the users’ actual tools [27]. Next, the socio-technological issues surrounding tool design for expert users complicate a traditional design approach, and require analyzing the political, social, and technological climate [37]. Contextual methods, applied early in the design stages of the work, acquaint the researcher with the user’s perceived and real workflows and may reveal design opportunities.

User-centered, participatory, and contextual research methods lead to designing the right thing for the right person. In this dissertation, I use the term design artifacts to mean physical representations that distill the results of the analysis to aid in discussions with users and stakeholders. Design artifacts can be storyboards, process diagrams, user interface drawings, or prototypes at any stage of fidelity.

Soon enough, the process of design becomes research in itself. According to Faste and Faste (2012) [24], design research includes a) studious design research: research practice that also generates design artifacts; b) formative design research: design of research techniques to evaluate user experience; c) diagnostic design research: research of the design process to advance the field of design techniques; and d) embedded design research: traditional design producing design artifacts, with a contribution to the field of design at large. The work presented here includes all of these types of design research.

Researchers process the notes, sketches, and thoughts resulting from design
research in a technique called design synthesis. Information synthesis in design can be described as “[the] practice of integrating, organizing, filtering, and evaluating external information” [30]. A synthesis technique called abductive reasoning suggests that new conclusions follow from known, yet incomplete data [40]. Schön calls the process of obtaining information and synthesizing the results into new data “reflective practice” [59]. Data synthesis from user-centered design techniques draws on grounded theory to ensure that results arise directly from the users’ data [26, 66].

In this dissertation design synthesis is used as an integrative and generative process commingled: in a single 60- to 90-minute design synthesis session, one or more collocated researchers physically integrated (e.g., by writing down on sticky notes) and generated (e.g., by grouping the sticky notes into themes) ideas from field work, resulting in grounded design artifacts.

Integration: First, one or more researchers aggregate field notes and observations. We discuss routine observations and scenarios, as well as things that we found surprising or contrary to our understanding of the domain. We use sticky notes, whiteboard diagrams, storyboards, and other visual aids to generate a shared understanding of the research. One technique to guide these discussions involves writing the ideas, observations, and topics on sticky notes, with one idea per note. Then, each researcher describes the observation and sticks the note to a large whiteboard. Over time, themes cluster on the whiteboard, which leads us to the generative stage of synthesis.

Generation: Then, we generate new themes from the shared understanding using techniques such as affinity diagramming and process flows. New themes emerging from the collated raw data can unlock new unanswered questions or lead the direction of design activities [59]. The results of the generation could be, for example, further questions, storyboards of problem areas in a user’s process, or an evolving statement of purpose based on the emerging information and synthesis.

Design synthesis drives the direction of design research. As more user sessions
and interactions follow, the design direction is subject to change. New information resulting from additional interviews, observations, or user tests can alter the direction of the project in design cycles.

The ethnographic thread of the research approach embedded me in the subjects’ community as a participant-observer, participating in the activities and interests and observing the community of link control operators both individually [39, 35] and as a socio-technical system together with their tools and equipment. I used ethnographic techniques [63] to understand link control operations work in the context of the observed culture [58], technology, and workflow.

3.2 How Automation Affects Awareness

Understanding the basics of how a user acts on, and interacts with, a system gives us a foundation on which to build more complex concepts, with an eye for designing interactions in a dynamic environment.

In this section, I describe how introducing automation into a complex collaborative system affects the operator’s involvement with the system and awareness of what the system is doing. I discuss the class of problems called human out of the loop and present design principles to mitigate these kinds of problems.

3.2.1 Norman’s Cycle of Perception and Action

Don Norman’s Task Action Sequence provides a framework for understanding how people interact with technology [45]. The sequence, shown in Figure 3.1, consists of two distinct phases: mental activity and physical activity. The user first perceives the environment and tries to extract meaning from it. The user’s interpretation is used to form goals about what to do next. Next, the user constructs intention for action based on the goals, and identifies a specific action sequence that can help achieve the goal. Finally, the user acts on the system by executing the action sequence. The results of
the actions are examined, interpreted, and evaluated against expectations, and necessary modifications in the goals or action sequences are made for the next round of interaction.

According to Norman, there are seven stages of action: four on the human-to-system side (execution) and three on the system-to-human side (evaluation); these then repeat to form a cycle [47]:

— Execution:
1. Forming the goal
2. Forming the intention
3. Specifying an action
4. Executing the action
— Evaluation:
5. Perceiving the state of the world
6. Interpreting the state of the world
7. Evaluating the outcome

Interface designers must pay particular attention to two critical places at which users may experience a disconnect from the system. The first, dubbed the Gulf of Execution, occurs when the user does not know (or cannot imagine how) to perform an action on the system. For example, a user may not know which button to press to result in the intended behavior, or may not know that the system knows about the intended behavior at all. The second, the Gulf of Evaluation, occurs when the user cannot perceive the result of the system’s action (the system does not display the result, or displays it unclearly), or cannot comprehend the result (the system displays the result, but the user cannot make sense of it).

Some authors show that the expert user skips steps in the task action sequence – taking in cues from the environment, reasoning about those cues, and projecting the
Figure 3.1: Norman’s task action sequence (cycle) [45]
expected state of the world given the not-yet-performed actions [2]. For example, when touch typing, the expert typist does not need to pause and make projections or reflect on expectations. Moreover, in a more complex task, the perception of the environment, largely secondary or even tertiary, may lead to decisions yet is hardly at the forefront of the mind.

Activity theory gives us another hierarchical framework for describing how people work. A user might have a high-level goal of an activity, or long-term project. The activity has an object (the main benefit), a motive (what motivates the user’s activity toward the object), and a tool (the thing the user uses to carry out and support the operations and actions). In order to complete an activity, a person has to perform actions. Actions have operations, directly influenced by the conditions of the environment. Generally, operations are easy; they require low cognitive load to complete. [43, 44].

3.2.2 Mental Models

As a user interacts with a system, she creates in her mind a model, appropriately called a mental model, of how the system works [46, 64]. This model reflects the user’s perception of how the system behaves and how it reacts to the user’s stimuli. A mental model becomes a sort of dynamic simulation of the system in the user’s mind, altered by new information from the external system, and acting as a basis for the user’s prediction of how the system will react to changing state and to the user’s input.

This simple concept has profound consequences. With an appropriate, robust mental model which matches well to the system studied, a user can make accurate predictions on how the system will behave to her interactions. For example, the reader’s mental model of a light switch is simple, and complete: flipping the light switch up causes electric current to flow, which lights a light bulb. Flipping the switch down opens the circuit, and the light bulb turns off. The reader can use this mental model to troubleshoot a faulty system. If you flip the light switch on, and the bulb fails to
light, you might diagnose the circuit or the bulb itself. However, an inappropriate or incomplete mental model makes diagnosing the system difficult or impossible. Norman describes his two-compartment, two-control refrigerator circuit, and the two corresponding mental models that could be formed from interacting with the system: one with two separate cooling units (one for each compartment), and one with a single cooling unit but a valve controlling the cold air flow between the freezer and food compartments. He claims it is difficult to tell which model is correct and which is faulty, and thus how to properly set the desired temperature in his refrigerator system [45, 48].

Mental models are difficult to elicit and record. In a complex, dynamic system which changes over time, mental models are inherently incomplete abstractions of the actual system, limited in fidelity and/or scope, difficult to pinpoint and complicated to study [12, 28]. However, authors agree that when a user has formed and maintains a correct mental model of a system, her work benefits: she understands how the system behaves and can make quick, accurate predictions of the system’s changing state.

For an interaction designer, creating an interaction experience that can leverage how an operator perceives and reasons about a system depends on a thorough understanding of how the operator creates and maintains an appropriate mental model.

### 3.2.3 Situation Awareness

The theory of situation awareness drives the interactions and designs described in subsequent chapters.

In order to form the context for the user’s activity, the user must understand the activities of other people and of the system. Dourish and Bellotti define this as awareness [11]. Mica Endsley created the field of situation awareness, defining a three-level hierarchy to the way people understand and reason about their environment. In Endsley’s model of situation awareness, the three levels are summarized in Table 3.1. Each subsequent level of situation awareness necessarily relies on information from the
previous level. A user is said to have a good situation model when operating at high levels of situation awareness – that is, when able to perceive and comprehend the state of the system, and predict upcoming situations and events. In a study of air traffic controllers, Endsley found that 88% of user errors were related to situation awareness. She further categorized these errors and attributed them to perception (Level 1, 72%), comprehension (Level 2, 22%), and projection (Level 3, 6%) [17]. In another study, Endsley found that 23% of voluntarily-reported incidents were situation awareness errors with a breakdown of 69% Level 1, 19% Level 2, and 12% Level 3 [23].

The user’s goals and objectives themselves contribute to the situation awareness. A user’s perception and comprehension of the situation can be helped by system user interface designs tailored to maximizing situation awareness, and projection of the future can be helped by certain automation mechanisms that free the user to think. The three levels of situation awareness are part of the larger decision-making system which, for Norman, was goal setting, action sequence forming, and execution. According to Endsley, situation awareness is one of three major components driving decision-making, with the other two being information about the system and the human-system interface (e.g., system capability, interface design, stress and workload on the user, complexity, and automation), and information about the human (e.g., information processing mechanisms, abilities, effects of training, long-term memory store and past experiences, and automaticity) [20].

Endsley & Rogers (1994) outline a procedure of creating a job task taxonomy, which allows researchers to create a thorough model of a specific job’s situation awareness requirements at each of the three levels [22]. To create the taxonomy, the researcher separates the subject’s top-level goal(s) into subgoals, then further into tasks and actions. At each decision point, the subject requires a piece of information, i.e., a bit of situation awareness.

The taxonomy gives researchers two powerful pieces of information. First,
Table 3.1: Endsley’s three-level model of situation awareness, with the percentage of overall errors for user errors attributed to situation awareness in two studies, and including examples of errors that could occur at each level [23, 17]

<table>
<thead>
<tr>
<th>Description</th>
<th>Level</th>
<th>Examples of Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perception of elements in the current situation</td>
<td>L1</td>
<td>Information display fails at articulating and organizing information</td>
</tr>
<tr>
<td>(Level 1) – 69–72%</td>
<td>L1</td>
<td>Pertinent information is missing or obscured</td>
</tr>
<tr>
<td></td>
<td>L1</td>
<td>Operator forgets information</td>
</tr>
<tr>
<td>Comprehension of the elements perceived</td>
<td>L2</td>
<td>Operator forms inappropriate mental model</td>
</tr>
<tr>
<td>(Level 2) – 19–22%</td>
<td>L2</td>
<td>Operator runs out of time to map the situation to the appropriate model</td>
</tr>
<tr>
<td></td>
<td>L2</td>
<td>Operator relies on default values</td>
</tr>
<tr>
<td>Projection of future states of the system</td>
<td>L3</td>
<td>Operator lacks or forms poor mental models</td>
</tr>
<tr>
<td>(Level 3) – 6–12%</td>
<td>L3</td>
<td>Operator over-projects the future of the system</td>
</tr>
</tbody>
</table>

Researchers can use the taxonomy to create interfaces that cater to specific situation awareness requirements. Second, the taxonomy provides a framework for creating probes (question sets) for situation awareness assessment. In this work, the job task taxonomy provided the requirements for the micro-display, as well as the questions asked of the operators during the user sessions.

3.2.4 Automation and Human Out of the Loop

Complex systems typically have some automation to mitigate the complexity of the system for the operator. Automation is an activity that the system performs for the operator. Sheridan, one of the pioneers in work in human-computer collaborative systems, separates operator’s work roughly into monitoring tasks, in which the operator watches the output of the system and its automated or semi-automated components; and control tasks, when the operator directly acts on the system [61]. Endsley and Kaber (1999) identified four basic domains in operations work (shown in Table 3.2): monitoring the state of the system, generating a solution set given a decision point, selecting a solution based on a specific set of rules, and implementing the solution. In
some cases, automation at one or more of these four domains can make it easier for
an operator to control a complicated, dynamic system. Those cases are a) multiple
competing goals for the operator, such as monitoring several telemetry channels while
responding to anomalies; b) multiple tasks competing for a user’s attention, each with
different relevance to system goals; and c) high task demands under limited resources
[21].

Automation may offer reduced mental workload, yet maintaining a complicated
mental model of how the system operates given the introduction of automation creates
cognitive overhead. Moreover, trusting that the automation will perform the correct
tasks at the correct times is a challenge for operators, especially when the automation
behaviors are hidden or obscured [51]. Under- and over-reliance on automation, in which
an operator might disregard automated activities or alerts or place undue trust in them
causes error in operating the system and has resulted in railroad accidents and pilot
crashes [51].

The greatest risk to the operator’s situation awareness is a class of problems
called the human out-of-the-loop phenomenon: when the user experiences a break in
understanding what the system is doing, and thus fails to properly understand or project
the future states of the system. For example, an operator at a control console with no
automation must pay constant attention to a display. In the case of error, she has full
awareness of what has happened to cause the error and what must be done. On the
other hand, an operator at a semi-automated facility can look away for long periods,
until alerted by an error. But because she has focused her attention elsewhere prior
to the error, she cannot begin troubleshooting right away: first, she must re-create a
mental model of the system, and then she may act on the system.

Experimentally, Endsley and Kaber found that interspersing mid-level automa-
tion with brief periods of manual control has positive effects on the user’s situation
awareness and feelings of being in the loop. That is, pausing the automation and drop-
Table 3.2: Endsley and Kaber’s main categories for automation opportunities [21]

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>MONITOR</td>
<td>Identify and perceptually highlight items of interest based on an algorithm predicting what’s interesting</td>
</tr>
<tr>
<td>GENERATE</td>
<td>Generate a solution set or partial list of potential solutions for selecting the next course of action</td>
</tr>
<tr>
<td>SELECT</td>
<td>Assist in performing a selection (e.g., by presenting confidence metrics) or selecting the best choice based on an algorithm</td>
</tr>
<tr>
<td>IMPLEMENT</td>
<td>Performing the selected action</td>
</tr>
</tbody>
</table>

ping the operator into the console for low-level commands (which could be, but are not, automated) helped operators maintain a mental model of the changing system. In these experimental setups, users’ performance increased, and failure recovery time decreased, with any automation – but the operators’ perceived workload stayed the same.

In Figure 3.2, I combine Norman’s task action sequence, and superimpose Endsley’s three levels of situation awareness and Endsley and Kaber’s categories of automation.

3.3 Visual Display of Information

Presentation of quantitative information is both an art and a science. It is an art because color, layout, font, and other elements of attractive and appealing representations lie in that realm. But it is also a science because cognitive science – the science of perception and cognition – provides a framework for discussing the effective display of information. What’s more, selection of data, along with what makes a representation attractive and appealing, are well-studied issues in visual presentation of data.

3.3.1 Data Ink, Chartjunk, and Clutter

Presenting information clearly, and presenting the right information, results in first-level situation awareness. In the work described in subsequent chapters, I embraced the design philosophies of Tufte and others when designing the intervention, and
Figure 3.2: Norman’s task action sequence, from Figure 3.1 [45] is shown here with the letter N. I map Endsley’s 3-level model of situation awareness onto Norman’s cycle with the letters SA [17], and Endsley and Kaber’s categories for automation with LOA (for Levels of Automation) [21].
I designed it to be a peripheral, glanceable display.

An attractive display presents the maximum amount of necessary information, and requires a minimum amount of the viewer’s cognitive function to process that information. Thinking about the design of such a display, how does one maximize the amount of information while minimizing the amount of thinking? Edward Tufte, one of the century’s leaders in visual display of information (having authored several books on the subject), defines the data-ink ratio as follows:

A large share of ink on a graphic should present data-information, the ink changing as the data change. Data-ink is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented. [71]

More specifically, non-data-ink refers to the area of a quantitative data presentation: it is “chartjunk.” Chartjunk unnecessarily detracts from the data presented, and, under certain philosophies, may confuse the reader or inhibit her comprehension of the data. This interference – either due to too many visual embellishments to the data, or through too much data presented, is called display clutter.

Clutter inhibits the user in perceiving and/or comprehending what a visual display is trying to communicate. It pulls the user’s focus away from the data and onto unnecessary embellishments on the display. There have been some efforts to measure display clutter and the density of features in a display [36, 56] in an effort to guide interface designers to creating less cluttered, more intuitive interfaces. Interestingly, cluttered displays have been found to affect experts more than novices – perhaps because they can process more information and thus spend precious cognitive cycles cogitating over the chartjunk more than do novices [38].

Four basic tenants of creating displays for human-computer interaction work are visibility, a good conceptual model, good mappings, and feedback. Visibility is the system’s exposure of its states and their transitions in a way that the operator may perceive and reason upon. In this way, the operator understands what the system is
doing (or not doing) just by looking. Good conceptual models match the system’s behavior to the operator’s understanding of how it works. The visible system state effectively communicates a correct and complete internal model to the operator. Clear and understandable mappings correlate the controls for the system and the actual system they control. Finally, the system provides clear and consistent feedback to the operator’s actions[47].

3.3.2 Partitioning Space and Designing for Glanceability

Explicit, physical partitioning of space can address some problems of display clutter. In fact, people naturally partition their visual space when working. If a single physical display (such as a monitor) on a computer or computing system is like a one-room house containing all of a person’s belongings, adding displays is like adding rooms to the house. The person now has a dedicated place to cook, to sleep, and to work. Grudin (2001) found that the partitions people typically apply to their displays divide the workspace into primary, time-based work (things that require constant attention) and event-driven work (such as notifications). Participants in his trials on visual partitions considered notifications distracting when they appeared on the primary display, and preferred to confine them to a secondary display, even if it was much smaller. Grudin found that an extra monitor carries the benefit of added visual real-estate – but more importantly, the partition of a second display provides a necessary partition: “space with a dedicated purpose, always accessible with a glance” [29]. Thus, we come upon the definition of a glanceable display: one that provides the user information at a glance. The user should, at a glance, be able to perceive and comprehend the information on such a display.

In a ubiquitous computing environment – where devices and their displays are all around us – it is impossible for a person to give attention to each device. At any one point in time, one device is primary and the others are secondary, or peripheral.
A peripheral display, using terms from activity theory [43], is “any information display that is (1) a tool in at least one activity of its user and (2) is used at the operation level” (that is, at low cognitive cost) [42]. That is, peripheral displays should be providing glanceable information to the user.

Peripheral displays support four kinds of activities. Imagine you are switching between several different tasks – such as making breakfast while waiting for an important phone call. Your primary activity is making breakfast. If you turn your attention away from the pan, your eggs will burn. Your secondary activity is unrelated to completing the primary one. That is monitoring the phone. A dormant, or set aside, activity could be checking your e-mail. Finally, horizon activities are monitored, but on hold temporarily as you focus on your current task. Horizon activities will soon become primary activities [42].

This classification of activity types gives us a framework for applying partitions to displays. Combining Grudin’s, Matthews’, and others’ work, we can assign a single display or display element to the primary activity, and offload secondary and horizon activities to a peripheral display. Matthews et al. identifies three dimensions for design of peripheral displays. a) the scope of the use, meaning the number and types of activities the display supports; b) the class or classes of supported activities, as defined above; and c) the criticality of the information for the user.
Chapter 4

Methods

The work presented follows the practice of a longitudinal case study as published by Yin (2014) [76], in which I studied contemporary events within the sample over more than 14 months of investigation. I researched, designed, prototyped, and evaluated an intervention for link control operators in their work environments. I performed a longitudinal case study using a two-track approach to the investigation: 

a) user-centered and participatory design research techniques with iterative cycles of design and evaluation as a participant, and

b) anthropologically-rooted ethnographic techniques to embed myself in the environment as an observer. I applied design techniques developed specifically for expert users in complex work environments.

Portions of the work described were a part of my job as an interaction designer and software engineer at the Jet Propulsion Laboratory. This included applied research investigating techniques to potentially improve operator performance under increased workload. Work toward this dissertation followed principles of user-centered design and design research to identify a design analogy, investigate cognitive science phenomenon contributing to improved operator performance, and design synthesis resulting in the creation of the intervention (postage stamp). The resulting intervention showed promise among operators who use (or anticipate) the postage stamp in their everyday work.
I chose to use user-centered [45] and participatory [60] design methods. Understanding the operator’s workflows and the kinds of decisions operators face lets the designer tailor the interface to the workflow, and may improve technology acceptance and infusion. Moreover, understanding the underlying cognitive science phenomenon allows the researcher to tap into research in the field to look for design suggestions and criteria for improving performance.

The design and research work followed these steps: domain selection, observation kick-off, several design research cycles converging on a specific design, and simultaneously involving operators on a daily basis in the design and decision-making process. I produced notes and sketches of the intervention both in pen-and-paper format (some as paper prototypes [62]) and digital format. Mixed media allowed interactions to be prototyped before being built. Figure 4.1 shows an early version of an interactive paper prototype.

4.1 Domain Selection

The following factors contributed to the domain selection. The system is a complex dynamic sociotechnical system operating in real time with partial automation and time-critical activities. The users are operators who are deep domain experts performing a high variety of monitoring and controlling tasks requiring timely decision-making. Error in the system or its operation results in grave losses. These factors are summarized in Table 4.1. The Deep Space Network, an international asset owned and operated by NASA, is such a system, providing real-time communications of irreplaceable data to the world’s deep space missions. The DSN contains the most sophisticated telecommunications equipment in the world. Operators of the Deep Space Network are highly trained both before arriving at the DSN and on-shift in an apprenticeship program. The work is intellectually complex and operators are under variable load, with little activity punctuated by bursts of high task load. Each operator must understand the workings of
Figure 4.1: Mixed media interactive postage stamp prototype implemented on paper and whiteboard, and showing the interactions with the postage stamp (center). When the user clicked on a portion of the postage stamp, a secondary display was launched.
Table 4.1: Characteristics of a complex dynamic system as defined in this dissertation, with references, for human Operator tasks and for the System

<table>
<thead>
<tr>
<th>CHARACTERISTIC</th>
<th>Oper.</th>
<th>Sys.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly complex technical tasks [18]</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Primary physical or perceptual tasks [18]</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Required timely decision-making [18]</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>High variety of task situations [61]</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>High task load [50]</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Required simultaneous secondary tasks [50]</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Highly trained operators</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>High-risk scenarios</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Partially automated [51, 52]</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Variety of sensor sources [8]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large aggregation of hardware, software, and human behavior [61]</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Great variety of states the system can assume [61]</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Causal relationships which may be distant in time and space [28]</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Feedback loops between system variables [28]</td>
<td>×</td>
<td></td>
</tr>
</tbody>
</table>

each component of the DSN in order to comprehend the states, state changes, and what it means for the multitude of interconnected variables. Mistakes in the system or in the operations of the system may result in loss of irrecoverable data beamed from multi-million-dollar space missions millions of miles away. For these reasons, the Deep Space Network operators were selected as the subjects for the work in situation awareness of complex systems in bursty environments.

4.2 Design Research, Validation, and Synthesis

Field work included two days of initial observations and shadowing with four operators, a station chief, and a deputy supervisor on-site at Goldstone. Further, I performed shadowing of the operations chief and track support specialist in the Darkroom at Jet Propulsion Laboratory, both of whom had supervisory monitoring roles overseeing operations at the three remote sites.

Iterative testing and feedback validated the design at each point in the study. I used two user-centered research techniques: general and targeted. In conducting general
design research activities, I sought to capture breadth of experience and context of work. Although broad in scope, the activity always began with a research question or some discrepancy in understanding of the subject’s workflow, technology, tooling, or socio-technical or political environment. More targeted techniques elicited specific answers to specific questions, while still situating the user research in the context of the subject’s work. See Table 4.2 for a list of the user research techniques used in this study.

Design research happened in one of four ways: a) in the subject’s domain – i.e., the researcher visited the subject’s workplace; b) in the researcher’s laboratory – i.e., the subject visited the researcher’s workplace; c) Over a teleconference call; or d) over written communication, like e-mail.

Where a single operator was the subject of a test, I used think-aloud and question-asking protocols where appropriate to elicit qualitative feedback about a system under test, or to understand what the subject was doing. Where two subjects were tested together, such as in the “Harold” design exercise, I asked subjects to think through the problem aloud together, and noted the conversation, prompting with questions where appropriate.

Simultaneous with the design research cycles, I maintained ongoing working relationships with five link control operators from three different sites and the Darkroom.

4.3 Bias in Data Collection

Data collection included extensive field notes and photographs. The organization did not allow audio or video recording of the interviews and shadowing sessions at Goldstone. Some video was taken at the user study sessions; otherwise, this report relied extensively on field notes, photographs, and e-mails.

Work with the Deep Space Network was done as part of my job tasks as an employee of the Jet Propulsion Laboratory in Pasadena, CA. My task was to use user-centered design methods to create and deploy a display that would help operators mon-
Table 4.2: User-centered design research techniques used in this study included general (G) and targeted (T) techniques to elicit answers to specific questions in the context of the subject’s work flow

<table>
<thead>
<tr>
<th>Technique</th>
<th>G</th>
<th>T</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shadowing</td>
<td>×</td>
<td></td>
<td>Observing the subject working in the environment; meanwhile, another operator narrates and translates what the subject is doing</td>
</tr>
<tr>
<td>Observation</td>
<td>×</td>
<td></td>
<td>Observing the subject directly, and asking questions when appropriate</td>
</tr>
<tr>
<td>Interview</td>
<td>×</td>
<td></td>
<td>Asking and discussing specific questions with the subject</td>
</tr>
<tr>
<td>Artifact walkthrough</td>
<td>×</td>
<td></td>
<td>Discussing an artifact (screenshot, spreadsheet, code snippet, or another workplace product from the subject’s domain) to understand the artifact’s use, usability, and importance in the subject’s workflow</td>
</tr>
<tr>
<td>Process flow analysis</td>
<td>×</td>
<td></td>
<td>Discussing an information or processes flow diagram compiled from earlier research with a subject.</td>
</tr>
<tr>
<td>Needs validation</td>
<td>×</td>
<td></td>
<td>Validating understanding of the subject’s problems using storyboards as a communication aid</td>
</tr>
<tr>
<td>On-task drawing</td>
<td>×</td>
<td></td>
<td>Asking participants to draw or illustrate new tools, or mark up existing tools to improve their workflow.</td>
</tr>
<tr>
<td>Written word</td>
<td>×</td>
<td></td>
<td>Analyzing subjects’ written detailed technical manuals, which illustrate a specific part of their work.</td>
</tr>
<tr>
<td>Design discussions</td>
<td>×</td>
<td></td>
<td>Presenting design concepts to the subject to gauge degree of conformity to the design vision, and solicit changes to any of the parts of the design artifacts</td>
</tr>
<tr>
<td>A/B testing</td>
<td>×</td>
<td></td>
<td>Asking pointed questions about specific discrete design choices</td>
</tr>
</tbody>
</table>
itor the status of the Deep Space Network.

As a participant-observer embedded in the research, I recognized and attempted to mitigate bias in data collection. Bias is “any process at any stage of inference which tends to produce results or conclusions that differ systematically from the truth” [57].

**Experimenter bias.** Sackett’s identification and categorization of observer bias provides a framework for self-checking the data collection approach: he identified seven stages at which a researcher’s observations may introduce bias into the data:

1. in reading-up on the field,
2. in specifying and selecting the study sample,
3. in executing the experimental manoeuvre (or exposure),
4. in measuring exposures and outcomes,
5. in analyzing the data,
6. in interpreting the analysis, and
7. in publishing the results.

**Subject bias.** When observing someone, the person might change his or her behavior as a result of being observed (i.e., Hawthorne effect). One might mitigate the subject’s bias by observing the subject over longer periods (so he or she becomes accustomed to being observed) or observing remotely (e.g., by security camera). There was no possibility of observing participants remotely due to organizational restrictions. Because all of the subjects in this study worked in an observable operations room. Visitors observed them working on a regular basis. Thus, the possibility that subjects would alter their behavior due to being observed was reduced due to the work environment.

The nature of ethnographic research is that the researcher becomes part of the study. The ethnographic component of the study carried a risk of altered subject behavior. This was probably not an issue because the subjects’ first loyalty was to
their job. It was clear in the research that if a design or intervention did not work, the operators would not use it. In performing the 14-month study, the researcher developed relationships with several subjects. While the relationships may have contributed to more candid discussions of workplace challenges, feedback on the design process or artifacts may have become more positive in order to refrain from hurting the researcher’s feelings.

Mitigating the risk of bias. The greatest sources of bias, from Sackett’s taxonomy, included specifying and selecting the study sample; design research of the intervention (executing the experimental manoeuvre); measuring the outcomes; and analyzing and interpreting the data. Reading-up bias was not an issue in conducting the research because standard documentation ensured the same exposure to the operations environment and toolset. Result publication follows traditional double-blind peer review to account for bias.

Some subjects selected themselves for participation by volunteering to provide information for the study. Stakeholders selected other study participants based on job performance. Yet other subjects were participants by virtue of scheduling – for example, they happened to be on shift during research observation activities.

As this was a qualitative study with a strong ethnographic component, design research and outcome measurement included inherent bias. In performing design research, one temptation could be to follow one outspoken subject’s recommendations. Frequent, methodical engagement with subjects and with primary and secondary material (such as technical manuals) helped mitigate the risk of bias in the largest part of data collection. I consulted multiple subjects, performing the same or different design techniques to elicit different perspectives on an issue. After performing design synthesis by myself or with other designers, I presented the work to multiple subjects for discussion and review. Moreover, I corroborated qualitative data, such as interview notes, with quantitative data, such as aggregate usage logs. Thus, I addressed the bias by
1. ensuring iterative design cycles;
2. including multiple subjects in each cycle;
3. creating design artifacts and reviewing the artifacts with the subjects to better understand the subjects’ needs, abilities, and opportunities for intervention;
4. situating the design artifacts in the user’s environment (simulated or otherwise) and observing the interactions;
5. performing design synthesis with other designers for alternate perspectives; and
6. consulting primary and secondary material.

4.4 Data Collection

Figure 4.2 shows the timeline of work. I was presented with the problem in October of 2013. Field observations occurred in November of 2013. In addition to the field observations, between November 2013 and January 2014 I performed in-context investigations with operators in the Darkroom, and developed an analogy to determine the theoretical work space. The analogy, that Deep Space Network operations is like air traffic control, drove the project in the direction of designing for situation awareness, and especially for an “is everything okay?” display for glanceable answers of whether a support (or a collection of supports) is behaving nominally. Data gathering and problem refinement continued between December 2013 and February 2014. I created the initial concept of the postage stamp following analogy creation, and iterated on the design and implementation of the stamp, interleaving each version with evaluation with operators, between December 2013 and September 2014. I used simulated data for the first batch of concepts, and real data between July and December 2014. Operators from the Goldstone complex came for two major user tests in June and September 2014, though I solicited their feedback and performed remote evaluations of the design concepts throughout the process. Major milestones for the work are presented in Table 4.3.
Table 4.3: Major milestones of data collection

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 23–25, 2013</td>
<td>Kick-off field study</td>
<td>Observation, shadowing, and interview of operators in-situ at Goldstone Deep Space Operations Complex</td>
</tr>
<tr>
<td>December 9, 2013</td>
<td>“Harold and the Purple Crayon”</td>
<td>Design activity with operators to draw displays as they become necessary at JPL</td>
</tr>
<tr>
<td>December 10, 2013</td>
<td>Initial darkroom tour and observation</td>
<td>Observation, shadowing, and interview of subjects overseeing activities related to tracking spacecraft at Darkroom</td>
</tr>
<tr>
<td>June 9, 2014</td>
<td>Goldstone operator visit and user test</td>
<td>Operators from Goldstone came to JPL for interviews and user testing of the design and prototype</td>
</tr>
<tr>
<td>September 9, 2014</td>
<td>Goldstone operator visit and user test</td>
<td>Operators from Goldstone came to JPL for interviews and user testing of the design and prototype</td>
</tr>
<tr>
<td>October 15, 2014</td>
<td>Brainstorming the future of remote monitoring and control</td>
<td>Canberra operators and others join in a design session to identify areas of opportunity for the future of Deep Space Network operations at JPL</td>
</tr>
<tr>
<td>October 22, 2014</td>
<td>Track support specialist observation</td>
<td>Observation, shadowing, and interview of subjects responsible for overseeing track support activities in all three remote sites at Darkroom</td>
</tr>
</tbody>
</table>
**Figure 4.2**: Timeline of work on understanding and addressing the problem presented in this dissertation, from October 2013 to November 2014.
Chapter 5

The Perfect Analogy: Why It Is (And Isn’t) an Issue of Situation Awareness

Air traffic control seemed like such a good analogy to link control operation for the Deep Space Network. There were tons of similarities, and all signs were pointing to a similar problem space. If there were a problem in one, there would necessarily be a problem in the other. I designed a solution that would address that problem, and it worked. It skyrocketed. And yet, in the end, retrospectively, I couldn’t find evidence that the problem even existed. What problem did I just solve?

When I observed link control operators, back in my first observation session in 2013, I thought about how similar link control operation was to air traffic control. Link control operators track a spacecraft through their portion of the sky. So do air traffic controllers, with aircraft, but on a smaller scale: spacecraft are much farther away from Earth and much faster than aircraft, and the portion of sky link control operators monitor is much larger than that of air traffic controllers. When a spacecraft goes out of range, link control operators transfer control to another station in a handover. So
Table 5.1: Deep space link control operators and air traffic controllers have a lot in common

<table>
<thead>
<tr>
<th></th>
<th>Air Traffic</th>
<th>Deep Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitor the progress of an aircraft/spacecraft through the sky</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Hand over control to another station when the aircraft/spacecraft leaves operation area</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Operator error results in huge losses</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Operators need a lot of training</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Operators monitor tons of data on lots of displays</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Situation awareness is really important</td>
<td>×</td>
<td>?</td>
</tr>
</tbody>
</table>

do air traffic controllers, when aircraft leaves the air traffic controller’s local area. And when a link control operator makes a mistake, the result can be a loss of irrecoverable science data beamed down from deep space. Air traffic control operators’ mistakes can result in delay, injury, or death. Table 5.1 summarizes these commonalities between air traffic controllers and operators.

According to Mica Endsley’s theory of situation awareness, you develop and maintain three levels of awareness: seeing what’s happening (perception), reasoning about what you see (comprehension), and anticipating the future given the reasoning (projection of future states). In my field work, I observed link control operators watching several hundred numeric and color-coded pieces of information, all of which update several times a second. In my research, I read about air traffic controllers struggling to perceive, process, and predict everything going on in their air space as vast volumes of data flow past on their information displays.

If link control operation is like air traffic control, and situation awareness is a problem in air traffic control, it follows that situation awareness may be an issue in link control operation. Certainly the complexity of the environment, the real-time nature of the data, and the serious repercussions in error situations require operator awareness. Since decreased situation awareness affects air traffic controller performance, decreased
situation awareness likely affects link control operator performance as well. Thus, I
decided to design, implement, and test an intervention for situation awareness: the
postage stamp.

When I finished, the stamp took off (described in Chapter 2). But my post-
mortem analysis of the stamp's trajectory revealed something odd.

If situation awareness had been a serious issue in link control operations, I
would have expected to see operators saying, things like, “boy, it sure is hard to keep
all this stuff in my head.” Operators might have recollected a time they did not see
something important happen, did not comprehend what it meant, or failed to predict
what would happen next.

Instead, operators were complaining about the work ethic and training level of
the maintenance staff. Or the humidity in the room. Or the comfort of the office chairs.

Here I had built the stamp expressly designed for situation awareness, but was
finding no evidence of situation awareness issues in the workflow. And yet the stamp
still caught on. Why? What issue did the design address? And what does design for
situation awareness have in common with whatever it is the stamp did for operators?

Operators keep a lot of stuff in their heads. Once, when talking with a link
control operator for over 20 minutes, the operator suddenly stopped, mid-sentence, and
walked over to his colleague’s empty terminal. He did something, then returned, and
completed the sentence. But I could not move on so easily. “What were you doing?” I
asked him. The operator replied, “Remember that 18 minutes ago he asked me to check
on his support? It’s been 18 minutes.” Operators keep perfect time.

What’s more, operators have encyclopedic memory. I asked an operator how he
“locks on” to a specific support and he explained that with this “bird,” it’s different for
each antenna. On one antenna, you will expect to see one behavior. But if it’s raining,
it will behave differently. Each spacecraft, and each subsystem, because they are crafted
rather than mass-produced, are unique. And their uniqueness must be studied, known,
remembered, and later, recalled. Operators keep all this stuff in their heads.

The future is coming. In the future of operating the Deep Space Network, operators will control equipment remotely, relying only on the telemetry sent over miles of wire. The other two sites will be offline, waiting for daybreak when the staff arrive. Keeping one site in an operator’s head may work for now, but maintaining the mental model of the whole network is a daunting, and risk-prone, task.

According to Endsley, maintaining situation awareness includes projecting the future. But the future depends on history.
Chapter 6

Field Work Resulting in Findings
About Operators’ Work

The Deep Space Network (DSN) provides communication between Earth and man-made satellites outside the moon’s orbit in deep space. Operators at three sites – Canberra (Australia), Madrid (Spain), and Goldstone (near Barstow, CA, USA) – handle the transmission, reception, and delivery of this crucial data by monitoring and commanding an array of radio antennas and associated equipment. The positioning of the sites, 120 degrees longitude apart, allows at least one site to see every patch of sky at all times, thus facilitating continuous coverage for any deep space spacecraft that partners with the DSN.

The DSN is operated by National Aeronautics and Space Administration (NASA) and European Space Agency (ESA). Its charter states, “[The DSN project] provides telecommunications products that support solar system exploration missions undertaken by the international community.” The Deep Space Network was formally born in 1965 and has been operating uninterrupted for 50 years.

The requirements for DSN operations force the sites to locate out of town, far from residential areas and areas of radio interference. Voyager 1, for example, is the
farthest-away man-made object, with a round-trip light-time of more than 37 hours. That means it takes 37 hours for light generated on Earth to reach Voyager 1 and bounce back. Where nearby spacecraft such as the ones orbiting Mars scream, Voyager whispers.\textsuperscript{1} High-power transmission (up to 20kW) and low-noise reception of its faint signals require specialized equipment and new techniques to continue supporting Earth’s communication to Voyager.

Inside each Deep Space Network operations facility, operators configure, operate, and troubleshoot equipment as needed for spacecraft tracking and operations. Around the clock two or three shifts of operators pay attention to hundreds of individual pieces of information in order to provide uninterrupted support to most of the world’s deep space missions. Once, one operator would manually control one antenna and related subsystems to provide support to one spacecraft. Now, however, semi-automated processes assist the operator in monitoring, controlling, and/or troubleshooting one or more links: antenna (single or multiple, and their subsystems) to spacecraft (single or multiple) pairs. An operator might have two links: one link consists of an antenna array and two spacecraft, and the other link is just a single antenna-spacecraft combination. The operator performs these monitoring, control, and troubleshooting tasks on his or her workstation inside the operations facility in a windowing environment. The operator’s workstation is shown in Figure 6.1.

The data received at each site funnels through high-speed data connections to the Deep Space Operations Center, affectionately called the Darkroom, at the Jet Propulsion Laboratory in Pasadena, CA, USA. Figure 6.2 shows a panoramic photograph of the inside of the Darkroom. Operations personnel have staffed this room uninterrupted for 50 years, routing information between Earth and satellites in deep space. The Darkroom has been dubbed the “Center of the Universe” – for all of the universal information (that we know about) goes through this room.

\textsuperscript{1}Mars orbiters and landers have a round-trip light-time of between 8 to 40 minutes.
Figure 6.1: Operator workstation in Goldstone Deep Space Communication Complex operations facility

Figure 6.2: The Deep Space Operations Center, called the Darkroom for its low lighting, routes all deep space traffic. In a way, the Darkroom is the informational center of the universe.
Operators ensure that the information transmitted to and from spacecraft reaches its destination. If lost, the data transmitted from spacecrafts in deep space can never be replaced. The signal for which operators are responsible carries critical science data from one-off experiments on a planetary scale; it carries engineering data about the health and safety of spacecraft which can never be repaired. Received signals come at –130dB to –160dB, requiring precise configuration and pointing, and the proper calibration of each and every subsystem. The operator’s job requires a profound understanding of how the Deep Space Network system works as a whole, how each subsystem functions on its own and when connected to other subsystems, an appreciation for radio engineering, an attention to detail, a precise sense of time, and, most of all, situation awareness attuned to the peculiarities of this dynamic system.

It is reasonable to believe that operators’ workload will only increase. Under an initiative called Follow the Sun, operators will staff one of the three stations only during the day shift rather than around the clock, and remote-control all three sites (including their local site). Moreover, each operator will be responsible for three simultaneous supports instead of two.

But what must an operator pay attention to? Currently, the Deep Space Network operations facility sees data rates on the order of tens of gigabytes of telemetry data per site per hour, and that accounts for less than one-fifth of the overall telemetry. Visually distributed, hierarchy-free data fills operator screens to capacity.

From the user-facing side, the lack of screen real estate leads to a loss of situation awareness, and is the only factor which threatens the Deep Space Network’s plan to double the number of supports per link control operator within the next decade.

Currently, operators have an encyclopedic memory of the equipment and spacecraft they support. This memory arises from the system’s unique nature: each component of the system has been individually crafted to fulfill a specific purpose, installed on Earth without interruption, or launched into space for continuous performance. The
operator remembers the previous times a specific set of equipment supported a specific spacecraft, the history of the track, and the potential implications on the future supports. The operator monitors hundreds of telemetry channels for each support, understanding what each channel means and how it relates to the other channels. But with triple the number of antennas under Follow the Sun, relying on perception and memory alone will be insufficient.

An intervention object – a micro-display called the Postage Stamp, or “stamp” for short – provides compact, curated, and low-level information about several key subsystems involved in a support. Following observation and extensive research, user-centered design and development of the stamp revealed its effectiveness at capturing operator mental models. Operators internalized the abstract representation of the stamp, and used the changing information fields to narrate the events being shown. One operator returned to work and built the stamp to the specifications he memorized. Six months later, he continues using the stamp to augment his support displays.

Figure 6.3 shows the information consolidated from the operator’s screen (Figure 6.1) into a single postage stamp, inset in the lower right. The stamp shows the operator consolidated and color-coded system and subsystem information.

In order to arrive at the postage stamp design, I conducted shadowing, interview, and artifact walkthroughs sessions to understand the operator’s work.

6.1 Five Phases of Operation

Following initial field work, I performed a flow analysis following the methods outlined by Endsley (2012) to provide a general framework of an operator’s work process [20]. I found that the bulk of an attended spacecraft support was composed of the following five phases of work, outlined in Figure 6.4.

**Prepare.** LCOs verify that necessary documents and equipment are provided and current. I observed that all LCOs examined a semantically marked schedule of
Figure 6.3: Where does the information come from? The operator station has “is everything okay?” information at a variety of places on the screen. The postage stamp consolidates this information.
events for the expected support. The semantic markings (colors) provided information about how to set up the support in the pre-track phase. Most LCOs printed the schedule of events and marked it with colored highlighters; some LCOs used computer software to create the colored markings. The preparation phase usually takes a few minutes, unless documentation and associated support files are not available. The LCO takes note of the support type offered: whether the support is one-way, two-way, or three-way; whether there will be an uplink, downlink, and/or ranging; whether to expect occultations, bit-rate changes, mode changes, or any other phenomenon that can affect the support.

**Pre-calibrate.** Operators configure and calibrate equipment by running a program that pre-populates hardware parameters from the event schedule and begins equipment warm-up and readiness activities. An important manual verification step ensures that the default parameters and event timings have been ingested correctly. If any real-time changes need to be made, a project representative calls the operator on a reserved radio network and communicates the change to the operator, who then either implements the change immediately in the automation engine, or takes note of the change for later implementation. The setup phase takes under an hour, with the actual time dependent on the support offered.

**In-track.** For most supports, tracking makes up the longest phase of work. The antenna follows the support as it moves across the sky. Expected occultations, unexpected weather, and other expected and unexpected factors can cause the support to drop for seconds or minutes at a time. For expected behavior, I observed during field work that the operators mentally mark the passing of the event (with a comment, such as, “Dropped lock because Mars is in the way – we’ll get it right back”). For unexpected behavior, however, I observed operators troubleshooting the behavior to try to regain lock as quickly as possible. A loss of data for the project requires a report to explain the missing data.
Post-calibrate. After the end of the track (usually as the spacecraft sets beyond the horizon), the operator runs a program to post-calibrate the equipment and stow the antenna.

Report. If any reports need to be filed, for example in the case of loss of project data, the operator helps provide details and delivers the report to a shift supervisor.

6.2 Operations Work Is “Bursty”

Observations at Goldstone and continued field work in the Darkroom showed signs of a “bursty” workload: in a nominal scenario, operators worked in long periods of low to medium-low activity punctuated by occasional bursts of medium- to medium-high activity. Most of the activity occurred in one of the following ways:

1. Forming a mental model of the upcoming support when preparing to track
2. Verifying documentation about the upcoming support with extensive (internal to the operator) historical knowledge of the kinds of support certain projects usually take on
3. Communicating with others about the upcoming support and noting any anticipated changes from the published data
4. Pre-calibrating the automation for human-machine collaborative work
5. If needed, communicating with the ops chief
6. Punctuated periods of more intense automation monitoring, or manual command entry, during mode changes, such as bandwidth, band, and other changes in the support
7. Monitoring the post-calibration of equipment and proper cool-down and stowing
8. If needed, reporting any problems or issues related to the track
Figure 6.4: Flow analysis for link control operator tasks show five distinct phases (steps) in providing support.

These ideas are presented in a generalized workload graphic, shown in Figure 6.5 below. In gauging my perception of the magnitude of the workload, I used behavioral cues. For example, carrying on a conversation, especially if stepping away from the operator’s console implied a low workload; stopping a conversation and focusing on a task for an extended amount of time implied a higher workload. In specific instances of the field work, I noted bursts of task attention when an operator would cut off a conversation mid-sentence and turn in his seat to attend to the support. When he returned to the conversation a few minutes later, he picked up the conversation exactly where he left off.

In periods of low activity, I observed operators learning new skills (such as programming), reading, and socializing. Interviews with other personnel revealed that historically, night shifts were especially prone to long periods of low activity and thus the operators brought extra tasks to keep themselves occupied.

6.3 The Sequence of Events (SOE) Forms Operator SA Needs

For an operator, the published sequence of events (SOE) is the heartbeat of a support, forming the basis of all the operator situation awareness, and determining
Figure 6.5: Perceived operator workload for one track, showing the “bursty” nature of the work. The $x$-axis shows the phases of work as outlined in the flow analysis (Figure 6.4), scaled roughly by the amount of time allocated to each phase of the work. The $y$-axis shows the perceived magnitude of the workload. The graph is not to scale.

situation awareness needs, for a nominal support.

Operator activities with the SOE can be summarized as reading, marking, and following along.

6.3.1 Reading Through the SOE To Construct a Mental Model

In the preparation phase of the required work, one of the main operator tasks was reading through the SOE. This task primed the operators with information necessary to construct a mental model of the support, including:

1. the activities and characteristics of the support (one-way vs. two-way, bit-rate changes, etc.);
2. timings (when events will occur and how long before the next key event);
3. risks (potential for alarms and anomalies);
4. workload (how much operator-system interaction will be required during the support, and whether any external assistance will be required);
5. project interactions (whether a representative of the project will provide more
information for potential changes to the support); and

6. tools (e.g., the required displays).

### 6.3.2 Marking the SOE to Verify, Project, and Follow Along

In field work at Goldstone, I observed operators meticulously marking, with multiple colors of pens, the SOE. Operators explained that different colors represent various events in the support. One color indicates lock events, another color highlights mode or bandwidth changes, and so on. Figure 6.7 shows an example SOE with such markings.

The precise method in which the SOE gets marked varies from operator to operator. One operator explained that the method depends on who trained you. As operators are trained on the job following a long apprenticeship period, trainees pick up the habits of their trainers. Some operators used electronic SOEs with custom color configurations to view the same information as those using the paper versions with colored markings. Marking the SOE served three major purposes, described below.

**Verify.** The SOE’s color-coding helped operators verify the ingested electronic data that would pre-populate the automation engine. For example, when the automation ingests lock times from the SOE, it prompts the operator to verify that the times are correct. The operator looks at the SOE page, scans to the appropriate color, and performs the verification quickly.

**Project.** The act of reading through and marking the SOE served to prime the operator’s situation model about the system. In reading the SOE, the operator formed a mental model of the expected support structure – much like a story of what to expect.

**Follow along.** Marking the SOE helped operators “find their place.” By matching the current time to the most nearly upcoming event, operators would have insight into what to expect the system to do next. Marking events as completed helped
operators find the next event in the SOE faster. Operators are required to split focus
across several goals, as outlined in the job task taxonomy (Figure 6.10). Moreover, oper-
ators currently support up to two (but will support three or more) links simultaneously.
Following along with a support based on the SOE allows the operator to offload some
of the working knowledge onto the page. Returning to a support, I observed operators
mark off events that had occurred and earmark upcoming events. At Goldstone, I ob-
served one operator take over at another operator’s station. The second operator was
going on break, and left the marked SOE on his desk. The operator taking over glanced
at the SOE and determined all events were up to date and accounted-for.

6.4 Displays Are Pre-Set Based on the SOE

I found that operators used a standard set of displays selected based on the
contents of the sequence of events document (SOE).

I observed that the operator used information in the sequence of events (SOE),
along with their experience with supporting certain tracks and equipment, to determine
which displays to use. Two operators providing support for the same activity may
use a different set of displays, based on their training, preferences, and history with the
activity. Furthermore, I observed that an operator brings up all necessary pre-calibration
displays simultaneously. When moving from the pre-calibration to the in-track phase of
support, I observed some operators closing these displays to make space for the in-track
displays; other operators minimized the pre-calibration displays or moved them behind
the more-relevant in-track displays. The same was true for post-calibration displays.

To further explore the issue of bringing up displays, I conducted a design
exercise with my colleague Jesse Kriss. The exercise was inspired by the children’s book
*Harold and the Purple Crayon*. In the book, Harold, a little boy, draws an adventure for
himself. Whenever Harold is in a bind, he draws something that helps him out of the
predicament and furthers the story. For example, when he is hungry, he draws several
pies for himself; when he is full and there are leftover pies, he draws a hungry moose to
eat the leftover pies so they are not wasted [34].

We hypothesized that link control operators could and would draw themselves
the tools they needed based on a given a sequence of events (SOE), in the order of
events that the sequence of events listed. We tested this hypothesis by asking operators
to draw their workspace. We provided a SOE, a blank canvas, and markers; and we
expected operators to read down the SOE, pausing when a new display was needed, and
drawing it on the canvas. We conducted two trials with two operator subjects each.
The subjects would work as a team, talking to each other and discussing the problem
together. We found this to be a more effective method of eliciting qualitative feedback
than the think-aloud protocol because the subjects could use jargon and discuss their
goals and tasks in more fidelity than when explaining to a researcher from outside the
domain. Moreover, the operators were used to working with, and talking to, each other,
and thus the conversation seemed to come naturally.

To set up, I created a drawing surface, and provided colored markers and a
sequence of events (SOE) file. Figure 6.6 shows the setup concept. For the first trial,
I used sheets of paper covering a table and vertical surface such that it was similar in
relative size and shape to the operators’ workstations; for the second trial, I used a
whiteboard. I selected the SOE for its relative complexity (two or more occultations
and/or bitrate changes). I instructed each set of two operator subjects to approach the
blank canvas as if it were their workspace, and to use the SOE as a guideline of how to
build the environment needed to support a track.

Contrary to the hypothesis, the link control operators from both trials a) first
ingested the entirety of the SOE, rather than reading one line and reacting to it, then
the next line, and so on; and b) next drew a standard set of tools, rather than drawing
a custom set of as-needed tools (given each line of support as read on the SOE).

We observed operators read through the entire SOE document prior to making
any markings on the blank workspace. The operators encoded certain keywords on the
SOE with various color of highlighter (see Figure 6.7). Only then, after understanding
the requirements for the support, the operators drew the required tools on the paper.
But even then, the operators drew the standard, rather than support-dependent, tools.
For example, in one trial, one subject reminded the other, “You always need a TDN.
Draw that in.” Drawing a box representing a standard downlink performance display,
the first subject explained, “You always need these, unless you don’t have a downlink
support – but I would have them up anyway.” ²

Figure 6.8 shows the result of the first trial for the Harold design exercise, and
Figure 6.9 shows the result of the second trial.

²One operator later corroborated this observation by writing the following: “Most spacecraft perform
the same activities every pass, every day, every week. We know the spacecraft so we can pull up displays
without an SOE.”
Figure 6.7: Marked up sequence of events as a result of the Harold design activity: colors indicate event types, and checkmarks indicate completed monitoring tasks.
Figure 6.8: Resulting environment from the Harold design activity (first trial): all displays identified by two operator subjects to support a single track given a sequence of events
Figure 6.9: Resulting environment from the Harold design activity (second trial): all displays identified by two operator subjects to support a single track given a sequence of events

6.5 Situation Awareness for Link Control Operators

In order to better understand the situation awareness requirements of Deep Space Network operators, I followed the procedure outlined in Endsley & Rodgers (1994) to create a job task taxonomy [22]. A job task taxonomy allows a designer to focus design efforts on exposing information for situation awareness at each level. I include the result of this exercise, the top-level goal-directed task analysis for Deep Space Network operations, as Figure 6.10.

6.5.1 Level 1 SA: Determining “Is Everything OK?”

I found that the high-level goal for operations was to determine, “is everything ok?”, and ensure that it was. This high-level goal requires certain information in order to understand whether it is being met.
Figure 6.10: Goal-directed task analysis for link control operations, used to expose situation awareness needs. Following Endsley & Rodgers (1994) [22].
Determining what’s happening today. After reading through the sequence of events (SOE), as described in Section 6.3, the operator knows the characteristics of the support: which equipment will be used, the times of various events, details about the support activities, and other factors.

Bringing up displays. Next, the operator brings up displays based on the support needs described by the SOE and validated by the operator.

Determining what to observe. Then, the operator determines the telemetry channels necessary to determine whether the support is behaving nominally.

Observing actuals. Finally, the operator observes the telemetry channels.

6.5.2 Level 2 SA: Comparing Actual, Expected, and Predicted Values

The field work showed most of the level-2 situation awareness needs supporting comprehension involved comparing actual telemetry channels (“actuals”) to their system-predicted values (“expected”) and to the operator projections (“predicteds”).

While both refer to an imagined future, it is important to differentiate between system expected values and operator predicted values. The former, system-expecteds, are values that the system predicts for a specific telemetry channel. For example, in an antenna system, the expected antenna azimuth and elevation can be inputs to the antenna to guide its pointing. On the other hand, operator predicted values come from the operator’s mental model of how the system should behave. For example, in the same antenna system, the operator might have the same – or a different – prediction to what the system expects.

For link control operations in the Deep Space Network, projects generate expected antenna azimuth and elevation values in advance. Thus, the system is oblivious to external factors such as the setting and releasing of brakes and changes due to inclement weather (e.g., wind, which could cause the antenna to oscillate). As the antenna traverses the sky, following its target as it rises and sets, the operator’s predictions and
the system’s expectations are aligned. But unless the system is aware of the effects of the brake on the expected azimuth and elevation of the dish, as soon as the operator sets the azimuth brake on the antenna, the operator’s prediction (the antenna will stop turning) differs from system’s expectation.

As an operator works with the system, she develops a robust mental model of how the system behaves, observing the telemetry values and assessing them against her knowledge of the system with expert ease.

**Actuals vs. system-expecteds.** Some system telemetry channels expose expected values. For example, a file of expected values dictates the motion of the antenna. Calculated times for planetary occlusions and other reasons for degraded or dropped signals are included in the expected events.

**Actuals vs. operator-projected.** Operators perform rudimentary data conditioning: All operators I observed disregarded the noise on telemetry channels which had some noise, characterized by bounce in the tenths and hundredths places of the value. Such a bounce was considered nominal. Thus, the operator’s mental smoothing of the data served as a form of data conditioning.

**System-expecteds vs. operator-projected.** One operator told the story of an antenna array supporting the acquisition of a distant spacecraft. Two antenna from the same facility were supposed to point at the same spacecraft to acquire a signal from it. Each antenna was loaded with predicts – system-expected values for the antenna’s azimuth and elevation – and each showed that the antenna had pointed correctly to the predicted values. The operator projected that the antennae should have similar azimuth and elevation, since they were, after all, pointing at the same object in the sky from just about the same location on the ground. When the operator verified the pointing in the antenna display, he noticed that the antennae were pointed in completely different directions! Troubleshooting the problem, he eventually found that one antenna’s date had been reset, and it was calculating the predictions as if it were the start of the UNIX
epoch rather than the current year. In this case, the operator’s projection of the system state differed from the system’s expectation – again, a mental calculation that may have saved the project’s data.

6.5.3 Level 3 SA: Predicting the Future

I observed multiple occasions across multiple operators when expected events resulting in dropped lock were observed and mentally tagged as nominal. For example, when a spacecraft dropped lock to change bit-rate, the operator observed this event and commented that the dropped lock was expected and would be re-acquired momentarily. Similarly, he noted, when a spacecraft is occluded (e.g., by Mars), a transient lock drop is expected. In off-nominal circumstances, this drop-lock event would be considered an alarm, signaling failure. Here, in nominal circumstances, it was welcomed as a positive affirmation that the support was progressing as planned.

6.6 Data Requirements

Acquiring data of sufficient fidelity and realism posed one of the main challenges in building the data-driven prototype. To start, I investigated the types of data operators interact with over the course of a support, using the tools typical for operators. I categorized the data available to operators on three primary displays that operators used for uplink, downlink, and antenna status monitoring.

- Numerical data, such as antenna azimuth and elevation, and system noise temperature
- On-off states, such as ranging enabled/disabled
- More complex state data, such as green/yellow/orange/red information for alarm management
- String data, such as status messages
Consolidating observations, multiple interviews, and analysis of the operators’ tools revealed that a data stream characteristics include data latency, frequency, fidelity, and subsampling. These are described below, and summarized in Table 6.1.

**Latency:** Data must be delivered within 5 seconds. One operator recalled a story of a previous display, called the Iris, that was rolled out at Goldstone. He said that operators tried to use it but abandoned its use shortly after rollout. This operator and the others would bring up the Iris display alongside their older displays, and see states and numbers changing on the older displays five to thirty seconds before the same state changed on the newer display. The operator called this latency “unusable.” This story indicated the importance of speedy delivery of data.

**Frequency:** Data channels must update every second or faster. In addition to long latency, a channel that has a long delay between data items was considered “unusable.” When presenting one operator with the possibility of 5-second data rates (with one data item arriving every 5 seconds per channel), he snarked, “I could go outside and look at the antenna in that time.” An operator creating his own data displays said his data points were sampled several times a second. Five seconds of delay between data points was considered too long; one second of data separation was acceptable.

**Fidelity:** The ability to see 3 to 5 decimal places for most numerical telemetry channels. Operators described the significance of the calibrated and supercooled equipment. One operator said, “The system feels every fluctuation. Every extra inch of wire adds noise to the system.” He described how the humidity in the operations room, when increased from 20% to 30%, affects the quality of the downlinked signal: the extra moisture creates noise in the lines, and the added noise makes it that much more difficult to extract the signal. When talking about signal strength and noise, every measured decimal place matters.
Table 6.1: Data requirements for Deep Space Network operations

<table>
<thead>
<tr>
<th>CHARACTERISTIC</th>
<th>REQUIREMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency</td>
<td>Less than 5 seconds</td>
</tr>
<tr>
<td>Frequency</td>
<td>Updating every second or faster</td>
</tr>
<tr>
<td>Fidelity</td>
<td>At least 3 to 5 decimal places</td>
</tr>
<tr>
<td>Subsampling</td>
<td>Minimal to none</td>
</tr>
</tbody>
</table>

**Subsampling: Do not restrict the available telemetry channels.** The multiple gigabytes per hour data stream available for data-driven prototyping was identified as being downsampled to one-fifth Hz – that is, for every five samples of telemetry generated by the equipment, only one sample was exposed and saved in the external database. Moreover, I identified several data channels that operators used for each support, yet they were not sampled and stored at all. It is difficult to predict which telemetry channels will be useful in a given support. And that is, in part, why the equipment reports such a huge amount of telemetry every second. For a given situation, a particular subset of telemetry will help an operator diagnose and troubleshoot a problem. Thus, sub-sampling the wide variety of telemetry channels presents a danger for operators. What if the data architect chose the wrong channel to sub-sample? Then the operator lacks sufficient information complete the task.
Chapter 7

Results

The micro-display called the postage stamp was integrated with real-time station telemetry data. The font was selected for readability at a distance (across the room) and in close quarters (on an operator’s display). The stamp can be tiled, multiplying the amount of information an operator sees, without looking busy or cluttered. The color scheme reflects the standard colors familiar to operators. Data driving the micro-display and its transitions were selected via intensive user-centered and participatory methods to identify key telemetry channels.

Figure 7.1 shows the postage stamp design progression culminating in the deployed version. The early concepts showed less information than the modern counterparts and used bright colors. The design was limited to a single spacecraft per antenna. The monospace and single-height font was selected for legibility at a wide range of sizes and distances from the viewer. In the intermediate version of the stamp, more information contributing to understanding support health had been identified; the colors were more muted so that the small multiples would be easy on the operator’s eyes. In the final version, the data were transposed and stamp re-designed so that it could accommodate advanced support scenarios such as two-way coherent communication, more than two spacecraft in a single antenna, and antenna arrays. In this setup, all antenna-spacecraft
Figure 7.1: The evolution of the design of the postage stamp. The primary version was found to contain insufficient information, and information that was preferred by operators to determine support health; the second version was correct in concept, but lacked the flexibility to accommodate more types of supports; the deployed version was capable of displaying all support combinations and was driven by data.

configurations for a specific support can be represented by the single micro-display. The green color was bolded to distinguish it from grey, which had other uses.

When operators from Goldstone came to the Jet Propulsion Laboratory in Pasadena, CA, I provided a quick 30-second overview of the fields on the micro-display. Then, I asked the operators to watch the stamp as data caused state changes, and to explain what was happening in the support. Each operator was able to do this. The procedure was repeated with operators from Canberra and the Darkroom with the same results. The ability to narrate the events on a support after a micro-tutorial speaks to the display’s intuitive interface.

After seeing the intermediate version (Figure 7.1b) work in the user sessions, one operator re-created the postage stamp for his own use at Goldstone (Figure 7.2), continuing to use it for months. The ability of the operator to remember the display,
its layout, behavior, and telemetry sources in the days after seeing it, and the operator’s continued use of the display speak to its potential for technology infusion. Once deployed, it is reasonable to expect easy infusion into an existing workflow at the operations facilities.

I observed the operators choosing to use the postage stamp exclusively for nominal support. In one user test, an operator filled one of his six screens with postage stamps, and moved all other displays off to the side. At the end of the trial, he said he didn’t even bother getting acquainted with any of the other displays. When I asked the operators how they would troubleshoot the anomalous situation that I presented in the simulation, they each said they would use the postage stamp until the anomaly was expected (by the operator) or apparent, and then launch necessary displays as needed.
Chapter 8

Discussion and Analysis

8.1 Opportunities for Automation

For the postage stamp to be useful, operators asserted that the postage stamp must be displaying differences between sequence of events data and station telemetry. That is, the postage stamp should show inferred differences, calculated in real time on the live stream. The postage stamp should be alerting operators about inconsistencies between predicted and actual values. For this, automation which processes real-time events and displays the results in a digestible fashion is key in providing the kind of data operators require.

According to Norman’s task action sequence shown in Figure 8.1, there are seven stages of action: forming the goal, forming the intention, specifying an action, executing the action, perceiving the state of the world, interpreting the state of the world, and evaluating the outcome.

By Endsley’s theory of situation awareness, there are three levels of situation awareness: perception of the environment, comprehension of the system state, and prediction of the future state of the system.

The first two levels of situation awareness, perception of elements in the envi-
Figure 8.1: Norman’s task action sequence [45]
ronment and comprehension of the current situation, are largely covered by Norman’s task action sequence as perceiving the state of the world and evaluating the interpretation of the perception (the world-to-human side of the loop). However, projection of future status is implied, not explicit. When talking about situation awareness, it is imperative to make projection explicit because the fidelity and correctness of a user’s projection in part determines whether the user’s situation model is accurate. An accurate situation model corresponds to good situation awareness.

Rearranging the position of the operator in Norman’s cycle and adding an explicit future state prediction, the cycle becomes as shown in Figure 8.2. There are two explicit cycles: one for interaction with the physical system, and one for interaction with the operator’s mental model of the system.

We move counter-clockwise through the diagram, and begin where Norman began: at goal-setting. The operator sets goals (1) and intends to act (2). The operator specifies actions that would lead to the completion of her goal (3). The operator acts on the system (down, 4) while mentally predicting the expected state of the system (up, 5) and understanding what the future means (6). This is the mental equivalent of level-2 situation awareness: predicting the future is only one step of the process. The user must understand what that prediction means in order to operate on it in the following steps. The system exposes its state information on the right side of the operator. The operator perceives (11) and cogitates (12) on the data received. Then, the operator mentally compares the actual system state to the expected or predicted state (11). This allows the operator to infer similarities and differences (12) between what the operator predicted and understood to be the future of the system (up-path) and the system response (down-path). Finally, the operator modifies her expectations about how the system behaves (13), thereby making adjustments to the mental model of how the system works or to her situation model of the current system state.

It is important to note that the same operator actions route to both the ac-
Figure 8.2: Operator as central between perception/comprehension and projection of future states. Follow the numbers to move around the diagram. N refers to Norman’s task action sequence; SA refers to Endsley’s three levels of situation awareness.

...and the predicted systems, and the operator reasons about the outputs of each system before proceeding. The projected future of the system is only as accurate as the operator’s mental and/or situation model of how the system works.

Visualizing the system in this way, it becomes evident that now there are two clear opportunities for automation. On the right side, helping the operator compare the system to the projected future, and on the left side, helping the operator make decisions or act on the system.

8.1.1 Comparing Actuals, Expecteds, and Predicteds

Norman may call the right side of the cycle the evaluation stage of human-system interaction, or Endsley may refer to as perception and comprehension. See Figure 8.3 for a visual description of a scenario in which automation is applied at the
evaluation stage. See Section 6.5.2 for fieldwork supporting this analysis.

For an operator, it is important to see the deviation between the system’s actual telemetry, the system’s expectations, and the operator’s own projections of how the system should behave. Thus, there are three opportunities for automation to benefit the operator.

**Actuals vs. system-expecteds.** According to Endsley, good design exposes second-level situation awareness information directly, supporting comprehension rather than perception alone: Exposing the comparison of actual, real-time telemetry to system-expected values removes a computation step from the operator by off-loading the computation onto the system.

**Actuals vs. operator-projecteds.** Imagine an automation step taking the place of operator prediction by performing the same sorts of projections and analyses that an operator might perform mentally. This may require the operator’s mental model of how the system works to be closely aligned with the reality. If the system monitors the operator’s input, the system could build a model of how the operator is reasoning about the system. For example, when an operator issues the command to move the antenna to point, the system can infer that the operator will expect the azimuth and elevation numbers to change in a prescribed manner.

**System-expecteds vs. operator-projecteds.** This automation may include the following methods.

- Data conditioning, such as to remove noise, interpolate missing data, or retransmit missed data
- Event processing to flag or transform live data streams
- Post-processing to determine edge cases and flag potential errors

Event processing could involve correlating actual telemetry data against mathematical relations or historic data to show expectations on the system. In this way, the complex-
ities and interconnectedness of the variables in the system are codified against previous values or against relations to each other or to time.

In Deep Space Network operations, operators must understand the relationships between telemetry channels in the real-time system. Imagine that an operator expects a certain telemetry value, such as number of packets transmitted, to be monotonically increasing. An event processor monitoring the real-time telemetry stream would be cognizant of this fact, and could show the operator that the expectation (of increasing packets with each time-tick) is being upheld. Furthermore, operators rely on extensive historical knowledge of supports — for example, the spacecraft’s interactions with equipment, and weather effects on data throughput. That is what operators do when they mentally cross-reference one day’s support with other supports of the same characteristics, drawing parallels and making predictions based on memory. Similarly, an event processor could be monitoring the noise floor for a specific support. The event processor would assess the history for the support on the specific equipment, and note that the noise floor typically stays within a certain range. If, then, the noise floor were to go beyond that range, the event processor would flag the noise floor as non-conforming.

8.1.2 Deciding What to Do

Norman refers to the left-hand side of Figure 8.4 as the execution stage. In this stage, the operator sets goals, intends to act, specifies the action sequence, and performs the action(s). In the model described here, the operator’s action sequence goes to two places: as executions, acting on the system (down), and as inputs into the operator’s mental model, with which the operator can make projections on the system’s response. The operator then receives responses from the system and compares those system responses against her own understanding of how the system should behave. This forms the basis of the operator’s next set of goals, intentions, and actions.
Figure 8.3: Automation opportunity: Comparing the user’s projection of the system response to the actual and/or expected real-time telemetry stream
Using Norman’s task action sequence, these – goals, intentions, and actions – are three distinct opportunities in which automation could help operations.

**Goals.** In the Deep Space Network, the overall goals of the operator are well-understood: Provide uninterrupted support to the world’s deep-space missions. However, the way the operator interacts with the system to achieve those goals could be improved with well-placed automation. For example, abstracting the command suite to goal-level commanding rather than commanding at the hardware level could release some of the operator’s working memory to other tasks. Aiding operators in issuing commands (for example, with auto-completion, fuzzy search, or suggestions) may help, for much of the operator’s situation awareness is off-loaded in training manuals, documents, logs, and notes.

**Intentions.** Endsley and Kaber (1999) [21] described a system that could give the user suggestions on what to do next. Operators would be given computer-generated suggestions along with probabilities of success. Or, the system could create and/or maintain a queue of the next commands to be executed.

**Actions.** In the Deep Space Network, operators rely on a tool that executes a time-separated sequence of commands, called blocks, in order. If a block completes with an error, the operator must manually re-issue it; if the system must be reconfigured part-way through a block’s execution path, the operator must halt the block and re-instantiate its execution. An opportunity for automation here would afford the tool the ability to detect when a block requires execution or re-execution; and rearrange commands in the sequence based on the state of the system in the case of necessary reconfiguration.

8.1.3 Why the Human Will Never Be Replaced

Combining the automation opportunities on the execution-side of the operator (right) with the ones on the evaluation-side of the operator (left), we arrive at the
Figure 8.4: Automation opportunity in which helping the operator with goal-oriented commanding, intention-setting, and action specification
Figure 8.5: Automation opportunities at both acting on the system and evaluation of the system state

scenario shown in Figure 8.5. In this scenario, the operator receives the calculated results of the actual, expected, and projected system state, and decides what to do next, acting on the system in collaboration with some automation.

In some cases, it seems dangerous to let an automation system decide what kind of feedback the user sees. It can be argued that the operator should see everything the system is doing, including actual telemetry values, expected values, and results of the calculations. Such a scenario is shown in Figure 8.6. The benefits to more information include a potentially more robust operator mental model of the system; however, showing the operator too much data may overwhelm the operator’s capacity to filter the important and relevant from the irrelevant or less-relevant data.

One possible outcome of this scenario is that the operator moves to a more supervisory role. In this case, the system – armed with a multitude of data, feedback,
access to its own history, and information about relationships between variables – monitors itself and acts on its own information in nominal situations. The operator serves as an override, identifying and/or acting in off-nominal cases.

Perhaps this allows us to take the human out of the loop entirely (Figure 8.7) and settle for a monitoring-only supervisory role, devoid of even the ability to cause system override. But then why have the operator at all? Figure 8.8 has none – the fully-automated system drives itself.

But can the human ever be replaced? The human operator is an expert at understanding and manipulating the complex system. As a system becomes more complicated, the human operator must become more skilled in order to tackle the nuances of the system's behavior. This is the fundamental conflict of balance at the core of every interaction problem between the human and a complex, dynamic system. The
Figure 8.7: Automation opportunities at both the decision- and data-gathering points, but with the operator in a monitoring-only role.
complexity of the system drives the necessary skill of the operators; and the skill of the operators enables a more complex system to exist.

The inverse relationship between the skill of the operator and the complexity of the system affects a) recruitment, training, and retention of operators; b) the design, validation, and acceptance of interfaces for interacting with the system; and c) the fragility or robustness of the system. That is, a system that is highly robust requires a less-skilled operator than one that is highly fragile; and reducing the skill level of the operator allows for a more complex system. A simple system – one lacking in interconnected variables, time dependence, and collaboration between agents – requires little from operators.

In order to accommodate operators of reduced skill, the complex system must present as sufficiently reduced in complexity. The artificial intelligence must be capable of a) capturing the intricacies of the complex, dynamic, collaborative system, b) data conditioning to re-route missed values and clean up noisy channels, c) machine learning algorithms to investigate the history of the system and compare it to the present case, and d) determining the important bits of information to prioritize and resolve, and e) performing the tasks in an efficient and error-free manner. Moreover, the system must learn and reason in the way an expert human operator learns and reasons about the system. Perhaps this is not unfeasible. However, one of the characteristics of complex systems is that it is too difficult for a single person to describe; and moreover, humans are collaborators in the system, adding a complex system as a collaborator into the complex system.

The role of the human operator will never be completely replaced. The definition of what constitutes a “complex” system may change with time, given advanced capabilities of computing machines, but the human operator will remain the single point of failure because of her better ability to reason, create analogies, and problem-solve beyond what an algorithm can do.
Figure 8.8: Automation drives the entirety of the system: No humans here; the system is fully-automated
8.2 On Simulation

The difficulty with designing for complex, dynamic, collaborative systems is that there is a unique chicken-and-egg problem: users require a robust, complete, and correct mental model of the way the system works in order to be able to predict how it will behave as a result of their interactions. Specifically, the best way to understand how the Deep Space Network operates is by operating the Deep Space Network. This is evidenced by the long apprenticeship process of training new operators, and by the experience level of the trained operators.

Users generate a mental model of how the system works through extensive interactions with the system to establish proficiency. Interaction designers must understand and emulate this mental model in order to find opportunities for improvement.

Through extensive field work, I found that the best way to mimic the user’s mental model of the system is by simulating the system, or a portion of it. Simulation works well when a) describing the real system is impossible or impractical, b) when we are only interested in a subset of the real system, c) when we are designing something and need to iterate quickly, d) when we have specific evaluation requirements, and/or e) for rapidly prototyping new ways of interacting with an existing system.

Simulation provides benefit to interaction designers in the following ways. These uses of simulators are not exclusive, nor is the list complete.

**Verification and validation:** Using simulation output to verify or validate a system under test. For example, having produced a hardware component, one could use a validated simulator to run in parallel with the hardware, validating the functionality of the component.

**Training, re-training, cross-training:** Education on complex, high-risk activities; getting (back) up to speed on less frequent activities, or learning a new role within a collaborative task. For example, pilots, astronauts, and doctors use simulators to learn or re-learn new techniques, or to achieve empathy with other roles.
User evaluation: Leveraging the controlled simulation environment to test user-centered hypotheses. For example, there is a huge body of work with air traffic controllers to explore a certain new user interface feature and validate a theory of situated situation awareness [6].

Designing new interfaces: Rapidly prototyping new ways of interacting with a system. The work presented here is an example of this use of simulation.

In the section below, I outline three pitfalls faced in creating the simulation for this body of work. Each is accompanied by a design suggestion of how to address the problem.

8.2.1 Create Experiences, Not Simulations

Pitfall 1: Data-driven simulation lacks emotion, cost, and risk. A well-designed simulation can approximate a complex, dynamic, collaborative system, but the simulation is only a subset of the experience an operator has when interacting with the system. A simulated system can only simulate risk; for an expensive, high-stakes system, simulated risk may not provide sufficient emotional context under which an operator works. The researcher should pay attention to the environment in which an operator works, identifying the important factors that must be replicated for a realistic simulation. Imagine inviting an operator who is accustomed to a busy, high-stress operations facility full of noisy machinery and shouting colleagues into the researcher’s quiet laboratory containing wall-to-wall carpeting and soothing pictures on the walls. Taking the operator out of the situation alters the operator’s ability to perform as usual. For situation awareness, the psychosocial environment of the operator’s workplace is just as important as the technical and computing environment. Researchers can address this by creating experiences rather than simulations. The distinction here is that an experience encompasses the data-driven simulation and the environment that hosts it. For example, by coupling data-driven simulation with user enactments [49], researchers
recreated a parent’s morning routine including the emotions, even in the lab: they asked the working parent to call the subject-parent and cancel childcare plans with no explanation. This caused the subject-parent sufficient exasperation to carry out the simulation activity.

8.2.2 Simulate the Data for Data-Driven Simulation

Pitfall 2: Real data is difficult to obtain, realistic data is difficult to create. A data-driven simulation is only as good as the data that drive it, and in most cases, it is better to simulate the data stream than to use real-time data. A simulated data stream allows the researcher to control the scenarios under test and ensure consistent data at the appropriate level of fidelity. Often, real-time data networks, such as the one for the Deep Space Network, do not allow client application access at the source. Client alternatives further downstream down-sample the data (for bandwidth and long-term storage considerations), reducing the fidelity of it and rendering the data inadequate for a real-time simulation. The clever researcher can simulate the data stream – a simulation driving a simulation – and control for test scenarios, level of fidelity, and user experience. This requires the researcher to fully understand the scenarios under test, and to vet the simulated data stream with operators at several points, including but not limited to a) at scenario definition, when the scenarios are still being discussed and considered for the application; b) at scenario refinement, when the scenario has been chosen and is being translated into instruction-level code for programming; c) at code refinement, when the scenario has been codified and now needs to be vetted by someone who fully understands the scenario; and d) at simulation readiness, when the fake data stream has been written and is being used to drive the simulated experience.

Starting with a simple data simulation and iteratively increasing its fidelity (with operator feedback as outlined above) gives the researcher greater confidence that she is accurately representing the scenario.
While waiting for a data stream to become available, the researcher may test concepts rather than designs. The difference is that a concept addresses an operator need, and a design embodies the concept. For example, an operator need may be to see all relevant subsystem telemetry collocated in a single display. The concept is a unified subsystem telemetry display; the design of the telemetry display may be its presentation, layout, and aesthetic elements. Techniques such as “speed dating” (also known as needs validation) [9] can help the researcher determine if the needs exist. Prototyping potential concepts on paper, and showing them to operators, quickly assesses whether those concepts provide benefit to operators.

The Wizard-of-Oz technique of prototyping allows the researcher to bypass the data stream altogether by simulating the data stream using a domain expert. Instead of attaching the simulation to the data stream (real or simulated) for an operator to test, the researcher asks a second operator to be the system “Wizard,” reacting to the operator’s inputs, as if the Wizard were the system and associated data stream. For example, when testing a Deep Space Network operator station but lacking a proper data stream, one operator sat at the station, and another – the Wizard – sat at the end-cap of the station at a secondary display. The Wizard could see everything the operator was doing, but the operator could not see the Wizard’s display. The operator interacted with the system normally. The adversary intercepted the operator’s inputs and modified the output channels to simulate an appropriate system reaction. In this way, the researcher tested the concepts and system interactions, and the Wizard provided the algorithms necessary to achieve a compelling conversation with the system.

8.2.3 Don’t Try To Make It Fun (It’s Not)

Pitfall 3: Real-time simulations are slow and boring. For most scenarios in a real-time dynamic collaborative system, especially with bursty workload, it’s like asking an operator to sit through a real-time simulation of watching a pot of water boil.
Part of the unique problem of real-time dynamic systems is the long periods with little happening. How do you design a test for an operator to sit through the long periods of little to do, yet still remain alert for the bursts of activity? The temptation to address this problem by either increasing the speed of the simulation or by increasing the number of things an operator must do (or pay attention to) is wrong. Increasing the speed of the simulation unfairly dilates time. Things that happen quickly in the simulation can be overlooked by the operator. Moreover, the long periods of low activity may be important for studying the operator’s development and maintenance of situation awareness. Similarly, increasing the number of things an operator must do may remove the realism of the simulation. Under one theory, the operator’s attention resources are malleable, meaning the attention capacity grows when there is more to do, and shrinks when there is less to do [78], so creating more for the operator to do just for the sake of simulation changes the operator’s methods of dealing with information. The researcher must ask herself about the tradeoffs between creating more for the operator to do versus the realism of the scenario. If an operator must monitor two items of interest simultaneously, increasing that number to six just to “keep it interesting” for the operator decreases the realism of the simulation. Being specifically cognizant of the operator’s typical working environment and the distractions therein may guide the researcher to an appropriate distractor task.

Using expert operators rather than “hallway sampling” subjects of convenience, and ensuring that each operator is invested in the design under test, may help operators remain motivated to be engaged with the scenario.

Prototyping the experience may help operators cope with the slow and boring nature of the task. Operators are used to the work they do, and are well acquainted with the bursty nature of the task. Some suggestions for providing environmental cues include providing any documentation the operator is used to having for the task, setting up the desk in the same way as in the operations facility, and distractions similar to the
operator would have at work.

Finally, rapidly testing scenarios in bite-sized pieces may reduce operator fatigue. For example, if testing the design for whether it allows operators to glanceably understand the state of the whole system, rapid testing of screenshots may be better suited to the test than asking the operator to sit through an eight-hour shift.

8.3 Measuring Performance, Workload, and Situation Awareness

The most straightforward method to measure performance is by measuring how long it takes a subject to perform a task. Additional metrics include measuring the number and type of user error, time to perception of an error (e.g., an error state that the user must find), misdiagnosing errors (including false positives and false negatives), and others. Further, subject performance and confidence metrics provide some information about a subject’s perception of how well they did, and how confident they feel in their performance, among other factors.

The NASA-TLX is a workload measurement scale [32] that has been in use for over 20 years [31]. The measurement tool asks subjects to rate, on a bipolar scale, the amount of perceived workload in six areas: mental, physical, and temporal demand; performance; effort; and frustration. Additionally, subjects sort the six areas in a systematic way, based on how much they feel the area contributes to their overall feelings of workload. The subject’s classification together with the ratings create a combined workload score.

Like eliciting mental models, measuring situation awareness is tricky. Subjective techniques, such as asking a subject about how much situation awareness they are feeling, are easy to administer, but pose problems because it is difficult for a subject to know what they don’t know. An example of a subjective technique is the the Situa-
tional Awareness Rating Technique (SART): a bipolar rating scale to gauge subjective situation awareness [68].

Objective techniques, such as probes or questions during a simulation or task, may be more compelling data, yet they remove the subject from the situation and thus may affect awareness. In order to test situation awareness, the researcher must possess a thorough understanding of the kinds of data that will result in level 1 (perception), level 2 (comprehension), and level 3 (projection) states.

In addition to becoming the main tool to assist in authoring interface requirements for situation awareness, the job task taxonomy [22] becomes an entry point for creating probes (i.e., questions) to test an operator’s situation awareness when using the interface. Endsley’s theories on situation awareness include a measuring system called Situation Awareness Global Assessment Technique (SAGAT) [19]. For conducting a SAGAT query, the researcher identifies key scenarios to test an operator’s situation awareness, and asks the operator to perform the necessary functions, e.g., in a simulator. The researcher stops or pauses the simulation and blanks the display at certain prescribed times to administer probes, asking the operator specific questions about the state of the system. The results are aggregated into a metric of situation awareness.

Critics of the SAGAT claim that pausing the simulation creates an unrealistic work simulation. Durso et al. describe the Situation Present Assessment Method (SPAM). Using SPAM, the researcher continues the simulation without interruption, but asks the subject whether they are ready to receive a question. And unlike the SAGAT, which presents many questions at the same time, the SPAM method presents one question at a time, thereby complementing the subject’s workflow rather than working against it [13, 14, 15].
8.4 Advancements in Situation Awareness Theory

Critics of situation awareness theory argue that it is impossible for people to keep all of the necessary information in their working memory alone. Evidence exists supporting the idea that users situate knowledge within their environment by relying on external artifacts such as displays and notes, and thus, the person’s memory represents only a partial situation awareness [14]. Situated situation awareness distributes the person’s cognition to the environment; with situated situation awareness, an individual does not need to maintain the actual knowledge in his or her head, but rather can rely on the structure of the environment. For example, rather than memorizing an airplane’s call sign, an air traffic controller may know where on an information display to look to find the information [7]. Endsley’s model of situation awareness, and the tests necessary to test for situation awareness, do not address situated situation awareness. For example, Endsley’s test for situation awareness would ask for the aircraft call signs, fuel levels, headings, and destinations specifically. The air traffic controller would need to answer these questions without looking at the information display, even if he or she knows where to find the call sign if required [14]. To use an example from software engineering, rather than memorizing a particular function call, a programmer might memorize where to look up the function call. The information is situated in the environment, and the programmer needs only access the environment to retrieve the data.

Unlike Endsley’s SAGAT test for situation awareness, which pauses the test and blanks the screen, relying on the operator’s memory, the Situation Present Assessment Method (SPAM) method of measuring situation awareness [15] does not break the simulation. With this method, the researcher asks the operator when he or she is ready to answer a question. The operator chooses an appropriate time, and the researcher presents a multiple-choice question to the user. Researchers measure response times. This method gives three kinds of results: quick response times give SA stored in an operator’s working memory; medium response times give SA stored in a situated agent;
and long response times give clues about information not in an operator’s SA. Chiappe et al. (2012) discuss the reasons SPAM works better under the situated situation awareness model: with SPAM, a researcher can separate the internal representation of situation awareness from the external, or situated, SA [7]. With this method, the researcher can identify what the operator keeps in working memory, which aspects of situation awareness are offloaded onto the environment, and which necessary knowledge items are completely missing from the operator’s situation awareness.

8.5 Malleable Attention and Underload

Young and Stanton defined, and provided evidence for, a theory about how operators' attention changes in response to how much attention demands of a system. This is called the malleable attention resources theory. A person’s dynamic attention capacity depends on the level of task demands [77, 78]. When an operator’s job requires more attention resources, her attention capacity grows to accommodate the need. Inversely, when her job requirements lessen, the attention capacity lessens as well. Under this theory, a diminished attention capacity means an operator cannot re-establish a robust mental model quickly and efficiently should the need arise.

The malleable attention resources theory has two important implications. First, under a dynamic attention pool, an operator experiences periods of high activity (when demands are high or complex) and low activity (when demands are low or trivial). This means that there exist periods of what Young and Stanton called mental underload, or periods of extremely low mental demands which could be responsible for degraded performance. In certain environments, this difficult-to-detect phenomenon may be more dangerous than mental overload, which is well-understood. Second, the theory implies workloads that change over the course of an operator’s shift. For example, imagine a security guard at an apartment complex. Most of the activity that the security guard must monitor occurs during peak commute hours: people going to work and school, or
returning from work. Sometimes people will return home for lunch. During the majority
of the day, however, and into the night, activity slows or even halts.

If operator performance depends on creating and maintaining robust and accu-
rate mental models of the system, an operator’s attention capacity varies proportionally
to the tasks’s mental demand, and the demand the system places on the operator varies
greatly throughout the day in “bursts” of high activity punctuating long periods of
inactivity, then interactions must be tailored to these specific cases of mental underload.
Chapter 9

Conclusion and Implications for Research and Practice

Complex systems pose specific operational challenges, related to the interrelationships of system variables, the large quantity of data generated by the system and its subsystems, the use of automation, and the dire consequence for error. Operating the Deep Space Network, a collection of telecommunication equipment situated around the globe and providing uninterrupted 24/7 support for deep space missions for over fifty years, presents the same challenges.

Situated situation awareness theory provides powerful knowledge we can use for design. What an operator uses to develop and maintain situation awareness separates into three discrete categories: information in working memory, with nearly instantaneous access for the operator; information situated in the environment, which the operator knows how to access and can access quickly; and information elsewhere, requiring lookup or investigation.

Early human-computer interaction work showed that people organize information in tiers. One example is the “pile” method of organization, which allows users to make virtual piles of information on a desktop [41]. More recently, this analogy can be
seen in file system research – another problem in which a lot of information must be managed by the end-user. My colleague Aleatha Parker-Wood studied emergent categorization, provenance, and metadata in making sense of a music library [53] and files in a file system [54]. Bergman, et al. studied how people find personal files and found that hierarchies are typically shallow, with a median of three subfolders inside a top-level folder [1].

Design researchers and human factors researchers have yet to reach a consensus on how to design for situation awareness. So far, though united by a common overarching goal and by a shared understanding of how the person’s mind works, every case of design for situation awareness is different. Design research techniques are evolving, psychologists are developing deeper understandings of how people work, and situation awareness theory, design, and measurement are hot topics in human factors research.

General design guidelines for human-computer interaction include visibility, a good conceptual model, good mappings, and feedback for an operator to develop and maintain a coherent mental model of a complex dynamic system [47]. Endsley presented many design principles and warned against pitfalls in her book, Designing For Situation Awareness; some principles include organize information around goals, support comprehension by presenting level-2 situation awareness information directly, provide assistance for level-3 situation awareness when operators project the future system state, support global situation awareness, support trade-offs between goal-driven and data-driven processing, make critical mental cues for schema activation salient take advantage of the operator’s ability to parallel process, and use information filtering carefully so that the operator has the proper amount of information to maintain situation awareness [20]. Design principles for creating small peripheral information interfaces are make information always present (careful filtering), minimize motion, make the information personal, make the notification system extensible, support quick drill-down and escape, and make it scalable [4]. Some theoretical factors influencing whether a person will place
information in working memory, situate it in the environment, or access the information by searching for it in the environment include confusability, ease of encoding, frequency of use, ease of access, likelihood of task interruption, level of expertise, working memory capacity, and stress and/or anxiety [5].

9.1 Implications for Design for Situation Awareness With Bursty Workload

The hallmark of a bursty workload environment is long periods of low activity punctuated by short-duration bursts of high activity. According to the malleable attention theory, an operator’s attention capacity changes with the demands of the task [77, 78]. The interface must occupy the operator and maintain the operator’s situation awareness when little is happening, yet be able to rapidly get the operator’s attention and help the operator grow the attention capacity necessary to keep up with the higher levels of task attention necessary. In periods of high activity, the interface must be tailored to provide necessary situation awareness information rapidly and efficiently. Finally, the interface must be able to shrink gracefully from the high-activity level to a low-activity level. This requires the designer to consider four specific cases in which to design for operator situation awareness.

It is crucial that the operator create and maintain a robust, up-to-date, and complex mental (internal) representation of the system at all times. The operator accesses this internal model for predictions about system behavior. In computer science, the term coherence implies an equivalence relation between the contents of the nodes in a multi-nodal memory system. When data are stored in multiple caches, the caches are considered coherent if their data are mutually consistent. In this case, the operator’s mental model, or internal representation of the system, including all expectations and projections of how the system will behave, must remain consistent with the actual sys-
tem so that the operator’s predictions about how the system will behave will be valid. A correct (and complete) mental model is not enough; it must also remain coherent, or aligned, with the actual system on which it is modeled. It is crucial that the interface help the operator maintain that alignment. The cases outlined below are shown in Figure 9.1.

*Case 1: Low activity.* Low activity periods (“underload”) may include monitoring-only tasks with only a few individual threads. When operating under periods of little activity, an operator might focus on other tasks for periods of time and would require the interface to alert the operator of what’s happened in the interim. The operator may look away from the interface for a period of time while focused on other tasks. Even when performing secondary tasks, the operator still has an active mental model of what the system is doing. Then, when the operator looks back to the interface to check on the state of the system, the operator updates her situation model, incorporating any new information and synchronizing the mental (internal) representation of the system with the real (external) system.

To accommodate this workflow, the interface must provide access to what’s happened in the period the operator was doing other things. The operator should be able to find the missed data easily. One way to show historical data is with sparklines, concise and visual representations of interval data showing the magnitude of the dependent variable (on the $y$-axis) against time (on the $x$-axis) in a small space [70].

The interface must make relevant changes prominent. The interface must answer the operator’s question, “What’s happened since I last looked up?” And the changes that the interface presents may be of a coarser grain. For example, small deviations indicative of rudimentary noise in a telemetry signal may be irrelevant at this scale; however, a larger deviation such as one indicative of electrical interference (if relevant) may be cause for highlighting the change.

*Case 2: High activity.* Periods of high activity may include monitoring and
control tasks in multiple individual threads, possibly with anomalous conditions. In these periods, the operator may struggle to keep up with the changing state of the system, or keep the internal situation model consistent with the external system.

To design for periods of high activity (“overload”), one might consider “calm” displays [75] or displays that encourage change blindness [33] for low-impact data. This would allow the operator to focus in the areas that require attention, and not distract the operator from irrelevant changing details. Calm displays do not curate the information displayed, yet they do curate the animation causing attention to be irrelevantly focused on the changes. If mis-curated, there is a danger of failing to focus attention to relevant variables.

Case 3: Grow attention capacity efficiently. Transitioning from low-activity periods to high-activity periods presents unique challenges because the operator’s attention capacity must grow to match the new demands. As described above, the operator must have already had a robust mental model that she can access. In this case, the challenge is that of maintaining alignment between the operator’s mental model and the system, in this rapidly-changing environment.

A common solution is to use alarms, alerts, or notifications. These displays provide immediate, invasive information to the operator. However, studies show that alarms are over-used, and as such, their utility suffers: operators stop paying attention to alarms when there are too many of them or when they are irrelevant [65]. A special display for notifications may have better effects [4] but may take the operator away from the interface requiring attention.

One design solution may be to use real-time data processing algorithms to monitor data and detect the beginnings of the high-activity periods. Alerting the operator that a burst is on the way may help the operator prepare mentally for the approaching period of high activity.

Case 4: Shrink gracefully. When a high-activity period ends, the operator’s
Figure 9.1: Four design opportunities for bursty environments: design for (1) low activity; (2) high activity; (3) low-to-high transitions in which operators must grow their attention capacity quickly; and (4) high-to-low transitions in which operators must shrink their attention capacity gracefully.

attention capacity shrinks back to low-activity levels. An interface should gracefully end the high-activity period. The operator should be able to perceive (at a glance) that the rate of activity is decreasing.

9.2 Design Guidelines

The following guidelines define effective micro-displays in bursty environments.

The purpose of a micro-display is twofold: to tell the operator whether everything is okay, and support operator mental model re-alignment following a period of being out of the loop, such as in low-to-high and high-to-low task load scenarios common in bursty workload environments.

1. Understand what operators mean by “everything.” Data-driven designs depend on determining data requirements. Design research plays a critical role in understanding what contributes to an operator’s assessment of overall system state. Is it the combination of subsystem states? Component availability? Variables within a range? The designer’s job is to understand what operators mean by
“everything,” and what goes into determining whether everything is “okay” for a specific system.

Operators of complex systems tend to secondary tasks, especially in periods of low task load. Operators must then return to the primary task and re-align the mental model of the system’s state with the actual system’s state. The micro-display can help operators re-establish their understanding of the system following periods of being out of the loop.

2. **Provide information only (do not determine results).** Highly-trained expert operators monitor and control complex systems. These operators have a wealth of knowledge due to training and experience. It is their job to determine whether the system is behaving nominally given the operators’ expectations. Provide the right information for an operator to make a reasoned decision about system operation.

3. **Work in parallel with other displays.** The micro-display displays information that helps operators determine whether the system is behaving nominally. An operator should still be able to drill down as needed for troubleshooting and detailed system information. Moreover, in order to preserve the operators’ screen real-estate, the micro-display must be compact to work alongside other displays.

4. **Use static position encoding for glanceability.** To enable quick access, the information in a micro-display must be presented in the same form and in a standard location on the display every time an operator accesses it. Over time, operators come to know where to look on the micro-display quickly. The data will change within the perimeter of the display, but the standard layout must be consistent to preserve glanceability and provide a dependable reference point for the operator.

These design guidelines work to support first- and second-level situation awareness requirements. Providing operators with the correct information allows them to
make reasoned judgements about the state of the system. However, for third-level situation awareness requirements, operators need to be able to project the future. Thus, the micro-display must answer the question, “will it all keep being okay?” for some definition, as described previously, of “all” and “okay.” In this case, the data requirements for the micro-display could be altered to use data sources that have pre-computed differences between actuals, predicted values, and expected values, and/or taken into account historical and environmental factors.
Appendix A

Technical Appendix: Building the Postage Stamp

To address the goal of developing and maintaining global situation awareness, I invented a display called the *postage stamp*, thus named for its relative dimensions, and designed with the purpose of capturing basic elements about a track.

NASA Deep Space Network operations depends on operator situation awareness. Operators in three sites around the world coordinate the uplink and downlink of critical science data for most of the world’s deep space missions. The semi-automated operator-equipment system lends itself to common situation awareness errors, including human-out-of-the-loop errors. Using Endsley’s theory of situation awareness, I present the results of an investigation of Deep Space Network operators’ workflow.

In this chapter, I present a user-centered design intervention for the operators, called the Postage Stamp: a scalable and tileable low-level summary display, based on the workflow analysis of the operators. I document the process of designing the Postage Stamp, ground the design in insights from user research, and describe how I received feedback on the design given the expert niche nature of the user population.
A.1 User-Centered Design of a Data-Driven Prototype

With the direct collaboration with Goldstone and JPL link control operators, I identified 12 elements that operators described as basic track information. I incorporated these elements into the postage stamp design, using colors familiar to LCOs, and choosing shapes and fonts that would be visible from a distance and at small scale for two reasons: First, because LCOs sometimes move about the room during attended supports, it was imperative that the stamp be visible from a distance. Second, I postulate that global situation awareness may include identifying the basic characteristics of non-attended supports (i.e., supports that other LCOs may be responsible for). The form-factor created by the postage stamp lends itself well to scaling and tiling, as in the case of small multiples [69].

A.1.1 Designing summary information display on paper

The basic characteristics of a support, which were incorporated into the postage stamp, include the following.

1. Name of support (usually, spacecraft moniker)
2. Stage of support (pre-cal, in-track, or post-cal; awaiting support; post support)

3. Supporting antenna number

4. Antenna state (stowed, slewing, on-point, stowing, stopped)

5. Support type (one-way, two-way)

6. State of uplink (radiating, in lock)

7. Uplink power (if available)

8. State of downlink (receiving, in lock)

9. Downlink conscan number (if available)

10. State of ranging mode (enabled or disabled) for uplink and downlink ranging

11. State of command mode (enabled or disabled)

12. Time to beginning or end of track

First, I created the postage stamp on paper. As an example of the postage stamp in use throughout a day, I printed a booklet called *A Day In the Life of a Postage Stamp* to identify and expose all state transitions through a 24-hour period for four tracks.

I selected the tracks from published sequences of events across Goldstone and Canberra, filtering them for their simplicity: Voyager 1 (VGR1) and Voyager 2 (VGR2) with predictable supports and limited capabilities, supporting only uplink and downlink; Cluster (CLU1) with a downlink-only support; and India’s Mars Orbital Mission (MOM) – which, at the time of this work, was in cruise to Mars – with support for uplink, downlink, ranging, and command mode. Special support class such as multiple spacecraft per aperture and Delta-Differential One-Way Ranging were excluded from the initial analysis.
I identified the state transitions for the fields of the postage stamp, including the phase of the track (Figure A.2), antenna state (Figure A.3), downlink (Figure A.4) and uplink (Figure A.5), command mode and ranging (Figure A.6), and one-way and two-way coherent communications (Figure A.7). The state transitions drove the visual component of the postage stamp.

A.1.2 Creating a command language for data-driven display

Next, the data-driven portion of the display needed to be written in software. I consolidated the four identified tracks into a single sequence of events file. Together with my colleague Bryan Duran, who wrote the postage stamp rendering engine in JavaScript, I wrote a rudimentary interpreter for the sequence of events file that would trigger the necessary state transitions on the parts of the postage stamp.

I learned that the sequence of events as published by the project was insufficient to capture the state of the system. The sequence of events files did not contain three main types of information: information about spacecraft, track state, antenna pointing; transitionary events, such as when a spacecraft is expecting downlink after a certain round-trip light time; or error events, including nominal events such as when a spacecraft drops one-way downlink lock to enable two-way coherent communication and anomalies such as a dropped lock for equipment failure or weather events. Additionally, I calculated any missing round-trip light time information in the sequence of events files. I created a command language for the interpreter to include the three types of events (non-sequential telemetry, transitionary events, and errors) and methodically added the events to the sequence. Finally, I added values to the number-specific portions of the stamp (e.g., downlink signal strength, transmitter power) with the help of several link control operators. I planned and prototyped the simulation, and verified its realism with subjects.

Two rounds of user testing, plus a pilot test, were used to determine whether
Figure A.2: Postage stamp state diagram for the track section of the stamp, which indicates the stage of support by background color. The text always indicates the name or mnemonic of the spacecraft being supported. Before the support begins, this portion of the stamp is idle or waiting for the beginning of the support. Beginning of activity (BOA) coincides with the initiation of the pre-cal stage of support. Beginning of track (BOT) corresponds to beginning the in-track stage of support. End of track (EOT) corresponds to the beginning of the post-cal stage of support. End of activity (EOA) corresponds to the end of support. Alerts are handled separately.
Figure A.3: Postage stamp state diagram for *antenna* section of the stamp, which indicates the antenna state by background color. The text always indicates the antenna number providing support. Before the support begins, this portion of the stamp is idle or waiting for the beginning of the support. As the antenna begins slewing to point, the color changes and the antenna state is slewing. Once on point, the antenna is said to be tracking, and moves along with its target. After the support, the antenna is stowing and moves to stow. Finally, the antenna is stowed and moves into the stowed or done state. Alerts are handled separately.
Figure A.4: Postage stamp state diagram for downlink section of the stamp, which indicates the downlink lock status by background color. The text indicates the downlink channel number. Before the support begins, this portion of the stamp is in the idle state. As downlink is acquired, the state changes to busy until lock is acquired; then the downlink portion state is in lock. If lock is dropped, the state changes to out of lock (an alert state). Downlink lock can be reacquired.
Figure A.5: Postage stamp state diagram for uplink section of the stamp, which indicates the uplink lock status by background color. The text indicates the transmitter power. Before the support begins, this portion of the stamp is in the idle state. As the transmitter begins radiating, the state changes to busy until uplink lock is acquired (if possible); then the uplink portion state is in lock. If lock is dropped, the state changes to out of lock (an alert state). Uplink lock can be reacquired.
Figure A.6: Postage stamp state diagram for command mode and ranging which indicates the command and ranging mode state by background color. If command mode is engaged, the command portion of the stamp highlights. When it is turned off, the command portion un-highlights. For ranging mode, a lock is possible. Until lock is being acquired, the ranging portion of the stamp displays a busy state. If ranging is turned off, the ranging portion of the screen un-highlights.

Figure A.7: Postage stamp state diagram for spacecraft support obtaining one-way lock and then two-way coherent lock. This portion of the postage stamp is a thin line around downlink and uplink signals.
the postage stamp would be an effective method of building and maintaining situation awareness. Unfortunately, all quantitative data were discarded because of the low number of participants. In order to participate in the user test, a trained link control operator working for the Deep Space Network had to travel to the Jet Propulsion Laboratory in Pasadena during a working day. Remote testing was in development, but was not yet available. At the time of this thesis, there are at most 54 individuals who fit this profile worldwide (18 in Goldstone, 16 in Canberra, and 20 in Madrid. Funding options limited the pool of candidate user testers to Goldstone. Further restrictions, such as seniority, scheduling and staffing, and union rules, allowed us to test with just two expert users.

For the first test, the same schedule interleaving spacecraft activities across four tracks was used to drive a computer-based prototype of the postage stamp. The result of the first test showed that the information feeding data to the displays was insufficient for operator consumption; the second prototype test instead used a full day of real telemetry data.

Individual test sessions involved observation activities which were halted at random intervals to administer situation awareness questions similar to SAGAT [16]. Operators were asked to think aloud and to narrate the state of the system.

A.2 First user test (June): Trust in data sources

Two operators from Goldstone Deep Space Communications Complex visited the laboratory June 3, 2014 to test the effectiveness of the postage stamp as a top-level summary status display. Two scenarios were identified with the help of two other subjects.

One operator at a time monitored eight postage stamps. Up to four stamps were active at a time, running through a total of two scenarios.
A.2.1 Nominal and off-nominal scenarios: MOM and Voyager

In anticipation of user testing with trained link control operators, I created two scenarios in which a track dropped lock, and then, some time later, re-acquired it.

In the nominal case, Mars Orbital Mission (MOM) received a downlink signal and locks on. The postage stamp showed the downlink lock as a green background in the downlink signal box, and the appropriate numerical value in the box. Then, the antenna began radiating an uplink signal, shown by lighting the uplink box green and placing the appropriate uplink power value in the box. One estimated light-time later,\(^1\) the spacecraft dropped its one-way operations (lock) in order to begin two-way coherent operations. The postage stamp showed the dropped lock as red in downlink, followed by green downlink a few seconds later and the two-way bar appearing.

In the off-nominal case, Voyager dropped downlink lock, then reacquired it five minutes later.

I changed the time of the supports to mutually coincide, so that I could investigate supporting three links per operator. Four additional tracks – Maven (MVN), Dawn (DAWN), Advanced Composition Explorer (ACE), and Cluster-4 (CLU4) – were selected from the list of supports in Goldstone and Canberra on the same day as the primary four supports. Figure A.9 shows a screenshot of the test scenario.

A.2.2 Distractor task

In field work, I observed operators frequently engaged in other work, discussing work and non-work topics with other operators, or even learning a new programming language. I hypothesized that a distractor task could mimic the nominal level of distraction that operators experience. Moreover, a distractor task would keep operators’ focus and test the glanceability of the postage stamp display.

\(^1\)The light-time between Earth and Mars is anywhere between six and 20 minutes. At the time of this user study, MOM was en route to Mars, so six minutes was used as a ballpark.
To select a distractor task, I investigated applications with the following qualities:

- Be sufficiently engaging, yet easy (not imposing cognitive load)
- Contain a penalty for losing focus on the distractor task. For example, it could be real-time: If the user looks away, the game should continue.
- Be monotonic in difficulty (i.e., not get more difficult over time) to maintain an even level of operator focus
- Be simple to learn and control

I chose *Asteroid Speedway* (2014), released by Chris Srch through the Apple iTunes store. In the game, the player controls a rocket with his or her finger or a stylus by dragging it around the screen, and maneuvers the rocket around asteroids falling from the top. Collision with an asteroid causes a crash. The player receives a score based on the distance travelled. Figure A.8 shows a screenshot from Asteroid Speedway.

**A.2.3 Results: Trust as related to data validity and reliability (and source)**

Overall, both operators liked the displays, and said they could see themselves using the displays in the field at Goldstone. The main concern was data validity. The source of the data is of paramount importance for reliability and situation awareness. Neither operator had trust for the timing information and sequences of events taken from the sequence of events files — “[sequence of events times] change when the predictions change. In one 24-hour period I’ve seen them change up to ten times.”

Both operators abandoned the distractor task within three minutes, choosing instead to turn from their consoles and talk to the researcher.
Figure A.8: Asteroid Speedway was used as a distractor task during user studies. In the game, the user maneuvers the rocket to avoid the asteroids and receives a score for the distance traveled before collision. Scores were recorded.

Figure A.9: Postage stamp setup for Goldstone link control operator visit, showing eight concurrent tracks. Only VGR1 and MOM were used for testing situation awareness and display effectiveness; the others were created for distraction.
Following the June user test, I changed the postage stamp design to use canned telemetry data coming directly from the equipment, rather than the sequence of events file which is subject to change. Bryan Duran implemented the changes.

### A.3 Incorporating real telemetry data into the postage stamp

To address the data reliability and validity concerns from the first user test, I removed the data driving the postage stamp and replaced it entirely with telemetry data.

I worked with an LCO to select a day of data that would have the following qualities:

1. Large amount of activity
2. Large selection of simple tracks
3. At least two visible anomalies

I selected the 228th day of the year (DOY) and obtained a day of telemetry data from all antennas. I defined the data mappings necessary to populate the postage stamp display, and Bryan Duran wrote an engine capable of processing the data, storing it in a database, and reading it back at the appropriate times. I obtained the sequence of events files for all spacecraft tracked on DOY 228, and I fed the events into the database for coordination with the telemetry data.

#### A.3.1 Wizard of Oz

One way to test an operator’s response to an off-nominal situation would be to have the operator view, and respond to, captured data containing an anomaly. Another way is with a “Wizard of Oz” testing setup. The Wizard is a researcher collaborating
with the subject in the subject’s experience with the tools. Using the Wizard of Oz technique allowed the project to test user experience capabilities that had not yet been developed.

I designed and built a Wizard of Oz display with the capability of pausing, resuming, and restarting the data stream, fast-forwarding the data stream to a specified time, and zooming to a specific event in the data stream by using the sequence of events. These capabilities allowed the researchers to target specific events, investigate them to become intimately familiar, and play back the events to operators to see their reactions.

Additionally, to test situation awareness with Endsley’s SAGAT method, the Wizard display could clear all values in the display while keeping the outlines of all visual elements. I kept the visual elements exposed because I hypothesized that operators look at the various displays as containers of information, or memory elements. Under this working hypothesis, the physical arrangement of the displays provide a valuable clue for easy recall. For example, if I ask an operator for a spacecraft’s downlink signal strength as part of the SAGAT test, the operator might look at the display, find the downlink area of the stamp, and use the visual clues of the stamp’s arrangement as a method of recalling the values the area contained.

A.4 Second user test (September): Postage Stamp Success

Figure A.10 shows the setup of the station for the user test.

Figure A.11 shows an operator at work during the user test.

A.4.1 Scenario Selection

To implement the postage stamp simulation, I chose a day of telemetry containing two off-nominal scenarios. These scenarios were finalists for inclusion in the user study. I studied each scenario thoroughly by reading the discrepancy report and associated documentation, discussing the scenario a subject, and viewing the log output.
Figure A.10: Link control operator setup for second user test (September 2014). Six displays, eight tracks. Sequence of events files printed, stapled, and collated by antenna.

Figure A.11: An operator monitors supports while solving math/memory problems.
produced during the anomalies. The two scenarios were as follows.

1. A hardware (station) anomaly with the Global Geospace Science WIND spacecraft resulted in loss of data. The problems were encountered prior to the beginning of the track, in pre-calibrating the 250W transmitter. The corrective action was to turn the transmitter off and on again, wait for it to warm up, and pre-calibrate it again.

2. A spacecraft anomaly with Mars Orbital Mission (MOM), in which the station failed to acquire the spacecraft’s signal. The problems were encountered about six minutes near the beginning of the track. The corrective action was to wait for the spacecraft to turn towards Earth, and verify the signal as soon as it comes into view.

I selected the scenario in which Mars Orbital Mission (MOM) had a late acquisition. The operator expected for MOM to come online at a specific time, which was already six minutes after the published time in the sequence of events, but the spacecraft was late. The anomaly was attributed to the project because the project provided an incorrect predicted time of acquisition. The situation resolved itself; however, the operator debugged the problem for several more minutes before seeing that the spacecraft was up and radiating.

**Distractor Task**

In accordance with the previous user test, I selected a distractor task. This time, the distractor task was an OSPAN test, designed to overwhelm operational memory [72, 73] and test situation awareness. I verified its use in the laboratory with colleagues and found it to be sufficiently engaging and occupying.
Table A.1: NASA-TLX metrics of mental demand (0=low, 10=high), temporal demand (0=low, 10=high), performance (0=good, 10=poor), effort (0=low, 10=high), and frustration (0=low, 10=high) show vastly different scores between the two operators across two tasks ($t_2$ and $t_3$)

<table>
<thead>
<tr>
<th></th>
<th>Operator 1</th>
<th></th>
<th>Operator 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ment</td>
<td>Temp</td>
<td>Perf</td>
</tr>
<tr>
<td>$t_2$</td>
<td>1</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>$t_3$</td>
<td>3</td>
<td>2.5</td>
<td>3.5</td>
</tr>
</tbody>
</table>

A.4.2 Testing Metrics

I used a SAGAT-style situation awareness test, delivered at two points in the user test: Approximately two minutes prior to an event, and two minutes after the event.

I used a modified NASA-TLX world assessment test to gauge perceived workload in using the intervention. I modified the workload assessment to include only mental demand, temporal demand, performance, effort, and frustration, as physical demand was irrelevant for our experiment: operators stay seated and have no physical component to this test.

A.4.3 Results

The major result pointed at the postage stamp as a successful design.

Situation Awareness and Perceived Workload

Reference Table A.1 for results from SAGAT and workload assessment. The workload assessment was delivered after the SAGAT worksheet.

Descriptive and inferential statistics hold no meaning for this small sample size of two operators. Qualitative results provide a deeper understanding of the effects of the system.
On the Visual Frames During Situation Awareness Test

I hypothesized that keeping the visual elements exposed would be handy clues for recall of the system state for the situation awareness test. During the test, the data presented are hidden from the user. However, neither operator looked up from the test to consult the displays, indicating that both operators completed the situation awareness test entirely from memory.

On the Distractor Task

The distractor task in the second user study differed significantly from the one selected for the first user study. In the first user study, Asteroid Speedway was abandoned as a distractor task by both operators within minutes. Here, the distractor task overwhelmed the operators with its insistence for timely attention. On completing the Task 2 worksheet, Operator 2 said, “I wasn’t really paying attention [to my tracks] because I was [performing the distractor task activity].”

A.5 State of the Future: Stamp Roll

Knowing what’s coming up next was important to operators. In user sessions, I provided schedules of events (SOEs) for the whole simulated day, but this was seen as unhelpful when looking over an entire day’s worth of tracking and three or more simultaneous supports. Operators suggested designers create and test an interactive, electronic SOE that allows operators to additively create the aggregate SOE that they feel is most relevant to them.

One way to develop and maintain a model of the projected future (level-3 situation awareness) can be operationalized by stamp roll, presented here in its initial form (Figure A.12). The postage stamp roll is a projected view of all of the day’s events for the attended tracks on a single roll or spool of miniature postage stamps. Initial
feedback from operators on the stamp roll was positive. One operator said the stamp roll would help him plan out the day, and see at a glance what to expect from the support.
Figure A.12: Stamp roll
Bibliography


